MECHANISTIC AND FUNCTIONAL CHARACTERIZATION OF *SULFOBOLUS SOLFATARICUS* PRIMOSOME COMPONENTS

by

Brian W. Graham

Bachelor of Arts, Hiram College, 2007

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

2014

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Brian W. Graham

It was defended on

August 11th, 2014

and approved by

Patricia Opresko, Ph.D., Associate Professor, Environmental and Occupational Health

Sunil Saxena, Ph.D., Professor, Chemistry

Stephen Weber, Ph.D., Professor, Chemistry

Dissertation Advisor: Dr. Michael Trakselis, Assitant Professor, Chemistry

MECHANISTIC AND FUNCTIONAL CHARACTERIZATION OF *SULFOBOLUS SOLFATARICUS* PRIMOSOME COMPONENTS

Brian W. Graham, Ph.D.

University of Pittsburgh, 2014

DNA replication is a complex process that involves an organism's ability to faithfully replicate and maintain its genomic code. One of the initial stages of DNA replication is the separation of double-stranded DNA into complementary single-stranded DNA. This action is performed minimally by the replicative DNA helicase. DNA helicases are present in all three domains of life and are essential for cellular survival. Mutations to helicases can lead to a predisposition for susceptibility to cancers and other diseases. The actual mechanism for DNA unwinding and processing at the replication fork, however is unknown. We chose a model archaeal system (*Sulfolobus solfataricus* - *Sso*) that is homologous to eukaryotes, but is simplified in its DNA replication machinery. The DNA replicative helicase in *Sso* is the homohexameric minichromosome maintenance protein (MCM) which is homologous to the eukaryotic heterohexamer Mcm2-7. I have discovered the novel mechanism for unwinding in *Sso* which we named the steric exclusion and wrapping model (SEW). The SEW model involves MCM encircling the DNA leading strand and sterically excluding the lagging strand which then physically wraps around the exterior of the helicase. We hypothesize this wrapping protects the single-stranded DNA from degradation to allow for elongation on the lagging strand. The binding path on the exterior of the helicase was further characterized through the mutation of two basic residues that we determined electrostatically interact with the single-stranded DNA. Additionally,

I have characterized a conserved single point mutation that disrupts hexamerization, implying additional functional significance for the motif in which it is located. Finally, I characterized interactions within the archaeal primosome including the atypical physical interaction between eukaryotic-like helicase MCM and bacterial-like primase DnaG and the functional relationship between MCM and single-stranded binding protein. My thesis provide a significant contribution to the overall understanding of the mechanism and function of proteins in the archaeal primosome, which will be applicable to more complex eukaryotic systems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Matriculating to Hiram College for my undergraduate degree, I did not want to declare a major to "keep my option open." Quickly, however, I realized that chemistry and science was my passion. One day Dr. Adrian Michael came to give a talk and spoke highly of Pitt's graduate program in chemistry. This glowing recommendation was echoed by my mentor and Pitt chemistry grad school alum Dr. John Cragel. Having taken their advice I applied to Pitt and discovered my curiosity for biochemistry after hearing Dr. Michael Trakselis present his research in a half hour talk. He seemed like someone I could both work for and with and most certainly someone from which I would gain much knowledge.

This thesis is the result of years of both smart and hard work throughout my graduate school tenure. I would like to take this opportunity to thank those who have helped me over the years. I would first like to thank my advisor and mentor Dr. Michael Trakselis. He has helped shape me into the researcher and scientist that I am today. Michael, as he is known by the graduate students, has provided keen insight and support throughout my time at Pitt. He has always been there to provide assistance to both myself and the other graduate and undergraduate students in our lab. Michael has cheerfully provided assistance too many ways to count and without his guidance and knowledge I would not be the person and scientist I am today. He has also taught me how to properly write and create figures for scientific papers. I thoroughly appreciate everything he has done both for me, our lab, and the University of Pittsburgh.

The current and former members of the Trakselis lab have also aided me in my quest for graduation. The people who work for Dr. Trakselis have helped me improve as a researcher and I hope that I have been able to help them as well. I will do my best not to miss anyone, Jonathan Bartko, Robert Bauer, Michael Begley, Sean Carney, Will Denq, Rodger Dilla, Elizabeth Jeffries, Hsiang-Kai Lin, Jessica Meyers, Andrey Mikheiken, Parry Patel, Cory Rodgers, Curtis Schauder, Anna Schreider, Erika Varner, Ian Wolff, and Zhongfeng Zuo. Additional thanks go to Robert Bauer, with whom first authorship was shared on the MCM-DnaG manuscript.

I would also like to thank our collaborators and their corresponding labs for enriching my scientific knowledge by providing access to equipment unavailable within our department. Dr.

Sanford Leuba and his graduate student Grant Schauer for their expertise in single pair fluorescence resonance energy transfer. Dr. Alan Marshall and Dr. Nick Young and their graduate student Yeqing Tao for collecting hydrogen deuterium exchange-mass spectrometry data. Dr. Guillermo Calero and his graduate student Christopher O'Neill Barnes for providing materials for 2-aminopurine fluorescence data collection. A general thanks to the faculty and staff of the University of Pittsburgh Chemistry Department. Finally, I would like to thank the members of my committee who have either taught me graduate level chemistry (both Drs. Weber and Saxena, albeit the latter only briefly) or helped me be able to resolve single and double-stranded DNA on polyacrylamide gels (Dr. Opresko). I appreciate their teachings and assistance for all of my time here at Pitt.

Additional thanks to my parents, Bill and Linda Graham and sister Aubrey, for their unconditional support and understanding when I could not attend family functions due to protein purifications or needing that fourth set of fluorescence anisotropy data so that I could have sexier error bars over the years. I would also like to thank my friends who have helped me keep my sanity and have a little fun over these past few years.

My most important thanks, however, goes to the woman who has transitioned from girlfriend to fiancé to wife during the course of my graduate school tenure. Elicia Schlosser Graham is the best friend and person that I could ask for and has helped me innumerable ways through graduate school. Although she typically does not provide scientific insight, she has helped and supported me throughout the years in far too many ways enumerate. She is the best.

1 INTRODUCTION

1.1 DNA REPLICATION

In 1953 Watson and Crick published the structure of DNA and they proposed that each strand could act as a template to copy genetic material [1]. It has been revealed that DNA replication is a vital, yet complex, process involving an organism's ability to faithfully replicate and maintain its genomic code. Initial stages of DNA replication are marked by the unwinding of double-stranded DNA (dsDNA) into single-stranded DNA (ssDNA): this action is performed by the DNA helicase. DNA helicases are present in each of the three domains of life and are absolutely essential for cellular survival. In humans, mutations of certain helicases can lead to a predisposition for susceptibility to cancer and/or other diseases [2, 3]. Aberrant DNA replication, common in cancer, is characterized by uncontrolled and upregulated minichromosome maintenance (MCM) activity [4-6]. Though research is ongoing to understand the mechanistic phenotypes of helicases involved in normal DNA replication and how their activities change in the presence of disease; the mechanism of unwinding by the replicative DNA helicase and its role in the DNA primosome is not well understood, and demands further research [7].

The biochemical characterization of nucleic acid processing in Archaea is convenient due to their streamlined replication machinery compared to that of homologous, yet more complex, Eukaryotes [8-11]. We are using crenarchaeal *Sulfolobus solfataricus* (*Sso*) as a model system for understanding DNA replication in the more complex Eukaryotic system [12-14]. In Figure 1.1, we show one half of the DNA replication fork for *Sulfolobus solfataricus* representing elongation. This figure shows MCM with a DNA primase physically bound to it unwinding DNA into two daughter strands. The primase *de novo* synthesizes short RNA primers which are subsequently elongated by a DNA polymerase. Replication occurs continuously on the leading strand and discontinuously on the lagging strand. The DNA polymerase is bound to its heterotrimeric

processivity clamp which was loaded onto the DNA by a heteropentameric clamp loader. Free single-stranded DNA is bound by the single-stranded binding protein (SSB) to prevent degradation. The four stages of DNA replication and the corresponding proteins involved are shown in Table 1.1: preinitiation, initiation, elongation, and maturation.

Table 1.1: DNA replication proteins and features in the domains Bacteria, Eukaryota, and the two major phyla of the Archaea domain.

DNA replication stage	Process or factors involved	Bacteria	Eukaryota	Archaea	
				Crenarchaeota	Euryarchaeota
Preinitiation	Origin(s) of replication	Single	Multiple	Multiple	Single[a]
	Origin recognition	DnaA	ORC (ORC 1-6)	Orc1/Cdc6	Orc1/Cdc6[b]
Initiation	DNA unwinding (helicase)	DnaB	MCM complex (MCM 2-7)	MCM complex	MCM complex
	DNA unwinding (accessory proteins)	DnaC DnaT	Cdc6	Orc1/Cdc6	Orc1/Cdc6[b]
			Cdt1		
			GINS complex (Sld5, Psf1-3)	GINS23/GINS51	GINS51[c]
			Cdc45	RecJ homolog?	RecJ homolog?
	Primer synthesis	DnaG	Pol α/primase complex	DNA primase (PriSL/DnaG[d])	DNA primase (PriSL)
	Single-stranded DNA protection	SSB	RPA	SSB	RPA
Elongation	DNA synthesis (polymerase)	Pol III (C-family DNA polymerase)	Polδ and Polε (B-family DNA polymerase)	B-family DNA polymerase	D Family DNA polymerase[e]
	DNA synthesis (processivity factors)	γ-complex (clamp loader)	RFC (clamp loader)	RFC (clamp loader)	RFC (clamp loader)
		β-clamp (clamp)	PCNA (clamp)	PCNA (clamp)	PCNA (clamp)
Maturation	Maturation (Okazaki fragment processing)	Pol I (A-family DNA polymerase)	Fen1/Dna2	Fen1	Fen1
		RNase H	RNase H	RNase H	RNase H
		DNA ligase	DNA ligase	DNA ligase	DNA ligase

Modified from [7].
[a] Except for the order *Halobacteriales* which has multiple.
[b] Unknown for members of the orders *Methanococcales* and *Methanopyrales.*
[c] GINS23 has only been found only in the order *Thermococcales.*
[d] DnaG has been shown to have *in vitro* priming activity in *Sulfolobus solfataricus*.
[e] B-family DNA polymerase is also essential in *Halobacterium* and may play a role in replication, however, currently it's exact function has not been determined.

In this introduction, we will focus on three of the proteins represented in the replication fork (Figure 1.1). Specifically, helicases, primases, and single-stranded binding proteins. In addition to functioning in the elongation process, these proteins participate in the initiation process. Collectively these proteins are members of the primosome. The primosome is responsible for opening DNA at an origin, unwinding the DNA into two separate strands and creating short RNA primers for use by the DNA polymerases during elongation. Figure 1.2 shows the number of homologs for key genes involved in archaeal DNA replication. Finally we will show our current view of the archaeal replisome.

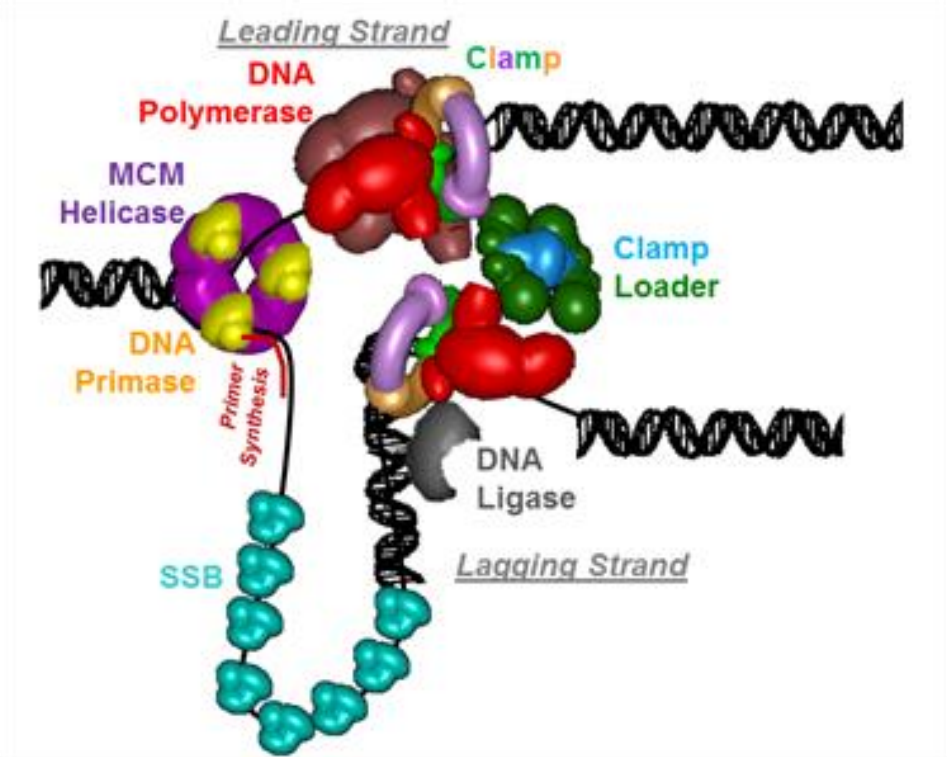
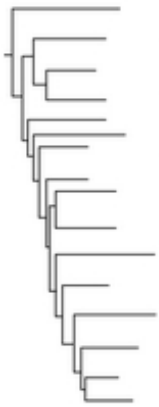


Figure 1.1: Archaeal DNA replication fork in *Sulfolobus solfataricus*
The MCM helicase (dark purple) unwinds DNA into complementary strands. A DNA primase (yellow) physically interacts with MCM and creates RNA primers for elongation by a DNA polymerase (red) in complex with the clamp (green, orange and light purple). The clamp had been loaded on by the clamp loader (green and blue). Continuous replication occurs on the leading strand, while discontinuous replication occurs on the lagging strand. Okazaki fragments created on the lagging strand are stitched together by a DNA ligase (grey). Free single-stranded DNA is bound and protected from degradation by the single-stranded binding (SSB) protein (cyan).



Order	Phylum	Number of genomes	Number of homologs for replication genes: ORC/CDC6	GINS15	GINS23	RPA	MCM	RFCS	PCNA	PolB	DP1	DP2
Nitrosopumilales	*Thaumarchaeota*	2	1	0 or 1	0 or 1	1	1	1	1	1	1	1
Thermoproteales	*Crenarchaeota*	14	1*	0	0	0*	1	1 to 3	1 to 3	2*	0	0
Desulfurococcales		9	2	0	1*	0 or 1	1	1*	3*	2*	0	0
Sulfolobales		15	3	1	1	1	1	1	2*	3	0	0
Not named	*Korarchaeota*	1	2	0	0	0	1	1	1	2	1	1
Nanoarchaeota		1	1	0	0	1	1	1	1	1	1	1
Thermococcales		11	1	1	1*	3	1*	1	1*	1	1	1
Methanopyrales		1	0	0	0	0	1	1	1	1	1	1
Methanobacteriales		7	1 or 2	0 or 1	0	1*	1	1	1	2	1	1
Methanococcales	*Euryarchaeota*	15	0	0	0	2 to 4	2 to 8	1	1*	1	1	1
Thermoplasmatales		4	1*	0*	0	1	1	1	1	1	1	1
Archaeoglobales		4	2 or 3	0	0	0*	1 or 2	1	1	1*	1	1
Halobacteriales		15	3 to 15	0	0	1	1 or 2	2*	1	1*	1	1
Methanomicrobiales		6	2*	0	0	1	1	2	1	1 or 2	1	1
Methanocellales		1	2	0	0	1	1	1	1	3	1	1
Methanosarcinales		9	2*	0	0	3	1*	1*	1	1*	1*	1

Figure 1.2: Phylogenetic chart for archaea with homologs for genes involved in DNA replication
This is reprinted from [15] and was distributed under the Creative Commons Attribution License which permits reproduction in any medium. Homologs for some key genes involved in archaeal DNA replication. Archaeal orders are phylogenetically organized following a rooted maximum likelihood tree of Archaea based on 53 concatenated ribosomal proteins [16]. The homology search was performed by RAST v4.0 (Rapid Annotation using Subsystem Technology), and the annotated data was viewed through the SEED viewer (*http://www.theseed.org/*). A total of 115 archaeal genome sequences were obtained from NCBI and uploaded into the RAST server. RAST annotation was performed using default parameters with the genetic code for Bacteria and Archaea. The replication proteins homologs were checked through the SEED subsystem DNA replication. Homology results of *Thermococcus kodakaraensis* KOD1, *Pyrococcus furiosus* DSM 3638, and *Thermofilum pendens* Hrk 5 were searched for replication protein homologs using BLAST (blastx v2.2.28+). Results with homology coverage of >80% and E-values less than 0.001 were considered as real homologs. The results were also supplemented with data from literature reviews and BLAST searches. When indicated in the figure, zero (0) homologs means that no homolog was found for a specific gene by using the described methodology. Asterisks indicate that one or two of the analyzed microorganisms possess an exception for that specific feature. Exceptions noted are (feature/order) ORC1/CDC6/*Thermoproteales*, *Thermofilum pendens* Hrk5 (3 homologs), and *Thermosphaera aggregans* DSM11486 (2 homologs); ORC1/CDC6/*Thermoplasmatales* and *Picrophilus torridus* DSM 9790 (1 homolog); ORC1/CDC6/*Methanomicrobiales* and *Methanoplanus petrolearius* DSM11571 (4 homologs); ORC1/CDC6/*Methanosarcinales* and DSM4017 (4 homologs); GINS15/*Thermoplasmatales* and *Thermoplasma acidophilum* DSM1728 (1 homolog); GINS23/*Desulfurococcales*, *Staphylothermus hellenicus* DSM12710 (no homolog), and *Staphylothermus marinus* F1 (no homolog); GINS23/*Thermococcales*, *Pyrococcus horikoshii* OT3 (2 homologs), RPA/*Thermoproteales*, *Thermofilum pendens* Hrk5 (3 homologs), and *Thermosphaera aggregans* DSM11486 (1 homolog); RPA/*Methanobacteriales* and *Methanothermus fervidus* DSM2088 (no homolog); RPA/*Archaeoglobales* and *Archaeoglobus fulgidus* DSM4304 (1 homolog); MCM/*Thermococcales* and *Thermococcus kodakaraensis* KOD1 (3 homologs); MCM/*Methanosarcinales* and *Methanosarcina acetivorans* C2A (2 homologs); RFCS/*Desulfurococcales* and *Hyperthermus butylicus* DSM 5456 (2 homologs); RFCS/*Halobacteriales* and *Haloquadratum walsbyi* DSM16790 (1 homolog); RFCS/*Methanosarcinales*, CG6 (2 homologs), and *Methanosaeta* thermophile PT (2 homologs); PCNA/*Desulfurococcales* and *Ignisphaera aggregans* DSM17230 (1 homolog); PCNA/*Sulfolobales* and *Metallosphaera cuprina* Ar-4 (1 homolog); PCNA/*Thermococcales*, *Pyrococcus horikoshii* OT3 (2 homologs), and *Thermococcus kodakaraensis* KOD1 (2 homologs); PCNA/*Methanococcales*, *Methanococcus maripaludis* S2 (2 homologs), and Kol5 (2 homologs). PolB/*Thermoproteales*, *Caldivirga maquilingensis* IC-167 (3 homologs), and *Pyrobaculum calidifontis* JCM 11548 (3 homologs); PolB/*Desulfurococcales*, *Staphylothermus hellenicus* DSM12710 (3 homologs), and *Staphylothermus marinus* F1 (3 homologs); PolB/*Archaeoglobales*, *Archaeoglobus fulgidus* DSM4304 (2 homologs); PolB/*Halobacteriales*, *Halorhabdus utahensis* DSM12940 (2 homologs); PolB/*Methanosarcinales*, *Methanosaeta concilii* GP6 (2 homologs); DP1/*Methanosarcinales*, *Methanococcoides burtonii* DSM6242 (2 homologs).

1.2 SINGLE-STRANDED BINDING PROTEINS

Proteins that bind single-stranded DNA are present in all three domains of life and in viruses. These ubiquitous proteins are involved with a variety of DNA related processes including DNA replication, DNA repair and recombination [17, 18]. They were originally discovered (T4 gene product 32) in *Escherichia coli* (*E. coli* or *Ec*) cells that had been infected with T4 bacteriophage [19]. There are generally two classes of single-stranded binding proteins: Bacteria utilize single-stranded binding (SSB) protein and Eukarya generally utilize replication protein A (RPA). Archaea, generally either have an SSB and/or an RPA homolog. However, recently, human SSB1 and 2 were discovered based on homology with archaeal SSB's and appear to play an important role in the DNA damage response and genomic stability [20, 21].

The common motif shared between SSB and RPA is the oligonucleotide/oligosaccharide binding fold (OB-fold) [22]. Bacterial SSB typically have one OB-fold per monomer, however *Deinococcus radiodurans* and *Thermus aquaticus* have two OB-folds. *E. coli* SSB has been characterized most thoroughly and forms a homotetramer in solution [23] while *Deinococcus radiodurans* and *Thermus aquaticus* form homodimers [24]. A crystal structure for *Ec*SSB tetramer is shown in Figure 1.3A [25]. Eukaryotic RPA is a heterotrimer composed of three subunits named after their corresponding molecular weights in kDa 70, 32, and 14 [21]. The crystal structure of two of the OB-folds of Human RPA70 is shown in Figure 1.3B [26]. Human RPA70 has four OB-folds, and Human RPA34 and Human RPA14 both have one [21]. However one of RPA70's OB-folds and the OB-fold for RPA14 are non-binding folds [21]. The active form for the previously mentioned bacterial and eukaryotic species harbors four total OB-folds. Archaea, on the other hand, have been shown to harbor both SSB and RPA depending on the phylum. *Sulfolobus solfataricus*, belonging to the crenarchaea phylum of Archaea, has SSB [27]. *Sso*SSB's crystal structure (Figure 1.3C) shows its OB-fold which is more similar structurally to RPA70, but

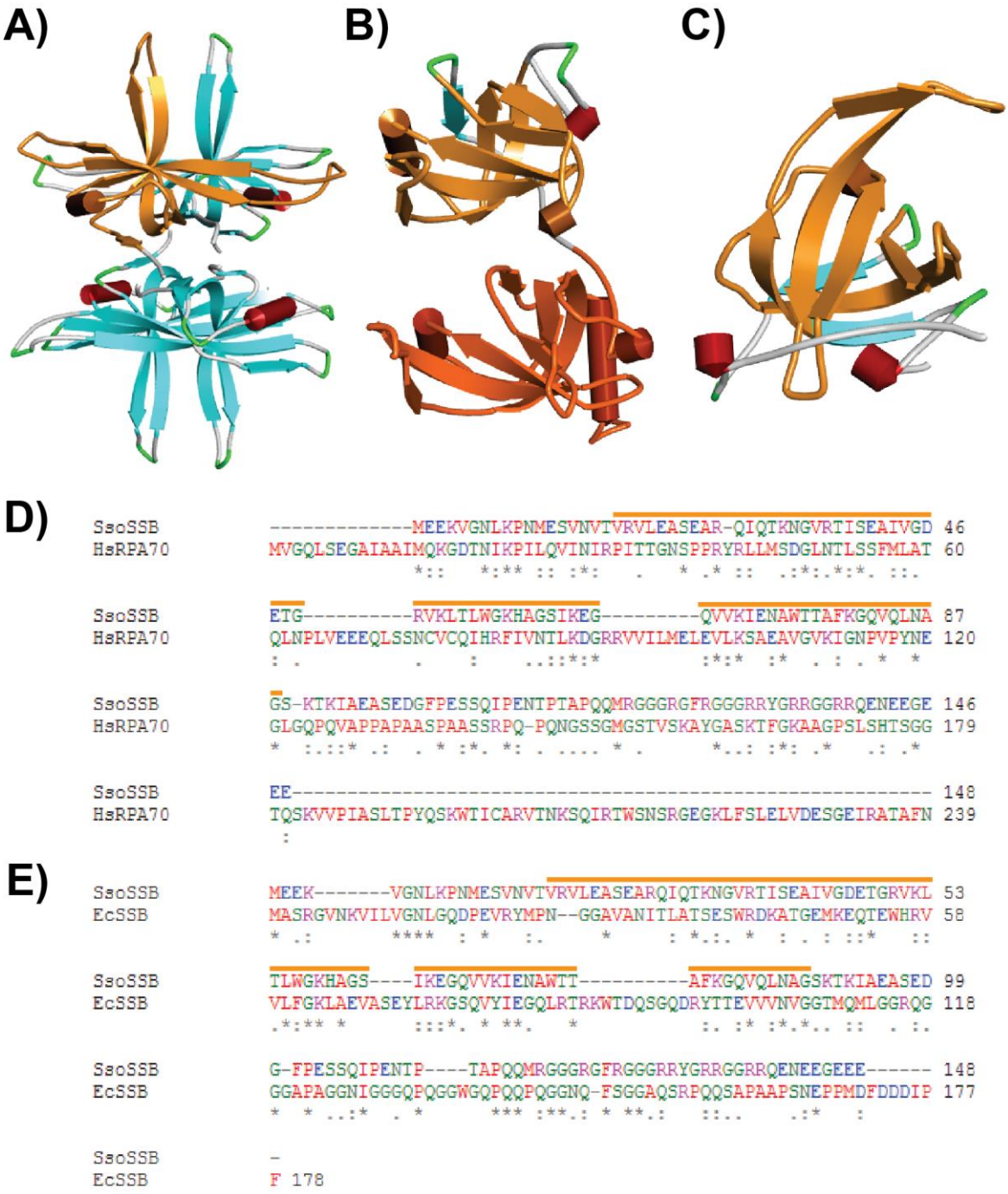


Figure 1.3: Single-stranded binding protein crystal structures and ClustalW2 alignments
A) The crystal structure of the *Ec*SSB tetramer residues 1-114 (PDB ID: 4MZ9), one OB-fold is colored orange. **B)** The crystal structure for *Hs*RPA70 residues 183-420 (PDB ID: 1JMC), the orange OB-fold is residues 181-290 and the dark orange OB-fold is residues is 300-420. **C)** The crystal structure (PDB ID: 1O7I) for *Sso*SSB residues 1-119, the OB-fold (residues 19-88) is colored orange. **D)** ClustalW2 alignment for *Sso*SSB and *Hs*RPA70. E) ClustalW2 alignment for *Sso*SSB and *Ec*SSB. The location of the *Sso*SSB OB-fold is denoted with an orange bar over the corresponding residues.

*Sso*SSB has more sequence homology to *Ec*SSB shown by ClustalW2 alignment in Figure 1.3D and E [15]. *Sso*SSB publications have shown a variety of oligomeric states by different groups including monomer [27-31], dimer [30], and tetramer [30, 32]. Three groups have shown that *Sso*SSB exists as a monomer in solution [27, 28, 31]. Upon binding DNA, however it likely forms a tetramer [27]. Overall, crenarchaea generally have SSB, while euryarchaea generally have RPA [33]. There are exceptions however, in the crenarchaea phylum thermoproteales have neither SSB nor RPA, but instead have ThermoDBP [34]. Genetically distinct phyla Thaumarchaea and Korarchaea harbor both SSB and RPA [33].

Both bacterial and crenarchaeal SSB's typically have acidic C-terminal tails. However, both Thaumarchaea and Korarchaea lack this acidic tail, for these phyla, the RPA homolog may bind proteins for DNA metabolism-related functionality [33]. Human RPA additionally interacts with at least 14 proteins including MCM3-7, but not MCM2 [35]. In bacteria, the acidic C-terminal tail has been implicated in binding at least 14 other proteins [24] including itself [36, 37], RecQ [38], and DnaG [39]. *Sso*SSB has been shown to be able to detect DNA damage [40], interact with reverse gyrase [41], *Sso*RadA during homologous recombination [42], and *Sso*MCM [9, 43]. Eukaryotic RPA's two non-binding OB-fold have been implicated in protein-protein interactions [17]. Similarly, other analogous bacterial systems have shown interactions between SSB and DnaB (another homohexameric DNA helicase) including *Escherichia coli* [44], *Mycobacterium tuberculosis* [45], and *Helicobacter pylori* [46]. In *E. coli*, SSB stimulates DnaB unwinding but also inhibits ssDNA-dependent ATPase activity [44].

The DNA binding mechanisms have been most thoroughly characterized in *Ec*SSB. It binds ssDNA through two mechanisms including $(SSB)_{35}$ where two of the four subunits contact the 35 nt of DNA and $(SSB)_{65}$ where all four subunits interact with 65 nt of DNA by wrapping around the exterior of the tetramer [23, 47-49]. Recent work has indicated that the binding mode is dependent on salt concentration [50]. Human RPA binding is modulated through RPA70 primarily, though for proper function in DNA replication the heterotrimeric complex must be present [17]. Human RPA has three binding site sizes of 8, 30, and 90 nt [51]. *Sso*SSB's DNA binding is modulated primarily through base-stacking interactions including residues W56, W75, and F79, similar to human RPA which also utilizes homologous residues to W56 and F79 [26, 52]. The binding site size for *Sso*SSB monomer is 4-6 nt per monomer [27, 29-31].

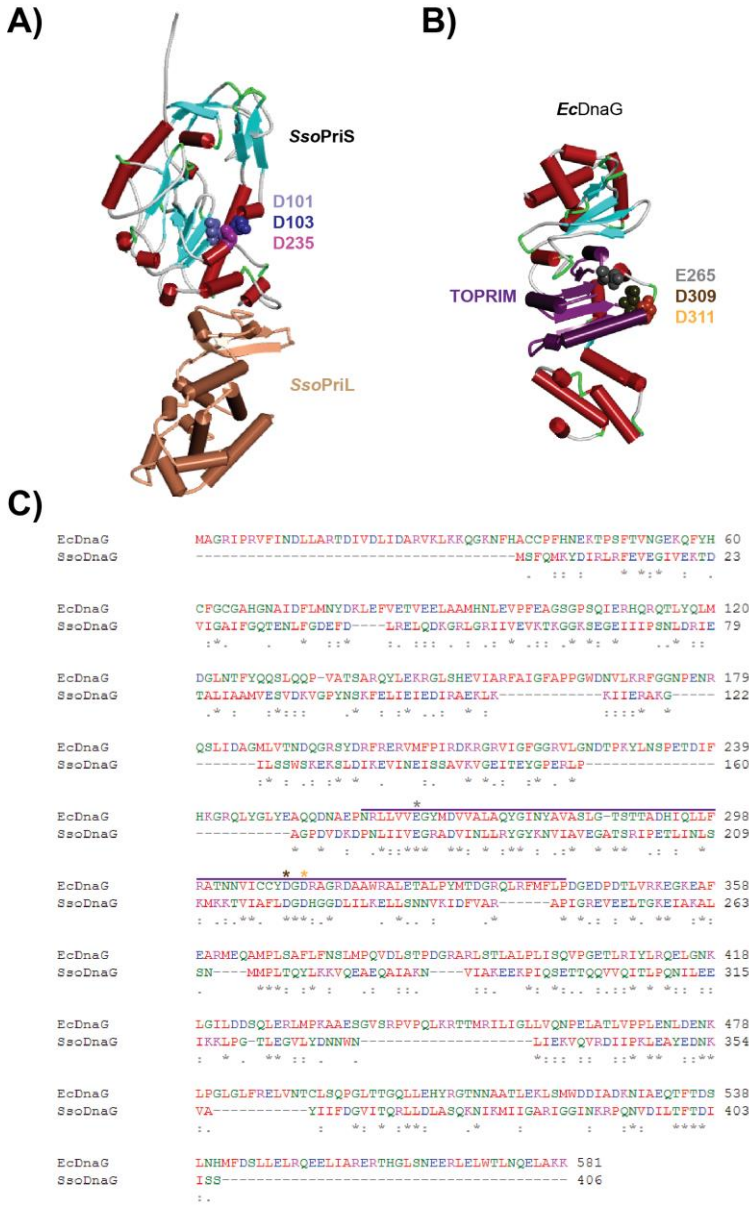
1.3 PRIMASES

Primases are also present in all three domains of life. DNA polymerases are unable to synthesize DNA *de novo*, and so they typically extend short RNA primers produced by the primase. The leading strand requires only one primer due to continuous synthesis, but the lagging strand requires primers for each Okazaki fragment produced which are ligated together forming a single continuous strand [53]. There are two families of primases: the DnaG family present in Bacteria and corresponding phages and the Pri family present in Eukaryotes [54]. Archaea, however, harbor both DnaG and Pri family members, Table 1.1 [15].

1.3.1 Archaeoeukaryotic PriSL-type

Eukaryotic systems utilize PriSL-type primases. The primase is a heterodimer consisting of PriS (small catalytic subunit – p48) and PriL (large regulatory subunit – p58), that interact with both DNA polymerase α (p180) and its accessory subunit B (p68) [54]. This forms the active primase in eukaryotes known as polymerase α/primase complex (pol-prim) [54]. Pol-prim complexes have been identified in humans, rats, mice, *Drosophila* and *Saccharomyces cerevisiae* among other species [54]. Crystal structures available are only from archaeal species including: PriSL from *Sulfolobus solfataricus* (lacking the C-terminal domain of PriL) [55] and PriS from *Pyrococcus furiosus* [56] and the primase from *Pyrococcus horikoshii* with UTP bound in its active site [57]. In Figure 1.4, we show the *Sso*PriSL structure. The conserved catalytic triad is shown (D101, D103, and D235) [58]. Even though PriL lacks catalytic activity, it is critical for the priming reaction, due in part to its C-terminal domain that aids in positioning the di-nucleotide to initiate the priming reaction [59]. The eukaryotic primer is actually a hybrid RNA-DNA molecule, which is initially created by PriSL with minimal template specificity [60] and is then handed off polymerase α to add deoxynucleotides forming a primer that is approximately 20-25 nucleotides [61]. Primer handoff then occurs to either DNA polymerase δ on the lagging strand, or DNA polymerase ε on the leading strand [62].

Archaeal PriSL can function without other polymerases or accessory proteins and have been characterized in number of species including *Pyrococcus* [63, 64], *Thermococcus kodakaraensis* [65], and *Sulfolobus solfataricus* (*Sso*) [55, 58, 66]. In addition to RNA primer



synthesis ranging from 2 to 1000 nt, these primases have surprisingly novel and unregulated DNA synthesis abilities producing oligonucleotide products greater than 7 kilobases [58]. Temperature and slight differences in the affinity of NTP or dNTP may direct function towards RNA or DNA synthesis, respectively. An indirect link between PriSL and the minichromosome maintenance (MCM) helicase is thought to be mediated by GINS23 within the GINS heterotetramer to coordinate priming and unwinding activities in archaea [67]. The DNA replication homology between archaea and eukaryotes would predict that like other functional homologs, PriSL will fulfill a DNA priming role in archaea as well [68, 69].

Figure 1.4: Crystal structures of *Sso*PriSL and *Ec*DnaG and ClustalW2 alignment for *Sso*DnaG and *Ec*DnaG
A) Crystal structure of *Sso*PriSL (PDB ID: 1ZT2). The catalytic residues in *Sso*PriSL are represented as space-filling CPK molecules where D101 is lavender, D103 is purple and D235 is fuchsia. **B)** Crystal structure of *Ec*DnaG (PDB ID: 1DD9). The TOPRIM domain is colored dark purple. The conserved catalytic residues are E265 (grey), D309 (brown), and D311 (orange). **C)** ClustalW2 alignment for *Ec*DnaG and *Sso*DnaG. The *Ec*DnaG TOPRIM domain and catalytic residues are denoted by a dark purple bar and colored asterisks as in B), respectively.

1.3.2 DnaG-type

Primase activity was initially discovered in *E. coli*'s bacteriophage T7 gene product 4 which was classified as a primase [70]. The bacterial DnaG primase, however has three domains: an N-terminal zinc ribbon binding domain, the catalytic TOpoisomerase PRIMase (TOPRIM) domain, and the C-terminal helicase binding domain [54]. The Zn^{2+} binding domain is a CHC2-type that lacks alpha helices which classifies it as a zinc ribbon motif implicated in DNA binding [71]. In Figure 1.4B, we show the crystal structure of *Ec*DnaG's catalytic core highlighting the TOPRIM domain (residues 259-341) and the conserved metal binding residues E265, D309, and D311 [72]. E265 is located in the consensus sequence EGxxD, while D309 and D311 is located in the consensus sequence DxDxxGxxA [54] TOPRIM is a signature motif in DnaG-type primases, topoisomerase, and other similar metal-binding phosphtransfer proteins [73, 74]. Binding to DNA is generally weak and transient but occurs along an elusive, positively charged region adjacent to the active site that orientates the DNA template [75]. The C-terminal domain of *Ec*DnaG is associated with the N-terminus of the DnaB helicase (*Ec*DnaB) to form the bacterial primosome complex that increases both priming and helicase activities [76-80]. This primase–helicase interaction and orientation have been seen in other bacterial and phage organisms including T7 [81], SV40 [82], *Bacillus subtilis* bacteriophage SPP1 [83], and *Bacillus stearothermophilus* [84] and are required for synthesis of primers on the lagging strand. Primers synthesized by *Ec*DnaG are typically 11 nt, but have been shown to range from 9 to 14 nt [85] and also 10 to 60 nt [86]. The mechanism of RNA synthesis and the protein interactions of *Ec*DnaG have been well characterized [72, 75, 87-95]. Primer length is controlled through coordination of the two subunits of bacterial DnaG with the Zn^{2+} binding domain to regulate DNA template binding [71, 72, 75, 78].

Interestingly, Archaea also contain a homolog to the bacterial-like DnaG primase, in Figure 1.4C we show a ClustalW2 alignment for *Ec*DnaG and *Sso*DnaG [73]. Archaeal *Sso*DnaG has a conserved and essential active-site glutamate (E175) required for synthesis of primarily 13mer RNA products [96]. In support of priming activity, *Sso*DnaG is able to *de novo* synthesize RNA primers with greater efficiency than archaeal *Sso*PriSL [96]. We have identified a physical interaction between *Sso*MCM and *Sso*DnaG localized to MCM's N-terminal domain [97]. At the same time, *Sso*DnaG is found as a structural component within the archaeal exosome, albeit with no initial associated enzymatic activity [98]. A recent report, however, showed in addition to

specifically binding to the Csl4- and Csl4-Rrp4-exomsomes, that it is a poly(A)-binding protein that enhances the degradation of adenine-rich transcripts [99]. The exosome complex is required for RNA degradation which is the opposite function that occurs during priming. However, since the *Sso*DnaG expression index is 7 to 20 times greater than either *Sso*PriS or *Sso*PriL; it would be possible to serve a dual role of both priming during replication and participating in the exosome [96].

1.4 HELICASES

The minichromosome maintenance (MCM) helicase is utilized during DNA replication and elongation in both archaeal and eukaryotic species, but not in the Bacterial domain [100]. MCM was originally discovered in *Saccharomyces cerevisiae* in 1984, while helicases, in general, were first characterized in *Escherichia coli* in 1976 [101, 102]. Eukaryotic MCM is composed of MCM 2, 3, 4, 5, 6, and 7 with one copy of each forming a heterohexamer, while archaeal MCM are composed of six copies of the same monomeric protein forming a homohexamer [103]. Similarly bacterial DnaB is also a homohexamer.

1.4.1 Superfamilies

Helicases have recently been classified into six superfamilies (SF), depending on a number of factors (Figure 1.5) [104]. To be classified as a helicase or translocase, the minimum requirement is the "core domains." These form RecA-like folds that either occur within the same polypeptide chains or between two subunits that both bind and couple NTP hydrolysis into conformational changes [105]. The core domains minimally consist of Walker A and B boxes [106] and an arginine finger [107]. SF1 and SF2 are either monomeric or dimeric and must have two RecA-like folds. All other helicases S3-6 are hexameric or double hexameric with single RecA folds [104]. Additionally helicases that unwind with a 3'-5' polarity are type A and 5'-3' are known as type B. SF1, SF2, and SF6 contain both A and B type helicases, while SF3 are all type A, and

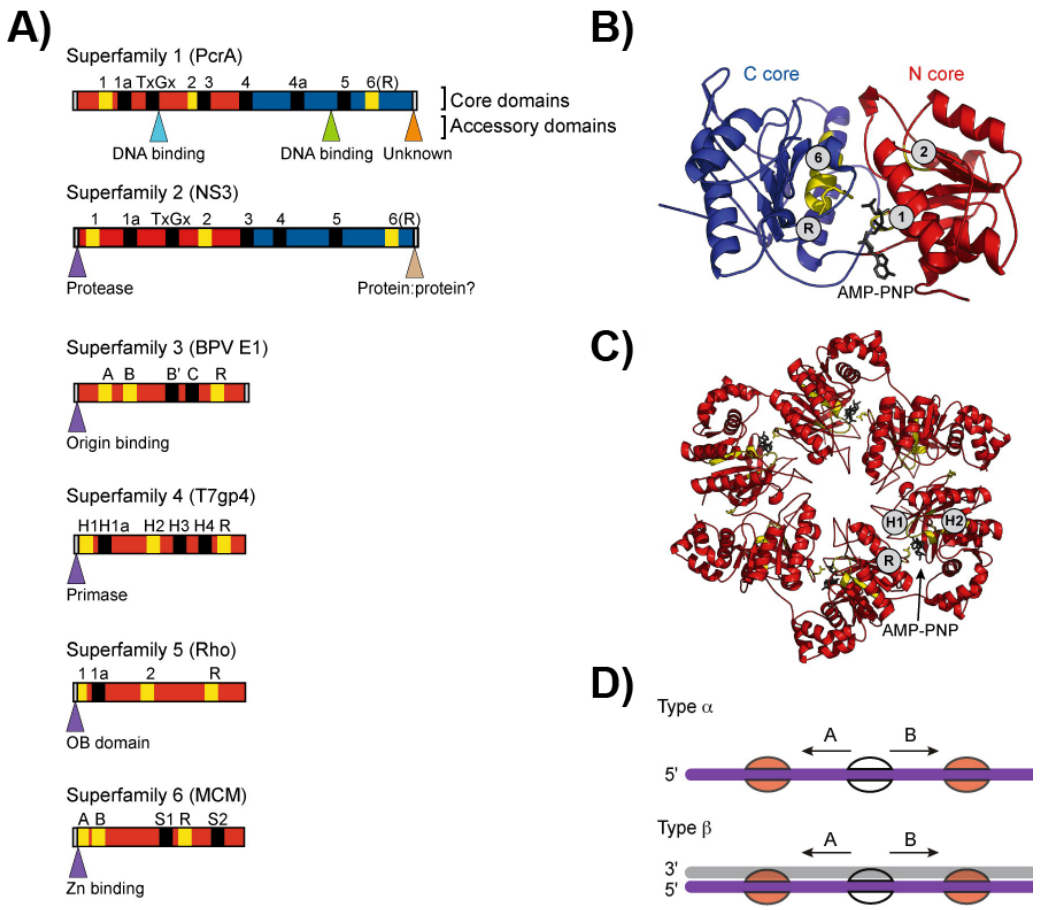


Figure 1.5: Classification of helicase and translocases.
Reprinted from [104]. **A)** This classification is based largely on the work of Gorbalenya & Koonin [108]. The name of one member of each of the six superfamilies, which is used as a structural example in the text, is given in parentheses. The "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** and **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.

SF4 and 5 are all type B. Finally, helicases that bind ssDNA are classified as type α and dsDNA are type β. Currently, all SF1 are type α, but the other SF's contains both type α and β.

SF1 is likely the best characterized SF and include both type A and B helicases. SF1A includes gram negative Rep and UvrD and gram positive PcrA while SF1B includes RecD, Dda, Pif1, and Rrm3. SF2 includes DEAD-box RNA helicases RecQ-like helicases, and Snf2-like helicases. The remaining superfamilies include hexameric helicases with examples shown in Figure 1.6. SF3 contains both Walker A and B boxes labeled A and B and motif C is specific to SF3. These helicases possess both origin recognition and unwinding abilities among other enzymatic activities [109]. An example SF3 helicase is bovine papilloma virus E1. SF4 were initially characterized in both bacteria and corresponding bacterial phages. The helicases contained in this SF associate with a primase either through protein-protein interactions for bacteria or exist as one entity for bacteriophages. Examples include *Ec*DnaB and T7 gene product 4. Five sequence motifs are present H1, H1a, H2, H3, and H4; where H1 and H2 are Walker A and B, respectively. The other motifs are distinctive to SF4. All SF4 have type B polarity. The *Geobacillus stearothermophilus* DnaB hexamer deviates from the normal six-fold symmetry which causes the arginine fingers to be in non-equivalent positions shown in Figure 1.7A. SF5 contains the Rho helicase which is closely related to SF4 helicases, but has been placed in its own SF due to sequential differences and its open washer crystal structure. SF6 contains the ATPases Associated with diverse cellular Activities (AAA^+) helicases. This includes both eukaryotic and archaeal MCM helicases. The crystal structure of homohexameric with six-fold axis symmetry. *Sso*MCM is shown in Figure 1.7B. RuvB is also a SF6 helicase that processes Holliday junctions. The core domains of *Sso*MCM (Walker A and B and arginine finger) are shown in Figure 1.8A and a ClustalW2 alignment with *Homo sapiens* MCM2 and *Xenopus laevis* MCM2 is shown in Figure 1.8B. *Sso*MCM displays significant homology with eukaryotic MCM.

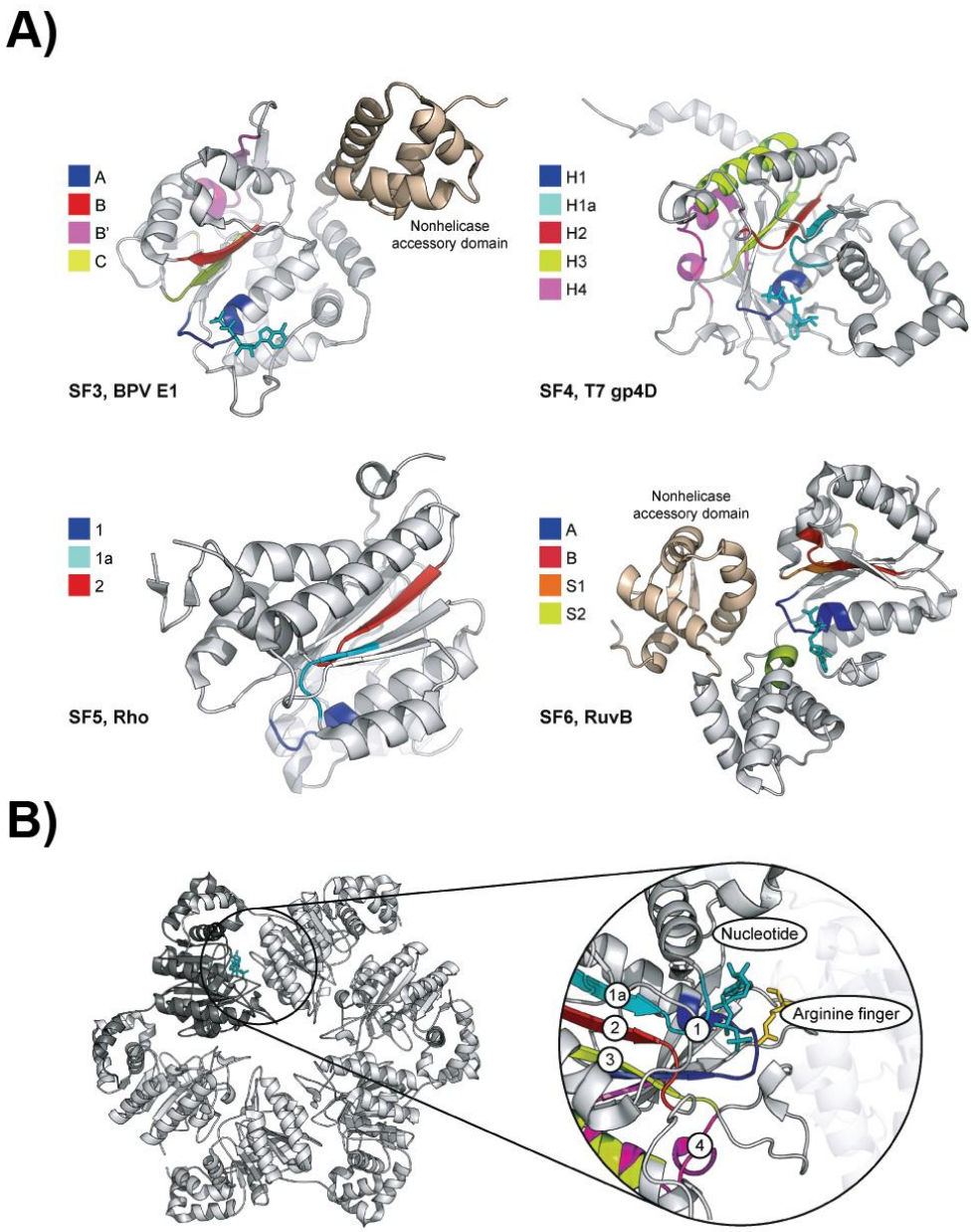


Figure 1.6: Motifs in Superfamilies 3-6.
Reprinted from [104]. **A)** Location of the conserved helicase motifs in each superfamily. A single monomer is shown, with a bound nucleotide where applicable. The Rho monomer has been truncated by removing the OB-fold domain for clarity. **B)** The subunit interface between two monomers in the gp4D structure, showing the relationship between the conserved motifs, nucleotide, and arginine finger from the adjacent subunit.

1.4.2 Archaeal MCM structure and motifs

Sso has a single MCM in its genome that has been shown to have helicase activity [9, 10]. *Sso*MCM was found to be an ATP-dependent homohexameric (77 kDa monomer, 462 kDa hexamer) helicase that preferentially unwinds forked DNA substrates with 3' to 5' polarity [9, 103, 110-112]. *Sso*MCM does not unwind blunt dsDNA or 5'-tail DNA [103]. The basal ATPase activity of *Sso*MCM is quite high, and interestingly, is stimulated 1.5-2 times by the presence of DNA [9, 97, 111]. Mutational studies of *Sso*MCM have found that at least three (six total) inactive ATPase sites can be endured before unwinding activity is diminished, thus it is likely that *Sso*MCM utilizes a semisequential model of ATP hydrolysis [112].

MCM helicases are composed of N- and C-terminal domains, which are connected by the N-C linker which is 40 residues in length that provides stability for the helicase [113]. The N-domain is composed of subdomains A, B, and C and the C-domain is composed of the α/β region, the α/β-α linker region, α domain (the three aforementioned regions also collectively are known as the AAA^+ core) and the winged-helix (WH) domains (Figure 1.9A and B) [10, 103, 113-115]. *Sso*MCM loads onto forked DNA substrates with the C-terminal domain facing the ssDNA-dsDNA junction [110, 111]. Wild-type *Sso*MCM is able to bind ssDNA (K_d 50 nM ± 22) and blunt dsDNA (K_d 85 nM) [114, 116].

Truncated forms of *Sso*MCM have been informative in elucidating the roles of each domain within the context of the hexamer. The C-terminal domain also binds blunt dsDNA (K_d 200 nM), while the N-terminal domain binds blunt dsDNA very weakly [116]. On the other hand, the C-terminal and N-terminal domains individually bind ssDNA with comparable K_d values (~250 nM ± 13 and 292 nM ± 13, respectively) [114]. The AAA^+ core, is dimeric in solution which allows for *trans* (active site from adjacent subunit) interactions that enable helicase activity; this form is able to unwind forked, 3'-tail, 5'-tail, and blunt dsDNA substrates [103]. If the N-terminal domain is added to the AAA^+ core alone, processivity increases, but only forked and 3'-tail DNA substrates are unwound [103]. The AAA^+ core contains residues capable of binding and hydrolyzing ATP, the mechanical energy provided for unwinding occurs within this domain [103, 116]. The N-terminal domain functions in selecting for 3'-tail for binding and stabilizes the hexamer binding to DNA [103, 116].

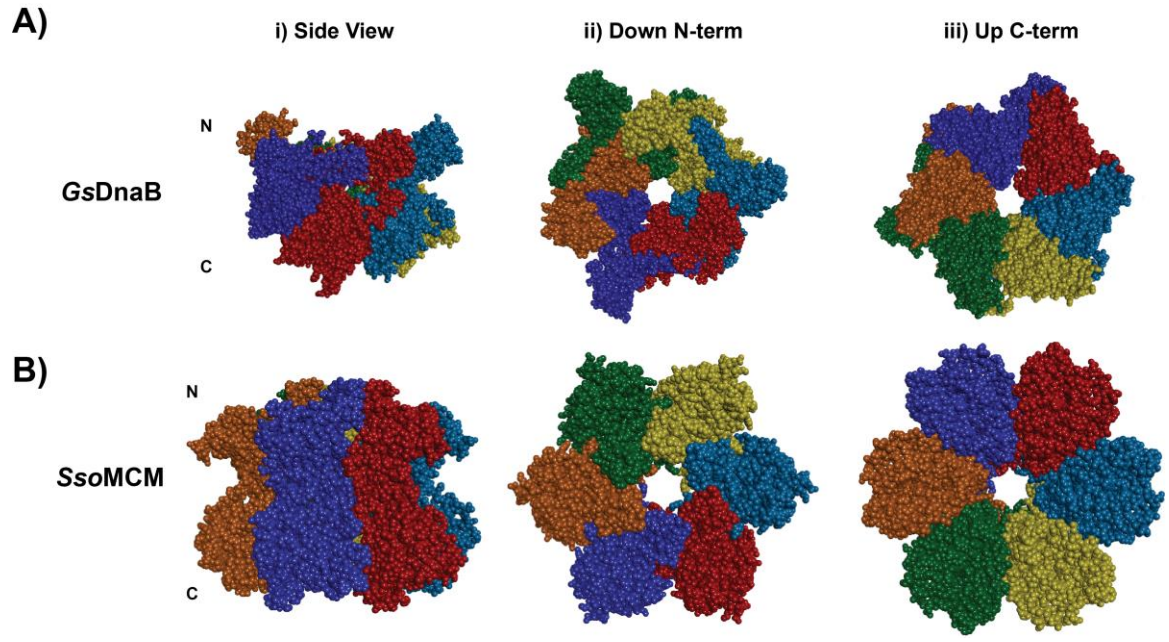


Figure 1.7: Crystal structures for *Gs*DnaB and *Sso*MCM.
A) The crystal structure for *Gs*DnaB (*Geobacillus stearothermophilus*) (PDB ID: 4ESV) is shown in i-iii. Each monomer is colored and the N and C termini are labeled with N and C, respectively. Three views are presented a side view (i), looking down the N-terminus (ii), and looking up the C-terminus (iii) **B)** The crystal structure for *Sso*MCM (PDB ID: 3F9V) is shown. Each monomer is colored and the three views are the same.

The N-terminal domain is composed of subdomains A, B, and C. The A subdomain provides the snowflake-like appearance and its removal minimally alters helicase activity [103, 116]. The B subdomain contains the Zn^{2+} binding motif (residues 144-174, Zn^{2+} is specifically bound by H144, C149, C171 and C174) and has been implicated in DNA binding via β-hairpins and hexamerization [116]. The Zn^{2+} motif is required for helicase unwinding in the euryarchaea *Methanothermobacter thermautotrophicus* (*Mth*) [117]. Contained within the C subdomain is the allosteric communication loop (ACL) (residues 199-211), whose deletion abrogates helicase activity, reduces ATPase activity, but does not affect DNA binding [118]. This loop facilitates communication between the N-terminal domain and AAA^+ core where ATP hydrolysis occurs [118]. The C subdomain also contains the NT hairpin (N-terminal) (residues 239-255), which has been implicated in DNA binding [111, 113]. Previous mutational studies within the NT (K246A and R247A) have shown that wild-type *Sso*MCM binds DNA eight times tighter than the NT mutant version, which results in decreased helicase activity [111].

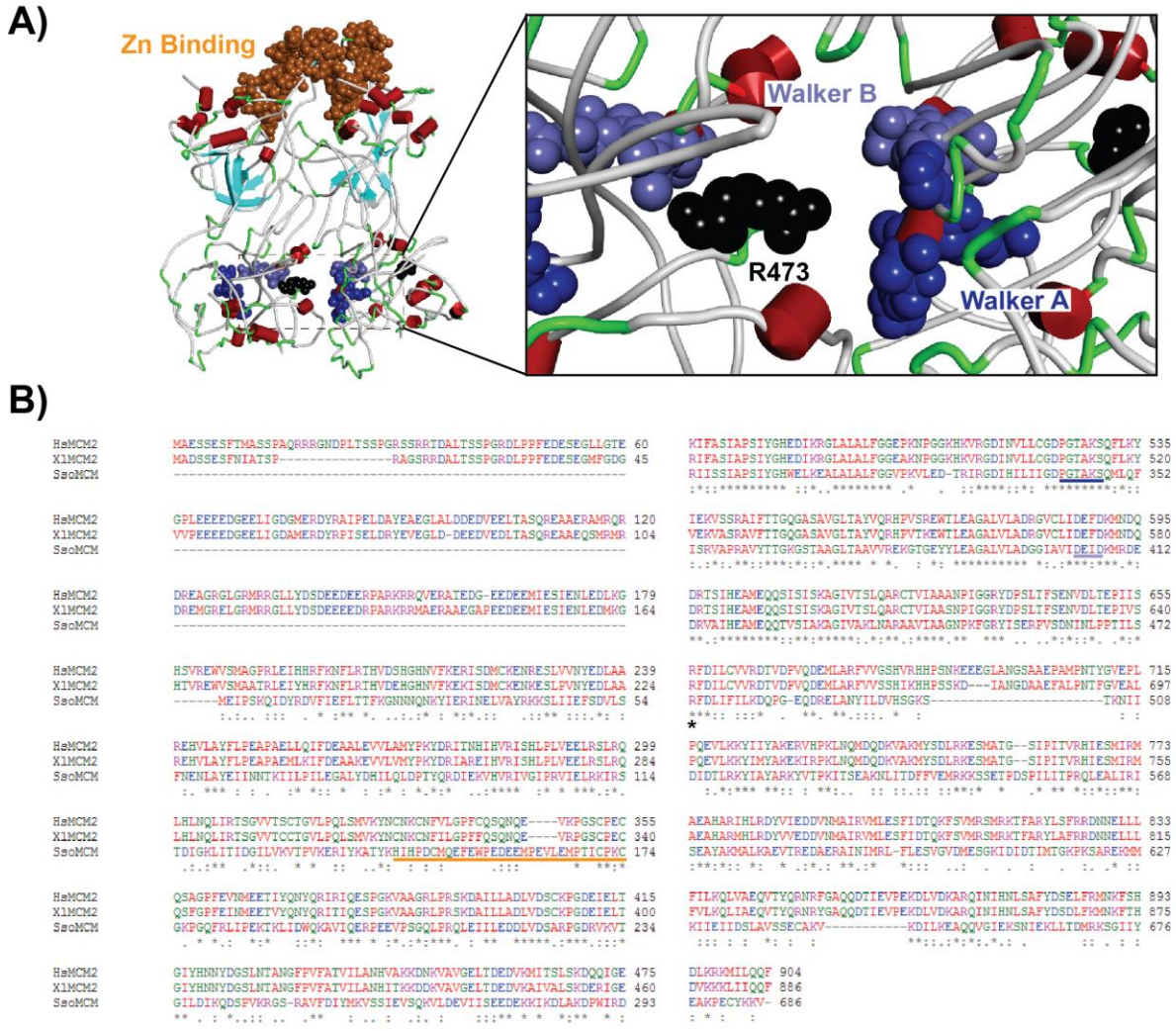


Figure 1.8: *Sso*MCM crystal structure monomer and ClustalW2 alignment.
A) *Sso*MCM near-full length crystal structure (PDB ID: 3E9V) showing a dimer with the Zn binding domain (orange), Walker A (blue), Walker B (lavender), and arginine finger (black) labeled with CPK molecules. The dashed box indicates the blown up at the right of the *trans* ATP hydrolysis site consisting of Walker A and B from the right monomer and the arginine finger from the left monomer. **B)** ClustalW2 alignment for *Homo sapiens* MCM 2 (*Hs*MCM2), *Xenopus laevis* MCM 2 (*Xl*MCM2) and *Sso*MCM. The Zn binding domain, Walker A, Walker B, and arginine finger are denoted by an orange bar, blue bar, lavender bar and black asterisk, respectively.

The C-terminal domain is composed of the AAA^+ core and WH domains. The AAA^+ core is where ATP hydrolysis occurs, which provides energy to unwind DNA [114]. The AAA^+ core contains the Walker A, Walker B, sensor-1, sensor-2, and arginine finger motifs [112, 113]. The Walker A motif PGTAKS (residues 342-347) binds to the β- and γ- phosphates of bound ATP, while the Walker B motif DEID (residues 404-407) binds a Mg^{2+} cation which allows for the hydrolysis of ATP [9, 106, 112, 119]. One study has shown that a Walker A mutant, K346A, is

still able to hydrolyze ATP, but compared to wild-type *Sso*MCM its ability is significantly reduced; this results in substantially decreased unwinding ability [9]. The sensor-1 (residues 441-448) and sensor-2 (residues 558-562) motifs facilitate ATP binding in the active site [112]. Sensor-2 specifically interacts with the γ-phosphate of ATP [112]. The arginine finger (R473) from one subunit protrudes into the active site of a neighboring subunit, which aids in ATP binding and oversees conformational changes related to ATP hydrolysis between the subunits [112, 120].

Additionally contained in the AAA^+ core are three β-hairpins: EXT, located on the exterior of the helicase (residues 319-333); H2I (helix-2 insert), the most N-terminal hairpin that may interact with DNA in the central cavity (residues 374-390); and PS1 (pre-sensor 1) (residues 424-439) [112, 113]. Previous mutational studies of PS1 (K430A) have shown that wild-type *Sso*MCM PS1 binds DNA only 2.5 times tighter than the mutant, but helicase activity is abrogated in the mutant [111]. Thus, the PS1 hairpin is implicated in translocating DNA during the unwinding process [111].

Due to the hexamerization of *Sso*MCM, some residues/motifs in the AAA^+ core function in *cis* (active site in the same subunit), while others function in *trans* (active site from adjacent subunit). Walker A and B, and sensor-1 motifs act in *cis*, primarily due to their close proximity, while the arginine finger, sensor-2 motif, and residues 331 and 423-424 act in *trans* [112]. Residues 331 and 423-424 acts as "levers" which communicate with neighboring active sites containing nucleotides to properly position the β-hairpins for unwinding [112].

The WH domain (also known as the helix-turn-helix domain) is not modeled in the crystal structure, primarily due to a lack of electron density [113]. Studies have shown, the WH domain alone, can neither bind DNA nor unwind DNA [103]. At the same time, when the WH domain is removed, DNA binding is comparable to wild-type, but unwinding is enhanced [103]. The AAA^+ core (in the absence of the WH and N-terminal domains) is able to unwind forked, 3'-tail, 5'-tail, and blunt duplex DNA substrates [103]. This implicates the WH domain in substrate selectivity, potentially through its ability to adopt multiple conformations [103, 113].

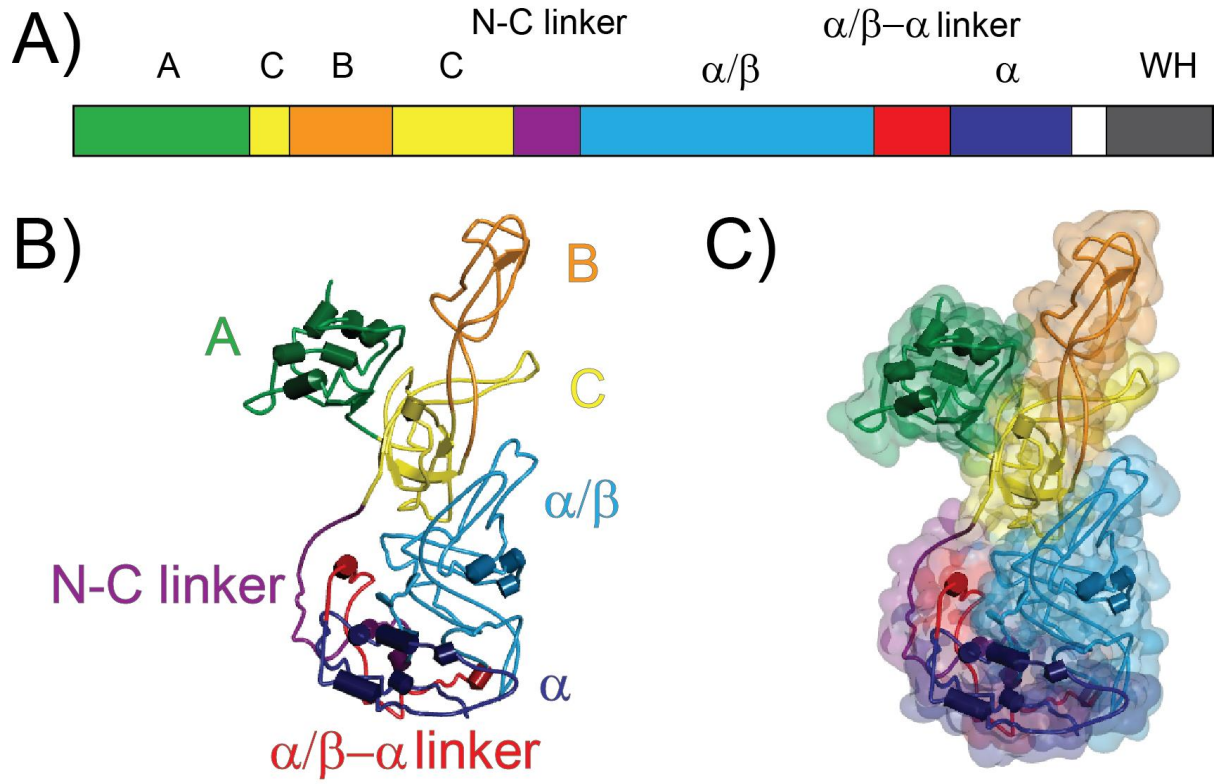


Figure 1.9: *Sso*MCM conserved motifs
A) *Sso*MCM domain organization, where the A domain is green, the B domain is orange, the C domain is yellow, the N-C linker is purple, the α/β domain is light blue, the α/β-α linker domain is red, the α domain is dark blue, the non-crystallized portion of the α domain is white, and the winged helix (WH) domain is grey. **B)** The color scheme and labels from A) are used again for the monomeric near-full length crystal structure of *Sso*MCM with each mutated residue represented as a CPK space filling model. **C)** The same monomer from B) with a CPK surface colored by domain.

1.4.3 Hexameric helicase unwinding models

Crystallographic information is useful for the determination of mechanistic information. Initially, the N-terminus of *Sso*MCM was crystallized and found to be very similar to the crystal structure of N-terminus of *Mth*MCM [116, 121]. More recently, a near full-length crystal structure has been determined for *Sso*MCM which has given more insight into how DNA may be organized within the context of the hexamer [113]. A variety of models have been presented for MCM2-7 unwinding and are shown in Figure 1.10 and include steric exclusion, the rotary-pump, T-antigen, ploughshare models [122]. We have contributed the Archaeal MCM unwinding model steric exclusion and wrapping model [123]. The steric exclusion model would involve MCM2-7 loading onto one

strand and physically excluding the other strand which is similar to *Ec*DnaB's unwinding mechanism [122]. The rotary-pump model would involve MCM immobilized to chromatin that is loaded on dsDNA and unwinding occurs through pumping in opposite directions. The T-antigen model is analogous to SV40's large T-antigen where a double hexamer loads onto dsDNA and unwinds DNA by pulling it into the interface region from which ssDNA is extruded. The ploughshare model involves binding on dsDNA and unwinding occurs through a protein at the posterior of the helicase that physically separates the strands similar to UvrB and RecBCD. The steric exclusion and wrapping model is similar to the steric exclusion model where translocation occurs on ssDNA, but here the non-translocating strand interacts with the exterior of the helicase minimally through electrostatic interactions [123]. In addition to preventing reannealing, the DNA is also protected from degradation prior to interacting with other proteins in the replisome.

1.4.4 CMG complex

In Eukarya, the MCM2-7 helicase has been shown to have helicase activity only recently [124]. The CMG complex consisting of Cdc45, MCM and GINS is thought to the active replicative helicase [122]. GINS is a heterotetramer that consists of four proteins Sld5 and Psf1-3 [15]. A similar complex forms in archaea. However archaea does not have a direct Cdc45 homolog but has a RecJ homolog [67]. Also GINS in *Sso* consists of two copies of GINS15 which is homologous to Sld5 and Psf1 and two copies of GINS23 homologous to Psf2 and 3. Collectively it has been shown that MCM, the GINS tetramer, RecJ and PriSL form a complex; potentially illustrative of how a primase is associates with the helicase in Archaea [67]. This complex will be discussed in greater depth in the next section.

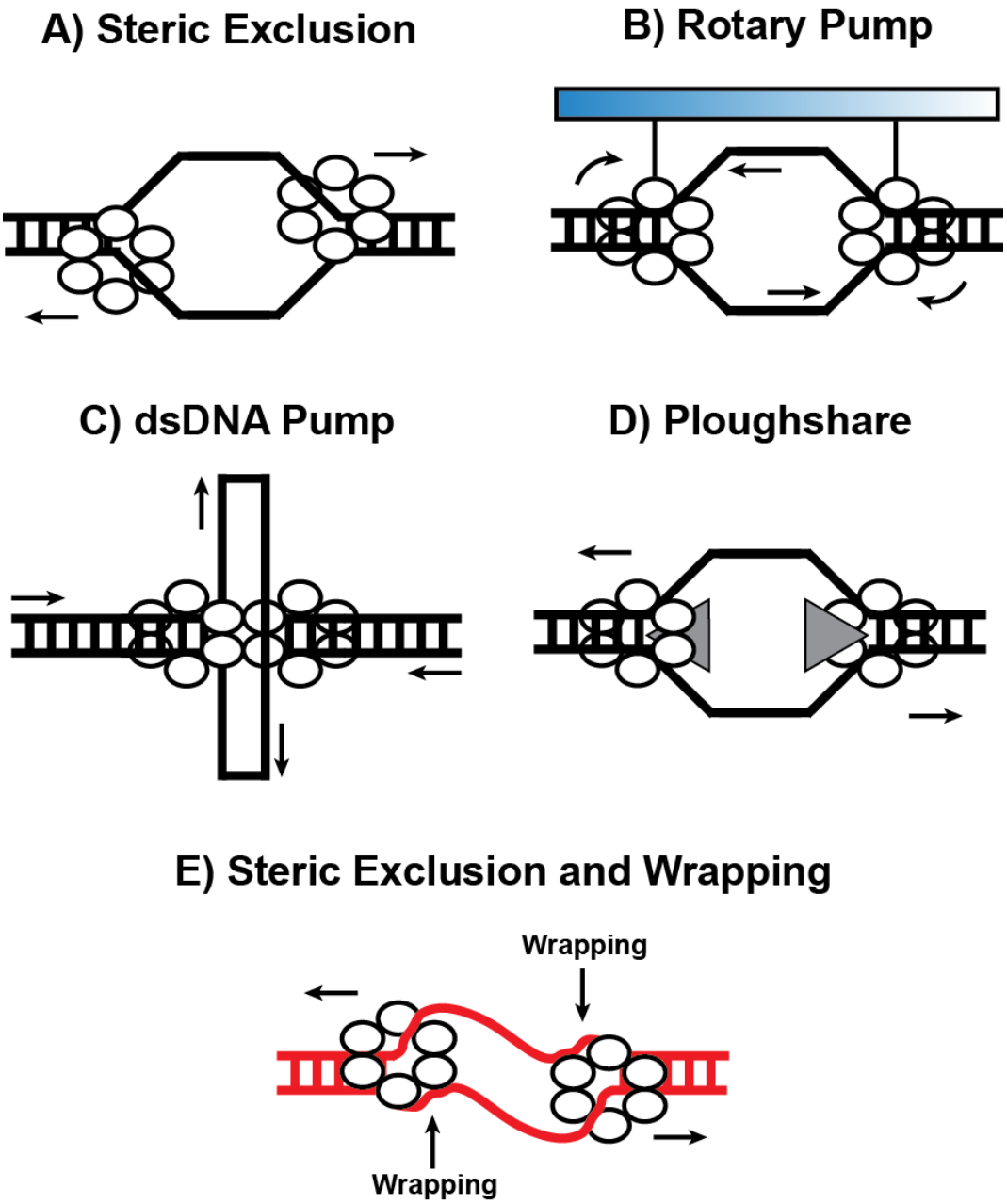


Figure 1.10: Hexameric helicase unwinding models.
This figure was adapted from [122]. **A)** Steric exclusion model. The helicase loads and translocates on the leading strand and unwinds by sterically excludes the lagging strand. **B)** Rotary pump model. MCM's loads onto dsDNA and rotate in opposing directions causing torsional strain which unwinds the DNA. The blue box above indicates that the helicases are attached to the nuclear matrix. **C)** dsDNA pump. The T-antigen model, where two physically connected helicases loaded on dsDNA pump DNA towards the interface connected the helicases. The unwound ssDNA is extruded laterally from positively charge channels. **D)** Ploughshare. MCM loads onto dsDNA and drags a separate ploughshare-like protein represented by the grey triangle that physically separates the DNA into two strands. **E)** Steric exclusion and wrapping. This model is similar to A) in that the helicase loads onto the leading strand and sterically exclude the lagging strand, however, the lagging strand physically interacts with the exterior of the helicase during unwinding. Arrows indicate the direction of travel by the helicase and/or the DNA strands.

1.5 ARCHAEAL PRIMOSOME

The best characterized primosome is that of *E. coli* which consists of six proteins: PriA, PriB, PriC, DnaB, DnaC, and DnaT [125]. Primosome formation begins with PriA and PriB assembling on a primomse assembly site (pas) that is coated with SSB. PriA and PriB form a complex with the ssDNA and then the other proteins assemble. ATP or GTP is required for assembly. Once assembled, DnaB is able to unwind the DNA, and DnaG is able to prime the recently unwound DNA.

In vivo, Eukaryotic MCM2-7 needs other co-factors and proteins to correctly load onto an origin. In eukaryotes, some of these proteins include: Orc1, Cdc6, and Cdt1 which together with MCM comprise the prereplicative complex [126]. Orc1 (origin recognition complex 1) recognizes and binds to an origin on DNA, at this point, two regulatory proteins Cdc6 (Cell division cycle 6) and Cdt1 (Chromatin licensing and DNA replication factor 1) are recruited in by Orc1 [126]. Cdc6 and Cdt1 are thought to actively load the MCM hexamer onto the DNA [126]. Finally, the MCM protein is able to bind and start unwinding the DNA so that replication can occur.

Sso and archaea in general do not have a Cdt1 homolog and instead have three Cdc6/Orc homologs, known as Cdc6-1, Cdc6-2, and Cdc6-3 [10, 127-131]. *Sso* has also been shown to have three origins of replication known as oriC1, oriC2, and oriC3 [131, 132]. The *Sso*Cdc6-1-3 proteins are similar in molecular weight (45-46 kDa). Cdc6-3 is notably different due to its inability to autophosphorylate or bind to oriC1 [127-130]. Their presence in the cell cycle is also notably different where Cdc6-1 and Cdc6-3 are present in G_1 and S, while Cdc6-2 is present in G_2 [10, 130]. This indicates that Cdc6-1 and Cdc6-3 are possible promoters of initiation, while Cdc6-2 is either a repressor or late S-phase activator of replication [10, 130]. Cdc6-1, 2, and 3 each have been shown to inhibit *Sso*MCM unwinding activity *in vitro* [127-130]. Thus after loading, these proteins are likely to dissociate prior to unwinding.

In Figure 1.11, we show our current view of *Sso* pre-initiation and initiation, culminating in the formation of the primosome. In *Sso*, we know that a Cdc6/Orc protein binds at the origin DNA (known as oriC) and recruits MCM. After MCM loads either through a dimer of trimers or a trimer of dimers and begins unwinding by the steric exclusion and wrapping model, the other members of the "archaeal" CMG complex assemble. Likely archaeal GINS binds first followed by RecJ, though these proteins may form a complex prior to binding MCM. PriSL then binds to

GINS23 and makes primers for the leading strand. We also have showed that *Sso*DnaG interacts with MCM [97]. Here, we propose that DnaG makes primers for the lagging strand in agreement with the dual primase system [96]. However, it is possible that DnaG primes on the leading strand and PriSL primes on the lagging strand. *Sso*DnaG, has a low solubility and is prone to aggregation [96]. If co-purified, however, with MCM and other primosome components of *Sso*, its solubility may increase or have a reduced aggregated fraction.

1.6 CONCLUSIONS

Although we have a significant amount of structural and biochemical data, the unwinding mechanism for the helicase and the components involved in the archaeal primosome remain elusive. Importantly, we have proposed and shown evidence for the steric exclusion and wrapping model. Residues affecting both exterior wrapping and the overall hexamer have been characterized allowing for the assignment of roles for conserved motifs among other Superfamily 6 helicases. Additionally, we have shown that a bacterial-like primase physically interacts with the eukaryotic-like helicase. Finally we have also characterized interactions between *Sso*MCM and *Sso*SSB. Overall, we have elucidated a number of mechanistic details towards the assembly of the replisome for archaeal *Sulfolobus solfataricus*. The model will allow for advances within the eukaryotic system as well as enhances the evolutionary understanding of DNA replication.

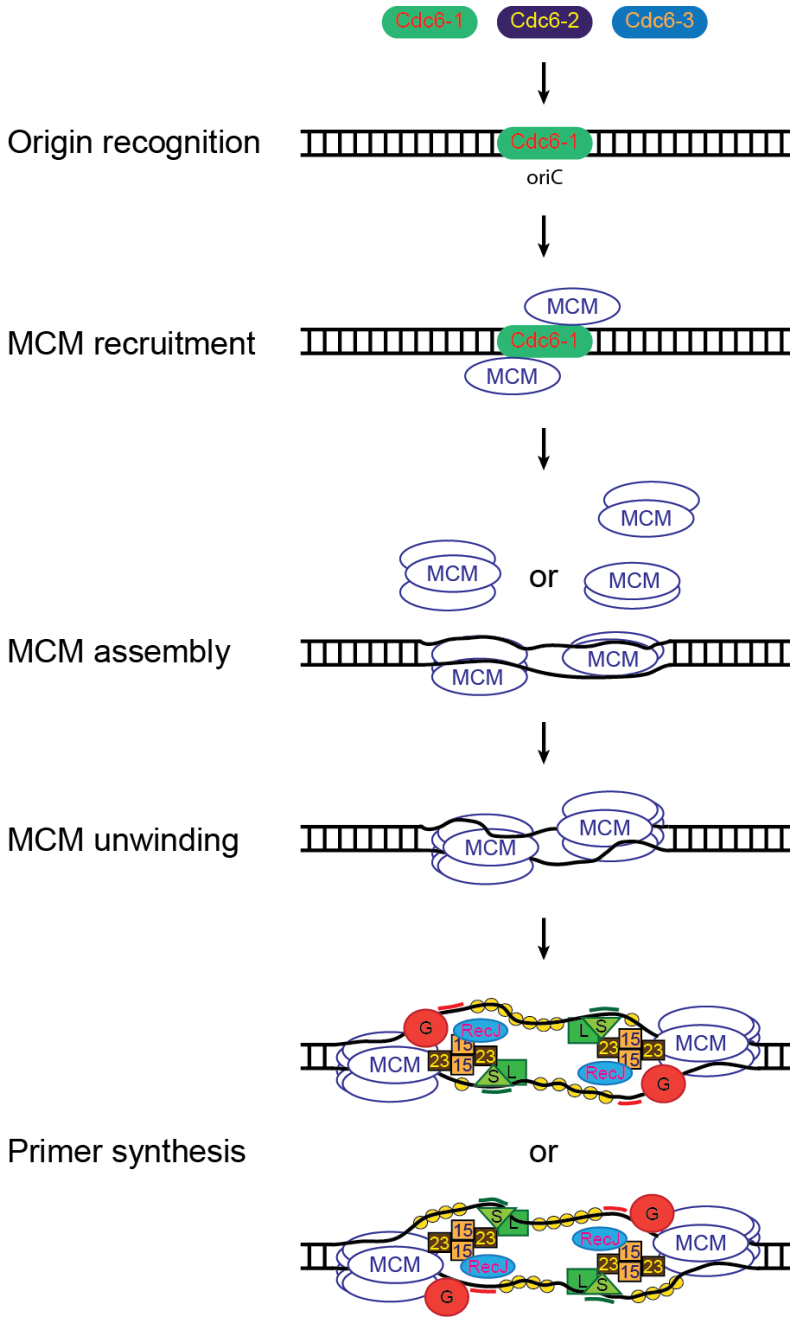


Figure 1.11: Cartoon of *Sso* initiation culminating in the formation of the primosome.
First, origin recognition occurs where one of the Cdc6 proteins binds at an origin denoted oriC depending on which origin is present. This signals MCM recruitment. Cdc6-1 leaves and bubble formation occurs allowing MCM to bidirectionally assemble either through a dimer of trimers or a trimer of dimers. Steric exclusion and wrapping unwinding then commences. This allows for partial primosome formation occurs where the GINS heterotetramer, consisting of two copies of GINS15 (orange square - 15) and GINS23 (brown square - 23) each, binds MCM through GINS23 [67]. Additionally, Rec J (blue oval - RecJ) interacts with GINS15, while PriS (light green triangle - S) and PriL (dark green square - L) both interact with GINS23 [67]. This portion of the primosome effectively positions PriSL to create primers for the leading strand. The other portion of the primosome consists of *Sso*MCM and *Sso*DnaG (red circle - G) whose interaction occurs at MCM's N-terminus. This positions DnaG to create primers for the lagging strand. The opposite orientation of DnaG priming on the leading strand and PriSL priming on the lagging strand is also shown. Free single-stranded DNA is bound by single-stranded binding protein represented by yellow circles.

2 STERIC EXCLUSION AND WRAPPING OF THE EXCLUDED DNA STRAND OCCURS ALONG DISCRETE EXTERNAL BINDING PATHS DURING MCM HELICASE UNWINDING[1]

2.1 SUMMARY

The minichromosome maintenance (MCM) helicase complex is essential for the initiation and elongation of DNA replication in both the eukaryotic and archaeal domains. The archaeal homohexameric MCM helicase from *Sulfolobus solfataricus* serves as a model for understanding mechanisms of DNA unwinding. In this report, the displaced 5′-tail is shown to provide stability to the MCM complex on DNA and contribute to unwinding. Mutations in a positively charged patch on the exterior surface of the MCM hexamer destabilize this interaction, alter the path of the displaced 5′-tail DNA and reduce unwinding. DNA footprinting and single-molecule fluorescence experiments support a previously unrecognized wrapping of the 5′-tail. This mode of hexameric helicase DNA unwinding is termed the steric exclusion and wrapping (SEW) model, where the 3′-tail is encircled by the helicase while the displaced 5′-tail wraps around defined paths on the exterior of the helicase. The novel wrapping mechanism stabilizes the MCM complex in a positive unwinding mode, protects the displaced single-stranded DNA tail and prevents reannealing.

[1] The material of this chapter is from **Graham B.W.**, Schauer G.D., Leuba S.H., and Trakselis M.A. Steric exclusion and wrapping of the excluded DNA strand occurs along discrete external binding paths during MCM helicase unwinding. *Nucleic Acids Research*, 2011. **39**(15): p. 6585-6595. Permission to reprint in this dissertation has been acquired from Nucleic Acids Research. Graham performed the mutant cloning, unwinding assays, footprinting assays, and fluorescence anisotropy assays.

2.2 INTRODUCTION

The eukaryotic minichromosome maintenance protein complex (MCM2-7) is essential for DNA replication initiation and elongation [126, 133] by participating in the licensing of chromatin [134] and subsequently functioning as a DNA helicase for unwinding [124, 135]. The archaeal MCM helicase generally exists as a homohexameric form of a single subunit product that is homologous to each of the six sequence-distinctive subunits in eukaryotes [113, 136]. All MCM helicases are members of the AAA^+ superfamily (ATPases associated with a variety of cellular activities) that share common ATP binding motifs [137, 138]. The archaeal MCM helicases from *Sulfolobus solfataricus* (*Sso*) and *Methanothermobacter thermoautotrophicus* (*Mth*) serve as simplified models for understanding mechanisms of ATP hydrolysis, DNA binding and unwinding for these classes of hexameric DNA replication helicases.

MCM hexamers share a common architecture consisting of a central channel that encircles DNA, an N-terminal tier and a C-terminal AAA^+ ATPase tier [113, 121, 139]. The unwinding polarity for the MCM helicase is 3′–5′, where critical residues at the tips of β-hairpins in the central channel contribute to DNA binding and ATP hydrolysis to drive unwinding [103, 111]. The N-terminal tier acts to increase the processivity of the helicase by modulating interactions with DNA throughout the hexamer [103]. The C-terminal AAA+ domain contains conserved motifs involved in ATP hydrolysis, demonstrated by a mutation in a lysine residue at the tip of a β-hairpin that abolishes unwinding activity [111]. The ATP binding site is positioned *in trans* between conserved domains from adjacent MCM subunits that couple ATP binding and hydrolysis to helicase activity [112]. Communication between the N- and C-terminal tiers is facilitated by a conserved allosteric loop that senses the presence of bound nucleotide and controls conformational changes between tiers coupled with ATP hydrolysis [118, 140] and presumably provides the underlying energy for unwinding.

Upon elucidation of the *Sso*MCM crystal structure, two limiting models of unwinding were proposed: steric exclusion and side channel extrusion [113]. The extrusion model is based on homology and similarities in DNA unwinding to the SV40 large T antigen, which forms a double hexamer structure where DNA is pumped out of the interface between the two hexamers [141]. In this model, double-stranded DNA (dsDNA) enters the central channel where it is separated into single-stranded DNA (ssDNA) and one strand is extruded out through a side channel. Interestingly,

*Mth*MCM has also been shown to form a double hexamer through self-association at the N-termini [139, 142], but the oligomeric state is easily modulated by the salt concentration, and the active form is thought to be hexameric, similar to *Sso*MCM [111, 143]. In the more classical representation of hexameric helicases, the steric exclusion model predicts that separation of the DNA strands occurs prior to entry into the central channel: one strand proceeds through the central channel, while the other is displaced away from the exterior surface of MCM [120, 122]. In either model, after separation of the two strands, the excluded or extruded 5′-strand was not known to interact further with MCM or to play any other role in the unwinding mechanism. Previously, the close spatial relationship between the ends of each separated ssDNA tail was observed by single-pair fluorescence resonance energy transfer (spFRET) [110]. It was hypothesized that after separation, the 5′-tail makes significant contact with the exterior of MCM, although the role of this putative interaction was unknown.

Helicase unwinding models have principally focused on defining the location and importance of the encircled DNA strand(s), while the role of the displaced strand has made relatively minor contributions to current models. In this study, we demonstrate that the 5′-tail is more than just a passive structure in the unwinding mechanism. We provide evidence that *Sso*MCM binds to and unwinds DNA using a modified steric exclusion mechanism that includes wrapping of the 5′-tail along specific paths on the exterior hexameric surface. Mutation of conserved *Sso*MCM surface residues alters the path of the 5′-tail and reduces its unwinding ability. We show that the length of the 5′-tail is important in the stabilization of the hexameric MCM structure on DNA. Furthermore, 5′-tails that are more than twice the longitudinal length of the MCM helicase are protected from nuclease degradation, suggesting that wrapping of ssDNA around the exterior of the MCM complex occurs during unwinding. We propose a steric exclusion and wrapping (SEW) model for MCM helicases, which the hexamer complex is stabilized by wrapping of the displaced 5′-strand around the exterior surface, resembling a spool of thread.

2.3 MATERIALS AND METHODS

2.3.1 Materials

ATP was obtained from Invitrogen. Mung bean nuclease, T4 polynucleotide kinase (PNK) and terminal transferase (TdT) were purchased from NEB. Optikinase was purchased from USB. All other materials were from commercial sources and were analytical grade or better. Helicase buffer is used in all unwinding and binding reactions and consists of 125 mM potassium acetate, 25 mM Tris acetate (pH 7.5) and 10 mM magnesium acetate.

2.3.2 Cloning and protein purification

K323A and R440A single and double mutants were cloned by overlap extension and insertion into pET30a (NdeI/XhoI). Mutations were confirmed using the DNA sequencing faculty at the University of Pittsburgh. Full-length wild-type (WT) and mutant *Sso*MCM were purified as previously described using 70 °C heat treatment as well as MonoQ, heparin and gel filtration columns to isolate the hexameric species [111].

2.3.3 DNA substrates

Oligonucleotides (Table 2.1) were purchased from IDT Corp and gel purified [144]. [γ-^{32}P]ATP and [α-^{32}P]dATP were purchased from MP Biomedicals and used with PNK/Optikinase or TdT to ^{32}P label the 5′- or 3′-ends of DNA, respectively. Fluorescent DNA was synthesized and HPLC purified by IDT. Complementary DNA was added in a ratio of 1.2:1. ^{32}P-labeled DNA substrates were heated at 95 °C for 5 min and then cooled to room temperature after turning off the heat block.

2.3.4 Unwinding reactions

Helicase unwinding reactions were incubated at 60 °C for 5 min and initiated upon addition of either *Sso*MCM or ATP. Final reaction conditions included helicase buffer, 2 mM ATP, 15 nM radiolabeled DNA and varying concentrations of helicase, totaling 10 µl/reaction.

Reactions were quenched with an equal volume of glycerol quench (0.5% w/v SDS, 50% v/v glycerol, 0.1% w/v bromophenol blue, 100 mM EDTA pH 8.0 and 150 nM trap ssDNA), and then stored on ice until loading. Denaturing gels [14% acrylamide (29:1 acryl:bisacryl), 8 M urea and 1× TBE buffer] or native gels [20% acrylamide and 1× TBE buffer] were used to monitor the unwinding of fork or tailed DNA, respectively, for enhanced resolution. The gels were exposed to phosphor screens, imaged using a Storm 820 Phosphorimager (GE Healthcare), and the fraction unwound was calculated.

2.3.5 Biotin–streptavidin unwinding assays

Biotin–streptavidin unwinding experiments were conducted as above with the exception that near the duplex region of the forked DNA, a thymidine residue was biotinylated on either the 3′- or 5′-strand (Table 2.1). Streptavidin concentrations were 100-fold higher than the biotinylated forked substrates to ensure a 1:1 streptavidin:biotin stoichiometry. Unwinding reactions were initiated by addition of *Sso*MCM and 20-fold excess biotin to trap the unwound DNA. Background streptavidin displacement from ssDNA was measure and corrected as described previously [145].

2.3.6 Nuclease footprinting

Stoichiometric concentrations (6:1) of *Sso*MCM (540 nM) were incubated with 90 nM DNA (spiked with ^{32}P-labeled DNA) for 5 min at 30°C, followed by addition of mung bean nuclease for 30 min before quenching with an equal volume of formamide quench (0.1% w/v SDS, 78% v/v formamide, 0.1% w/v bromophenol blue, 100 mM EDTA pH 8.0 and 900 nM trap ssDNA).

Reactions were boiled for 5 min at 95 °C, followed by 2 min on ice, before resolving and imaging as above.

Table 2.1: DNA substrates

DNA[1]	Total Length	Duplex	Sequence[2,3]
3'-tail-30 nt	66	36	5'-CACCTCTCCCTACGCTTCCCACCCACCCCGACCGGCATCTGCT ATGGTACGCTGAGCGAGAGTAGC
5'-tail-0 nt	36	36	5'-GCCGGTCGGGGTGGGTGGGAAGCGTAGGGAGAGGTG
5'-tail-20 nt	56	36	5'-GAGTCGCATGGTATCGTCTAGCCGGTCGGGGTGGGTGGGAAG CGTAGGGAGAGGTG
5'-tail-30 nt	66	36	5'-CGATGAGAGCGAGTCGCATGGTATCGTCTAGCCGGTCGGGGT GGGTGGGAAGCGTAGGGAGAGGTG
5'-tail-40 nt	76	36	5'-CGATGAGAGCCGATGAGAGCGAGTCGCATGGTATCGTCTAGC CGGTCGGGGTGGGTGGGAAGCGTAGGGAGAGGTG
5'-tail-50 nt	86	36	5'-CGATGAGAGCCGATGAGAGCCGATGAGAGCGAGTCGCATGGA TCGTCTAGCCGGTCGGGGTGGGTGGGAAGCGTAGGGAGAGGT G
5'-tail-80 nt	100	20	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTCGATGAGAGCGAGTCGCATGGTATCGTCTAGCCGGTCGGGG TGGGTGGGA
3'-tail-30 nt (IntB)	66	36	5'-CACCTCTCCCTACGCTTCCCACCCACCCCGACCGGCA<u>B</u>CTGCT ATGGTACGCTGAGCGAGAGTAGC
5'-tail-30 nt (IntB)	66	36	5'-CGATGAGAGCGAGTCGCATGGTATCGTC<u>B</u>AGCCGGTCGGGGT GGGTGGGAAGCGTAGGGAGAGGTG
3'-tail-30 nt (5'Cy3)	66	36	5'-<u>3</u>CACCTCTCCCTACGCTTCCCACCCACCCCGACCGGCATCTGC TATGGTACGCTGAGCGAGAGTAGC
3'-tail-31 nt (5'-B)(3'Cy3)	49	18	5'-<u>B</u>TGGCGACGGCAGCGAGGCTTTTTTTTTTTTTTTTTTTTTTTT TTTT<u>3</u>T
5'-tail-31 nt (5'Cy5)	49	18	5'-<u>5</u>TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGCTCGCTGCCGTC GCCA
5'-tail-51 nt (5'Cy5)	69	18	5'-<u>5</u>TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTGCCTCGCTGCCGTCGCCA
5'-tail-71 nt (5'Cy5)	89	18	5'-<u>5</u>TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTGCCTCGCTGCCGTCGCCA

[1]nt – nucleotides; [2]Modifications are underlined; [3]B – Biotin, 3 – Cy3, 5 - Cy5

2.3.7 Fluorescence anisotropy

Anisotropy experiments were performed using a Fluoromax-3 fluorometer (HORIBA Jobin Yvon) as described previously [146]. Cy3-labeled DNA was annealed to unlabeled DNA, forming forked, 3'- and 5'-tailed substrates. Anisotropy values were collected with an integration time of 0.5 seconds for eight consecutive readings. Final values from at least three independent experiments were averaged, normalized, and fit to a single binding equation:

$$v = \frac{A_{max} * [MCM]}{K_d + [MCM]} \quad (1)$$

where A_{max} is the maximal anisotropy and K_d is the dissociation constant.

Off-rate anisotropy experiments were monitored as a function of time after addition of 100-fold excess salmon sperm DNA to preformed MCM/fork DNA-Cy3 complex. Anisotropy values were collected every ~12.6 seconds for 2000 seconds with an integration time of 0.5 seconds. Results were fit to a double exponential decay equation.

$$v = A_{min} + A_1 e^{-k_1 t} + A_2 e^{-k_2 t} \quad (2)$$

where A_{min} is the final anisotropy value, A_1 and A_2 are the change in anisotropy, k_1 and k_2 are the observed rate constants. The final values are the average of at least three separate experiments.

2.3.8 Stopped flow fluorescence resonance energy transfer

Stopped flow experiments were performed on an SX.18MV (Applied Photophysics). *Sso*MCM was fluorescently labeled at the N-terminus with Alexa 488 or Alexa 555 succinimidyl esters (Invitrogen) as previously described [111]. Preformed Alexa 488-*Sso*MCM bound to fork DNA with various length 5′-tails were rapidly mixed with Alexa 555-*Sso*MCM at different concentrations and the fluorescence sensitization was monitored over time using a 570 nm cut-off filter. Changes in fluorescence were fit to single or double exponential equations using the included software.

2.3.9 Single-pair fluorescence resonance energy transfer

Cy3 and Cy5 fluorophores were placed on a 30 base 3′-termini and variable length (30, 50 and 70 bases) 5′-termini, respectively, of forked DNA substrates (Table 2.1). A 5′-biotin on the Cy3-labeled strand was used to immobilize the DNA onto a PEG-passivated quartz slide as described previously [147]. Experiments were performed on a prism-based total internal reflection

fluorescence microscope [148, 149] employing a 532 nm diode laser. Donor and acceptor emission signals were separated by a 610 nm dichroic longpass mirror, a 580/40 nm bandpass filter and a 660 nm longpass filter and subsequently imaged using an EM-CCD camera (Andor Technologies). Images from >10 regions (~50–200 molecules/region) were acquired at 10 Hz for several minutes and corrected for thermal/mechanical drift [150, 151]. Regions surrounding individual peaks (7×7 pixels) were identified by goodness of fit to a 2D Gaussian, and time-dependent intensity traces, corrected for local background, were extracted. Apparent E (E_{app}) was measured according to:

$$E_{app} = I_A/(I_A + I_D) \quad (3)$$

where I_A and I_D respectively represent acceptor and donor intensities. Histograms were produced from manually identified regions of traces displaying anti-correlated donor/acceptor signals and single-step dye photobleaching. Imaging buffer included helicase buffer and an oxygen-radical scavenging system consisting of 0.1 mg/ml glucose oxidase, 0.02 mg/ml catalase, 0.4% wt/v β-D-glucose and 2 mM Trolox. For experiments including *Sso*MCM, protein was infused into the flow cell containing tethered fork substrates and incubated for 5 min before imaging.

2.4 RESULTS

2.4.1 The 5′-tail is excluded from the central channel of *Sso*MCM during unwinding

We examined the helicase activity of *Sso*MCM on DNA substrates with different length 5′-tails (0, 30 or 50 bases). Consistent with previous results [9, 103, 111], the presence of any length 5′-tail resulted in efficient unwinding, while the absence of any DNA bases on the displaced strand reduced the unwinding efficiency (Figure 2.1A and B; Table 2.2). In order to differentiate between the two limiting models of helicase unwinding (steric exclusion and side channel extrusion), we utilized biotin/streptavidin as a physical block to unwinding that can identify which strand(s) proceeds through the central channel. Similar experiments have been performed previously with

other hexameric helicases, *Escherichia coli* DnaB and yeast MCM4/6/7, and the results were consistent with the steric exclusion model for unwinding [152].

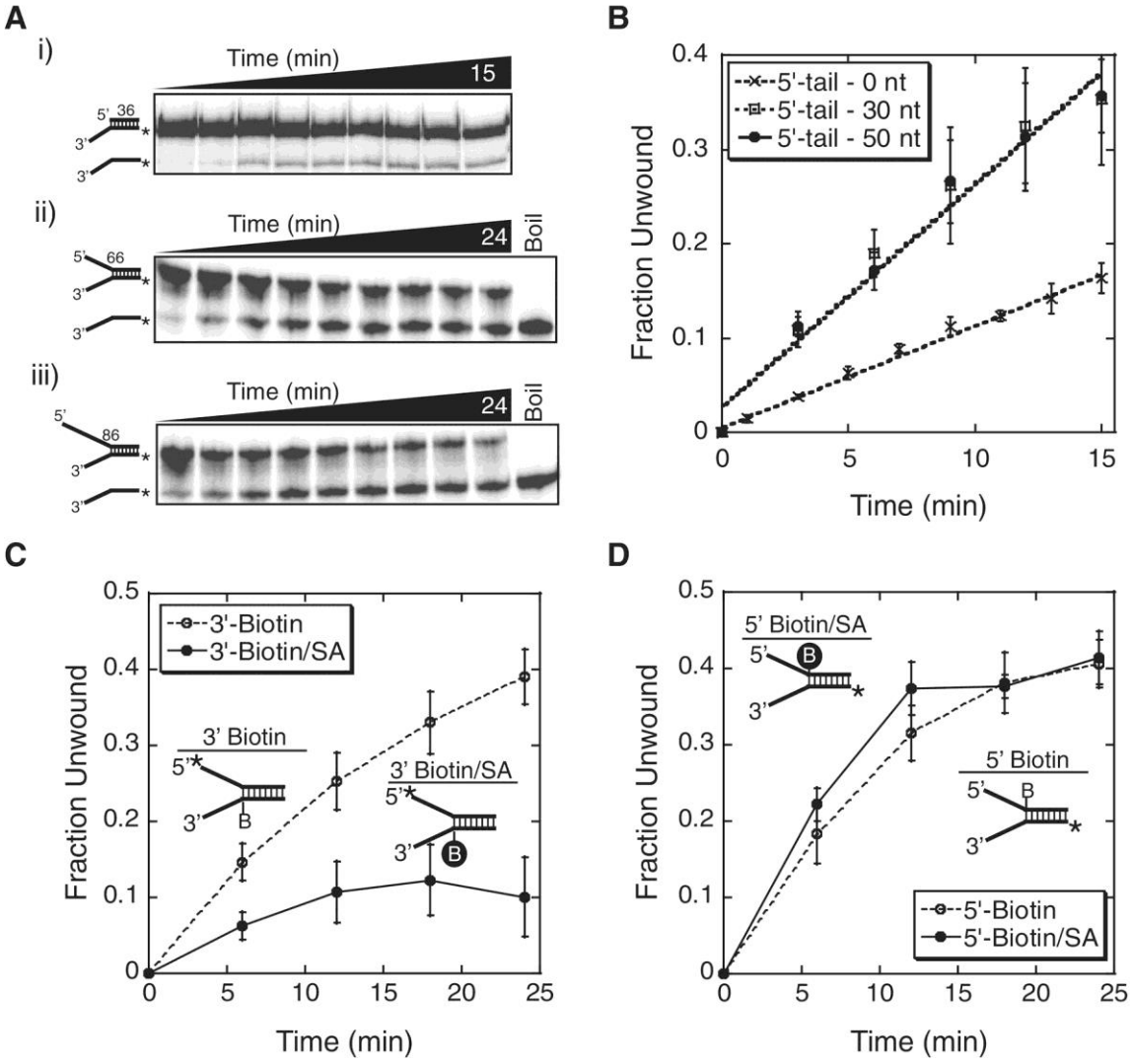


Figure 2.1: MCM unwinding of DNA with variable length 5′-tails and physical blocks on either the 3′- or 5′-strand.

A) DNA unwinding by WT *Sso*MCM (700 nM hexamer) on different length 5′-tail forked DNA substrates with (i) no, (ii) a 30 base or (iii) a 50 base 5′-tail and a constant 30 base 3′-tail. **B**) Quantification of the average of three separate unwinding reactions for each DNA substrate in A). The rates were linear for the first 12 min and equal to 0.012 ± 0.001 min^{-1} for 0 nt 5′-tail (times symbol), 0.027 ± 0.002 min^{-1} for 30 nt 5′-tail (open square) and 0.026 ± 0.002 min^{-1} for 50 nt 5′-tail (open triangle). The effect of biotin (open circle) or biotin/streptavidin (SA) (closed circle) blocks on the **C**) 3′-tail or **D**) 5′-tail strands on MCM unwinding. Cartoon DNA inset represents the ^{32}P-labeled DNA template used for each experiment. Data were from at least three independent experiments and corrected for experimentally determined streptavidin displacement as described in 'Materials and Methods' section and shown in Figure 2.2.

Biotin was incorporated on the ssDNA tail two nucleotides downstream of the ssDNA–dsDNA junction. Binding of streptavidin to this site creates a sizeable steric block to unwinding. When a biotin/streptavidin block was included on the 3′-tail, the rate of *Sso*MCM unwinding was significantly reduced (Figure 2.1C). Conversely, an identical block on the 5′-tail did not affect the unwinding rate (Figure 2.1D). We also found that *Sso*MCM was able to background displace streptavidin from a biotin-labeled ssDNA template under our experimental conditions, artificially increasing the unwinding rate when on the 3′-strand (Figure 2.2). This has already been noted for other DNA helicases and been suggested as a means to remove bound proteins from the DNA template during unwinding [145]. The background rate of streptavidin displacement was used to correct the raw unwinding data to yield the results shown in Figure 2.1C. These results are consistent with *Sso*MCM sterically excluding the 5′-strand from the central channel during unwinding.

2.4.2 The 5′-tail stabilizes the MCM hexamer on DNA

We then decided to measure specific kinetic parameters associated with DNA binding to assess the role of the 5′-tail, including a measure of *Sso*MCM subunit exchange and the off-rate (k_{-1}) of the *Sso*MCM complex from DNA. Previously, individual MCM subunits labeled at the N-terminus with fluorescent dyes were used to examine the subunit arrangement of the *Sso*MCM helicase on DNA assembled from a subunit exchange mechanism [111]. We took advantage of the ability of individual *Sso*MCM subunits to exchange in solution to kinetically monitor the stability of the *Sso*MCM hexamer on DNA templates with various 5′-tail lengths using stopped-flow FRET.

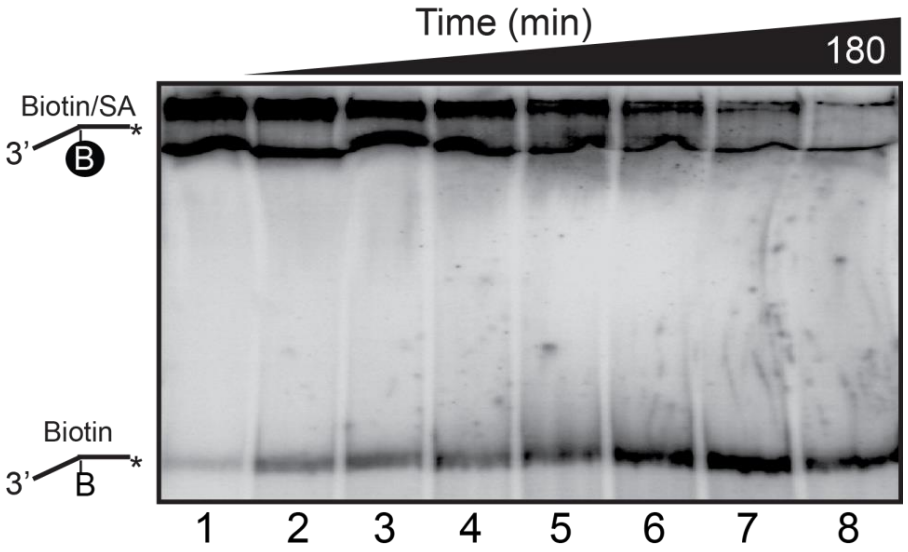


Figure 2.2: Streptavidin displacement by *Sso*MCM.
Streptavidin displacement by *Sso*MCM related to Figure 1.1. 3'-tail-30nt (IntB) (Table 2.1) was 5' radiolabeled as described in Materials and Methods. ^{32}P-DNA (15 nM) was preincubated for five minutes at 60 $^{\circ}$C with helicase buffer, ATP and 750 nM streptavidin. *Sso*MCM followed immediately by 20-fold excess biotin was added to initiate the reaction. Time points were taken from 0-180 minutes. A linear rate of 0.0067 ± 0.0002 min^{-1} fraction of streptavidin displaced was calculated over 60 minutes. This background value was subtracted and used to create Figure 2.1C and D.

Table 2.2: WT *Sso*MCM unwinding activity, binding affinity, and hexamer stability on fork DNA with different length 5' tails.

DNA Type	Helicase Activity[1]	K_d (nM)[2]	MCM Subunit Exchange[3]		MCM Off-rate[4]	
			k_{obs1} (10^{-3} s^{-1})	k_{obs2} (10^{-3} s^{-1})	k_{-1} (10^{-3} s^{-1})	k_{-2} (10^{-3} s^{-1})
No DNA	n/d	n/d	192 ± 3	23.7 ± 0.2	n/d	n/d
ssDNA (66nt)	n/d	26 ± 5	n/d	n/d	40 ± 6	1.7 ± 0.2
0 nt 5'-tail[5]	++	56 ± 6	196 ± 5	23.8 ± 0.5	14 ± 3	1.9 ± 0.5
30 nt 5'-tail[5]	+++	60 ± 13	55.6 ± 1.1	3.5 ± 0.1	12 ± 2	1.3 ± 0.2
50 nt 5'-tail[5]	+++	75 ± 13	47.4 ± 1.0	3.5 ± 0.1	7 ± 1	0.6 ± 0.1

[1] +++ represents full activity, ++ and + are 2-fold and 4-fold decreases, respectively.
[2]Measured from equilibrium fluorescence anisotropy. Fits are shown in Figure 2.5.
[3]Averaged from at least seven individual stopped flow FRET experiments;
[4]Averaged from at least three separate fluorescence anisotropy off-rate experiments.
[5]Using a 30mer 3'-tail. n/d – not determined.

Rapid mixing of acceptor-labeled *Sso*MCM with preformed donor-labeled *Sso*MCM on fork DNA resulted in an exchange of subunits and an increase in FRET (Figure 2.3A). The first exponential rate was concentration dependent and therefore attributed to the direct exchange of *Sso*MCM subunits. Doubling the concentration of acceptor *Sso*MCM while holding donor *Sso*MCM and 30-mer 5′-tail fork DNA constant gave observed rates of $0.125 \pm 0.005\ \mathrm{min}^{-1}$ and

$0.0039 \pm 0.0004\ min^{-1}$, while halving the concentration of acceptor *Sso*MCM resulted in observed rates of $0.031 \pm 0.002\ min^{-1}$ and $0.0029 \pm 0.0003\ min^{-1}$. The second exponential was unchanged with concentration and likely results from a conformational rearrangement between subunits to form the final hexameric state. Interestingly, as the 5′-tail length increased from 0 to 30 to 50 nucleotides while holding the MCM concentration constant, the subunit exchange rate decreased (Table 2.2). DNA substrates lacking 5′-tails had very similar and more rapid subunit exchange kinetics to that of *Sso*MCM alone.

Analogous experiments were performed to monitor the dissociation (k_{-1}) of the entire *Sso*MCM complex from DNA substrate with variable 5′-tail lengths after addition of a high concentration of unlabeled DNA trap using fluorescence anisotropy. Again, the presence of a longer 5′-tail reduced the off-rate of the MCM complex from DNA (Figure 2.3B and Table 2.2). As above, the change in anisotropy with time also fit better to a double exponential equation and is indicative of at least a two-step process, whereby faster dissociation of multiple subunits precedes the slower removal of any remaining MCM subunits. The second exponential rate was roughly 10-fold slower than the first rate (Table 2.2). Comparing the first and second observed off-rates between substrates with or without a 50 base 5′-tail shows that there is an equal 2- to 3-fold reduction for each rate. Therefore, stabilization of the *Sso*MCM hexamer on DNA through interaction with the 5′-tail is a concerted process that affects both steps and includes subunit dissociation as a mechanism for disassembly.

2.4.3 Mutations of conserved external residues on the surface of MCM abrogate unwinding

The recent availability of the *Sso*MCM crystal structure [113] allowed us to identify several conserved residues (Figure 2.4A) on the exterior surface that could interact with the displaced 5′-tail to stabilize the complex. Mutation of K323 and/or R440 to alanine resulted in a reduction of unwinding activity (Figure 2.4B and C; Table 2.3). These residues make up a positively charged patch on the exterior waist of the *Sso*MCM hexamer. K323 resides on the exterior of a previously identified exterior hairpin (EXT-hp) proposed to be involved in contacting DNA during unwinding [153]. Consistent with previous results [153], mutation of K323A resulted in reduced unwinding over a variety of concentrations (Figure 2.4C). R440 is positioned on the external surface at the

base of the presensor-1 hairpin (PS-1 hp). Alanine mutations of the lysine at the tip of the PS-1 hp have been shown to reduce DNA binding affinity, and eliminate unwinding [111]. Both the R440A mutation and the double mutant (K323A/R440A) reduced unwinding activity similar to K323A (Figure 2.4B and C; Table 2.3).

Table 2.3: *Sso*MCM mutant DNA unwinding and binding parameters.

*Sso*MCM	Helicase[1]	K_d (nM)[2]	Oligomeric State
WT	+++	75 ± 13	Hexamer
K323A	+	93 ± 10	Hexamer
R440A	+	92 ± 13	Hexamer
K323A/R440A	+	90 ± 9	Hexamer

[1] +++ represents full activity, ++ and + are 2-fold and 4-fold decreases, respectively.
[2]Measured using fork DNA with a 30mer 3'- and 50mer 5'-tail. Fits are shown in Figure 2.5B.

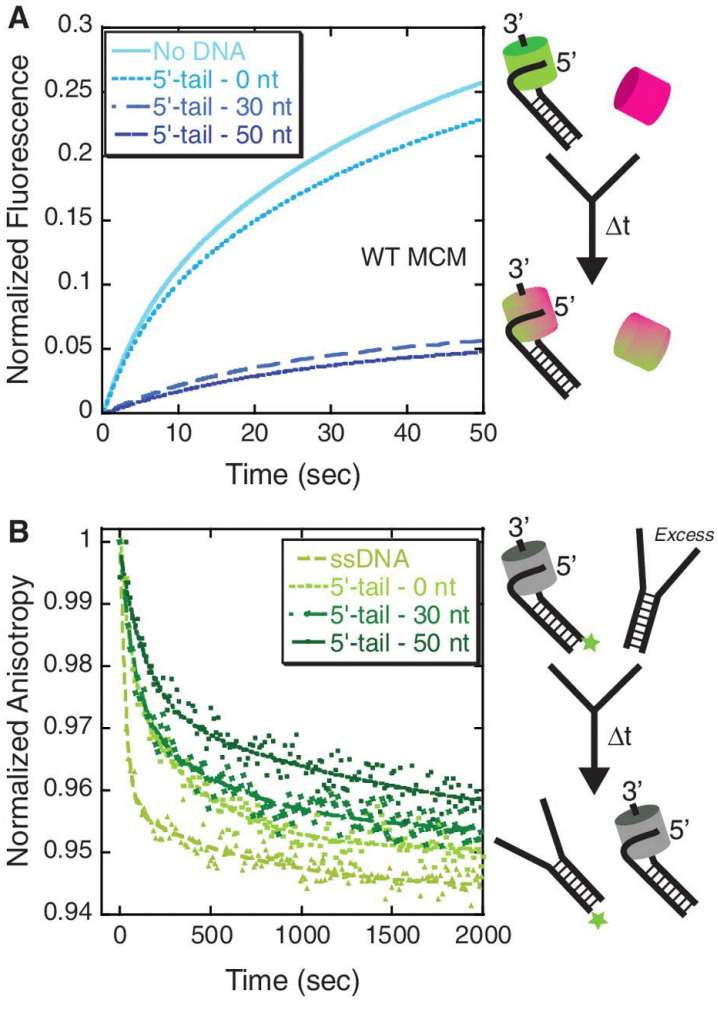


Figure 2.3: The presence of a 5′-tail stabilizes the MCM hexamer.
A) Stopped flow FRET experiments were performed to detect an exchange of donor (Alexa488) and acceptor (Alexa555)-labeled MCM subunits on DNA forks with different length 5′-tails through an increase in FRET. Concentrations of *Sso*MCM (1.2 μM) and DNA (200 nM) were stoichiometric and were held well above the K_d value to promote the DNA-bound state. The increase in FRET was fit to two exponentials for no DNA, 5′-tail, 0 nt; 5′-tail, 30 nt; and 5′-tail, 50 nt; reported in Table 2.2, and attributed to the exchange of MCM subunits from solution. The cartoon shows the result of the exchange of a donor-labeled MCM bound to DNA with a free acceptor-labeled MCM complex giving rise to a mixed donor and acceptor MCM hexamer and an increase in FRET. **B)** Kinetic anisotropy experiments monitoring the off-rate of the MCM complex from the fluorescently labeled DNA templates with different length 5′-tails upon addition of excess unlabeled DNA. The data was fit to Equation 2, and the individual rates are reported in Table 2.2. The cartoon shows the result of the dissociation of the MCM complex after trapping with unlabeled DNA leading to a decrease in the fluorescence anisotropy.

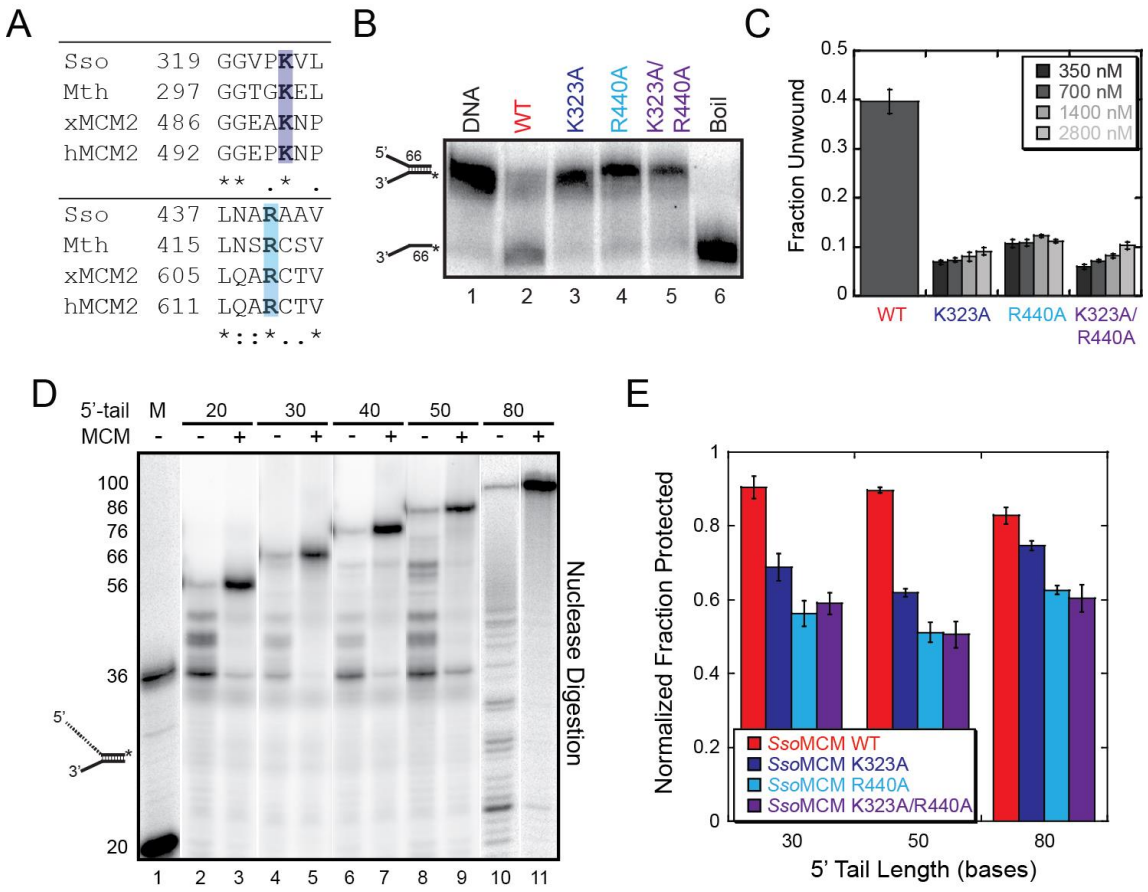


Figure 2.4: External mutations on the surface of MCM disrupt unwinding and protection of the 5′-tail.
A) Alignment of proposed exterior surface residues on MCM that interact with ssDNA using CLUSTAL W2 (http://www.ebi.ac.uk/Tools/clustalw2). Aligned are MCM exterior surface residues proposed to bind ssDNA from *Sulfolobus solfataricus* (*Sso*), *Methanothermobacter thermoautotrophicus* (*Mth*), *Xenopus laevis* MCM2 (*x*MCM2) and human MCM2 (hMCM2). **B**) DNA unwinding assays comparing wild-type and mutant MCM activities at 700 nM hexamer. Fork DNA with 30 base 3′- and 5′-tails were examined for unwinding at 60°C for 30 min as described in 'Materials and Methods' section. **C**) Quantification of fraction unwound in B) for WT at 700 nM and the three mutants at four separate concentrations (350, 700, 1400 and 2800 nM) from at least three independent experiments. **D**) Nuclease assays were performed in the presence and absence of *Sso*MCM with different length 5′-tails as described in 'Materials and Methods' section. DNA was labeled at the 3′-end with [α-^{32}P]dATP. DNA markers (M) are shown in lane 1. The length of the 5′-tail was varied from 20, 30, 40, 50 and 80 bases. The duplex region (36 bases) and 3′-tail (30 bases) were identical for lanes 2–9. The duplex region for lanes 10–11 were 20 bases and 3′-tail were 30 bases. **E**) Quantification of the fraction protected from at least three independent mung bean nuclease assays comparing WT *Sso*MCM to mutants (K232A, R440A and K323A/R440A) with 30, 50 or 80 base 5′-tails and shown and reported in Table 2.4.

Table 2.4: Quantification of nuclease protection of 5' tails, related to Figure 2.4

	*Sso*MCM Variant - Fraction Protected[1]			
DNA Type[2]	WT	K323A	R440A	K323A/R440A
30 nt 5'-tail	0.906 ± 0.030	0.689 ± 0.037	0.563 ± 0.035	0.590 ± 0.030
50 nt 5'-tail	0.896 ± 0.008	0.619 ± 0.010	0.512 ± 0.027	0.506 ± 0.036
80 nt 5'-tail	0.828 ± 0.023	0.747 ± 0.013	0.627 ± 0.012	0.604 ± 0.036

[1]Fraction of DNA (90 nM) protected in the presence of 10 units of mung bean nuclease where *Sso*MCM concentration is 540 nM. Averaged from at least three separate mung bean nuclease protection assays. Data is taken from Figure 2.4.
[2]Using a 30mer 3' tail.

It is possible that the reduced helicase activity observed with these mutations is a consequence of reduced DNA binding ability. In fact, the K323A mutation has been shown previously to have slightly reduced binding affinity as determined from EMSA experiments [153]. We instead chose to use fluorescence anisotropy assays to more accurately quantify the binding affinities of wild-type (WT) and mutant *Sso*MCMs under equilibrium solution conditions. The measured binding affinity for WT *Sso*MCM is slightly tighter than previously measured [10, 111, 113, 153], perhaps owing to our more quantitative fluorescent experimental approach. Changing the length of the 5′-tail from 0 to 30–50 nt did not significantly change the binding affinity of WT *Sso*MCM (Table 2.2 and Figure 2.5A) suggesting that the majority of binding energy results from encircling the 3′-strand. Interestingly, we detected only slight decreases in the DNA binding affinities of the mutants compared to WT *Sso*MCM using a forked DNA substrate with a 30 base 3′-tail and a 50 base 5′-tail (Table 2.3 and Figure 2.5B).

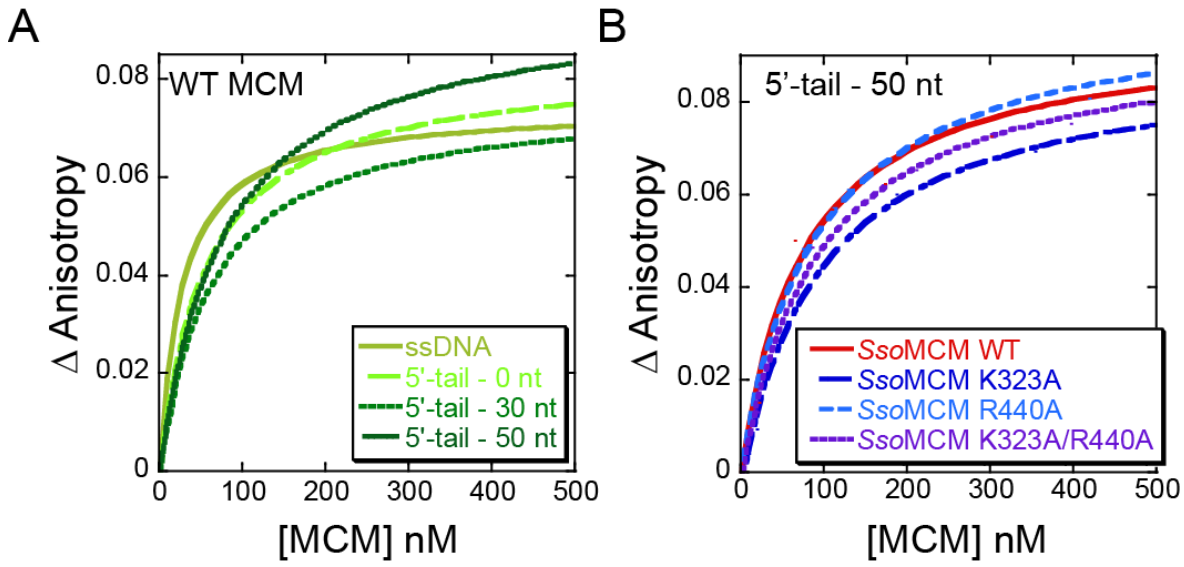


Figure 2.5: *Sso*MCM variants binding to various DNA substrates monitored by fluorescence anisotropy. **A)** Change in fluorescence anisotropy for *Sso*MCM WT binding to four different DNA substrates: ssDNA, 5'-tail – 0 nt, 5'-tail – 30 nt, and 5'-tail – 50 nt, as described in the Materials and Methods. The Kd values were calculated from the average of at least three sets of data using Equation 1. The values are shown in Table 2.2. **B)** Change in fluorescence anisotropy for *Sso*MCM WT, K323A, R440A, and K323A/R440A when binding to fluorescently labeled forked DNA (3'-tail – 30 nt and 5'-tail - 50 nt with duplex region of 36 bp). K_d values were calculated from the average of at least three sets of data using Equation 1 and reported in Table 2.3.

2.4.4 Long 5′-tails are protected from nuclease digestion

The stability of the interaction of the 5′-tail on the exterior of the *Sso*MCM hexamer was probed using a nuclease footprinting assay. Mung bean nuclease was selected since it is a single-strand-specific DNA endonuclease. We designed forked DNA substrates with a constant 30 base 3′-tail and 5′-tails ranging from 20 to 80 nt (Table 2.1). The 3′-end of the 5′-tail was labeled with [α-^{32}P]dATP to visualize the length of the digested 5′-tail.

In Figure 2.4D, a mung bean nuclease mapping experiment was performed in the absence and presence of WT *Sso*MCM. Varying lengths of the 5′-tail (20, 30, 40, 50 and 80 nucleotides) were probed for protection from nuclease digestion. The stoichiometry of protein and DNA were

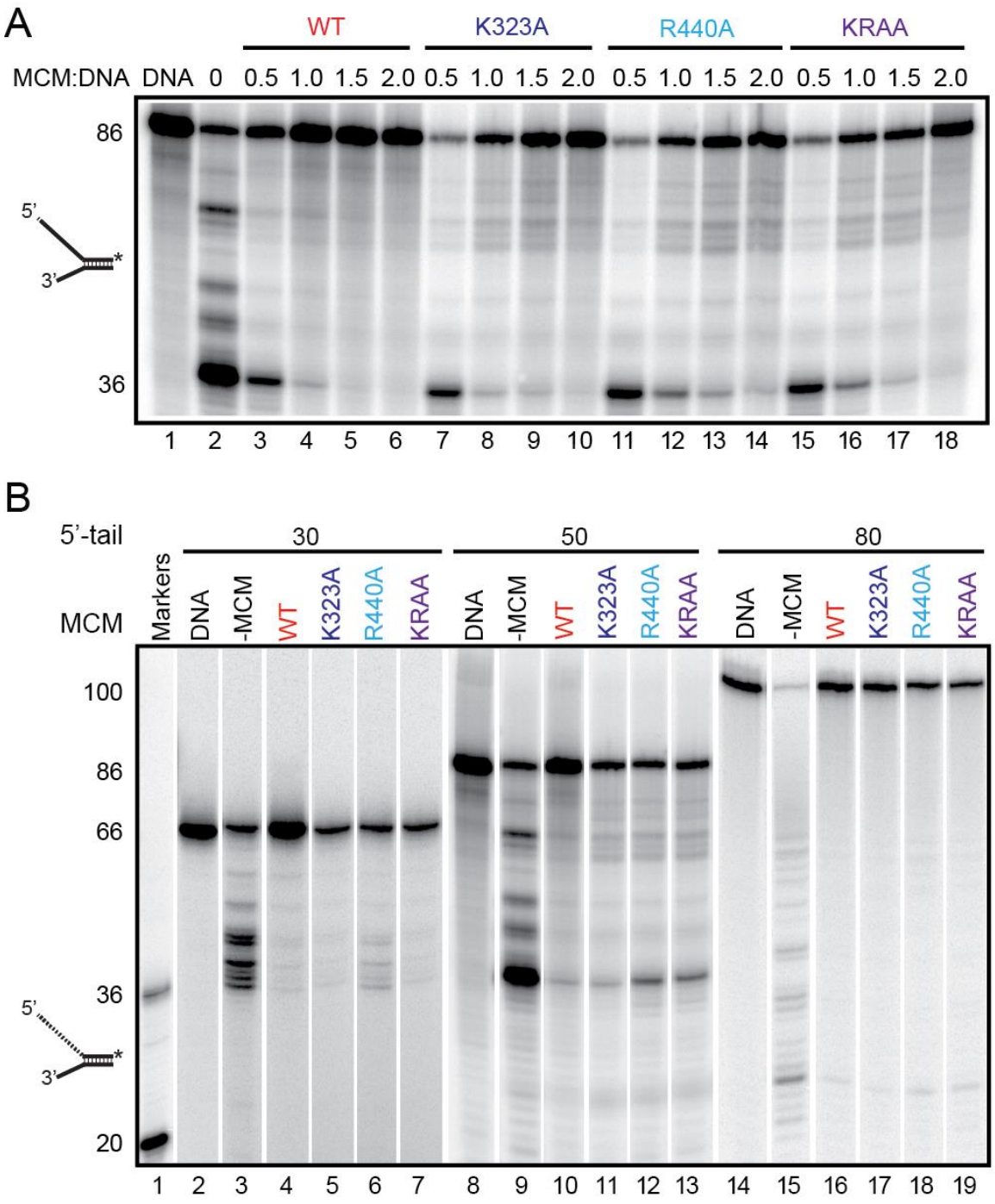


Figure 2.6: Mung bean nuclease 5'-tail protection assay gels for *Sso*MCM variants.
A) Nuclease assays were performed in the absence and presence of increasing *Sso*MCM (WT, K323A, R440A and KRAA) concentration to titrate stoichiometry. 90 nM of fork 50mer 5'-tail and 30mer 3'-tail DNA was utilized with MCM hexamer:DNA 0.5 (45 nM), 1.0 (90 nM), 1.5 (135 nM), and 2.0 (180 nM). **B)** Nuclease assays were performed in the presence and absence of 540 nM *Sso*MCM WT, K323A, R440A and K323A/R440A (KRAA) with different length 5'-tails as described in Materials and Methods. The 5'-tail strand was labeled at the 3'-end with ^{32}P. DNA markers (Markers) are shown in lane 1. DNA alone is shown in lanes 2, 8, and 14. The length of the 5'-tail was varied from 30, 50, and 80 bases. The duplex region (36 bases) and 3'-tail (30 bases) were identical for lanes 2-13. The duplex region for lanes 14-19 is 20 bases and 3'-tail is 30 bases. Quantification of this data is shown in Table 2.4.

carefully controlled to allow for 1:1 binding. To be certain, a titration of WT *Sso*MCM (270–1080 nM) with constant DNA (90 nM) was examined using these nuclease assays (Figure 2.6). In the absence of protein, there are digested bands of varying sizes down to the ssDNA-dsDNA junction (either 20 or 36 nt), whereas in the presence of protein, the DNA becomes almost fully protected (Figures 2.4D and 2.6). A limited amount of digestion of the 5′-tail occurred at the ssDNA-dsDNA junction in all cases when *Sso*MCM was bound.

We also measured the protection of the 5′-tail with the external ssDNA binding site mutants (K323A, R440A, K323A/R440A) in identical assays (Figure 2.6B). The fraction of mung bean digested 5′-tail was quantified and compared between WT and mutant *Sso*MCMs for 30, 50 and 80 base 5′-tails (Figure 2.4E and Table 2.4). There is a significant increase in 5′-tail digestion for each of the *Sso*MCM mutants compared to WT. The change in the digestion pattern is most significant for the 50 base 5′-tail. In other cases, there is a more distributed nuclease digestion pattern observed along the length of the 5′-tail for the mutants.

2.4.5 External surface mutations disrupt the path and stability of binding the 5′-tail of ssDNA

In order to characterize the 5′-tail binding path on the *Sso*MCM exterior in the absence of ensemble averaging, we utilized spFRET. Previous results from spFRET experiments indicated a dynamic interaction of the 5′-tail with the exterior of the *Sso*MCM hexamer [110], but the roles of 5′-tail length or specific *Sso*MCM residues involved in this interaction were not investigated. Based on the perturbation of unwinding by K323A/R440A shown above, we predicted that the path of the 5′-tail would be altered compared to that of WT. Here, we report results from experiments in which we monitored spFRET from the free ends of DNA with variable length 5′-tails in complex with WT or K323A/R440A *Sso*MCM.

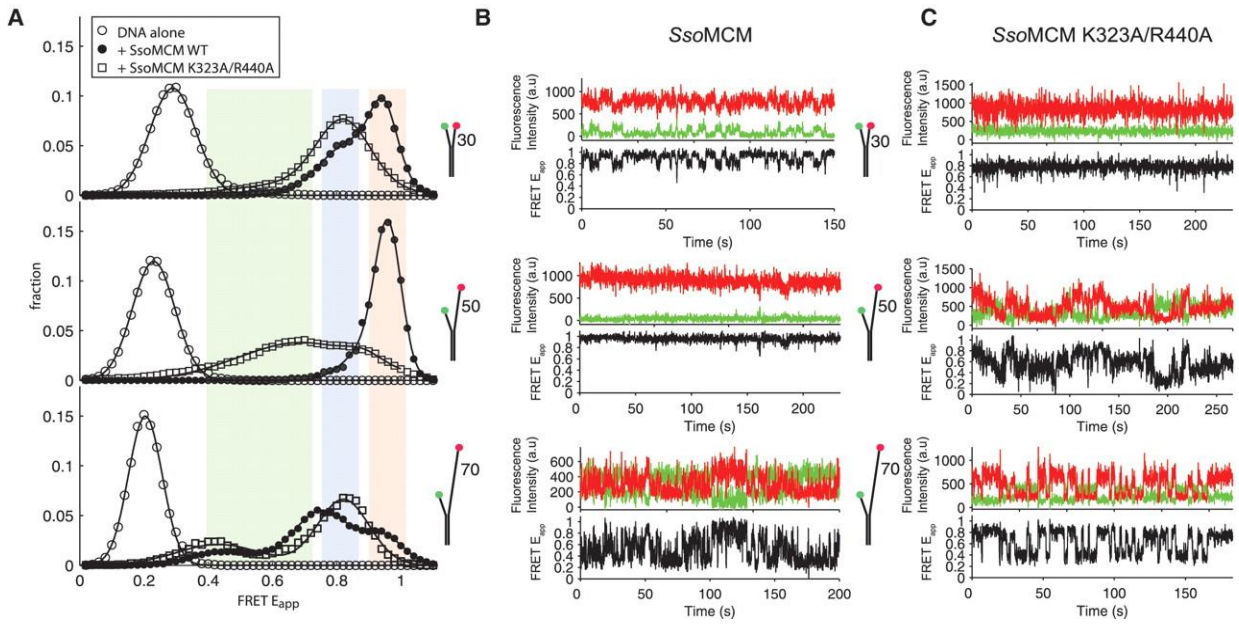


Figure 2.7: Single-pair FRET experiments monitoring the change in position of the 3′- and 5′-tails upon binding wild-type *Sso*MCM or external surface mutant *Sso*MCM (K3232A/R440A).
A) Shows the single-molecule distributions of the FRET populations of DNA alone (open circle), DNA bound to WT *Sso*MCM (filled circle) or DNA bound to *Sso*MCM (K323A/R440A) (open square) for a constant 30-mer 3′-tail and variable length 5′-tails (30, 50 or 70 bases). For all histograms, points represent binned data and solid lines represent composite Gaussian fits. Colored panes denote the high (red), medium (blue) and low (green) states as discussed in the text. Representative kinetic single-molecule fluorescence and FRET traces for each DNA substrate for **B**) WT *Sso*MCM or **C**) *Sso*MCM (K323A/R440A), indicated as donor (green), acceptor (red) and FRET (black).

Forked DNA substrates with Cy3 on the end of a 30 base 3′-tail and Cy5 on the 5′ end of different length 5′-tails (30, 50 and 70 bases) were attached to the surface of a flow chamber through a biotin–streptavidin interaction. As expected, the FRET values of the forked substrates alone decreased with increasing 5′-tail length (Figure 2.7A). Subsequently, WT or K323A/R440A *Sso*MCM was flowed into the chamber containing tethered fork substrates. In all instances, FRET values dramatically increased after *Sso*MCM was added, consistent with the 3′- and 5′-tails coming together in closer proximity. A titration of WT *Sso*MCM is shown in Figure 2.8A illustrating the FRET increase. In complex with WT *Sso*MCM, different length 5′-tails (30, 50 and 70 bases) displayed a consistent high FRET state (~0.95) (Figures 2.7A and B; 2.8B; 2.9). The complex with the 50-mer 5′-tail exists almost entirely in this high FRET state, with most traces rarely or never visiting the medium FRET state (highlighted in Figure 2.7B by a brief excursion to the medium state), whereas this high FRET state is transiently, but stably visited during the course of the

trajectory for the case of the WT *Sso*MCM bound to 30- and 70-mer 5′-tail forked substrates (Figure 2.7B). Deconvoluted Gaussian fits for each FRET population are shown in Figure 2.9.

In contrast, in the presence of K323A/R440A *Sso*MCM, the FRET states were shifted to either medium or lower, more broadly distributed FRET states, suggesting conformations different from those seen for WT (Figure 2.7A). In all cases, the extremely stable high FRET conformation observed with WT *Sso*MCM is completely absent when the external binding site is mutated. The FRET states of the surface mutations for each of the forked substrates showed a dramatically different dynamic behavior between WT- and K323A/R440A-bound complexes (Figures 2.7B and C; 2.8B and C; and 2.9). Lifetimes of the FRET populations for each experiment also show that mutation of an external binding site in *Sso*MCM alters the kinetics and dynamics of 5′-tail binding (Table 2.5). The bimodal switching behavior between high and medium states observed with 30-mer 5′-tail in complex with the WT *Sso*MCM collapses to a stable, unimodal medium FRET state when complexed with K323A/R440A. Furthermore, the stable high FRET state observed with the 50-mer 5′-tail and WT *Sso*MCM is shifted to a broadly distributed state upon addition of the double mutant. Addition of K323A/R440A MCM to the 70-mer 5′-tail forked substrate resulted in a bimodal distribution between a low and medium FRET state, in contrast to the highly dynamic FRET state observed with WT *Sso*MCM.

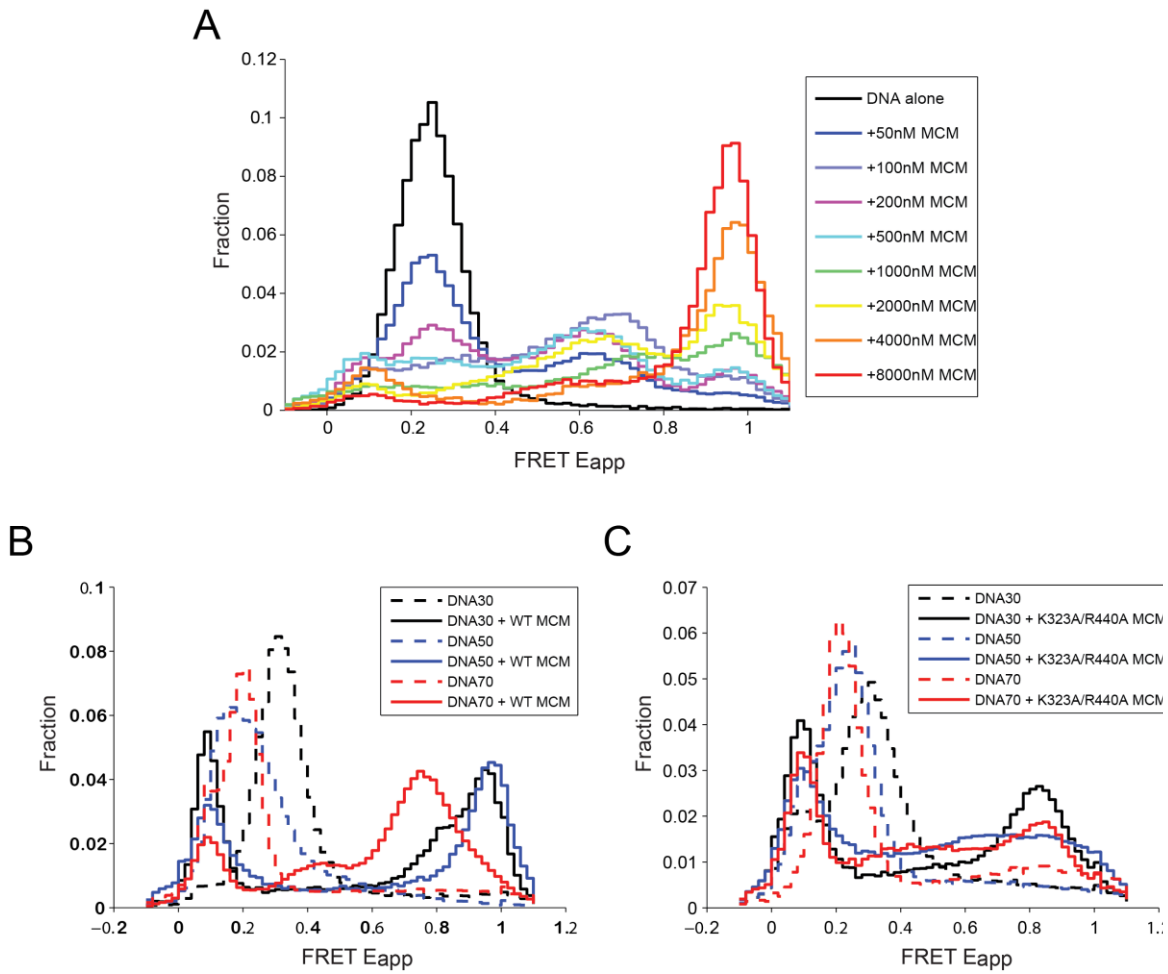


Figure 2.8: spFRET titration of *Sso*MCM to visualize hexamer formation.
A) DNA forked substrates with 50mer 5'tails were tethered to a flow cell, followed by incubation with WT *Sso*MCM helicase at the concentration indicated in the legend. Histograms represent data from all time points of every identified spFRET trace (i.e., regions of traces were not hand selected to eliminate signal from photobleached species, etc.). Example unfiltered datasets for **B)** WT or **C)** K323A/R440A *Sso*MCM. Raw spFRET histograms derived from all time points of every identified spFRET trace are indicated. Dataset represents the data displayed in Figure 4 before hand-picking regions in order to eliminate spurious data such as data without clear anticorrelated donor/acceptor signal as well as the "zero-peak" arising from photobleached species or species without acceptor. Forked substrate is indicated by DNAXX, where XX represents the length of the 5'tail in bases.

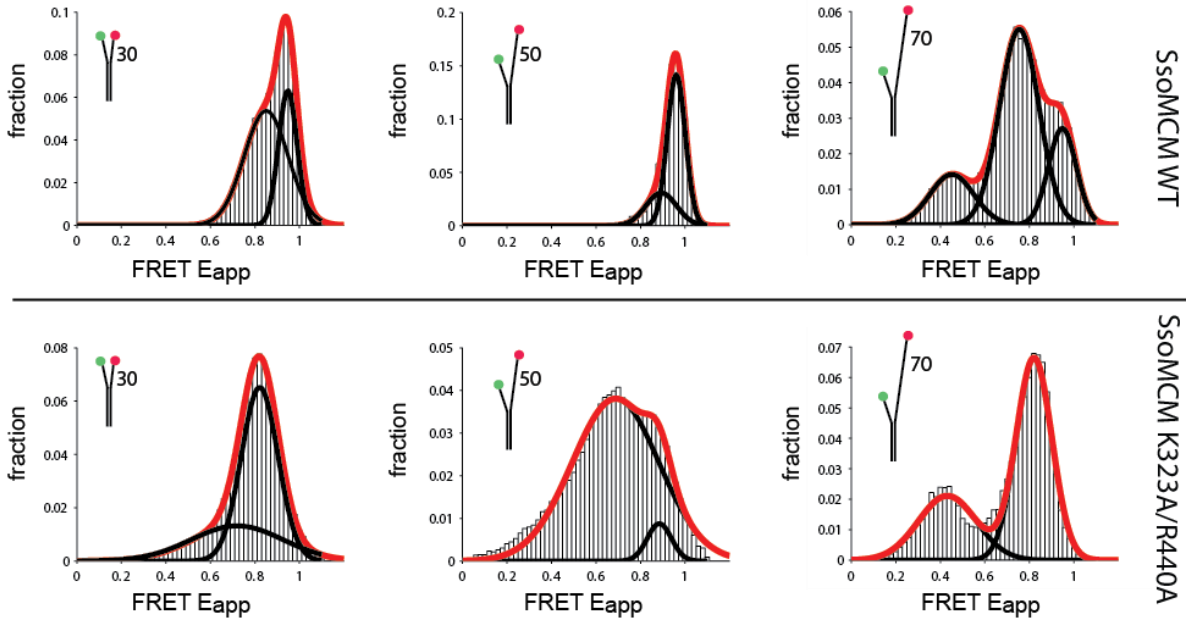


Figure 2.9: Deconvolution of composite spFRET histograms from Figure 2.7.
For each experiment indicated with 30, 50 or 70 base 5'-tail and either WT or K323A/R440A *Sso*MCM, the histogram is shown as a bar graph. The composite curve fit is shown in red, and the individually identified peaks are shown in black. Also see Table 2.5, where individual curves were numerically integrated and expressed as a percentage of total.

Table 2.5: Individual spFRET peak values.

*Sso*MCM Variant	5'-Tail Length (nt)	E_{app}[1]	Percent of Total Population (%)
WT	30	0.95 ± 0.14	67
		0.84 ± 0.06	33
	50	0.96 ± 0.06	83
		0.89 ± 0.10	17
	70	0.95 ± 0.09	18
		0.75 ± 0.12	61
		0.45 ± 0.14	21
K323A/R440A	30	0.82 ± 0.12	67
		0.72 ± 0.28	33
	50	0.88 ± 0.08	6
		0.69 ± 0.28	94
	70	0.82 ± 0.11	66
		0.43 ± 0.18	34

[1] Mean ± STDEV of individual deconvoluted peaks.

2.5 DISCUSSION

2.5.1 *Sso*MCM sterically excludes the 5′-strand during unwinding

Homohexameric MCMs from archaea serve as models for more complex eukaryotic MCMs for determining enzymatic mechanisms of DNA binding, ATP hydrolysis and DNA unwinding. Several models for DNA unwinding by hexameric helicases have been proposed [104, 122, 154]. Although steric exclusion models for unwinding by hexameric helicases have been supported by results from *E. coli* DnaB and Yeast MCM4,6,7 [155], SV40 Large T antigen is proposed to operate as a dsDNA pump displaying 'rabbit ear' like protrusions of DNA out of a double hexamer complex [141, 156, 157]. Support for the DNA extrusion model of unwinding for MCM helicases is based on homology to the SV40 large T, visualization of side channels that can accommodate ssDNA [113, 158] and the functional characterization of β-hairpins that are important for DNA binding and unwinding [103, 111, 113, 153].

Our biochemical experiments show a deceleration of DNA unwinding only when a biotin–streptavidin block is included on the encircled 3′-strand and not on the displaced 5′-strand. *Sso*MCM was able to displace some background amounts of streptavidin from the template similar to other non-hexameric DNA helicases [159, 160], perhaps owing in part to our elevated reaction temperatures. Nevertheless, there is clear reduction in the unwinding by *Sso*MCM when a biotin-streptavidin block is included on the 3′-strand suggesting that only the 3′-tail enters the central channel and the 5′-tail is excluded.

*Sso*MCM has been shown to be more efficient at unwinding fork substrates than 3′-tail only substrates [111]. The presence of any 5′-tail, regardless of length, increases the unwinding ability of *Sso*MCM. Tails as short as 9 nt have been shown to be just as efficient as longer 30 nt 5′-tails [9]. Longer 5′-tails (50 versus 30 nt) used did not elevate the unwinding efficiency. On the contrary, the side channel extrusion mechanism should be almost as efficient at unwinding substrates without 5′-tails as efficiently those with tails if separation occurs within the central channel. Clearly, the unwinding efficiency is primarily determined by the ability to preseparate DNA before entry into the central channel. DNA templates without a 5′-tail will consequently end up encircled in the central channel of the MCM complex in a futile unwinding attempt.

2.5.2 External surface residues on *Sso*MCM direct the binding of the displaced 5′-tail of ssDNA

We have identified two universally conserved residues (K323 and R440) on the exterior of *Sso*MCM that significantly contribute to its unwinding efficiency. These residues reside in close proximity on the external MCM surface and contribute to the structure at the base of separate β-hairpins. They are not directly adjacent to the side channels identified in the crystal structure [113] but are roughly 20 Å away. Collectively, they create a positively charged patch that likely directs binding of the displaced 5′-strand and contributes to unwinding. We measured an 6-fold decrease in unwinding activity for the K323A mutation but did not observe a decrease in binding affinity as detected previously [153]. Instead, we utilized a fluorescent anisotropy DNA binding assay and found that the K_d is similar for WT and mutant MCMs (Table 2.2). Thus, the decrease in unwinding ability of the mutants is most likely due to a reduced stabilization of the separated 5′-tail on the exterior surface. This external stabilization does not significantly add to the binding affinity in an equilibrium state, but instead contributes to strand stabilization to promote unwinding. Conserved DNA hairpins in the central channel involved in coupling ATP hydrolysis to unwinding [111] would then act in concert with the exterior interactions to direct the helicase forward.

Single-pair FRET experiments performed with the K323A/R440A mutant showed an increase in the distribution spread of the FRET signal between the 3′- and 5′-tails, reflecting a change in 5′-tail binding dynamics. This observation was most prevalent when a 50-mer 5′-tail was used, where the surface mutations caused the distribution to change from a fairly discrete high FRET state (0.95) that was stable on the order of 100 s to a more diffuse population that visited a broad range of FRET states. We consider medium (~0.8) and lower (~0.4–0.7) FRET states to represent alternate surface-bound and free conformations dependent on the 5′-tail length. Collapse of the bimodal peak observed for the WT complex with the 30-mer 5′-tail to a medium peak with the double mutant suggests that the 5′-tail is in an unbound, displaced conformation. Upon mutation, the stronger surface interactions observed with the 50-mer 5′-tail, afforded by an ideal binding path are weakened, causing a metastable interaction of the 5′-tail with the surface.

Although the 70-mer 5′-tail complexed with WT MCM occasionally visited the stable, high FRET state, it also displayed a wide range of FRET states that were unstable and displayed both rapid and slower transitions, possibly representing free Brownian motions of the unbound 5′-tail

and dynamic 5′-tail interactions with the surface, respectively. The decreased stability observed with the 70-mer 5′-tail may indicate a competition between the free energy of solvation and that of surface binding. For K323A/R440A, the surface binding energy is diminished, resulting in a free tail that rapidly switches between two extended conformations.

2.5.3 The 5′-tail stabilizes the hexameric MCM complex on DNA

Single-stranded DNA wrapping around the exterior also stabilizes the MCM hexamer complex. The off-rate (k_{-1}) of the hexamer from DNA is reduced in the presence of a 5′-tail. Shorter (30-mer) 5′-tails still allow for dissociation to occur by sliding off the end of the DNA. This off-rate reduction is seemingly in contradiction to the similar binding affinities measured for different length tails, however, a slower on-rate (k_1) of binding due to increased complexity of organization for 5′-tailed substrates would compensate for the overall dissociation constant ($K_d = k_{-1}/k_1$). As an aside, we measured the fastest off-rate (k_{-1}) for ssDNA, which also possessed the tightest binding. The ease with which ssDNA can bind in multiple conformations, either on the exterior surface or in the interior channel of *Sso*MCM, can be explained by a fast on-rate (k_1).

Subunit exchange of MCM subunits from solution also provides a possible assembly mechanism, especially on pre-separated DNA substrates with single-stranded regions. Once loaded into a competent complex with DNA, further assembly or exchange of MCM subunits would be reduced through interactions with the 5′-tail. We suspect that binding of the 5′-tail propagates conformational changes throughout the hexamer creating a tighter, more efficient DNA unwinding complex that restricts subunit exchange. The increased stability of the MCM–DNA complex would result in a more competent helicase, able to more efficiently unwind DNA. We speculate that initial binding to fork DNA is directed by a subunit exchange mechanism primarily centered around the 3′-tail, while more subtle exterior interactions with the 5′-tail stabilize the hexamer complex to promote MCM towards an active unwinding state.

2.5.4 Novel SEW model of DNA unwinding for *Sso*MCM

Previously, it was shown that the displaced 5′-strand dynamically interacts with the exterior surface of *Sso*MCM [110]. This result helped distinguish between the steric exclusion and side

channel extrusion models of unwinding, confirming that a 5′-tail can interact in a way only possible with an external binding event. For this study, we used different 5′-tail lengths to more specifically map the exterior binding site using spFRET. Discrete high FRET populations (~0.85 and ~0.95) exist that are present with all 5′-tail lengths. If the 3′-tail is fixed within the central channel of MCM, then the distance to the 5′-tail must be identical for 30, 50 and 70 base 5′-tails for those FRET states. Though a stable high FRET state could also be consistent with side channel extrusion, we consider it unlikely, since the 5′-tail dynamics for the 30-mer substrate would be more restricted due to the displaced strand first passing through the central channel and then out a side channel.

The longitudinal length of the *Sso*MCM hexamer is roughly 100 Å [115], approximately equal to 24 bases of linear ssDNA (4.1 Å/base). If the encircled 3′-tail follows a helical conformation similar to that detected for the hexameric papillomavirus E1 helicase binding to ssDNA [161], this would place the 3′-end towards the N-terminal end of *Sso*MCM, but not exposed [111]. Similar vertical displacement of the 30-mer 5′-strand on the exterior surface would position the donor and acceptor fluorophores in a similar location. However, if the longer 50-mer 5′-tail were to take the same longitudinal path, then the 5′-Cy3 would project past the N-terminal tier of the MCM complex. The 70-mer 5′-tail would project even further and would display an even smaller FRET value. Clearly, the linear length of the 50- and 70-mer 5′-tails precludes the appearance of such identical high FRET signals if the 5′-strand was traversing along a linear path on the MCM hexameric exterior.

Only a model that includes some wrapping of the 5′-tail around the exterior of the MCM complex could explain similar FRET states for different length 5′-tails. Complete wrapping of the 5′-tail around the exterior surface would produce an identical, unimodal high FRET state irrespective of the length. In addition to this state, we also observed similar discrete FRET states for each 5′-tail. The symmetry of the hexameric *Sso*MCM helicase dictates that identical repeating structural features make up the exterior surface. Therefore, wrapping of the ssDNA could interact with one, two, or more subunits before exiting past the N-terminal tier. In support of a wrapping model, electron microscopy studies have identified an external MCM hexamer binding site for dsDNA that is also proposed to wrap DNA around the exterior surface of the hexameric MCM complex [162]. Although this dsDNA binding path resides along a helix–turn–helix motif in the N-terminal tier, it is possible that the unwound ssDNA 5′-tail may transverse the waist and also interact along this dsDNA binding path before being released from the helicase.

Wrapping of the displaced 5′-tail would also provide a structural mechanism to prevent reannealing. The wrapped strand would include a larger number of DNA bases interacting with *Sso*MCM than the encircled strand. The resulting displacement in complementary sequence between the two strands would act to prevent reannealing behind the helicase. A significant fraction of even the longest 5′-tail (80-mer) is protected from nuclease digestion when WT *Sso*MCM is bound. Some specific cleavage can be detected at the 5′-single-strand/double-strand junction, liberating ssDNA with a length equal to the tail. Other low-level, non-specific cleavage sites can be detected along the length of the 5′-tail. Therefore, the majority of the 5′-tail resides are in a locally protected environment on the external surface of the MCM hexamer. Presumably, rapid fluctuations between high FRET (bound) and low FRET (unbound) states would not provide a stable enough substrate for nuclease digestion.

Mutations in the positive binding patch on the surface of the MCM hexamer result in a more disperse nuclease digestion pattern suggesting a more loosely bound 5′-tail. This is also consistent with the spFRET experiments that displayed a much broader FRET distribution with the mutant MCM. It is therefore likely that the K323A and R440A mutations at each MCM subunit reduce the binding affinity of the 5′-tail, indicating these residues probably serve as positively charged guides for wrapping negatively charged DNA on the MCM surface. Other external binding patches are most likely present, as suggested from incomplete total 5′-tail digestion and the presence of distinct FRET states for the K323A/R440A mutant, which can loosely bind the 5′-tail.

Based on these results, we have developed a new model for DNA unwinding by MCM termed the SEW model, resembling a spool of thread for wrapping the 5′-tail around the exterior surface (Figure 2.10). This model includes steric exclusion and subsequent wrapping of the 5′-tail around the exterior of the MCM hexamer, which promotes DNA unwinding, MCM–DNA stabilization, protection of ssDNA, and strand separation. There are distinct channels on the external surface of the MCM hexamer dotted with positive charge that could direct the displaced 5′-strand around the hexamer. After significant wrapping of the 5′-tail has occurred, further DNA bases will proceed past the N-terminal tier and be released from the hexamer.

The displaced 5′-tail cannot wrap the MCM complex too tightly; it must be allowed to slide over the external surface of the hexamer for efficient unwinding. The spFRET dynamics for binding the 5′-tail detected here provides the basis for this dynamic binding mechanism. Therefore,

binding of the 5′-tail is a compromise between adequate protection of the ssDNA, efficient stabilization of unwound DNA that contributes to the unwinding mechanism, and stabilization of the entire hexameric MCM complex.

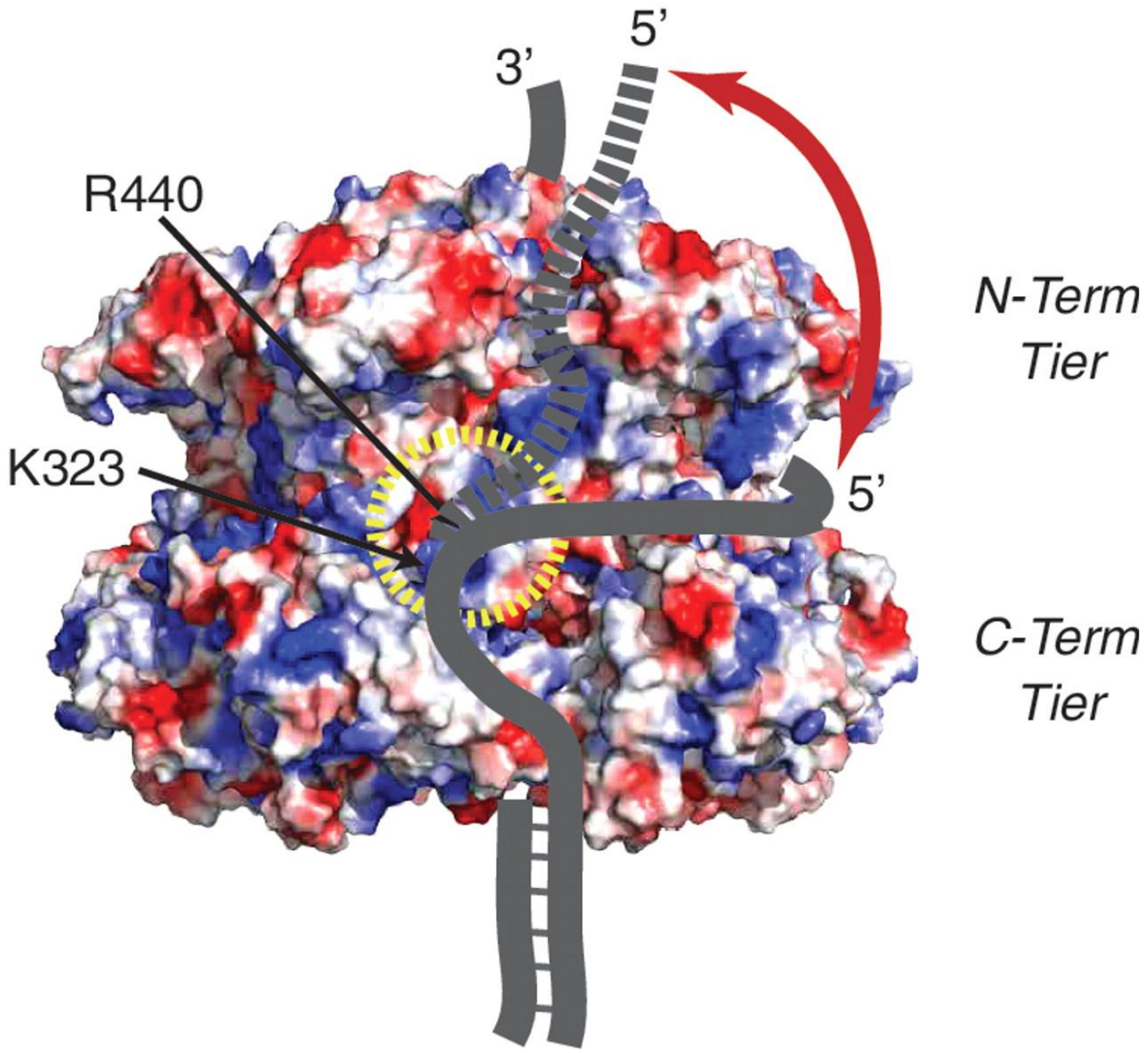


Figure 2.10: Proposed steric exclusion and wrapping (SEW) model for DNA unwinding.
At least two paths for wrapping ssDNA on the exterior surface of MCM are highlighted. Shown is the surface electrostatic potential representation of *Sso*MCM (PDBID:3F9V), where blue and red patches, respectively, represent positive and negatively charged residues, highlighting ssDNA binding residues (K323 and R440) that direct the path of the 5′-tail around a ssDNA protection channels in the waist (solid line) before exiting past the N-terminal tier (dashed line).

2.6 FUNDING

Department of Chemistry, University of Pittsburgh (M.A.T.); National Institutes of Health (GM077872 to S.H.L.). Funding for open access charge: Department of Chemistry, University of Pittsburgh.

2.7 ACKNOWLEDGEMENTS

We thank Xiaojiang Chen for providing the coordinates for the *Sso*MCM symmetrical hexamer model.

3 *SSO*MCM Y519 IS CRUCIAL FOR HEXAMERIZATION WITH INTERACTION PATCHES DETERMINED BY HDX-MS[2]

3.1 SUMMARY

In an attempt to further identify surface residues that interact with DNA, we have biochemically characterized six point mutations in *Sso*MCM. In particular, Y519A located in the α/β-α linker region of the C-domain disrupts the hexamer. Y519A also shows decreased unwinding and ATPase rates, while exhibiting an enhanced secondary structure that can be modulated by temperature. This is the first single site tyrosine mutation that provides insight on the α/β-α linker region's purpose which we postulate is for hexamerization. We also used HDX-MS to map the helicase binding locations for the 3' and 5'-strands of DNA and the hexameric interfaces.

[2] This work is in collaboration with Alan G. Marshall's laboratory at the National Magnet Lab at FSU, and is being prepared submission at this time. **Graham B.W.***, Tao, Y*, Young N.L, Marshall A.G., and Trakselis M.A. H/DX-MS insights into the α/β-α linker region critical for hexamerization. *Graham and Tao share co-first authorship. Graham performed the mutant cloning, protein purifications, unwinding assays, ATPase assays, EMSAs, and CD spectroscopy.

3.2 INTRODUCTION

Sulfolobus solfataricus minichromosome maintenance protein (MCM) is essential for DNA replication and initiation and elongation [10]. The archaeal replicative helicase generally exists as a homohexamer that shares homology with each of the six sequence-distinctive subunits in eukarya (MCM 2-7) [113, 136]. The archaeal MCM helicases from *Sulfolobus solfataricus* (*Sso*) and *Methanothermobacter thermoautotrophicus* (*Mth*) serve as simplified models for understanding mechanisms of ATP hydrolysis, DNA binding, and unwinding for more complex eukaryotic hexameric DNA replication helicases.

A near full-length crystal structure has recently been determined for *Sso*MCM [113]. Here, they characterized the various structural domains (Figure 3.1A-C). The N-domain has A, B, and C subdomains and is connected by the N-C linker to the C-domain. The A region imparts the snowflake like appearance, while the B region contains the Zn-binding motif and the C region has one β-hairpin. The C-domain consists of an α/β region which is connected to the α region by the α/β-α linker. The α/β region has the characteristic AAA^+ motifs including the Walker A and B sites and the arginine finger. The α region is made up of three alpha helices. The α/β-α linker is 47 residues in length and is composed of two long alpha helices connected with a loop. It interlocks with the N-C linker which is 40 residues in length and provides stability and connectivity. The C-terminus also contains a flexible winged helix domain.

We previously elucidated *Sso*MCM's unwinding mechanism and have designated it the steric exclusion and wrapping (SEW) model, in which the 5'-strand wraps around *Sso*MCM's exterior and interacts with the surface along discrete paths [123]. A double mutant (K323A/R440A) disrupts the ssDNA wrapping on the exterior surface through electrostatic interactions. In this study, we more fully characterized the interaction on the exterior surface by exploring the potential role of tyrosines that may base-stack with ssDNA providing stability in addition to electrostatics. Here, we have made three mutants (Y40A, Y61A, and R91A) in the N-domain's A subdomain and three mutants (Y516A, Y519A, and Y523A) in the C-domain's α/β-α linker region.

Surprisingly, we found that Y519A disrupts hexamer formation. This prompted us to characterize this mutant's unwinding, ATPase, DNA binding, secondary structure and solvent accessibility through hydrogen/deuterium exchange mass spectrometry. Y519A is present in

solution as dimeric, trimeric, and tetrameric species. It also has reduced unwinding and ATPase rates. EMSAs confirmed the oligomeric states present in solution also occur when binding forked DNA. CD spectra have shown that Y519A has an altered secondary structure compared to WT MCM, but increasing temperature toward physiological for *Sso* indicates an increase in secondary structure. In addition, we used hydrogen/deuterium exchange mass spectrometry (HDX-MS) to identify specific contacts with either the encircled or excluded DNA strands and probe the interfacial residues within the *Sso*MCM hexamer.

3.3 MATERIALS AND METHODS

3.3.1 Materials

ATP was obtained from Invitrogen. Optikinase was purchased from USB. All other materials were from commercial sources and were analytical grade or better. Helicase buffer is used in all unwinding and binding reactions and consists of 125 mM potassium acetate, 25 mM Tris acetate (pH 7.5) and 10 mM magnesium acetate.

3.3.2 DNA substrates

Oligonucleotides were purchased from IDT Corp and gel purified [144]. One forked substrate with a 36 nt duplex region was used consisting of two single stranded DNAs known as, 30 nt 3'-tail (5'-CACCTCTCCCTACGCTTCCCACCCACCCCGACCGGCATCTGCTATGGTACGCTGA GCGAGAGTAGC, and 50 nt 5'-tail (5'-CGATGAGAGCCGATGAGAGCCGATGAGAGC GAGTCGCATGGATCGTCTAGCCGGTCGGGGTG GGTGGGAAGCGTAGGGAGAGGTG). A 3'-tail only substrate was also used for H/DX in which 30 nt 3'-tail was annealed to 5'-tail-0 nt (5'-GCCGGTCGGGGTGGGTGGGAAGCGTAGGGAGAGGTG). [γ-^{32}P]ATP was purchased from Perkin Elmer (Waltham, MA) and Optikinase was used to ^{32}P label the 5′-end of 30 nt 3'-tail. 50 nt 5'-tail was added in a ratio of 1.2:1 and the either DNA duplex was heated at 95 °C for 5 min and then cooled to room temperature after turning off the heat block.

3.3.3 *Sso*MCM variants cloning and purification

*Sso*MCM Y40A, Y61A, R91A, Y516A, Y519A, and Y523A were created using a standard QuikChange protocol (Agilent, Santa Clara, CA) with KAPA HiFi DNA polymerase (KAPA Biosystems, Woburn, MA). Primer sequences are in Table 3.1. Mutations were confirmed using the DNA sequencing faculty at the University of Pittsburgh. Full-length wild-type (WT) and the point mutants and K323A/R440A *Sso*MCM were purified as previously described using 70 ^{o}C heat treatment as well as MonoQ, heparin, and gel filtration columns to isolate the hexameric species [111].

Table 3.1: DNA primer sequences

DNA Primers	Sequence (5'-3')
MCM Y40A FWD	5'-ATATATCGAGAGGATAAACGAGCTCGTAGCGGCTAGGAAAAAAAFTCT TATAATAGAATT
MCM Y40A REV	5'-AATTCTATTATAAGACTTTTTTTCCTAGCCGCTACGAGCTCGTTTATCCT CTCGATATAT
MCM Y61A FWD	5'-ATGTACTTTCGTTCAATGAAAATTTGGCTGCAGAGATAATAAATAATAC CAAAATTATTC
MCM Y61A REV	5'-GAATAATTTTGGTATTATTTATTATCTCTGCAGCCAAATTTTCATTGAAC GAAAGTACAT
MCM Y516A FWD	5'-ATATTATAGATATAGATACATTAAGAAAAGCGATCGCATATGCAAGGA AATACGTTACAC
MCM R91A FWD	5'-CTTGCAATTGGATCCTACATATCAAGCAGATATCGAAAAAGTTCATGTT AGAATTGTAGG
MCM R91A REV	5'-CTACAATTCTAACATGAACTTTTTCGATATCTGCTTGATATGTAGGATC CAATTGCAAG
MCM Y516A REV	5'-GTGTAACGTATTTCCTTGCATATGCGATCGCTTTTCTTAATGTATCTATA TCTATAATAT
MCM Y519A FWD	5'-ATTATAGATATAGATACATTAAGAAAATATATAGCAGCTGCAAGGAAA TACGTTACACCA
MCM Y519A REV	5'-TGGTGTAACGTATTTCCTTGCAGCTGCTATATATTTTCTTAATGTATCTA TATCTATAAT
MCM Y523A FWD	5'-AGATACATTAAGAAAATATATAGCATACGCAAGGAAAGCCGTTACACC AAAAATTACTAG
MCM Y523A REV	5'-CTAGTAATTTTTGGTGTAACGGCTTTCCTTGCGTATGCTATATATTTTCT TAATGTATCT

3.3.4 Analytical gel filtration chromatography

Analytical gel filtration chromatography was performed as previously described [163]. Briefly, the analytical gel filtration column (Superdex 200 10/30) was calibrated with protein standards including: thyroglobulin (669 kDa), catalase (250 kDa), conalbumin (75 kDa), myoglobin (17.6 kDa), and vitamin B12 (1.2 kDa). *Sso*MCM variants containing ~14 nmoles of protein, supplemented with vitamin B12 as an internal standard, were applied to the column at 0.1 mL/min in buffer containing 100 mM NaCl, 25 mM Tris-HCl pH 7.5, and 5 mM β-mercaptoethanol.

3.3.5 Unwinding assays

Unwinding assays were performed as previously described [123]. Briefly, *Sso*MCM variants were incubated with 15 nM 5'-radiolabeled DNA for 5 min at 60 ºC and initiated with ATP. Reactions were quenched with an equal volume of glycerol quench (0.5% SDS w/v, 50% v/v glycerol, 0.1% w/v bromophenol blue, 100 mM EDTA pH 8.0, and 300 nM trap ssDNA. Quenched reactions were stored on ice until the addition of 4 µL Proteinase K 20 mg/mL (Thermo Fisher). Samples were incubated at 37 ºC for 1 hr to allow for digestion. DNA was resolved on the indicated percentage acrylamide gels with 0.1% SDS to remove any remaining bound protein. Gels were then phosphorimaged.

3.3.6 ATPase assays

*Sso*MCM WT or Y519A was incubated in the absence or presence of unlabeled forked DNA as previously described [97]. Briefly, 30 µL reactions were incubated at 60 ºC for 5 min and ATP was added to initiate the reaction. Samples were quenched at 5, 10, and 15 min after initiation in equal volumes of 0.7 M formic acid. A total of 0.8 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), and then phosphorimaged.

3.3.7 Electrophoretic mobility shift assays

EMSAs were performed as previously described [97]. *Sso*MCM variants were incubated with ^{32}P-labeled forked DNA in helicase buffer. Briefly, 10 µL reactions were incubated for 10 min at 60 ºC to promote native thermodynamic complex formation and 2 µL of loading buffer (30% v/v glycerol) was added prior to being resolved on 5% native polyacrylamide gels. The gels were incubated on a rocker in fixing solution (70% ddH_20, 20% methanol, and 10% glacial acetic acid all v/v) for 20 min. The gels were then dried for 1 hr and phosphorimaged. Four regions were quantified, unbound DNA, hexamer (bound region at top of gel), non-hexamer (bound region between hexamer and unbound), and bound (non-hexamer and hexamer). Both bound and hexamer were fit to the Hill equation, equation (1)

$$Y = \frac{A \times [MCM]^n}{K_d^{\,n} \times [MCM]^n} \quad (1)$$

where A is the amplitude, K_d is the dissociation constant for the specified species and n is the Hill coefficient.

3.3.8 Circular dichroism spectroscopy

Circular dichroism spectroscopy was performed as previously published [164]. Measurements were performed on an Olis DSM17 CD spectrophotometer (Bogart, GA) in 0.1 cm path-length quartz cuvettes. 300 µL reactions were prepared of *Sso*MCM variants in helicase buffer in the absence and presence of forked DNA. Wavelength scans were collected at 20 ºC from 200-260 nm with 5 s integration times, 1 nm step size, and 2 nm bandwith. Spectra were corrected for a buffer blank collected in the same cell and residual baseline molar ellipticity at 260 nm. Raw CD data were smoothed by the Savitzky–Golay method as implemented in GraphPad Prism (GraphPad Software, La Jolla, CA). Molar ellipticity (θ) was calculated with equation (2)

$$[\theta] = \left(\frac{\theta_{samp} - \theta_{blank} - \theta_{avg}}{10 \times [M] \times 0.1\ cm \times \#_{res}}\right) \times 10^{-3} \quad (2)$$

where samp, blank, and avg are the molar ellipticities for the sample, blank, and avg is the average molar ellipticity which is calculated by equation (3)

$$[\theta_{avg}] = \theta_{255-260\ samp} - \theta_{255-260\ blank} \quad (3)$$

where 255-260 is the average molar ellipticity from 255-260 nm for either the sample or the blank. Thermal denaturation experiments were carried out by monitoring molar ellipticity at both 222 and 261 nm from 20–95 °C in 3 °C intervals with a 0.5 °C dead band, 5 s averaging time, and 2 min equilibration time at each temperature. Molar ellipticity was calculated similarly with equation (4)

$$[\theta] = \left(\frac{\theta_{samp\ @\ 222} - \theta_{blank\ @\ 222} - \theta_{blank\ @\ 260}}{10 \times [M] \times 0.1\ cm \times \#_{res}}\right) \times 10^{-3} \quad (4)$$

where the molar ellipticities are monitored at either 222 or 260 nm for either the sample or buffer blank. All parameters reported from CD measurements are the average of two or three independent experiments.

3.3.9 Hydrogen/deuterium exchange

Hydrogen/deuterium exchange (HDX) was performed as previously published by the Marshall lab [165]. Stock solutions of ~9 μM *Sso*MCM WT prepared in helicase buffer either alone, or in the presence of 3'-tail DNA or forked DNA. We also analyzed Y519A, K323A/R440A alone or with 3' tail DNA. *Sso*MCM and DNA are in an equivalent molar ratio (1:1). Similar buffer conditions were applied for the preparation of the corresponding D_2O buffers. Hydrogen/deuterium exchange (HDX) experiments were optimized and automated with an HTC Pal autosampler (Eksigent Technologies, Dublin, CA). 5 μL of *Sso*MCM was mixed with 45 μL of corresponding buffer in D_2O to initiate each H/D exchange period. For the blank control, the initial dilution was made in H_2O buffer. Reactions were performed in triplicate at 1-2 °C to reduce back-exchange for HDX incubation periods of 0.5, 1, 2, 4, 8, 15, 30, 60, 120, and 240 min, each followed by acid quench and proteolysis for 3 min. Each sample was quenched by rapid mixing with 25 μL of 200 mM

TCEP, 8 M urea in 1.0% formic acid, and 25 μL of a five-fold dilution of saturated protease type XIII (Sigma-Aldrich) in 1.0% formic acid (final pH ~2.3). On-line LC electrospray ionization 14.5 tesla FT-ICR MS with an LTQ Velos front end yielded high mass resolving power ($m/\Delta m_{50\%}$ = 200,000 at m/z 400). Data were analyzed by an in-house Predator software package [166]. An equivalent molar ratio (1:1) of *Sso*MCM and DNA for HDX. The average relative deuterium uptake difference (ARDD) was calculated by equation (5)

$$ARDD = \sum_i \frac{A(t_i) - B(t_i)}{A(t_i)} \quad (5)$$

where A is the deuterium uptake for sample A at a specified time (t_i) and B is the deuterium uptake for sample B at a specified time (t_i) [167].

3.4 RESULTS

3.4.1 *Sso*MCM mutants

Previously, we characterized an interaction between the 5'-tail of a forked DNA substrate and the exterior surface of *Sso*MCM [123]. In order to more completely characterize the binding path on the exterior surface, we created tyrosine mutants that would potentially disrupt base stacking interactions with ssDNA. Using the near full-length crystal structure [113] as a guide (Figure 3.1A-C), we identified surface tyrosine residues that are located in the A domain and the α/β-α linker region. Specifically, five surface tyrosine to alanine mutants (Y40A, Y61A, Y516A, Y519A and Y523A) were created. Additionally, another basic residue R91A also located in the A domain, was probed. Figure 3.1D shows a ClustalW2 alignment in which Y61, R91, Y516, and Y519 are conserved with eukaryotes, while Y40 and Y523 are not [168, 169].

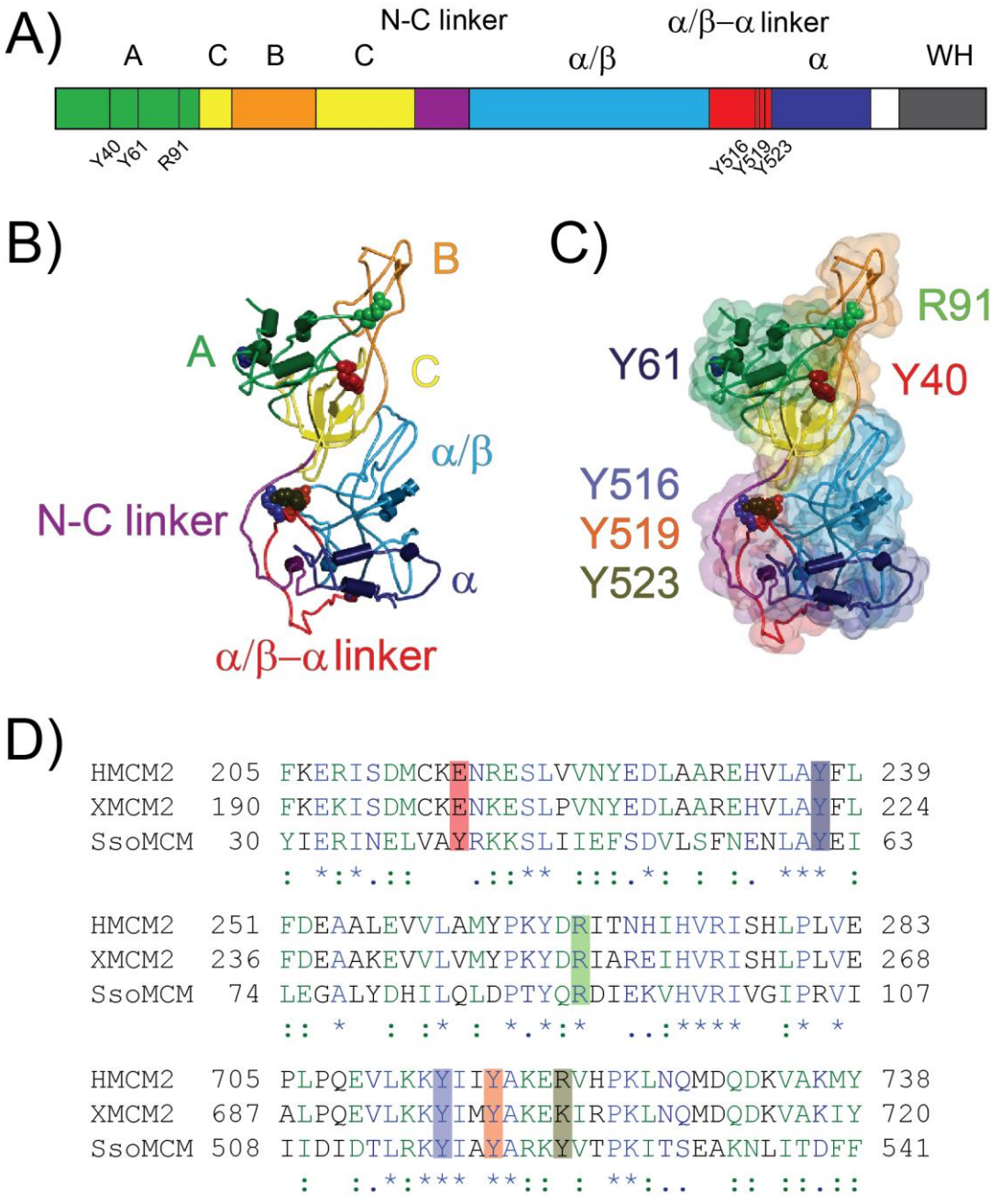


Figure 3.1: MCM domains, mutant locations, and sequence homology.
A) *Sso*MCM domain organization, where the A subdomain is green, the B subdomain is orange, the C subdomain is yellow, the N-C linker is purple, the α/β domain is light blue, the α/β-α linker domain is red, the α domain is dark blue, the non-crystallized portion of the α domain is white, and the winged helix (WH) domain is grey. The six new mutants are shown where Y40, Y61, and R91 are located in the A domain and Y516, Y519 and Y523 are in the α/β-α linker domain. **B)** The color scheme and labels from A) are used again for the monomeric near-full length crystal structure of *Sso*MCM (PDB ID: 3E9V) with each mutated residue represented as a CPK space filling model. **C)** The same monomer from B) with a CPK surface colored by domain and the mutated residues are labeled with Y40 colored red, Y61 colored dark blue, R91 colored green, Y516 colored purple, Y519 colored orange, and Y523 colored brown. **D)** ClustalW2 alignment of hMCM2, XMCM2, and SsoMCM where h – human, X – *Xenopus laevis*, and *Sso* – *Sulfolobus solfataricus.* Y61, R91, Y516, and Y519 are conserved residues, while Y40 and Y523 are not conserved.

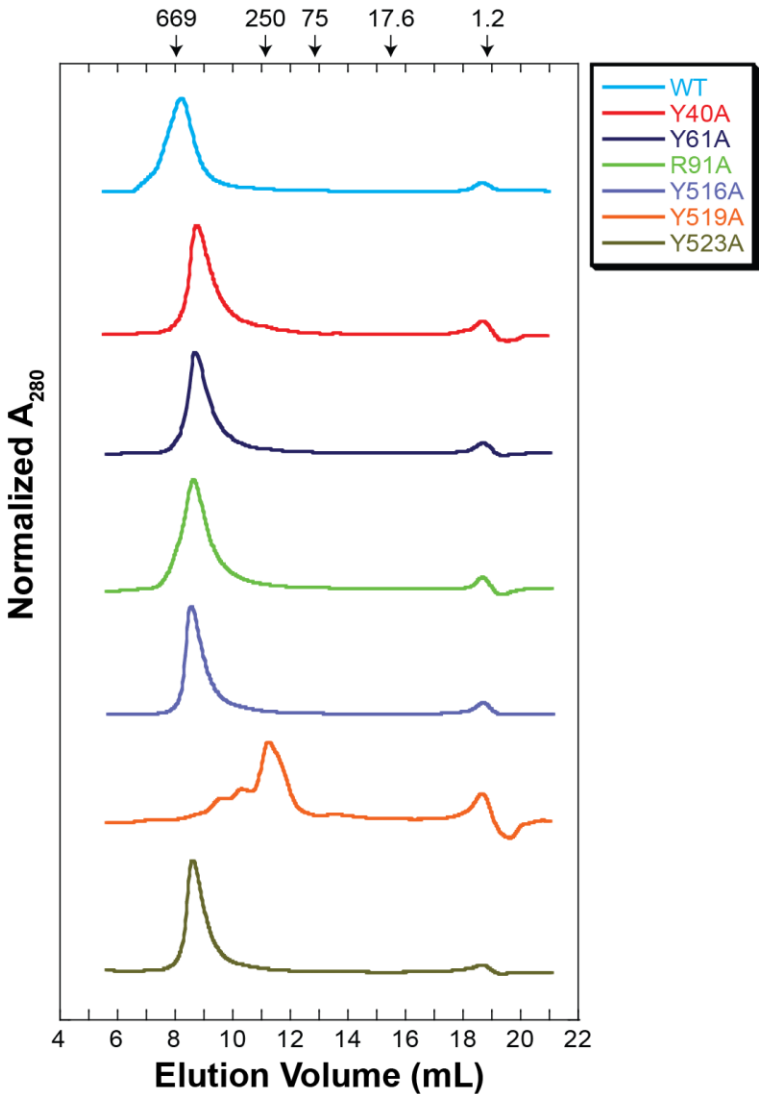


Figure 3.2: *Sso*MCM Y519A elutes as non-hexamer.
Normalized A_{280} (milli absorbance units – mAu) versus elution volume (mL). Seven *Sso*MCM variants (WT, Y40A, Y61A, R91A, Y516A, Y519A, and Y523A) were analyzed by analytical gel filtration chromatography. WT is cyan, Y40 is red, Y61 is dark blue, R91 is green, Y516 is purple, Y519 is orange, and Y523 is brown. Standard molecular weights (in kDa) are shown at the top with arrows pointing to their corresponding elution volumes. Vitamin B12 (1.2 kDa) was added to each *Sso*MCM variant to account for any drift.

3.4.2 *Sso*MCM variants are hexameric in solution except for Y519A

WT *Sso*MCM forms a hexamer in solution even in the absence of DNA and ATP [103, 111]. To determine whether the mutants also primarily form hexamers, we used analytical gel filtration chromatography. The analytical gel filtration column was calibrated with protein standards

including: thyroglobulin (669 kDa), catalase (250 kDa), conalbumin (75 kDa), myoglobin (17.6 kDa), and vitamin B12 (1.2 kDa). The molecular weights and their corresponding elution volumes are shown at the top of Figure 3.2A. *Sso*MCM monomer is 77.4 kDa, while the homohexamer is 464.4 kDa. WT and all point mutants except for Y519A elute between 8 and 9 mL indicating hexameric species. Y519A, however has three overlapping peaks eluting after the hexamer and based on size are likely dimer, trimer, and tetramer.

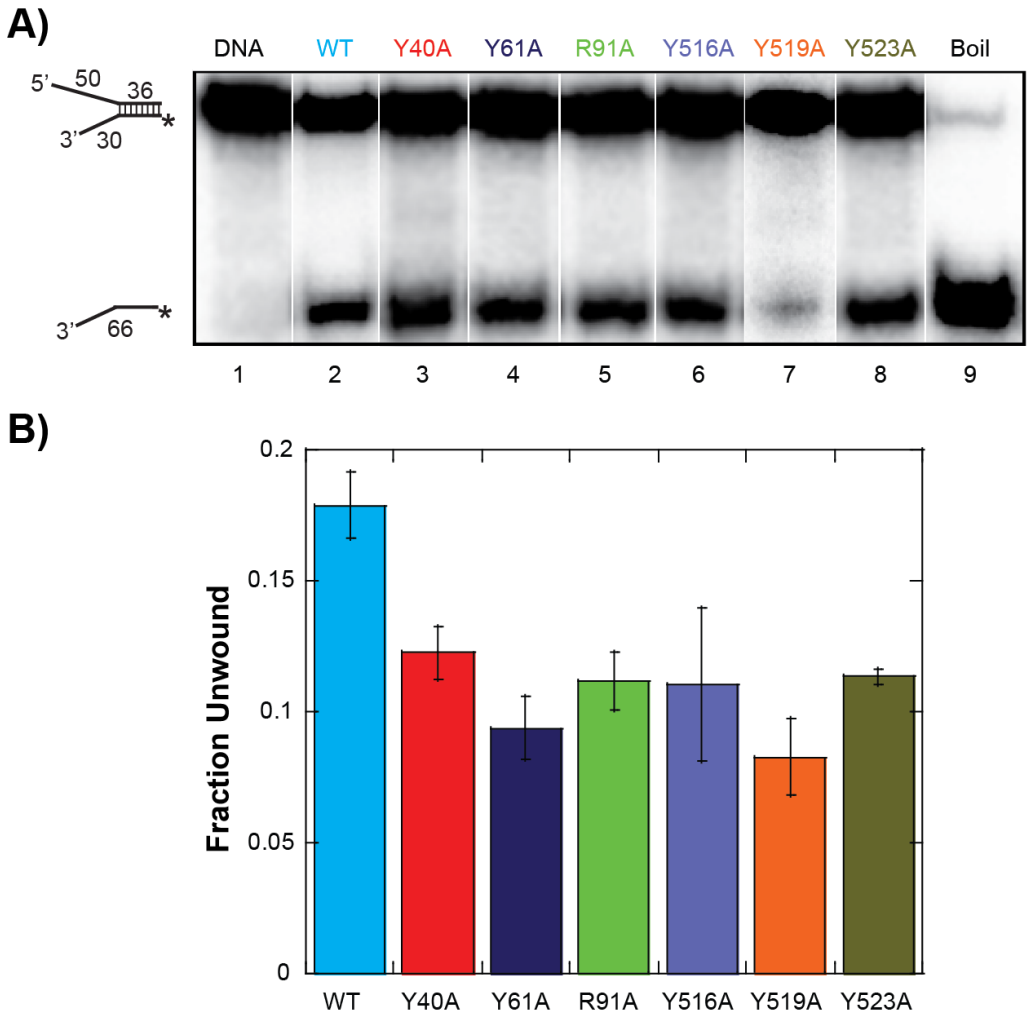


Figure 3.3: Unwinding comparison for the *Sso*MCM variants.
A) Forked DNA unwinding by *Sso*MCM. *Sso*MCM variants were 2 µM and all reactions contain 15 nM ^{32}P-forked DNA (3'-tail-30 nt annealed to 5'-tail-30 nt). Unwinding reactions were for 5 minutes at 60 ºC prior to quenching and Proteinase K treatment as specified in the 'Materials and Methods.' Bands were resolved on 14% polyacrylamide gels containing 0.1% SDS and phosphorimaged. DNA alone is shown in lane 1 and boiled DNA is in lane 9. **B)** Quantification of fraction unwound of A). Error bars represent the standard error from at least three independent unwinding experiments.

3.4.3 Unwinding decreased for Y519A compared to WT

We then tested the unwinding ability for each of the mutants. Figure 3.3 shows that the fraction unwound for WT is the highest (0.18 ± 0.01), while the mutants all show decreased unwinding compared to WT, with Y519A showing the lowest fraction unwound (0.08 ± 0.01). We compared the unwinding kinetics for 2 μM WT or Y519A in Figure 3.4A and B. We also titrated Y519A (1, 4, and 8 μM) to evaluate potential changes in equilibrium oligomeric states. Y519A exhibits reduced unwinding compared to WT likely due to its non-hexameric nature (Figure 3.4C). WT unwinds 0.022 min^{-1}, while Y519A unwinds three to four-fold slower (0.0067 min^{-1}). There is essentially no increase in rate with increasing concentration of Y519A suggesting that the oligomeric state may be saturating.

3.4.4 ATP hydrolysis rate decreased for Y519A both in the absence and presence of DNA

We have previously characterized the ATP hydrolysis rate of *Sso*MCM in the absence and presence of DNA [97], and our results were in agreement with previous work [111]. In the presence of DNA, *Sso*MCM's ATPase rate is stimulated approximately two-fold, allowing for unwinding activation or translocation [97, 111, 170]. Here, we compared WT and Y519A at 1, 2, 4, and 8 μM both in the absence and presence of DNA (Figure 3.5A). Representative TLC plates for ATPase experiments are shown in Figure 3.5B and C for WT and Y519A, respectively. *Sso*MCM WT ATP hydrolysis rate is enhanced in the presence of DNA, as previously shown, nearly two-fold (1.9 ± 0.1) on average. Y519A's ATP hydrolysis rate in the presence of DNA is enhanced less (1.7 ± 0.1 fold). Overall, the ATP hydrolysis rate for WT is approximately two fold greater than Y519A either in the absence or presence of DNA. It is also likely that Y519A is minimally a dimer which would allow for a complete ATP hydrolysis site, due to the ATP hydrolysis site located at the interface between two subunits [112, 171].

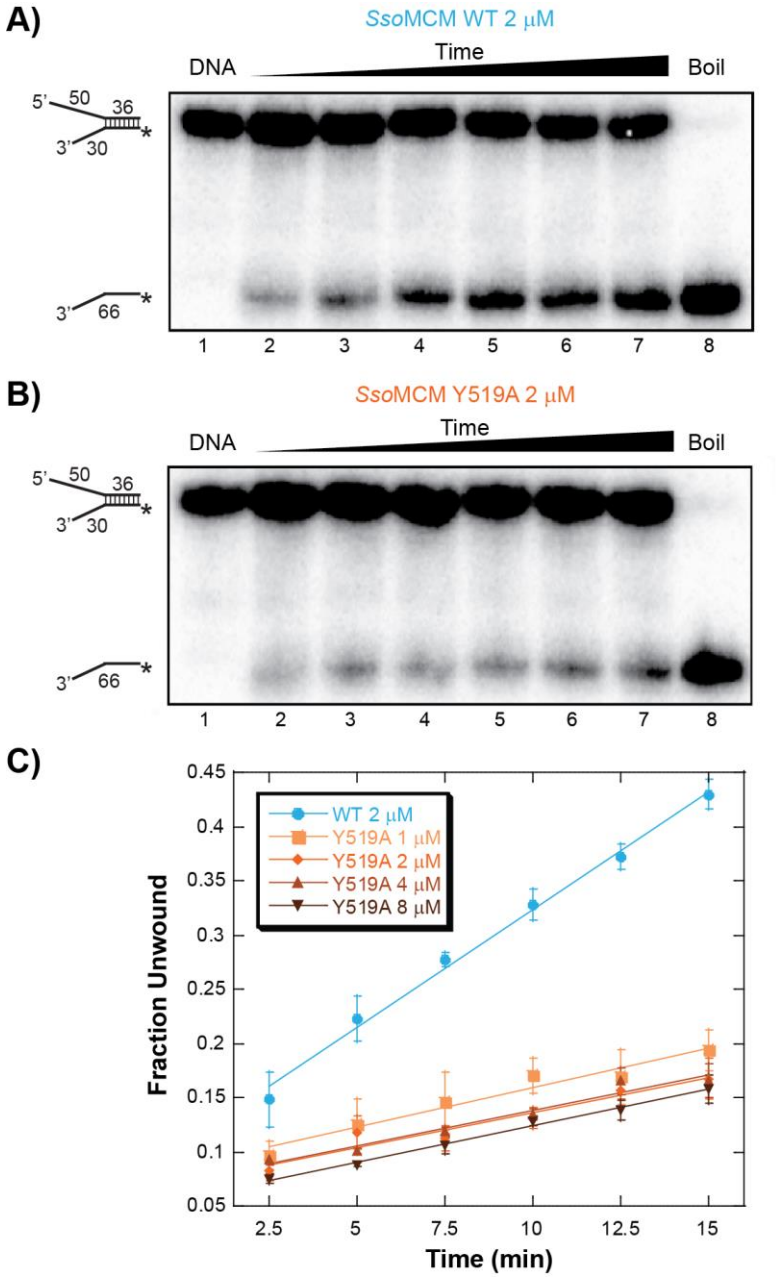


Figure 3.4: Unwinding time course comparing WT and Y519A.
A) Representative forked DNA unwinding by *Sso*MCM WT. *Sso*MCM WT (2 μM) was incubated in 15 nM ^{32}P-forked DNA (3'-tail-30 nt annealed to 5'-tail-30 nt). Unwinding reactions were quenched at 2.5, 5, 7.5, 10, 12.5, and 15 minutes at 60 ºC prior to Proteinase K treatment as specified in the 'Materials and Methods.' Bands were resolved on 14% polyacrylamide gels containing 0.1% SDS and phosphorimaged. DNA alone is shown in the first lane and boiled DNA is shown in the last lane. **B)** Representative forked unwinding by *Sso*MCM Y519A (2 μM) as A). **C)** Quantification of fraction unwound by WT (2 μM) and Y519A (1, 2, 4, and 8 μM). The data was fitted with a linear equation and the rates were calculated. Error bars represent the standard error from at least three independent unwinding experiments.

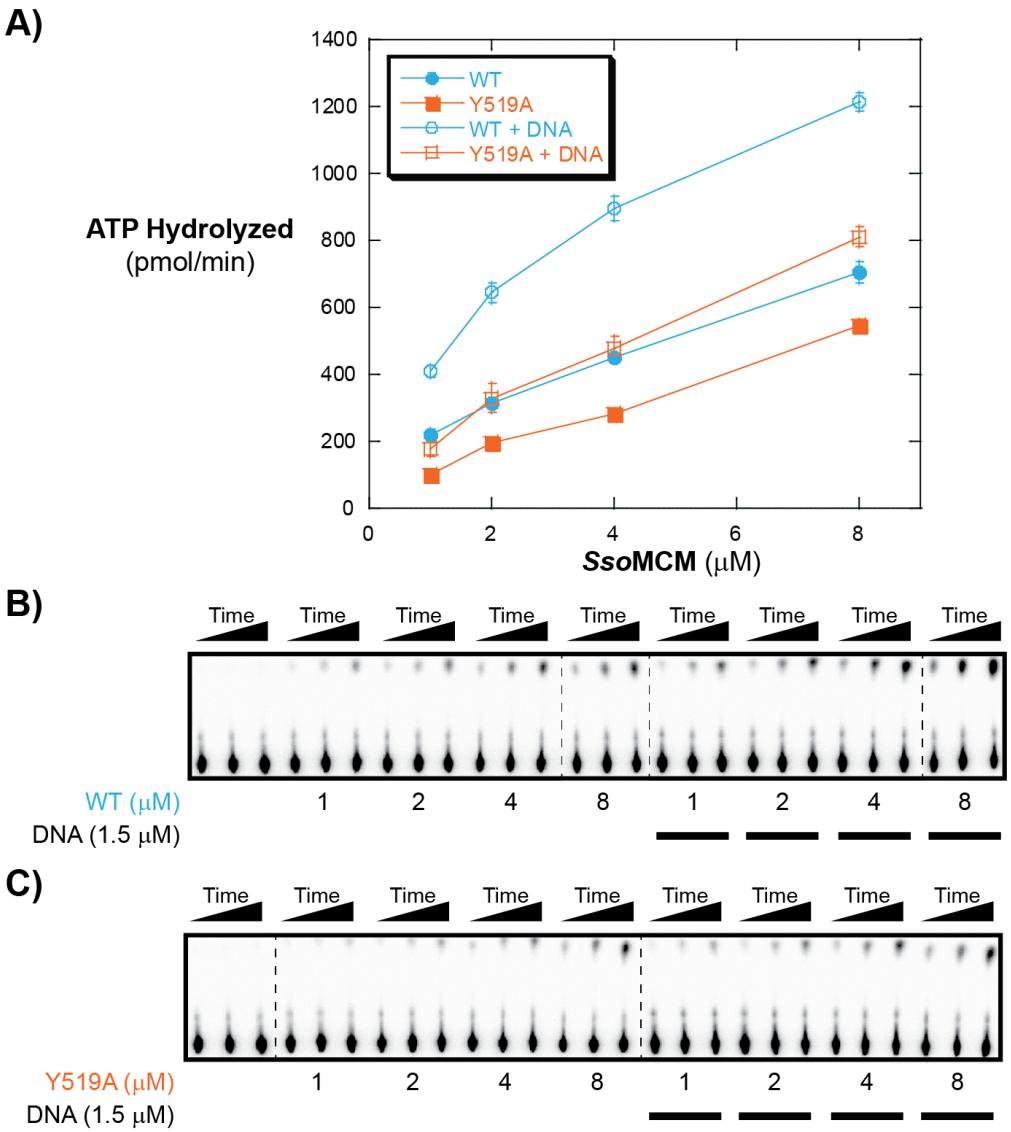


Figure 3.5: ATPase rates of WT and Y519A in the absence and presence of DNA.
A) ATP hydrolyzed (pmol/min) versus *Sso*MCM (μM) at 60 ºC in the absence and presence of DNA. *Sso*MCM WT or *Sso*MCM (1, 2, 4, or 8 μM) were incubated in the absence or presence of 1.5 μM forked DNA (3'-tail-30 nt annealed to 5'-tail-30 nt). The ATPase rate resultant from *Sso*MCM WT in the absence and presence of 1.5 μM DNA is shown with closed and open cyan circles, respectively and *Sso*MCM Y519A in the absence and presence of 1.5 μM DNA is shown with closed and open orange squares, respectively. Error bars represent the standard error from at least three independent ATPase experiments. **B)** Representative TLC plates for WT. **C)** Representative TLC plates for Y519A. The ATP hydrolysis rate shown in A) was calculated as a function of time (5, 10, and 15 minutes) shown in B) and C).

3.4.5 EMSAs show differential oligomeric states for WT and Y519A

On the basis of the ATPase results, we wanted to observe the oligomeric state of Y519A when binding DNA. Electrophoretic mobility shift assays provided a means to identify *Sso*MCM binding in various oligomeric states. Figure 3.6A and B shows *Sso*MCM WT and Y519A binding forked DNA, respectively. We completely quantified each gel with four fractional binding states; unbound, non-hexamer, hexamer and bound. We classified hexamer binding as the region at the top of the gel, unbound is the region of DNA alone, non-hexamer is the region between hexamer and unbound, while bound is both non-hexamer and hexamer binding. In Figure 3.6A, there are two distinct species notated by the arrows, while in Figure 3.6B there are four distinct species. Figures 5.2 and 5.5 indicate Y519A is minimally a dimer. Therefore, we classified the three non-hexamer bands as dimer, trimer, and tetramer. *Sso*MCM WT's non-hexamer band is in the same location as the trimer band for Y519A. This indicates that Y519A assembles either as three dimers or two trimers, while WT assembles as two trimers. The quantifications for WT and Y519A are shown in Figure 3.6C and D, respectively. Surprisingly, the hexamer K_d for WT (1.7 ± 0.2 μM) and Y519A (1.6 ± 0.4 μM) were essentially the same. The Hill factors for hexamer were 1.3 ± 0.2 for WT and 1.3 ± 0.2 for Y519A again indicative of positive cooperativity. The bound K_d value differed at least 2-fold between WT (0.49 ± 0.02 μM) and Y519A (0.20 ± 0.01 μM). The Hill factors for bound were 2.5 ± 0.2 for WT and 2.0 ± 0.2 for Y519A; both indicative of positive cooperativity. The calculated value for WT binding is comparable to previous results acquired through fluorescence anisotropy (0.45 ± 0.08 μM) [123]. Y519A binding is 2.4-fold tighter than the corresponding WT binding. The altered α/β-α region and/or its interactions with the intertwined N-C linker region may enhance DNA binding or simply dimer affinity for DNA is higher than trimer.

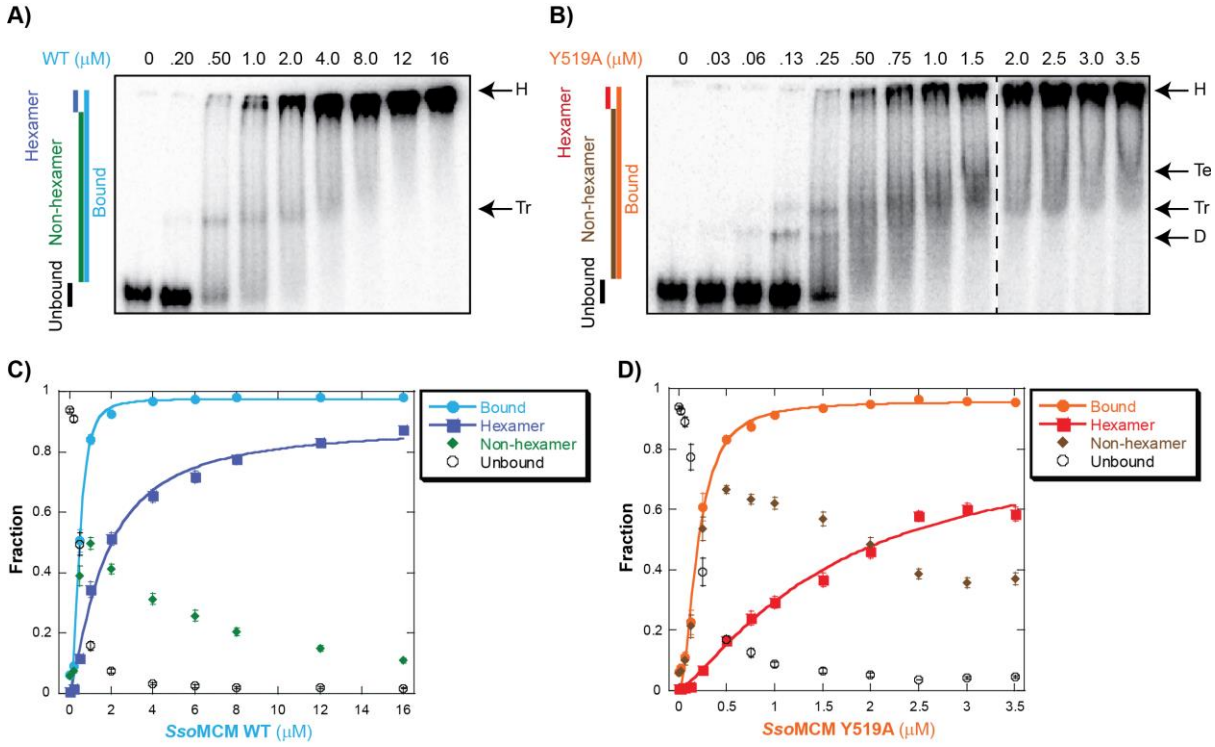


Figure 3.6: *Sso*MCM EMSAs – WT versus Y519A.
A) Representative EMSA of *Sso*MCM WT titrated in the presence of 15 nM forked DNA with the following concentrations (0.20, 0.50, 1.0, 2.0, 4.0, 8.0, 12, and 16 μM). Bars at the left show the various species where unbound DNA is black, non-hexamer is green, hexamer is purple, and bound is cyan. Arrows at the right show the locations of trimer (Tr) and hexamer (H). **B)** Representative *Sso*MCM Y519A titrated in the presence of 15 nM forked DNA with the following concentrations (0.03, 0.06, 0.13, 0.25, 0.50, 0.75, 1.0, 1.5, 2.0, 2.5, 3.0, and 3.5 μM). Bars at the left show the various species where unbound DNA is black, non-hexamer is brown, hexamer is red, and bound is orange. Arrows at the right show the locations of dimer (D), trimer (Tr), tetramer (Te), and hexamer (H). **C)** *Sso*MCM WT quantifications of the fraction bound closed cyan circles, hexamer closed purple squares, non-hexamer closed green diamonds, and unbound open black circles. **D)** *Sso*MCM Y519A quantifications of the fraction bound closed orange circles, hexamer closed red squares, non-hexamer closed brown diamonds, and unbound open black circles. For both C) and D) Hill equation fits are shown for bound and hexameric species. Error bars represent the standard error from at least three independent EMSA experiments.

3.4.6 CD spectra show that *Sso*MCM Y519A and K323A/R440A show different characteristics than WT

Circular dichroism (CD) provides relative secondary structure information for proteins. Alpha helices, beta sheets and random coils exhibit different molar ellipticity values. We analyzed *Sso*MCM WT, K323A/R440A, and Y519A both in the absence and presence of forked DNA. A wavelength scan from 200-260 nm for each condition is shown in Figure 3.7A. Here, both WT and K323A/R440A in the absence and presence of DNA have essentially the same spectra, while

Y519A deviates both to higher wavelengths and a decreased overall molar ellipticity. This indicates that Y519A has more secondary structure compared to WT and K323A/R440A. Shifts to longer wavelengths indicate more alpha helical character. Mutating tyrosine to alanine will extend the alpha helix because of alanine's propensity for forming alpha helices [172] or possibly because formation of the hexamer from three dimers can also result in increased secondary structure [173, 174]. Thermal melts were also performed to assess the protein stability. Figure 3.7B shows that DNA does not appear to increase stability for any of the variants tested. K323A/R440A has a lower thermostability than WT. Y519A displays an interesting thermal melt profile in which increasing the temperature to a physiological range (~70-72 ºC for archaea) induces enhanced secondary structure prior to melting at a similar temperature as WT, for this reason we favor the latter interpretation of hexamer formation from three dimers prior to melting as WT.

3.4.7 Hydrogen/deuterium exchange comparison for *Sso*MCM WT, K323A/R440A, and Y519A

We then utilized hydrogen/deuterium exchange mass spectroscopy (HDX-MS) to study *Sso*MCM hexamer and its DNA binding locations. HDX-MS yields solvent accessibility data for proteins. This approach successfully mapped the conformational dynamics of another hexameric helicase of dsRNA bacteriophage Φ8's P4 [175, 176]. In a collaboration with Alan Marshall's laboratory at Florida State University, we analyzed WT *Sso*MCM by HDX. Figure 3.8 shows that after digestion by protease XIII, there was 98.4% coverage for WT *Sso*MCM. This extraordinary coverage was because of the extremely high field instrument used. We probed WT bound to 3'-tail only DNA and forked DNA. Figure 3.9 shows the difference between each DNA binding state over time averaged over every residue. Even though the 5'-tail binds the exterior surface stably, the data is difficult to interpret due to each subunit being identical and the ssDNA interacting with only 1-2 subunits. This results in very small differences that may be averaged away. To combat this averaging, we analyzed the data based on work by the Engen lab [177]. Instead, we utilized absolute deuterium values for each peptide and their corresponding standard error values. The fraction deuterated per peptide was calculated for 576 unique peptides. If the standard error is >0.5, those time points were omitted. This yielded 497 peptides, which was further pared by averaging

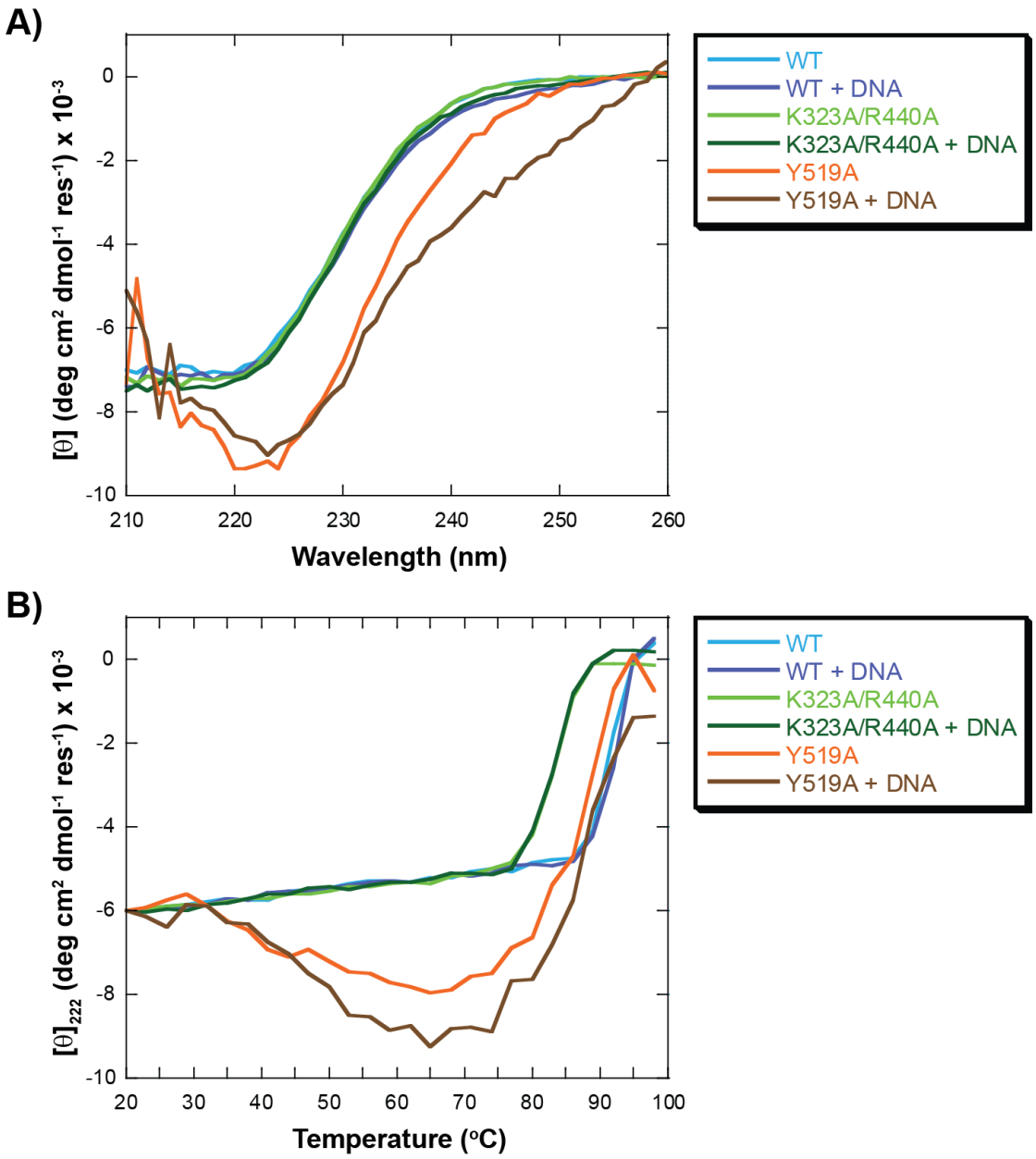


Figure 3.7: CD Spectra for WT, K323A/R440A, and Y519A in the absence and presence of DNA.
A) Molar ellipticity (deg cm^2 $dmol^{-1}$ res^{-1}) × 10^{-3} versus wavelength (nm) in the absence and presence of DNA for WT *Sso*MCM (4 µM) cyan and purple, respectively, K323A/R440A (4 µM) light green and dark green, respectively, and Y519A (2 µM) orange and brown, respectively. B) Molar ellipticity at 222 nm (deg cm^2 $dmol^{-1}$ res^{-1}) × 10^{-3} versus temperature (°C) in the absence and presence of DNA for WT *Sso*MCM (4 µM) cyan and purple, respectively, K323A/R440A (4 µM) light green and dark green, respectively, and Y519A (2 µM) orange and brown, respectively.

the fractions deuterated for peptides with multiple charge states. The remaining data was pared to 238 peptides by utilizing the latter nine time points (60, 120, 240, 480, 900, 1800, 3600, 7200, and 14400 seconds) only for peptides that contained data for each point. The 238 peptides were then plotted as butterfly plots in which one data set was plotted with positive values and a different data set was plotted with negative values. Regions that do not mirror one another have differences in deuterium uptake indicative of either structural changes or DNA binding. Figure 3.10A shows the 238 peptides used for analysis of *Sso*MCM in Figure 3.10, still exhibiting significant coverage. Figure 3.10B-D are the butterfly plots showing the relative fraction exchange for WT versus 3' DNA, WT versus 3' 5' (forked DNA), and 3' DNA versus 3' 5' (forked). In Figure 3.10B, comparing WT versus 3' DNA there are altered deuterium uptakes on peptides 56, 191, 193, 220, and 222 for which the corresponding residues are 183-194, 522-532, 522-534, 592-595, and 590-599, respectively. Residues 183-194, 522-532, 522-534, and 590-599 show reduced deuteration for the 3'-tail, while 592-595 displayed increased deuteration. The regions of reduced deuteration exhibit putative locations in which the 3'-tail binds *Sso*MCM. The increase in deuteration for 592-595 likely excludes these residues from participating in 3'-tail binding. Next in Figure 3.10C, we compared WT versus 3' 5' (forked) and found there are two additional peptides 6 (residues 30-33) and 153 (residues 425-440) that show reduced deuteration. The new regions of difference indicate putative binding locations for the 5'-tail. Peptide 153 includes R440 which we previously identified as a critical residue for 5'-tail binding. Peptide 6 includes Y30, conserved as an aromatic residue and R33 which is a conserved as a basic residue (Figure 3.1D). Figure 3.10D confirms the binding locations determined in the previous two plots. In order to visualize the binding locations, we mapped them to the crystal structure in Figure 3.11A and B. Here residues 183-194 interact within the central channel of the hexamer, indicative of loading onto the 3'-tail. Additionally, there are regions of altered deuteration near binding the ss-dsDNA junction (590-599). The other region (522-534) is located at the end of the α/β-α linker region and primarily in the α region of the C domain. Both of these regions have not been well characterized in the literature, but here we show that the α/β-α linker and α regions are likely altered upon MCM loading onto DNA allowing for a change in deuteration.

Finally, in the absence of DNA, we compared WT to both Y519A and K323A/R440A using average relative deuterium uptake difference (ARDD). Comparisons between WT and the corresponding mutated MCM are shown in Figure 5.12. Y519A (Figure 5.12A-B) shows increased

deuteration for residues 36-46, 128-139, 327-337, and 510-516 and decreased deuteration at residues 59-74, 109-124, 168-186, 214-220, and 318-323. K323A/R440A shows much more increased deuteration 36-46, 59-74, 128-138, 224-237, 396-404, 510-518, and 566-578 and decreased deuteration at 138-148, 262-269, and 462-474. There is some overlap in increased deuteration including residues 36-46 and 128-138 located in the A subdomain and B/C subdomains, respectively. Finally, we see an increase in deuteration for Y519A for 59-74, but a decrease in deuteration for the same residues in the K323A/R440A mutant.

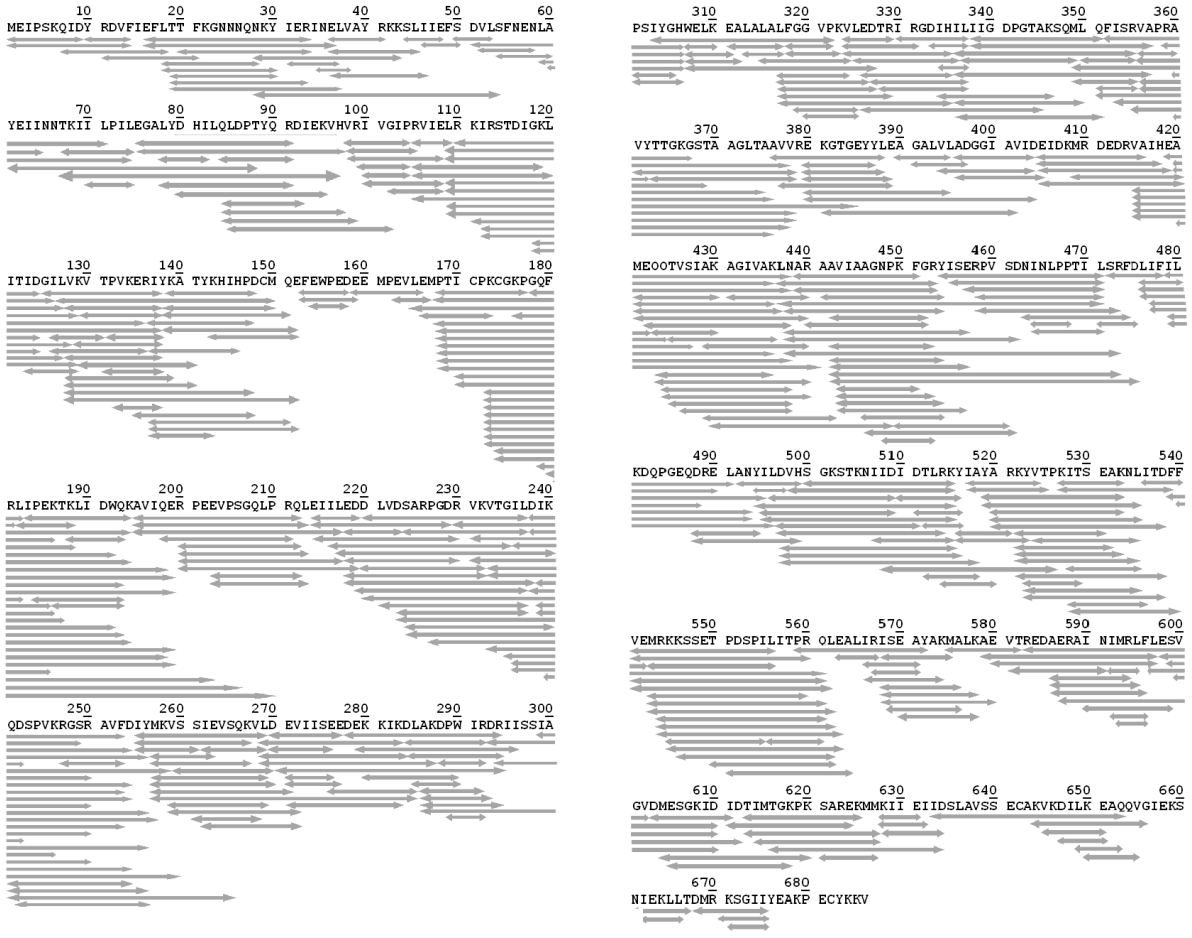


Figure 3.8: Protein sequence coverage for *Sso*MCM WT H/D exchange experiments.
*Sso*MCM WT was digested with protease XIII. We characterized 576 peptides (5-30 residues in length), providing 98.4% sequence coverage.

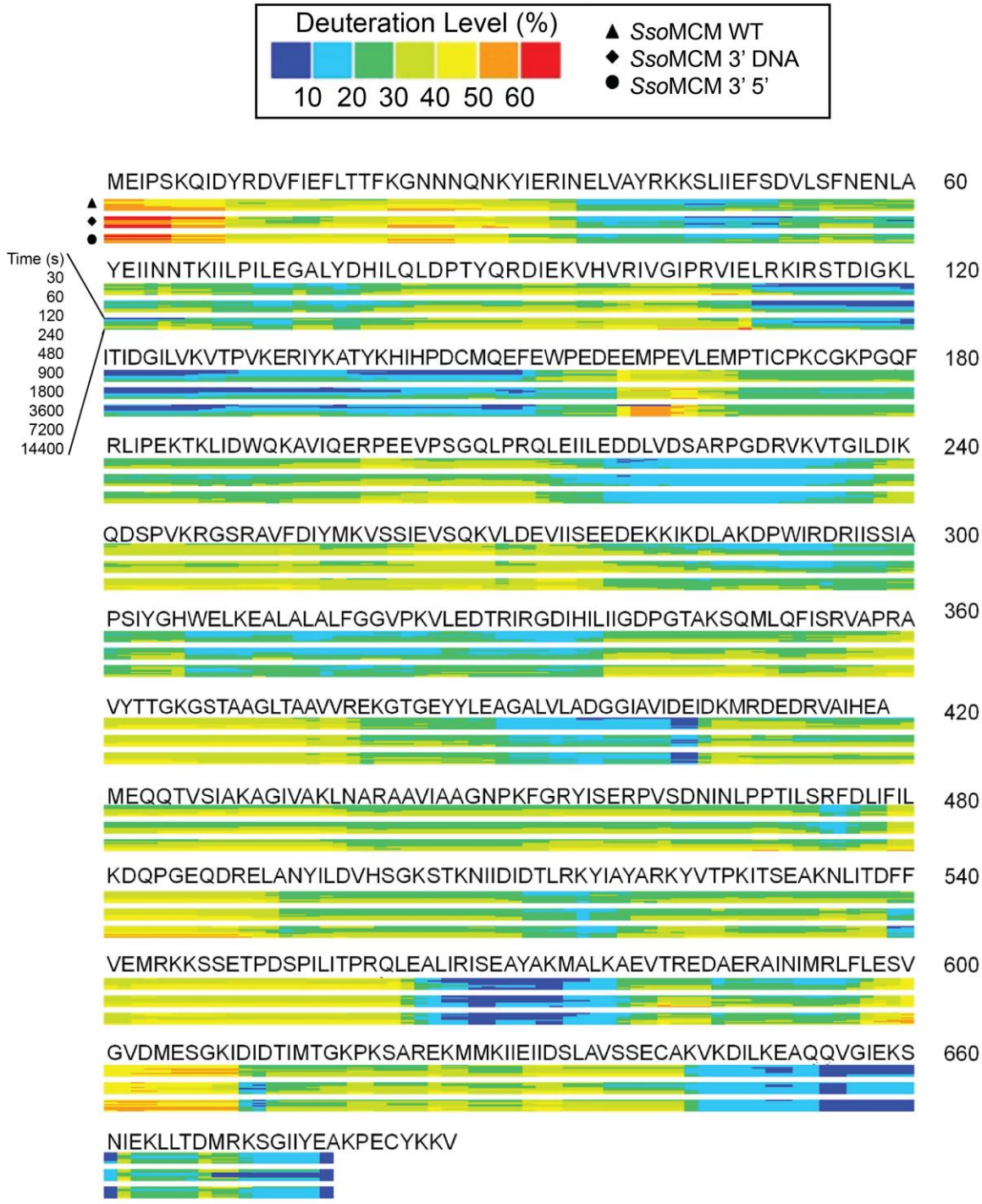


Figure 3.9: Heat map comparing deuteration levels for *Sso*MCM WT, bound to 3'-tail DNA, and bound to forked DNA.

The deuterium uptake level (in percentage) is calculated by averaging the deuteration of the peptides that contain each amino acid. Changes in deuterium uptake levels indicate conformational changes upon binding DNA.

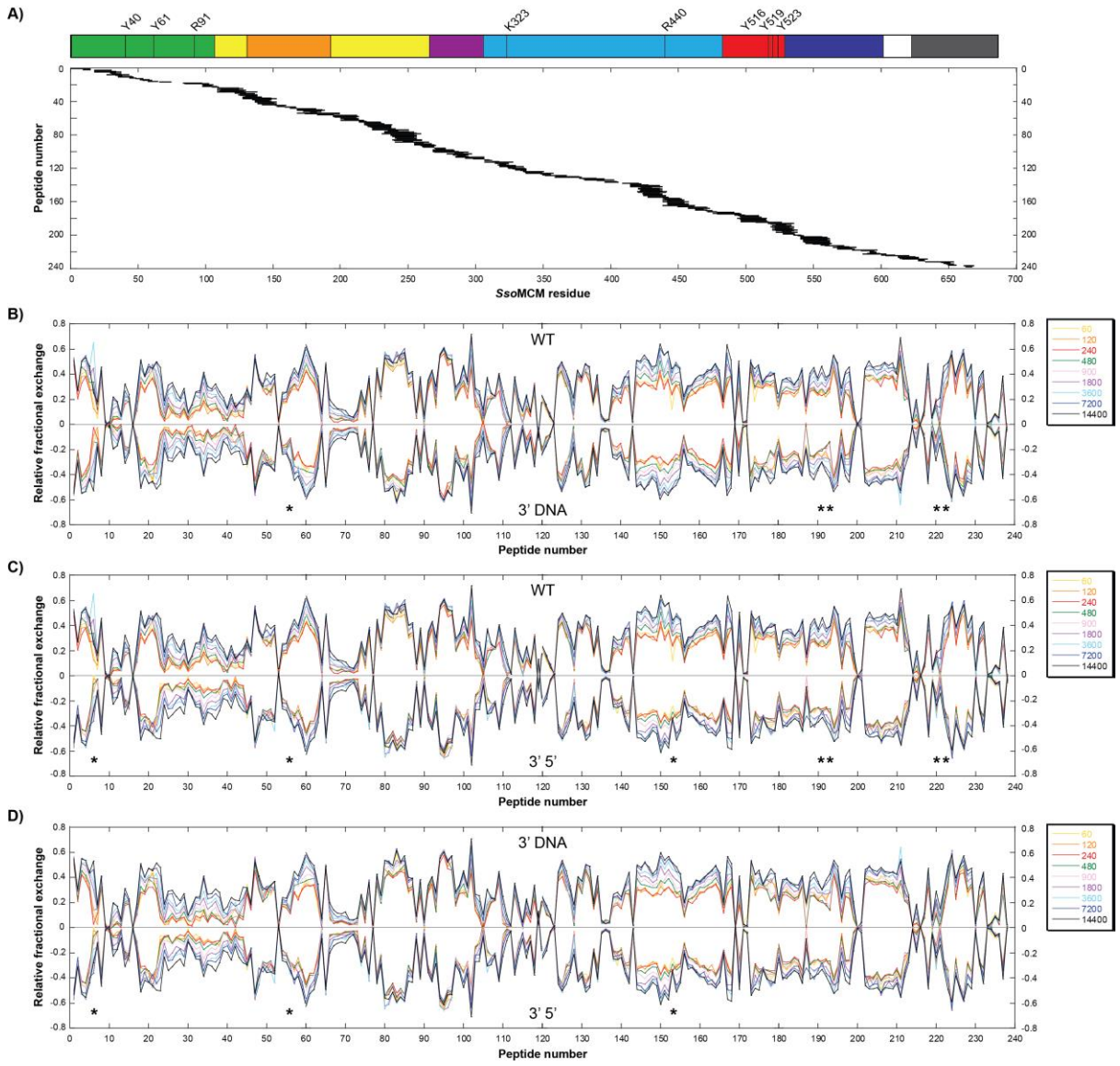


Figure 3.10: Relative fraction comparison for 238 peptides
A) Linear MCM at the top colored as Figure 3.1. Plot showing sequence coverage of 238 peptides for *Sso*MCM. **B)** Butterfly plot of WT (positive values) and 3' DNA (negative values) relative fractional exchange versus peptide number. **C)** Butterfly plot of WT (positive values) and 3' 5' (negative values) relative fractional exchange versus peptide number. **D)** Butterfly plot of 3' DNA (positive values) and 3' 5' (negative values) relative fractional exchange versus peptide number. For B-D nine time points are used 60 (yellow), 120 (orange), 240 (red), 480 (green), 900 (lavender), 1800 (purple), 3600 (cyan), 7200 (blue), and 14400 (black) seconds.

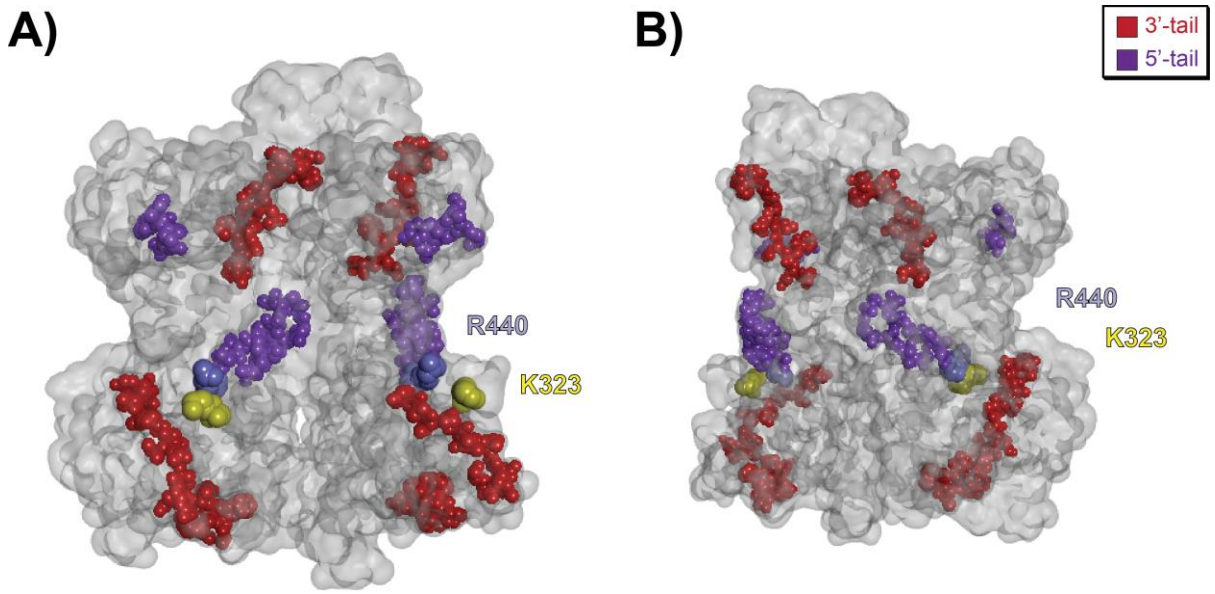


Figure 3.11: *Sso*MCM WT DNA binding sites by H/DX-MS.
A) Exterior view of two *Sso*MCM monomers displaying the binding sites of 3'-tail and 5'-tail. K323 and R440 are labeled in yellow and lavender CPK, respectively. The regions of altered deuteration for 3'-tail and 5'-tail binding are shown with red and purple CPK residues, respectively. **B)** Interior view of A).

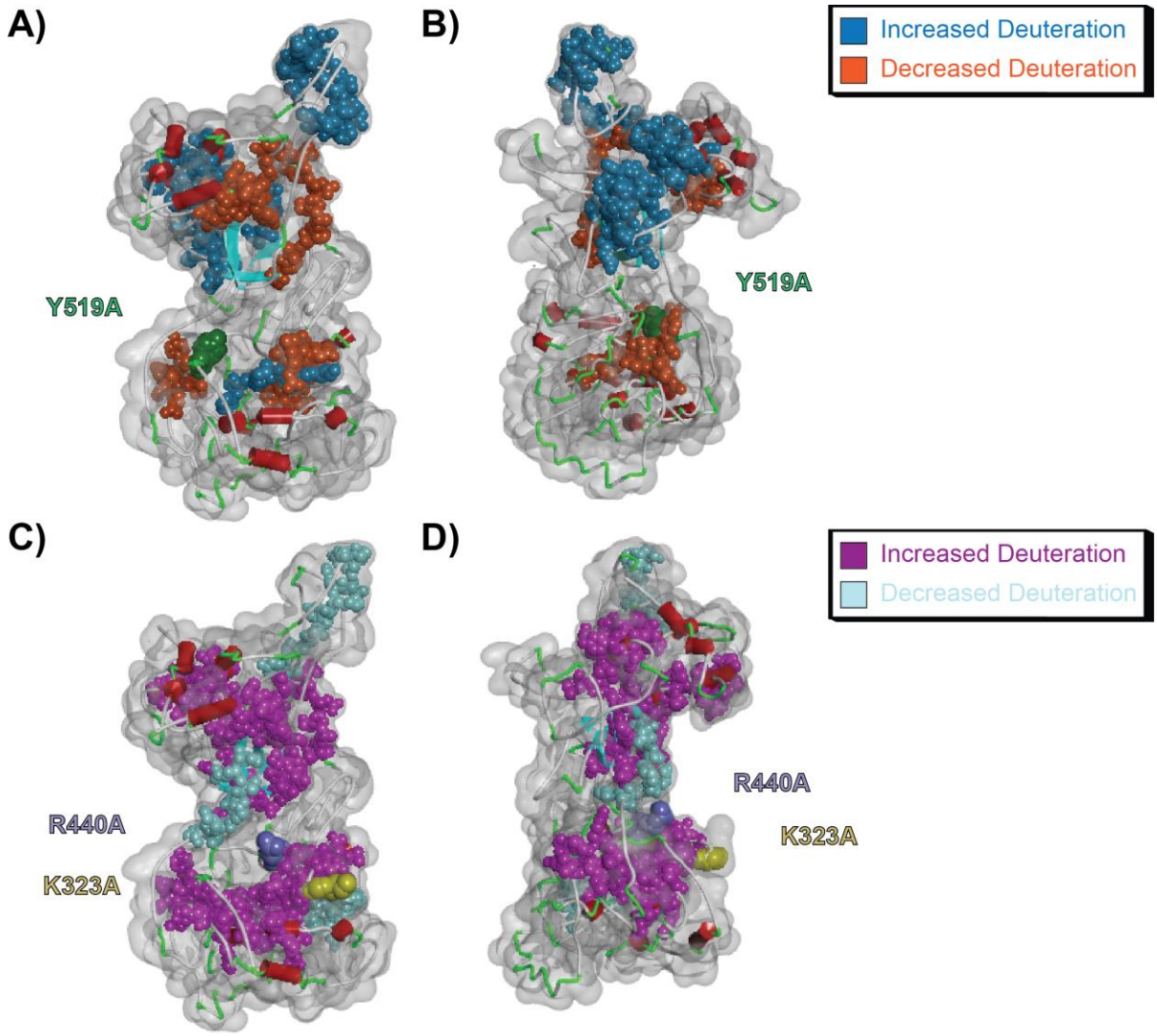


Figure 3.12: *Sso*MCM H/DX-MS - WT versus Y519A and WT versus K323A/R440A.
A) *Sso*MCM WT versus Y519A. Exterior view of an MCM monomer shows residues in which deuteration has increased or decreased in orange and cyan CPK, respectively. Y519 is shown as an enlarged orange CPK residue. **B)** Interior view of A). **C)** *Sso*MCM WT versus K323A/R440A. Exterior view of an MCM monomer shows residues in which deuteration increased or decreased in fuchsia and light blue CPK, respectively. K323 and R440 are shown as enlarged yellow and lavender CPK residues, respectively. **D)** Interior view of C).

3.5 DISCUSSION

The exterior binding site of *Sso*MCM for ssDNA has been shown to include basic residues K323 and R440, however it is likely that more residues are involved in DNA binding. In order to analyze potential base-stacking interactions between surface tyrosine residues and ssDNA [178], we made five tyrosine to alanine mutants (Y40A, Y61A, Y516A, Y519A, and Y523A) and an additional arginine to alanine mutant (R91A). These mutants were tested for their potential disruption of the interaction with ssDNA. Interestingly, during purification, we identified a tyrosine residue that is critical for hexamerization (Y519). By mutating Y519 to alanine, we effectively changed the hexameric equilibrium of the complex, however, this point mutation still formed dimers, trimers, and tetramers both alone in the absence of DNA (Figure 3.2) and on forked DNA (Figure 3.6). As a consequence, this mutant displayed a reduction in unwinding and ATPase rates. *Sso*MCM WT's ATPase rate is enhanced approximately two-fold when in the presence of DNA; however Y519A was reduced to nearly to 1.6-fold. The biochemical evidence suggests that this may have been due to 2-3 fold tighter binding by the Y519A bound species compared to WT (200 nM versus 490 nM).

Previously, the crystal structure of the near full-length *Sso*MCM provided interface mutations to test [113]. Four mutations were made to probe the interface between subunits: a double mutant in the C subdomain of the N-domain L189D/D191R, a double mutant in the α/β region of the C-domain A416R/A420R, a quintuple mutation in the α region of the C-domain TPDSP550GGGGG, and a triple mutant also in the α region of the C-domain ILI555DSD. All of the mutants resulted in monomeric species primarily with a small hexameric component, except for ILI555DSD which was entirely monomer. Another study determined that D488A, structurally located in an alpha helix at the extreme C-terminus, disrupts the hexamer and in this study they postulated that the one peak from the gel filtration column was either a dimer or trimer [112]. D488A, however, had increased helicase activity, but reduced ATPase activity compared to WT, unfortunately its DNA binding ability was not probed [112]. D488A and Y519A, both located in the α/β-α linker region, are the only two *Sso* mutants to date that have disrupted hexamers through single point mutations.

Both *Sso*MCM's N and C domains have been implicated in hexamerization, while *Mth*MCM's N-terminal domain is capable of forming a double hexamer [103]. Y519A alters the α/β-α linker region of MCM which in turn disrupt the hexamer likely through hydrophobic space

between subunits locally including residues 510-516 which showed increased deuteration compared to WT. This mutation likely also disrupts interactions between the interlocked N-C linker and α/β-α linker. By altering the stability of the N-C linker, communication between the N and C domains is reduced. Therefore, the reduced helicase activity combined with a primarily dimeric complex suggest that Y519 is critical for both oligomeric assembly and controlling communication between the two tiers. *Sso*MCM's N domain has been shown to act as a molecular clamp that binds the ssDNA that passes through the central channel [116]. By inhibiting communication between the two domains (C and N); Y519A has uncontrolled hyper DNA binding activity evidenced by the EMSAs quantifications in Figure 3.6C versus Figure 3.6D.

The CD spectra show that *Sso*MCM Y519A has enhanced secondary structure compared to WT and K323A/R440A (Figure 3.7A). This may be due to hexamer assembly from dimeric subunits. This assembly process appears to be temperature dependent according the CD melt data which shows more secondary structure as the temperature increases (Figure 3.7B). K323A/R440A shows similar secondary structure to WT in Figure 3.7A, but has a decreased thermal stability compared to WT. HDX-MS results show altered deuterium uptake values comparing both WT to Y519A and K323A/R440A. For Y519A, this is expected due to the lack of hexamerization and open subunit interfaces. However, K323A/R440A also shows a rather different deuterium uptake profile, in spite of its existence as a hexamer in solution. Interestingly, a recent study showed that *Pyrococcus furiosus* MCM can perpendicularly bind ssDNA within its central channel [179], two residues were implicated in this binding R124 and R186 in the binding of ssDNA. In *Sso*, a ClustalW2 alignment shows this corresponds to K129 and V196. Both Y519A and K323A/R440A identified K129 as having increased deuteration compared to WT. This basic residue will also be studied further.

Our EMSA experiments are also able to show that *Sso*MCM WT exists in an equilibrium between of trimer through hexameric species indicating a possible loading mechanism on DNA. Interestingly, other oligomeric states for Y519A were shown; predominately dimeric. This provides a plausible equilibrium oligomeric assembly mechanism for MCM onto bubble substrates that lack a free ssDNA tail. Therefore, assembly would likely occur through recognition of ssDNA by a single trimer which then recruits and interacts with a second trimer to load a hexamer around ssDNA. After loading, we have identified patches through H/DX-MS that are implicated in binding either the 3' strand that is encircled in the central channel or the 5' strand that interacts on the

exterior. Interestingly, we identified a novel patch (residues 30-33) that interacts with the 5'-tail which contains both a tyrosine (Y30) and arginine (R33). The ssDNA exterior binding deficient mutant K323A/R440A showed increased deuterium uptake for residues 224-237 compared to WT. This region is in the C region of the N domain and is located between loop 207 and the N-terminal hairpin. Loop 207 is also known as the allosteric communication loop due to its intersubunit allosteric communication with the N-terminal hairpin allowing for ATP hydrolysis and subsequent unwinding [118]. A report for *Mth*MCM demonstrated that mutations D212N and G218A (corresponding to *Sso*MCM conserved residues D229 and G235) demonstrated reduced ssDNA binding while P210G, G211A, and P210G/G211A (P227 and G228) displayed enhanced ssDNA binding both compared to WT [180]. The ATPase rates were also stimulated >10-fold for P210G, G211A, and P210G/G211A [180]. This region likely plays a role in the ssDNA binding on MCM's exterior. Finally, comparing Y519A to K323A/R440A there is increased deuterium uptake for residues 59-74 for Y519A but decreased deuterium uptake for these residues in K323A/R440A. This patch is located in the A region of the N domain and contains Y61 and the solvent accessibility is altered in the presence of these mutants relative to WT. This residue, when mutated to alanine had a reduction in unwinding similar to Y519A (Figure 3.3), although it retained the hexamer conformation (Figure 3.2). In future studies we will examine the 5'-tail interaction with the aforementioned residues through further mutational analyses and functional characterizations.

3.6 CONCLUSIONS

Y519 is a residue that is critical for *Sso*MCM hexamerization. We have shown that a mutation to alanine both disrupts unwinding and ATPase rates, while Y519A binds DNA tighter than WT, likely due to a lack of communication between N and C domains. Y519A also shows enhanced secondary structure that can be modulated by temperature. We also determined DNA binding patches for both the 3' and 5'-tails through HDX-MS and identified further future contacts to be tested.

3.7 ACKNOWLEDGEMENTS

We would like to thank Conor Haney and the W. Seth Horne for providing assistance and access to the Olis spectrometer, respectively.

4 NOVEL INTERACTION OF THE BACTERIAL-LIKE DNAG PRIMASE WITH THE MCM HELICASE IN ARCHAEA[3]

4.1 SUMMARY

DNA priming and unwinding activities are coupled within bacterial primosome complexes to initiate synthesis on the lagging strand during DNA replication. Archaeal organisms contain conserved primase genes homologous to both the bacterial DnaG and archaeo-eukaryotic primase families. The inclusion of multiple DNA primases within a whole domain of organisms complicates the assignment of the metabolic roles of each. In support of a functional bacterial-like DnaG primase participating in archaeal DNA replication, we have detected an interaction of *Sulfolobus solfataricus* DnaG (*Sso*DnaG) with the replicative *S. solfataricus* minichromosome maintenance (*Sso*MCM) helicase on DNA. The interaction site has been mapped to the N-terminal tier of *Sso*MCM analogous to bacterial primosome complexes. Mutagenesis within the metal binding site of *Sso*DnaG verifies a functional homology with bacterial DnaG that perturbs priming activity and DNA binding. The complex of *Sso*DnaG with *Sso*MCM stimulates the ATPase activity of *Sso*MCM but leaves the priming activity of *Sso*DnaG unchanged. Competition for binding DNA between *Sso*DnaG and *Sso*MCM can reduce the unwinding ability. Fluorescent gel shift experiments were used to quantify the binding of the ternary *Sso*MCM–DNA–*Sso*DnaG complex. This direct interaction of a bacterial-like primase with a eukaryotic-like helicase suggests that formation of a unique but homologous archaeal primosome complex is possible but may require other components to stimulate activities. Identification of this archaeal primosome complex broadly impacts evolutionary relationships of DNA replication.

[3] This material from this chapter is from Bauer R.J.*, **Graham B.W.***, and Trakselis M.A. Novel interaction of the bacterial-like DnaG primase with the MCM helicase in archaea. *Journal of Molecular Biology*, **425**(8): p. 1259-1273. Permission to reprint in this dissertation has been acquired from the Journal of Molecular Biology. *Bauer and Graham are joint-first authors. Graham was responsible for the purification of MCM constructs, unwinding assays, ATPase assasys, fluorescent protein labeling, and fluorescent EMSAs.

4.2 INTRODUCTION

DNA replication is a highly coordinated, yet dynamic, process that includes assemblies of multiprotein complexes to form the active replisome. After separation of the duplex DNA at an origin of replication by the concerted efforts of the DNA helicase and its accessory proteins, DNA replication begins with the *de novo* synthesis of an RNA primer by the DNA primase. Primer synthesis occurs at defined initiation sites on the single-stranded DNA (ssDNA) [73] template both to initiate leading strand synthesis and repeatedly for each Okazaki fragment on the lagging strand. DNA primases are thought to exist within protein subcomplexes to both control and coordinate activities at the replication fork in all domains of life.

Two separate DNA primase families exist: one group (DnaG family) contains all the primases from bacteria and their phages, while the other group consists of archaeo-eukaryotic primases [73, 181]. The eukaryotic primase is a four-subunit complex composed of a small catalytic subunit (p48) and a large regulatory subunit (p58) that modulates binding and activity. These subunits are almost always found in complex with two other proteins, DNA polymerase α (p180) and polymerase B (p68), to form the polymerase α–primase complex (pol–prim) [54]. This complex can synthesize RNA primers initially with minimal template specificity [60] and can then extend them by incorporating dNTPs by pol α and pol B. Eukaryotic primases may control primer length (7–10 bases) through p58 regulation of binding to the ssDNA template that directs RNA synthesis by closing a hinge between p49/p58 subunits[59, 182]. The handoff of RNA primers larger than 7 nt from p58 to pol α occurs by direct handoff within the complex for further extension into hybrid RNA–DNA products [183, 184]. The clamp loader complex [replication factor C (RFC)] plays an important role in displacing pol α after roughly 30 nt for replacement with a more processive DNA polymerase holoenzyme complex that includes proliferating cellular nuclear antigen and either pol δ or pol ε [185, 186].

Archaeo-eukaryotic primases from archaea contain only two of the four subunits of the pol α primosome (small catalytic, PriS, and large regulator, PriL, subunits) and have been characterized in *Pyrococcus* [63, 64], *Thermococcus kodakaraensis* [65], and *Sulfolobus solfataricus* (*Sso*) [55, 58, 66]. In addition to RNA primer synthesis ranging from 2 to 500 nt, these primases have surprisingly novel and unregulated DNA synthesis abilities producing oligonucleotide products greater than 7 kilobases. Temperature and slight differences in the

affinity of NTP or dNTP may direct function towards RNA or DNA synthesis, respectively. An indirect link between PriSL and the minichromosome maintenance (MCM) helicase is thought to be mediated by GINS23 to coordinate priming and unwinding activities in archaea [67]. The DNA replication homology between archaea and eukaryotes would predict that like other functional homologs, PriSL will fulfill the DNA priming role in archaea as well [68, 69].

The bacterial DnaG primase has three domains: the N-terminal Zn^{2+} binding domain, the catalytic topoisomerase primase (TOPRIM) domain, and the C-terminal helicase binding domain. TOPRIM is a signature motif in DnaG-type primases, topoisomerase, and other nucleases that encompasses an acidic metal binding active site required for catalysis [73, 74]. Binding to DNA is generally weak and transient but occurs along an elusive, positively charged region adjacent to the active site that orientates the DNA template [75]. The C-terminal domain of *Escherichia coli* DnaG (*Ec*DnaG) is associated with the N-terminus of the DnaB helicase (*Ec*DnaB) to form the bacterial primosome complex that increases both priming and helicase activities [76-79]. This primase–helicase interaction and orientation have been seen in a variety of other bacterial and phage organisms [80-84] and are required for synthesis of primers on the lagging strand. Primers synthesized by *Ec*DnaG are typically 11 nt but can range from 2 to 14 nt [85]. The mechanism of RNA synthesis and the protein interactions of *Ec*DnaG have been well studied [72, 75, 87-95]. Primer length is controlled through coordination of the two subunits of bacterial DnaG with the Zn^{2+} binding domain to regulate DNA template binding [71, 72, 75, 78]. The influence *Ec*DnaG and *Ec*DnaB have on each other's activities is only just starting to be revealed.

Interestingly, archaea also contain within their genome a homolog to the bacterial-like DnaG primase [73]. We have found that archaeal *Sso*DnaG has a conserved and essential active-site glutamate required for synthesis of primarily 13mer RNA products [96]. In support of priming activity, *Sso*DnaG is able to *de novo* synthesize RNA primers with greater efficiency than archaeal *Sso*PriSL. To date, there have been no reported interactions of archaeal *Sso*DnaG with any other members of the replisome. Instead, *Sso*DnaG is found as a structural component within the archaeal exosome, albeit with no associated enzymatic activity [98]. The exosome complex is required for the degradation of RNA and is actually contrary to the DNA priming or synthesis function of the primase.

To support a possible role for *Sso*DnaG in archaeal DNA replication, we have detected and verified a specific interaction with the *Sso*MCM helicase responsible for separation of duplex DNA

ahead of the replication fork. The interaction site has been mapped to the N-terminal domain of the *Sso*MCM helicase analogous to other primase–helicase interactions found in bacteria. The priming activity of *Sso*DnaG is unaffected by this interaction, the unwinding ability of *Sso*MCM is inhibited, and the ATPase activity of *Sso*MCM is stimulated. Active-site mutations in *Sso*DnaG verify conservation of acidic metal binding residues required for priming and DNA binding and are used to highlight *Sso*DnaG's interaction with both DNA and *Sso*MCM. DNA unwinding, ATPase, and fluorescent electrophoretic mobility shift assays (EMSAs) suggest that *Sso*DnaG binds to an *Sso*MCM–DNA complex forming a ternary conformation. This work has broad evolutionary consequences for RNA priming in archaea and suggests that formation of an analogous bacterial-like primosome complex may also be important for coordinating activities at the replication fork in archaea.

4.3 MATERIALS AND METHODS

4.3.1 Materials

Oligonucleotide substrates (sequences are in Tables 4.1 and 4.2) were purchased from IDT (Coralville, IA) and gel purified [144]. Fluorescent HPLC-purified DNA was from IDT. All radiochemicals were purchased from MP Biochemicals (Santa Ana, CA) or Perkin Elmer (Waltham, MA). Cy5 succinimidyl ester was from Invitrogen (Carlsbad, CA). Commercial enzymes were from NEB (Ipswich, MA). All other chemicals were analytical grade or better.

4.3.2 Protein purifications

The *Sso*DnaG gene was PCR amplified from pET30a–*Sso*DnaG [96] and cloned into pGEX-6P2 using SmaI and XhoI restriction sites included in the primer sequences. Active-site mutant constructs (pET30–*Sso*DnaG–D179A, D220A, and D222A) with and without a 6 × His tag were created using a standard QuikChange protocol (Agilent, Santa Clara, CA) with KAPA HiFi DNA

Table 4.1: DNA primer sequences

DNA Primers	Sequence (5'-3')
DnaG D179A FWD	5'-ATTTAATAATAGTAGAAGGAAGAGCTGCAGTAATAAATCTACTCAGA TATGGCTAC
DnaG D179A REV	5'-GTAGCCATATCTGAGTAGATTTATTACTGCAGCTCTTCCTTCTACTATTA TTAAAT
DnaG D220A FWD	5'-GACAGTAATAGCGTTTTTAGCCGGTGACCACGGTGGAGA
DnaG D220A REV	5'-TCTCCACCGTGGTCACCGGCTAAAAACGCTATTACTGTC
DnaG D222A FWD	5'-GCGTTTTTAGACGGAGCCCACGGTGGAGATCTG
DnaG D222A REV	5'-CAGATCTCCACCGTGGGCTCCGTCTAAAAACGC
DnaG RemoveStop F	5'-CCGATATTATTTCTTCTGTCGAGCACCACCACCAC
DnaG RemoveStop R	5'-GTGGTGGTGGTGTCGACAGAAGAAATAATATCGG
DnaG GST For	5'-CACCCATATGAGCTTCCAAATGAAATATGATATAAGG
DnaG GST Rev	5'-ATTACTCGAGAGAAGAAATAATATCGGTAAATGTC
DnaG pGADT7 For	5'-ACGTCGACCCCGGGATGAGCTTCCAAATGAAATATGATAT
DnaG pGADT7 Rev	5'-ATTACTCGAGAGAAGAAATAATATCGGTAAATGTC
MCM FWD	5'-ATTAGGATCCATGGAAATTCCTAGTAAACAGATTGAC
MCM 106 FWD	5'-TTTGGATCCATTAATGGGTAAACTAATAACTATTGATGG
MCM 267 FWD	5'-TTTGGATCCATTAATAAAGTATTAGATGAGGTAATCATCTC
MCM 267 REVStop	5'-ATTACTCGAGCTATTTTTGTGAAACTTCTATACTAC
MCM 612 REVStop	5'-ATTACTCGAGCTAATCTATATCTATTTTTCCACTTTCC
MCM pGBKT7 For	5'-ATTAGGATCCATGGAAATTCCTAGTAAACAGATTGAC
MCM pGBKT7 Rev	5'-ATAGATGTCGACCTAGACTTTTTTGTAACATTC

Table 4.2: DNA substrates

DNA[1]	Sequence[2,3]
ssDNA (5'Cy5)	5'-5GCTACTCTCGCTCAGCGTACCATAGCAG
3'-tail-30 nt	5'-CACCTCTCCCTACGCTTCCCACCCACCCCGACCGGCATCTGCTATGGTAC GCTGAGCGAGAGTAGC
5'-tail-30 nt	5'-CGATGAGAGCGAGTCGCATGGTATCGTCTAGCCGGTCGGGGTGGGTGGG AAGCGTAGGGAGAGGTG
3'-tail-30 nt (5'Cy3)	5'-3CACCTCTCCCTACGCTTCCCACCCACCCCGACCGGCATCTGCTATGGTAC GCTGAGCGAGAGTAGC
5'-tail-0 nt	5'-GCCGGTCGGGGTGGGTGGGAAGCGTAGGGAGAGGTG

[1]nt – nucleotides; [2]Modifications are underlined; [3]3 – Cy3, 5 - Cy5

polymerase (KAPA Biosystems, Woburn, MA). WT and His-tagged *Sso*DnaG constructs and *Sso*PriSL were expressed and purified as described previously [96]. Overexpressed GST-*Sso*DnaG was purified using a HiTrap glutathione Sepharose column (GE Healthsciences) and a Superdex 26/60 gel-filtration column. GST-*Sso*DnaG was stored in binding buffer (10 mM Na_2PO_4, 1.8 mM K_2HPO_4, 140 mM NaCl, 2.7 mM KCl, pH 7.3, and 10% glycerol).

Full-length *Sso*MCM was purified as previously described using 70 ºC heat treatment, MonoQ, heparin, and gel-filtration columns to isolate the hexameric species [123]. Truncated forms of *Sso*MCM with included stop codons (1–267, 106–612, 267–612, and 1–612) were cloned

into pET30a using NdeI and XhoI restriction sites, eliminating any affinity tags. Truncated proteins were expressed using an autoinduction protocol at 37 ºC [187] and purified similarly to WT.

*Ec*DnaB was autoinduced using pET11b-DnaB in BL21DE3 Rosetta 2 cells [187] and purified essentially as described previously [188, 189]. Briefly, *Ec*DnaB was purified using an ammonium sulfate precipitation and MonoQ and heparin columns with 0.5 M NaCl gradient elution. Positive fractions were pooled and stored in *Ec*DnaB storage buffer [20% glycerol, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 5 μM ATP, 5 mM $MgCl_2$, 20 mM Tris–HCl, pH 8.0, 50 mM NaCl, and 2 mM DTT]. The *Ec*DnaB concentration was determined with UV–Vis using an extinction coefficient of 29,870 M^{-1} cm^{-1}. All protein preparations were verified to have no significant background DNA contamination by UV (260/280 ratio < 0.9) and direct labeling with ^{32}P-γ-ATP in a polynucleotide kinase reaction.

4.3.3 Fluorescent protein labeling

Both *Sso*DnaG and *Sso*MCM constructs were labeled at the N-terminus with Cy5 succinimidyl ester in labeling buffer (50 mM Hepes, pH 6.8, 100 mM NaCl, 5 mM β-mercaptoethanol, and 10% glycerol) as described previously [110, 111]. The lower pK_a of the N-terminus compared to lysine residues preferentially labels at this location under these conditions. Proteins were reacted with 3 × molar excess of dye for 30 min at room temperature. Excess fluorophore was removed through extensive dialysis in labeling buffer. The labeling efficiency was determined by UV–Vis spectroscopy.

4.3.4 Yeast two-hybrid assay

To generate the yeast two-hybrid plasmids, we cloned *Sso*MCM into the SalI and NdeI sites of pGBKT7 (Clontech) and cloned *Sso*DnaG into the XhoI and NdeI sites of pGADT7 (Clontech). The yeast strain PJ69-4A (*MATa leu2-3,112 ura3-52 trp1-901 his3-200 gal4Δ gal80Δ GAL2-ADE2 lys2*::*GAL1-HIS3 Met2*::*GAL7-LacZ*) was transformed with the appropriate plasmids (see figure legends) according to the manufacturer's instructions using the lithium acetate procedure (Clontech Matchmaker manual). Liquid cultures were grown overnight in media lacking

tryptophan and leucine. The cells were serially diluted and spotted on either SD/-Trp/-Leu or SD/-Trp/-Leu/-His medium and incubated at 30 ºC for 2–3 days.

4.3.5 GST pull-down assays

GST-*Sso*DnaG (20 μM) was incubated at room temperature in the presence of 30 μM *Sso*MCM for 30 min. The sample was then immobilized on 200 μL of glutathione Sepharose 4B resin (GE Healthsciences), washed with 200 μL of GST binding buffer (10 mM Na_2PO_4, 1.8 mM K_2HPO_4, 140 mM NaCl, and 2.7 mM KCl pH 7.3), and centrifuged for 10 s at 6000 ***g***; washing was repeated at least seven times. Proteins were eluted with 600 μL of GST elution buffer (50 mM Tris–HCl, pH 8.0, and 10 mM reduced glutathione). Eluted samples were concentrated, separated using 8% or 10% SDS-PAGE gels, and stained with Coomassie. Fluorescent GST pull-down SDS-PAGE gels were performed similarly and then imaged using a Typhoon phosphorimager (GE Healthsciences).

4.3.6 DNA priming assays

Priming reactions with or without *Sso*MCM were performed in primase reaction buffer (50 mM Tris–HCl, pH 8.5, 1 mM $MnSO_4$, 100 mM NaCl, and 0.2 mM NTPs, including 0.025 μCi/μL [α-^{32}P]GTP) as described previously [96]. The concentration of *Sso*MCM is indicated in each figure legend and either 4 nM M13 ssDNA, 200 nM forked, 200 nM 3′-tail, or 200 3′-tail-30 nt ssDNA at 70 °C was used as the template. Forked DNA is 3′-tail-30 nt annealed to 5′-tail-30 nt; 3′-tail is 3′-tail-30 nt annealed to 5′-tail-0 nt (Table 4.2) as previously described [123]. Aliquots of the priming reaction were quenched in an equal volume of stop solution (88% formamide, 10 mM EDTA, and 1 mg/mL bromophenol blue) at times as indicated in each figure legend. The ^{32}P-labeled RNA primers were then resolved on a 20% denaturing polyacrylamide gel, phosphorimaged using a Storm 820 (GE Healthsciences), and quantified with ImageQuant software (version 5.0) to calculate reaction rates.

4.3.7 DNA unwinding assays

*Sso*MCM helicase unwinding reactions with and without *Sso*DnaG or *Sso*PriSL (concentrations indicated in the figure legends) were performed as described previously using 15 nM ^{32}P-labeled forked DNA [123]. Briefly, reactions were incubated at 60 ºC for 5 min and initiated upon addition of either *Sso*MCM, *Sso*MCM and *Sso*DnaG, or ATP. Reactions were quenched with an equal volume (1.6% w/v SDS, 50% v/v glycerol, 0.1% w/v bromophenol blue, 100 mM EDTA, pH 8.0, and 150 nM trap ssDNA complementary to the unlabeled strand) and then stored on ice until loading. ^{32}P-DNA products were separated on denaturing acrylamide gels and analyzed as described above.

*Ec*DnaB unwinding reactions were performed with 4.2 μM *Ec*DnaB (monomer) with and without *Sso*DnaG and incubated with 15 nM ^{32}P-labeled forked DNA in assembly buffer [50 mM Hepes, pH 7.6, 100 mM NaCl, 0.2 mg/mL bovine serum albumin, 5 mM β-mercaptoethanol, 10 mM $Mg(OAc)_2$, and 100 nM AMP-PNP] at 37 °C for 5 min to assemble hexameric *Ec*DnaB on DNA. *Ec*DnaB unwinding reactions were initiated with 2 mM ATP, quenched after 10 min, and processed as described above.

4.3.8 ATPase assays

ATPase reactions were incubated at 60 °C for 5 min and initiated upon addition of ATP. Reaction conditions were determined empirically by titrating ATP, DNA, and *Sso*DnaG concentrations while keeping *Sso*MCM constant at 4.2 μM (700 nM hexamer). Optimal reaction conditions included helicase buffer, 2.0 mM ATP, 1 μM cold forked DNA (if present), 4.2 μM *Sso*MCM, and 0.7 μM *Sso*DnaG (if present) totaling 20 μL per reaction. Samples were quenched at 5, 10, and 15 min after initiation into equal volumes of 0.7 M formic acid. A total of 0.8 μL of quenched reaction was spotted on Analtech Cellulose PEI F, allowed to dry, resolved in 0.6 M potassium

phosphate (pH 3.5) buffer, and then phosphorimaged to calculate ATPase rates. Data from titration with DNA were fit to a simple Michaelis–Menten equation

$$v = \frac{A \times [M]}{K_m + [M]} \quad (1)$$

where A is the amplitude, M is the concentration of titrant, and K_m is the Michaelis constant, while data from titrations with ATP required a mixed inhibitor model:

$$v = \frac{A \times [M]}{K_m + \left([M] + \left(1 + \frac{[M]}{Ki} \right) \right)} \quad (2)$$

where K_i is an inhibitor equilibrium constant. p values were calculated comparing the different conditions for ATPase experiments using Student's t test in Excel. A p value of < 0.05 was considered statistically significant.

4.3.9 Homology modeling of the *Sso*DnaG core domain

Local and global sequence alignments were performed using ClustalW2 analysis‡. The homology model of *Sso*DnaG was created by threading the global alignment of *Sso*DnaG with *Ec*DnaG onto the structure of *Ec*DnaG (PDB ID: 3B39) [75] using SWISS-MODEL [190] and overlayed and represented using PyMOL§.

4.3.10 Anisotropy DNA binding assays

Anisotropy assays were performed in *Sso*DnaG reaction buffer with 4 nM 28mer ssDNA (5′Cy5) and titrating *Sso*DnaG (WT, D179A, D220A, or D222A) as indicated. Anisotropy values were obtained and quantified as detailed previously [164]. The change in anisotropy was fit to a single binding equation:

$$r = \frac{A \times [P]}{K_d + [P]} \quad (3)$$

where A is the amplitude, P is the concentration of *Sso*DnaG, and K_d is the dissociation constant, or a cooperative binding equation:

$$r = \frac{A \times [P]^n}{{K_d}^n + [P]^n} \quad (4)$$

where n is the Hill coefficient.

4.3.11 Electrophoretic mobility shift assays

EMSAs were performed in 10 μL volumes by the stepwise addition of 100 nM 66mer ssDNA [3′-tail-30 nt (5′Cy3)], *Sso*MCM (as specified), and 200 nM Cy5-labeled *Sso*DnaG (if present). Reactions were incubated at 60 °C for 10 min to promote native thermodynamic complex formation and resolved on 5% native polyacrylamide gels (in 1 × TBE). The gels were imaged using a Typhoon phosphorimager and the Cy3 and Cy5 intensities were quantified using ImageQuant. The Cy5-*Sso*DnaG volume was corrected for background Cy3-DNA intensities and plotted *versus Sso*MCM concentration to determine the K_d of the ternary complex according to Equation 3, where P is the concentration of *Sso*MCM.

4.4 RESULTS

4.4.1 Detection and verification of a direct *Sso*MCM–*Sso*DnaG interaction

After screening for interacting *Sso* replication proteins with *Sso*DnaG using a yeast two-hybrid approach, we found that the *Sso*MCM helicase allows for growth on selective plates. In this experiment, *Sso*MCM is the bait constrained with the DNA binding domain within pGBKT7. *Sso*DnaG was cloned into pGADT7 containing the activation domain and acts as the prey. As observed in Figure 4.1A, growth on selective media (SD/-Trp/-Leu/-His) is only allowed when both *Sso*MCM and *Sso*DnaG are included. Empty vectors or strains containing only single baits or preys (*Sso*MCM or *Sso*DnaG) showed no background growth on SD/-Trp/-Leu/-His but grew as expected on media (SD/-Trp/-Leu) selecting for plasmids alone.

In order to verify the yeast two-hybrid interaction between *Sso*MCM and *Sso*DnaG, we performed a glutathione *S*-transferase (GST) pull-down assay. Using an immobilized GST-tagged *Sso*DnaG (hereinafter GST-*Sso*DnaG) construct, we are able to verify that *Sso*MCM specifically interacts with *Sso*DnaG (Figure 4.1B). After addition and flow through of *Sso*MCM, the column was washed seven times, before eluting with glutathione. The difference in molecular mass between GST-*Sso*DnaG (73 kDa) and untagged *Sso*MCM (78 kDa) is not large, but *Sso*MCM can be clearly seen above GST-*Sso*DnaG in lane 6. Mock pull downs in the absence of GST-*Sso*DnaG show no background *Sso*MCM binding (lane 9).

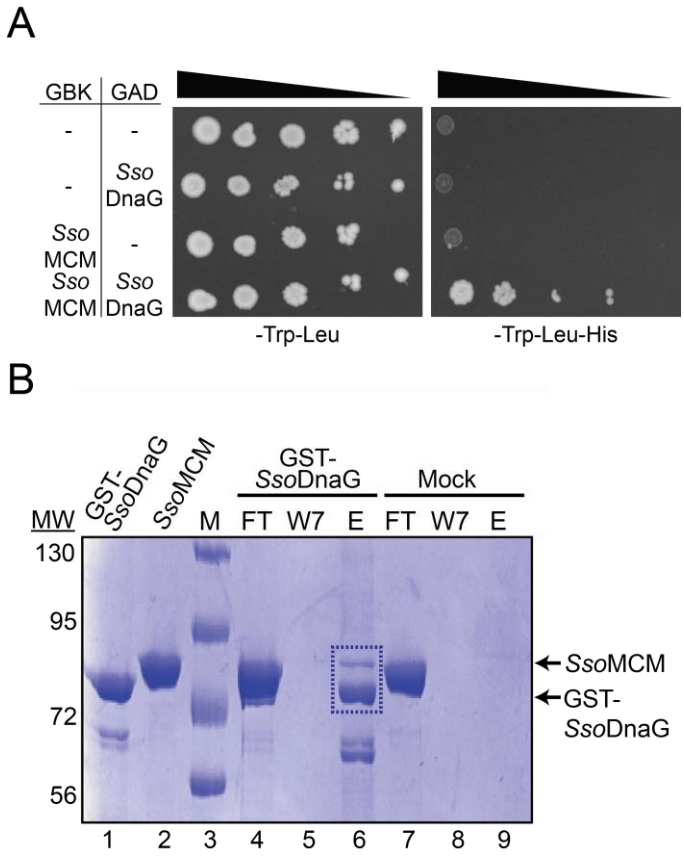


Figure 4.1: Physical interaction between full length *Sso*MCM and *Sso*DnaG.
A) Yeast two-hybrid analysis of the *Sso*DnaG and *Sso*MCM interaction. *Sso*MCM was cloned into the GAL4 DNA binding domain vector, pGBKT7 (GBK), and *Sso*DnaG was cloned into GAL4 activation domain vector, pGADT7 (GAD). Cultures were serial diluted on either SD/-Trp/-Leu or SD/-Trp/-Leu/-His media. **B)** Coomassie stained SDS-PAGE gel showing purified proteins**,** GST-*Sso*DnaG (lane 1), *Sso*MCM (lane 2) and molecular weight markers (lane 3). GST pull down with (lanes 4-6) or without (lanes 7-9) immobilized *Sso*DnaG. *Sso*MCM flow through (FT, lanes 4 & 7), seventh wash (W7, lanes 5 and 8), and elution with glutathione (E, lanes 6 & 9). Interaction between GST-*Sso*DnaG and *Sso*MCM is seen in lane 6 outlined in a box.

4.4.2 Effect of *Sso*MCM–*Sso*DnaG interaction on priming and unwinding activities

In order to determine if the interaction of *Sso*MCM influences the priming activity of *Sso*DnaG, we performed priming assays in the presence of *Sso*MCM (Figure 4.2A). No significant effect on *Sso*DnaG's priming rate on M13 was observed with *Sso*MCM concentrations up to 20 μM (Figure 4.2 and data not shown). The priming rate of *Sso*DnaG was 17.0 ± 1.0 pmol min^{-1} in the absence of *Sso*MCM and 16.5 ± 0.4 pmol min^{-1} in the presence of 10 μM *Sso*MCM. We note that the priming rates observed in these experiments were roughly 10-fold higher per unit enzyme than we reported previously [96] and attribute this to purification of a more active unaggregated enzyme. Additional priming experiments were performed on short DNA templates (forked, 3′-tail, and ssDNA) with similar results, illustrating that even with substrates traditionally utilized for *Sso*MCM binding and unwinding experiments, no further stimulation of priming activity was observed (Figure 4.3).

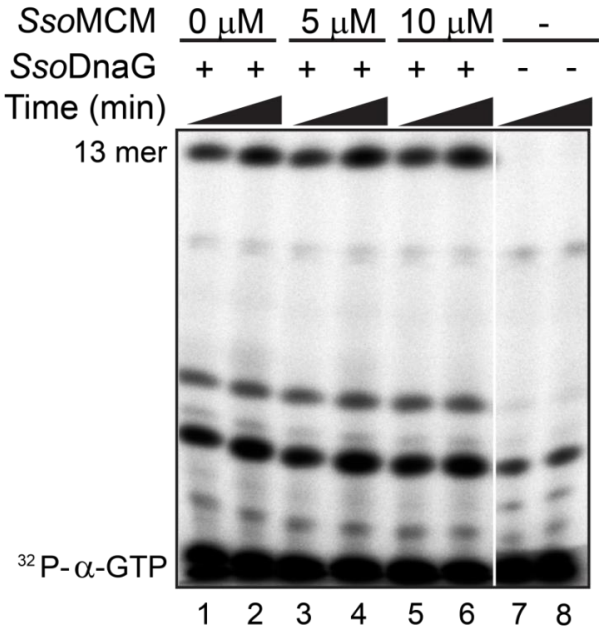


Figure 4.2: *Sso*MCM does not significantly increase RNA primer synthesis by *Sso*DnaG.
RNA primer synthesis by 1.3 μM *Sso*DnaG on M13 single strand DNA substrate at 60 °C showing 13mer and shorter products from ^{32}P-α-GTP as a function of *Sso*MCM concentration. Times are 30 and 60 minutes. Lanes 7-8 are no enzyme controls.

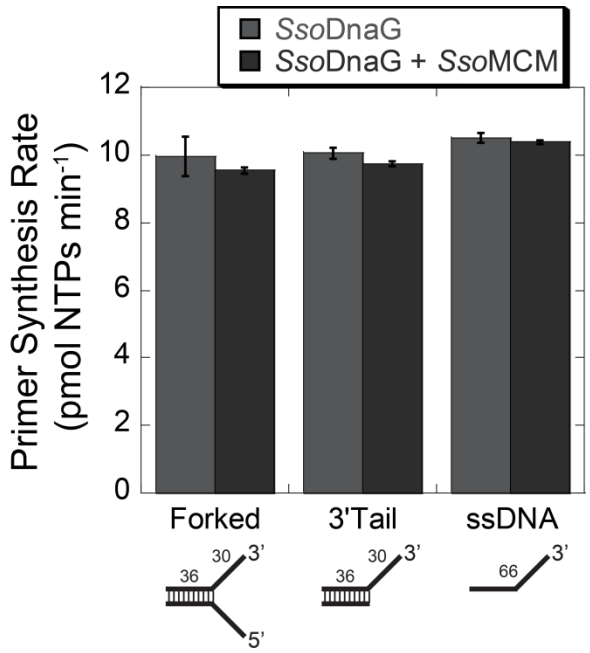


Figure 4.3: *Sso*DnaG priming reactions quantified.
*Sso*DnaG priming reactions were performed in the presence and absence of 3.2 μM *Sso*MCM (533 nM hexamer) on forked, 3'-tail and ssDNA. The plot quantifies the formation of RNA primer products by the *Sso*DnaG on these DNA templates. Error bars are the standard error for each reaction.

To test the effect of *Sso*DnaG on the unwinding activity of *Sso*MCM, we monitored the *in vitro* unwinding activity of a forked DNA substrate. *Sso*MCM (4.2 μM) (700 nM *Sso*MCM hexamer) was tested for unwinding in the absence and presence of 700 nM *Sso*DnaG. The presence of *Sso*DnaG inhibits *Sso*MCM unwinding 6-fold (Figure 4.4A and B). Experiments were initiated by addition of *Sso*MCM, but other order-of-addition experiments where ATP or primase/ATP was added to initiate showed no difference (data not shown). Importantly, inclusion of the other archaeal primase (up to 4.2 μM) from *Sulfolobus*, *Sso*PriSL, had no effect on unwinding as the unwinding rate is identical with *Sso*MCM alone and similar to that reported previously [67]. The K_d of *Sso*PriSL binding to DNA [191] is estimated to be larger than that of *Sso*DnaG, but in these experiments, both primases are well above their individual K_d values. Therefore, a specific interaction between *Sso*MCM and *Sso*DnaG on DNA is responsible for the unwinding inhibition, whereas no effect is seen from the purported *Sso*PriSL primase.

The ATPase activity of hexameric helicases is generally stimulated upon binding DNA to activate unwinding or translocation [111, 170]. ATPase experiments were used to monitor any

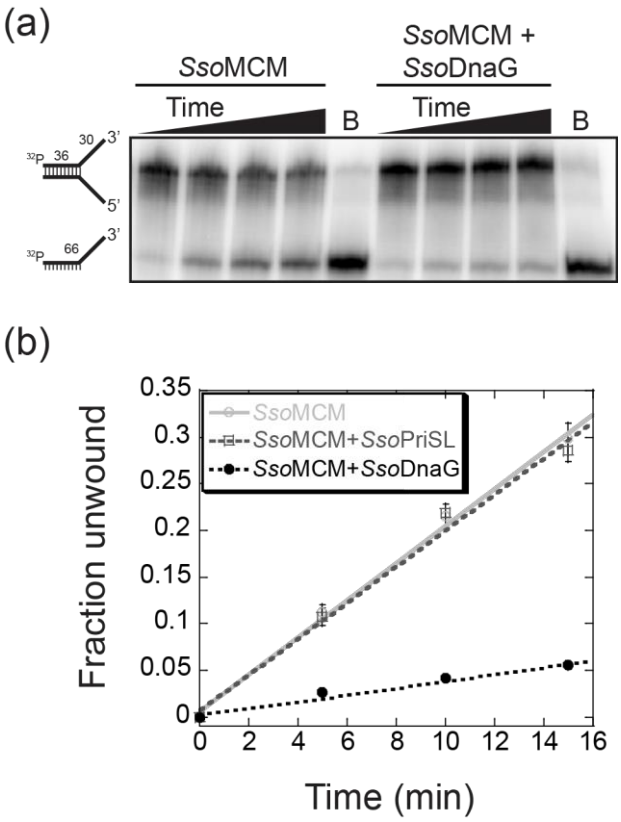


Figure 4.4: Inhibition of *Sso*MCM DNA unwinding upon interaction with *Sso*DnaG.
A) Representative DNA unwinding assay and denaturing acrylamide gel for 4.2 µM *Sso*MCM (700 nM hexamer) alone and in the presence of 4.2 µM *Sso*DnaG monitoring unwinding of 15 nM forked DNA. B stands for boiled only DNA samples. **B)** Kinetics of DNA unwinding as a function of time for *Sso*MCM alone (-o-) or in the presence of *Sso*PriSL (-□-) or *Sso*DnaG (-●-) at 60 °C. Error bars represent the standard error from at least three independent unwinding experiments.

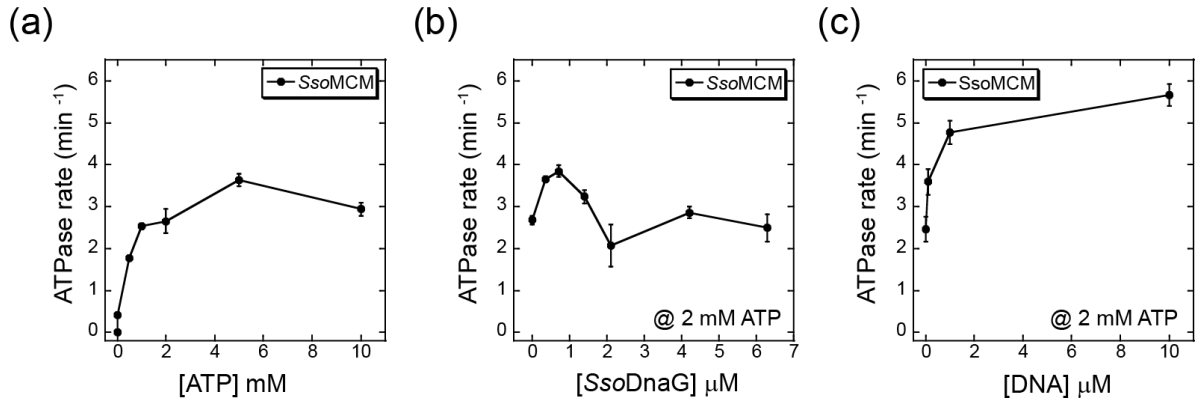


Figure 4.5: ATPase activity of *Sso*MCM under different conditions.
ATPase activity of *Sso*MCM as a function of **A)** ATP, **B)** *Sso*DnaG, and **C)** forked DNA concentration. ATPase activity of 4.2 µM *Sso*MCM was measured at 60 ºC. Multiple time points were averaged to calculate the number of ATP molecules hydrolyzed per minute per MCM monomer. The error bars represent the standard error from at least three independent experiments.

effect on ATP hydrolysis by *Sso*MCM when *Sso*DnaG was included. Titrations were performed to determine the optimal concentrations of ATP, S*so*DnaG, and DNA in the ATPase assays (Figure 4.5). Increasing concentration of ATP increased the ATPase rate for *Sso*MCM before leveling off and decreasing above 2 mM due to slight inhibition at high concentrations. Interestingly, when *Sso*DnaG was titrated up to 1 μM in the reaction, there was stimulation in *Sso*MCM's ATPase rate, while at higher *Sso*DnaG concentrations, the ATPase rate of *Sso*MCM returned to basal levels. When DNA was included, the ATPase rates of *Sso*MCM alone increased as expected, but when *Sso*DnaG was added, the ATPase rate was inhibited. Higher concentrations of DNA (> 5 uM) compete for the inhibition of *Sso*DnaG and restore *Sso*MCM's ATPase rate.

Under optimal reaction conditions, 4.2 μM *Sso*MCM (700 nM hexamer) alone has an ATPase rate of 217 ± 17 pmol min^{-1} (Figures 4.6 and 4.7). Inclusion of 0.7 μM *Sso*DnaG significantly stimulated the ATPase activity of *Sso*MCM alone 1.5-fold (326 ± 17 pmol min^{-1}). Addition of 1 μM forked DNA to *Sso*MCM stimulated the ATPase activity almost 2-fold (414 ± 28 pmol min^{-1}), consistent with previous results [103, 111]. The addition of DNA with *Sso*DnaG/*Sso*MCM had no further stimulatory effect (327 ± 22 pmol min^{-1}) over *Sso*DnaG–*Sso*MCM alone and is significantly less than that for *Sso*MCM–DNA. ATPase experiments with *Sso*DnaG alone or with DNA had only background levels of hydrolysis.

4.4.3 Characterization of the conserved acidic active-site mutants of *Sso*DnaG primase

Previously, we have shown that the mutation of a conserved glutamate (E175Q) abolished the priming activity of *Sso*DnaG [96]. The core TOPRIM domain of *Sso*DnaG also includes other conserved acidic aspartates (D179, D220, and D222) that are proposed to define metal binding in the active site. Mutation of these homologous aspartates in *Ec*DnaG resulted in catalytically inactive enzymes with decreased metal binding affinities [192, 193]. A homology model of the core domain of *Sso*DnaG was created by threading the alignment onto the crystal structure of the TOPRIM domain of *Ec*DnaG [Protein Data Bank (PDB) ID: 3B39] and illustrates these conserved acidic residues (Figure 4.8A). We individually mutated *Sso*DnaG D179A, D220A, and D222A and purified the mutant recombinant proteins (Figure 4.8B). The DNA binding ability of each of

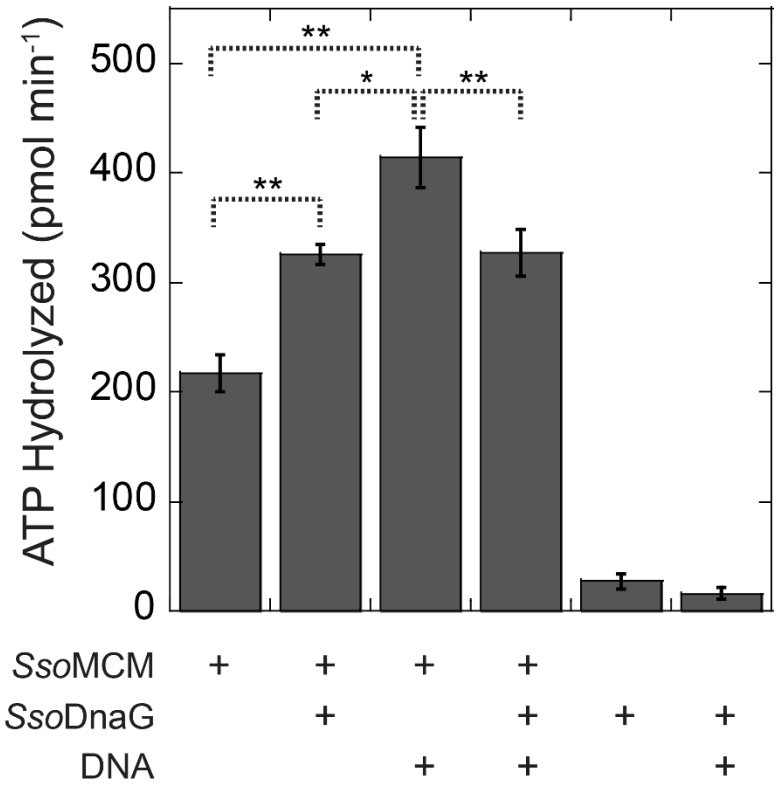


Figure 4.6: *Sso*DnaG stimulates ATPase activity of *Sso*MCM.
ATPase activity of *Sso*MCM without and with *Sso*DnaG measured at 60 ºC in the absence and presence of DNA. 4.2 µM *Sso*MCM (700 nM hexamer), 4.2 µM *Sso*DnaG and 1 µM forked DNA were used. DNA stimulated ATPase activity of *Sso*MCM 1.9-fold. In the absence of DNA, *Sso*DnaG stimulated the ATPase activity of *Sso*MCM 1.5-fold. However, the presence of DNA did not stimulate the ATPase activity of *Sso*MCM further in the presence of *Sso*DnaG. As a control, *Sso*DnaG alone and in the presence of DNA is shown to have no ATPase activity above background. P-values: * = $p < 0.05$; ** = $p < 0.01$. Error bars represent the standard error for each reaction.

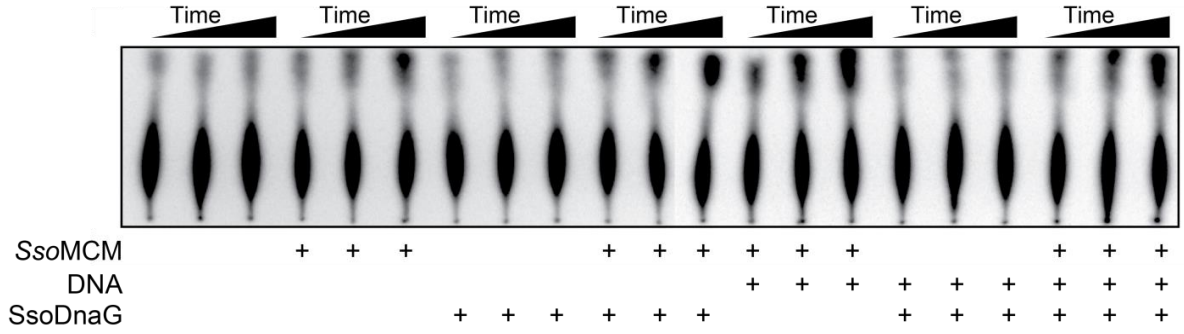


Figure 4.7: Example *Sso*MCM ATPase TLC plate.
Representative raw data from TLC separated ATPase assay at 60 °C. Quantification of the rate of ATP hydrolysis as a function of time (5, 10, and 15 minutes) is shown in Figure 4.6.

these mutants to 28mer ssDNA (5′Cy5) template was determined using fluorescence anisotropy. The D179A mutant had a 10-fold reduction in ssDNA binding affinity (K_d) for DNA (1.2 ± 0.1 μM) compared to wild-type (WT) *Sso*DnaG (0.10 ± 0.01 μM) (Figure 4.8C). The K_d values for the D220A and D222A mutants are significantly larger than that for D179A but cannot be calculated accurately with the experimental concentration range used (0.05–9 μM) as high concentrations of *Sso*DnaG promote aggregation, rendering the enzyme inactive.

The priming ability of *Sso*DnaG D179A (17.2 ± 1.0 pmol min^{-1}) was not significantly affected compared to WT (21.6 ± 0.5 pmol min^{-1}), but mutations at D220A (6.3 ± 0.5 pmol min^{-1}) and D222A (6.6 ± 0.5 pmol min^{-1}) had a 3-fold reduction in priming activity (Figure 4.8D and E). The composition of primers generally consists of two major products, a 13mer and a tetramer. However, the primer product distribution differs for D222A, where the amount of tetramer (4mer) is drastically reduced in favor of formation of a pentamer (5mer) product (Figure 4.8D).

4.4.4 Direct interactions of *Sso*DnaG with both DNA and *Sso*MCM inhibit unwinding

To further test the mechanism of *Sso*MCM unwinding inhibition by *Sso*DnaG, we titrated the *Sso*DnaG mutants in unwinding reactions. WT or the three metal binding mutants of *Sso*DnaG (D179A, D220A, and D222A) were preincubated with radiolabeled forked DNA for 5 min at 60 °C and then *Sso*MCM was added to initiate the reaction (Figure 4.9). In all cases, increasing concentrations of *Sso*DnaG inhibited *Sso*MCM unwinding, but the trends were different for the DNA binding mutants (Figure 4.10A). *Sso*DnaG concentrations less than 1 μM show little or no inhibition of *Sso*MCM. Unwinding inhibition is strongest at or above 2 μM where the *Sso*MCM:*Sso*DnaG ratio is 2:1. To show specificity, little to no unwinding inhibition by *Sso*DnaG was seen with the *Ec*DnaB helicase compared to *Sso*MCM (Figure 4.11).

Interestingly, each of the aspartate mutants of *Sso*DnaG has less of an inhibitory effect on MCM unwinding than WT. At 4 μM where *Sso*DnaG:*Sso*MCM is essentially 1:1, the mutants inhibit 2-fold, while WT *Sso*DnaG inhibits 4-fold. We also tested the effect of increasing the DNA concentration on unwinding of *Sso*MCM in the absence and presence of equal molar amounts of *Sso*DnaG (Figure 4.10B). For these experiments, *Sso*MCM and *Sso*DnaG were added simultaneously to initiate the unwinding reaction. At the lowest concentration of forked DNA,

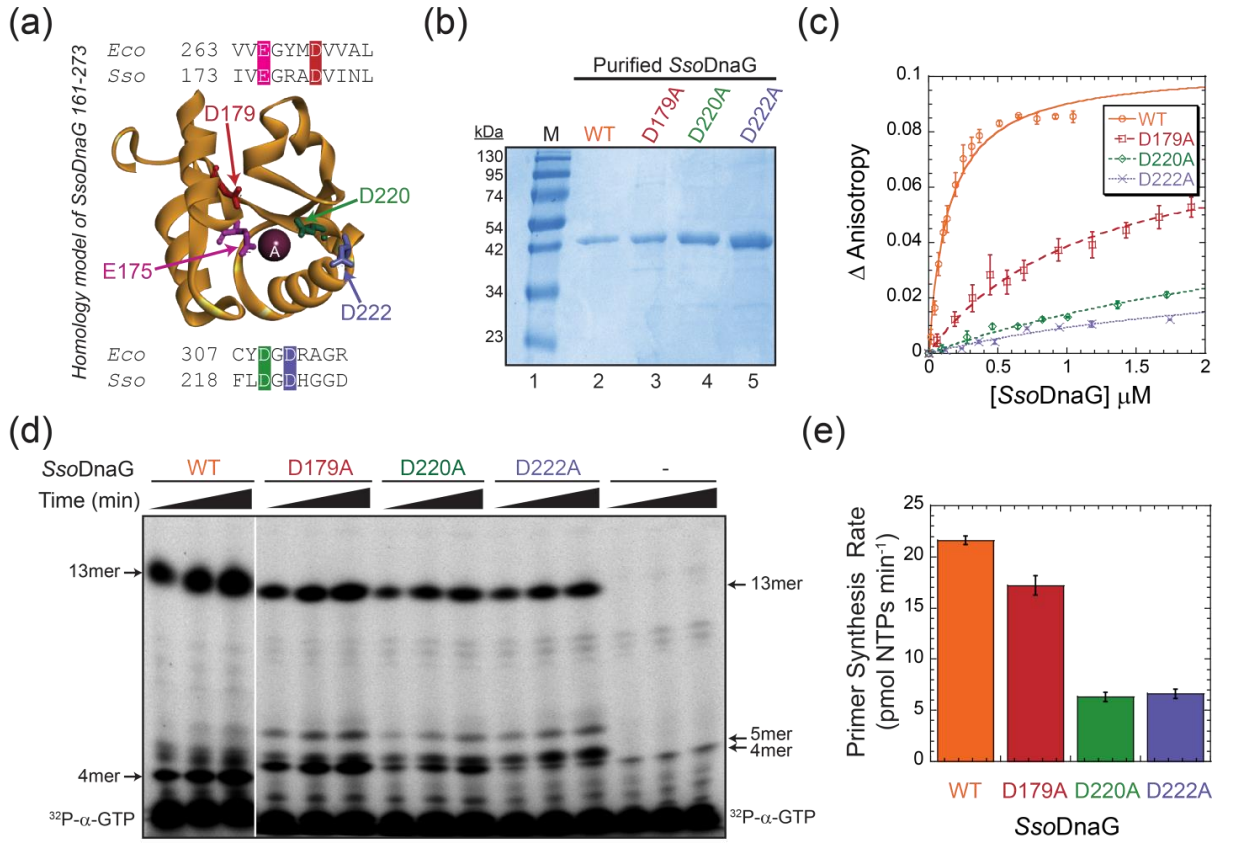


Figure 4.8: Conserved *Sso*DnaG active site residues disrupt DNA binding and priming activity.
A) Homology model of the TOPRIM domain of *Sso*DnaG (residues 161-273) highlighting conserved catalytic E175 (pink) and proposed metal binding residues D179 (red), D220 (green), and D222 (purple). The divalent metal A is positioned from alignment with the *Ec*DnaG structure (PDB ID: 3B39). **B)** Coomassie stained SDS-PAGE gel showing wild-type (WT) and mutant forms of *Sso*DnaG. **C)** Fluorescence anisotropy DNA binding experiments of each *Sso*DnaG construct to a 28 mer ssDNA (5'Cy5). Dissociation constants obtained from a fit to Equation 1 are 0.10 ± 0.01 and 1.2 ± 0.1 μM for wild-type (WT, orange -o-) and D179A (red -□-). The K_ds for D220A (green -◊-) and D222A (purple -x-) are much larger and not quantifiable over the concentration range tested (0.05 - 9 μM). **D)** DNA priming assays were separated on a 20% denaturing acrylamide gel and phosphorimaged showing primarily 13mer, 4mer, and dimer products synthesized from ^{32}P-α-GTP and **E)** quantified. The error bars represent the standard error from at least three separate experiments.

there is a significant reduction in the fraction unwound. As the DNA concentration increases towards 700 nM, where there is equal molar *Sso*MCM hexamer:DNA, there is a reduction in *Sso*DnaG's ability to inhibit unwinding consistent with competition of *Sso*MCM and *Sso*DnaG for binding DNA analogous to the ATPase assays. The individual binding affinities (K_d) for hexameric *Sso*MCM is 4-fold lower (26 nM) [123] than that for *Sso*DnaG (Figure 4.8). Therefore, at low DNA concentrations, *Sso*MCM binding to DNA will be favored followed by secondary binding of *Sso*DnaG to the *Sso*MCM–DNA complex. At higher DNA concentrations, the

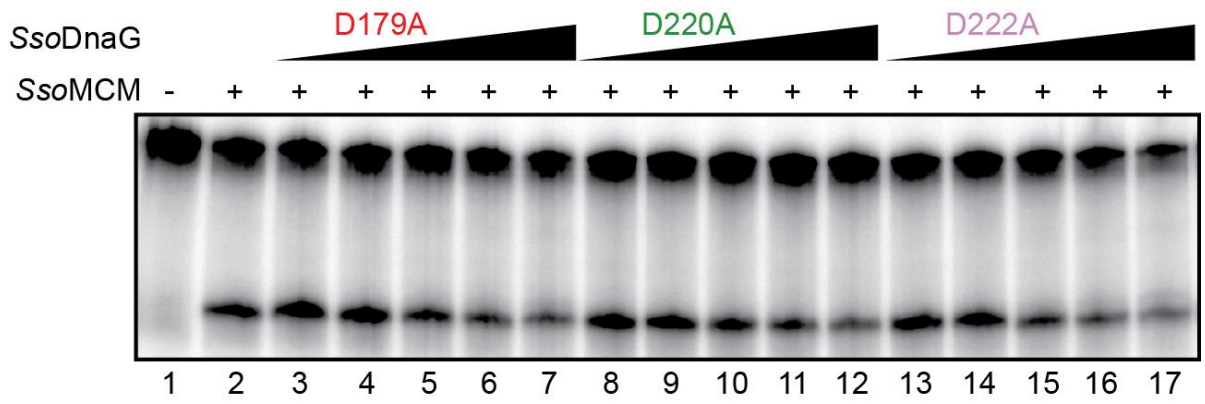


Figure 4.9: DNA unwinding assays for *Sso*MCM in the absence or presence *Sso*DnaG DNA binding mutants. Representative DNA unwinding assay on a denaturing acrylamide gel for 700 nM *Sso*MCM hexamer alone and in the presence of *Sso*DnaG mutants (350 nM, 700 nM, 2.1 µM, 4.2 µM and 6.3 µM for each set). Reactions are with 15 nM forked DNA at 60 ºC for 15 minutes.

equilibrium of the *Sso*MCM–DNA–*Sso*DnaG ternary complex will rearrange in favor of individual binding of each enzyme to DNA.

In order to clearly identify the presence of a stable ternary complex, we used fluorescent EMSA experiments to show and quantify binding of *Sso*DnaG to *Sso*MCM–DNA (Figure 4.12A). Unlabeled *Sso*MCM was titrated into a constant amount of 3′-tail-30 nt (5′Cy3) ssDNA (100 nM) in the presence and absence of N-terminally Cy5-labeled *Sso*DnaG (200 nM; hereinafter Cy5-*Sso*DnaG). Cy5-*Sso*DnaG in the absence of *Sso*MCM does not significantly form a stable complex with DNA (Figure 4.12A, lane 2). Titration of *Sso*MCM promotes the ternary complex formation as indicated by an increase in the Cy5-*Sso*DnaG signal towards the top of the gel (Figure 4.12A, lanes 3–5). Notably, the concentration of Cy5-*Sso*DnaG does not change across the gel; rather, it becomes more concentrated near the top of the gel only as *Sso*MCM is titrated. *Sso*DnaG alone does not efficiently enter the gel under these experimental conditions; nor does it form a stable complex with DNA at these concentrations [96]. At the higher concentrations of *Sso*MCM (especially lanes 3–5), there is a significant amount of ternary complex formed presumably by sequestering *Sso*DnaG within the complex for entry into the gel. Importantly, titration of unlabeled *Sso*MCM to Cy3-labeled ssDNA is required before *Sso*DnaG is able to bind and form the ternary complex. When *Sso*DnaG is absent in the reaction, binding of *Sso*MCM to DNA shifts the ssDNA towards the top of the gel similarly to form a binary complex (Figure 4.13B, lanes 3–10). The molecular mass of a binary *Sso*MCM–DNA complex alone is ~ 500 kDa and cannot be resolved from larger ternary complexes that include *Sso*DnaG. Background Cy5 fluorescence

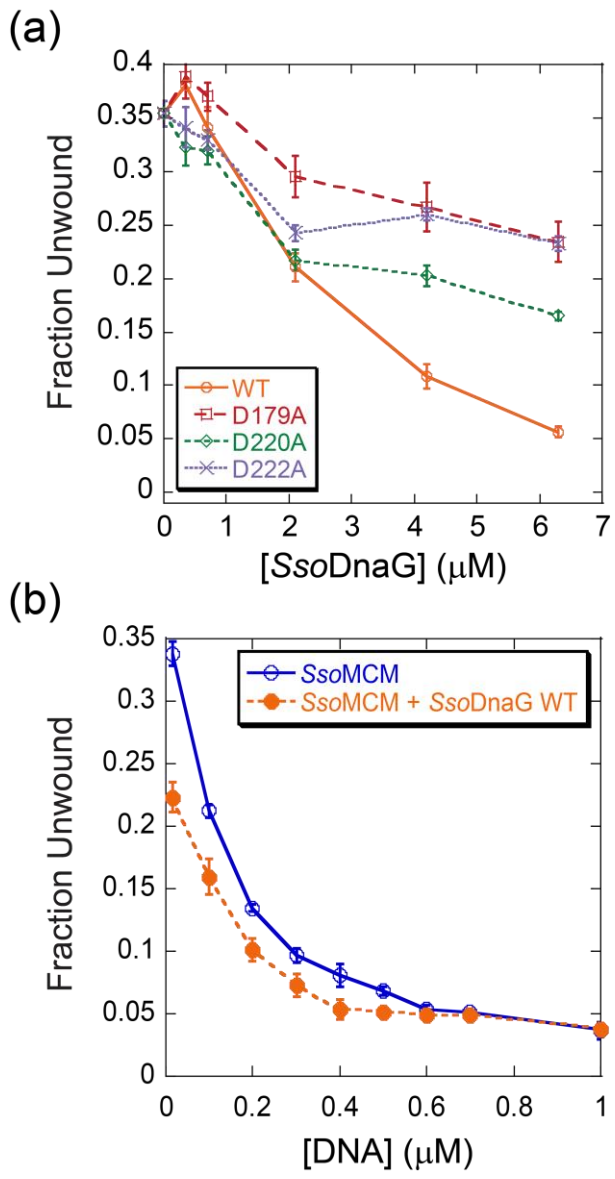


Figure 4.10: *Sso*MCM unwinding activity as function of increasing *Sso*DnaG or forked DNA concentration.
A) Fraction unwound of forked DNA by *Sso*MCM hexamer (700 nM) as a function of *Sso*DnaG concentration at 60 °C for 15 min. The concentration dependence inhibition by wild-type (WT, -o-), D179A (-□-), D220A (-◊-), or D222A (-x-) *Sso*DnaG are shown. Error bars represent the standard error from at least three independent unwinding experiments. **B)** DNA titration of *Sso*MCM hexamer (700 nM) unwinding in the absence and presence of *Sso*DnaG (4.2 μM). Fraction unwound of forked DNA by *Sso*MCM in the absence (-o-) and presence (-●-) of *Sso*DnaG at 60 °C for 15 min. Error bars represent the standard error from at least three independent unwinding experiments.

intensity due to overlapping excitation spectra in Cy3-only lanes was subtracted from the Cy3–Cy5 lanes from parallel reactions (Figure 4.13). A plot of the Cy5-*Sso*DnaG fluorescence intensity in the complex band towards the top of the gel averaged from at least three independent reactions gives an apparent K_d of 360 ± 60 nM for the *Sso*MCM–DNA–*Sso*DnaG ternary complex (Figure 4.12B). This value is estimated to be at least an order of magnitude lower than that of an *Sso*MCM–*Sso*DnaG binary complex in the absence of DNA isolated in the GST pull-down assays above and data not shown. More important than absolute quantification, these results show that binding of *Sso*DnaG is dependent on the presence of both *Sso*MCM and DNA to form a ternary complex.

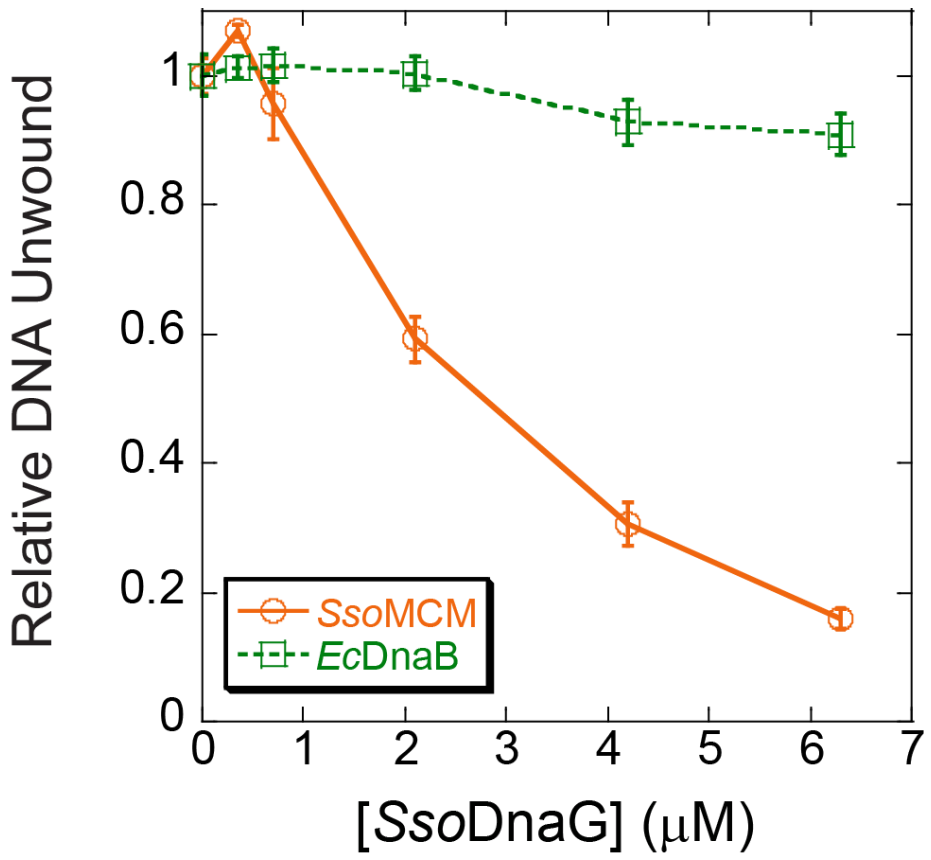
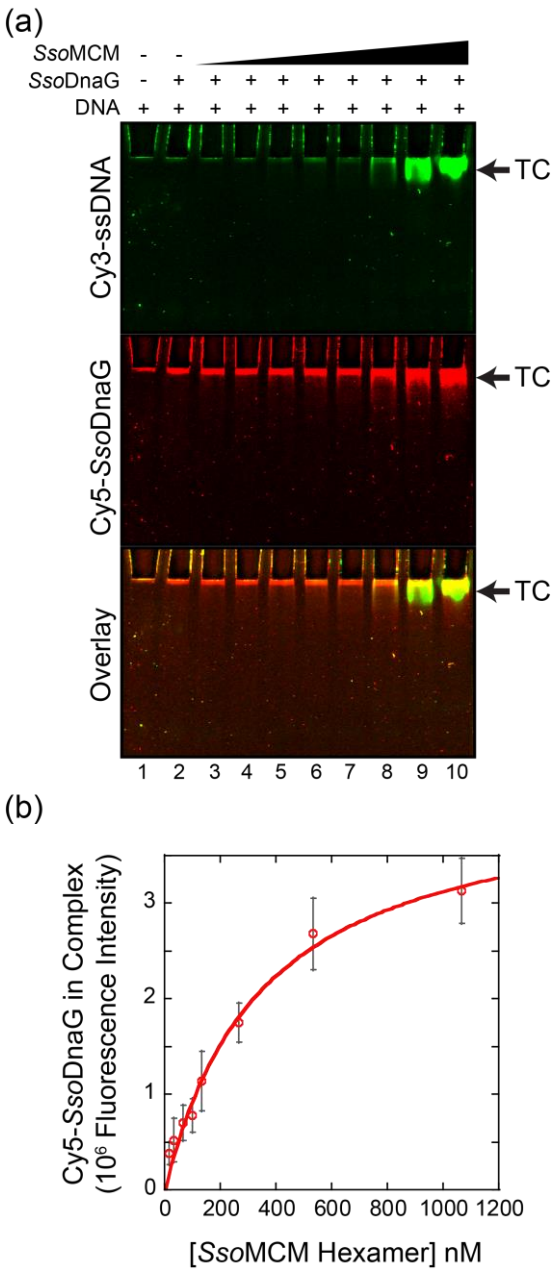


Figure 4.11: *Sso*MCM vs *Ec*DnaB Unwinding in the presence of *Sso*DnaG. *Sso*MCM (orange, -o-) or *Ec*DnaB (green, -□-) unwinding activity as function of increasing *Sso*DnaG concentration. Relative fraction unwound of forked DNA by 700 nM *Sso*MCM hexamer (@ 60 °C) as in Figure 6a or 700 nM *Ec*DnaB hexamer (@ 37 °C) for 15 minutes for *Sso*MCM and 10 minutes for *Ec*DnaB as a function of *Sso*DnaG concentration.

4.4.5 Mapping the site of interaction on *Sso*MCM using fluorescent GST pull-down assays

In order to determine the location of the interaction on the helicase with *Sso*DnaG, we utilized a fluorescent GST pull-down assay. Fluorescent pull downs provided greater sensitivity of detection in the presence of small amounts of contaminating degradation products from GST-*Sso*DnaG. Truncated (1–267, 106–612, 267–612, and 1–612) and full-length (1–686) *Sso*MCM constructs containing combinations of the identified domains (Figure 4.14A) were cloned, expressed, and purified without any potentially interfering affinity tags (Figure 4.15). The truncated *Sso*MCM constructs had oligomeric states generally consistent with His-tagged versions previously published [103] with 267–612 in a monomer–dimer state and 1–612 and 106–612 as hexamers (data not shown). The only subtle difference is that 1–267 is a dimer–trimer instead of a monomer–dimer [103] as identified by gel filtration (data not shown), but this is likely a concentration-



dependent effect. Each *Sso*MCM construct was labeled specifically with Cy5 at the N-terminus, and the ratios of *Sso*MCM:Cy5 were essentially 1:1 in all cases. GST-*Sso*DnaG was incubated with various Cy5-labeled *Sso*MCM truncation constructs and immobilized to a glutathione Sepharose column. After extensive washing, we observed an interaction between *Sso*DnaG and all *Sso*MCM constructs except 267–612 after elution with glutathione (Figure 4.14B). Mock experiments in the absence of GST-*Sso*DnaG showed no significant background *Sso*MCM binding for any construct (Figure 4.16). Because both 1–267 and 106–612 showed an interaction with GST-*Sso*DnaG, we can confine the interaction site to the B/C domain of *Sso*MCM (residues 106–266). Mapping residues 106–266 onto the hexameric model of the crystal structure of *Sso*MCM [113] identified the extreme N-terminal tier as the most probable *Sso*DnaG interaction site (Figure 4.14C, yellow).

Figure 4.12: Fluorescent EMSA demonstrating the ternary complex (TC) of *Sso*MCM-ssDNA-*Sso*DnaG. **A)** Separate channels of a fluorescent EMSA showing signals for 100 nM 66mer 3'-tail-30 nt (5'Cy3) ssDNA, Cy5-DnaG, and the overlay for a titration of *Sso*MCM (16-1066 nM hexamer) in the presence Cy5-*Sso*DnaG (200nM) (lanes 3-10). Lane 1 is Cy3-ssDNA alone and lane 2 is *Sso*DnaG and Cy3-ssDNA. **B)** A plot of the Cy5-*Sso*DnaG intensity contained in the complex band versus *Sso*MCM hexamer concentration. A fit of the data to Equation 2 gives a K_d = 360 ± 60 nM. Error bars represent the standard error from at least three independent EMSA experiments.

4.5 DISCUSSION

The DNA replisome is a large dynamic assembly of multiple proteins required for efficient synthesis on the leading and lagging strands. The interactions and coordination of protein subunits and complexes within the bacterial replisome have been well characterized structurally and biochemically. The eukaryotic replisome is much more complex, adding an additional dimension of regulation to DNA replication. Therefore, the archaeal DNA replisome is quickly becoming an appropriate core model for the more complex eukaryotic replisome, but it suffers from incomplete evolutionary linearity. For example, DNA priming in archaea is complicated by the fact both bacterial (DnaG) and eukaryotic-type (PriSL) primases are conserved throughout this domain with similar *in vitro* priming activities [96]. Here, we report the surprising result that the

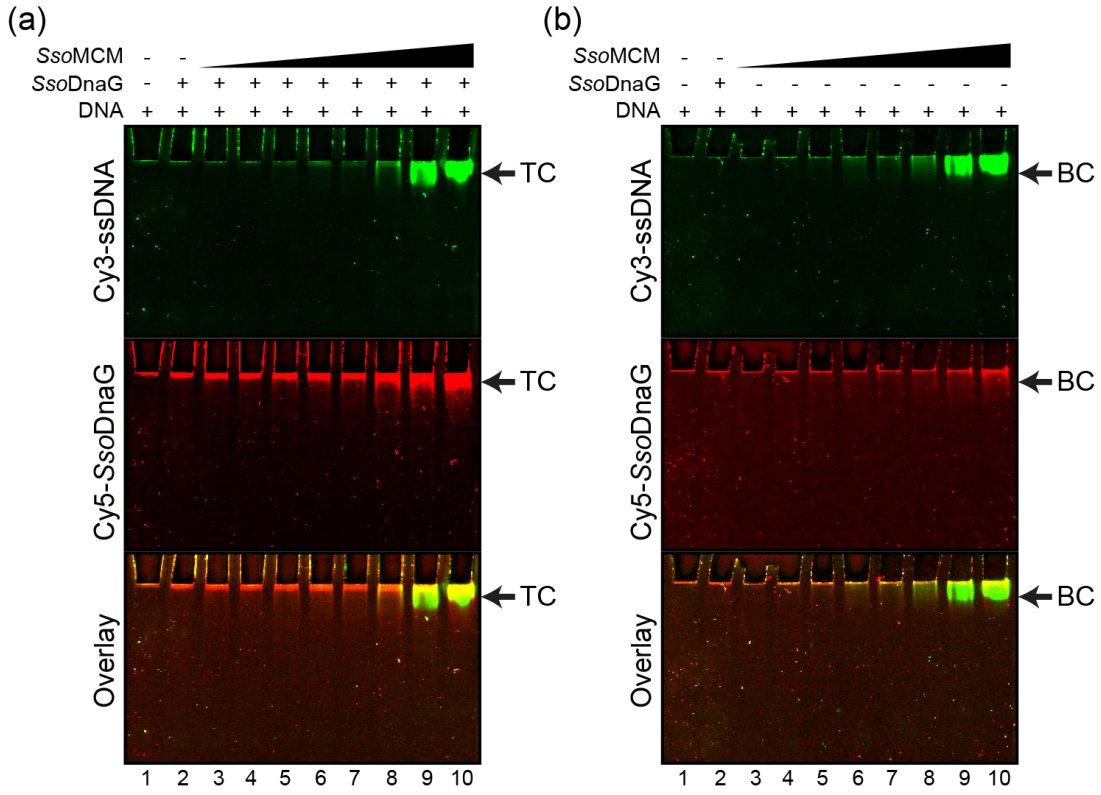


Figure 4.13: Fluorsecent EMSA in the presence and absence of *Sso*DnaG.
Fluorescent EMSA demonstrating the (a) ternary complex (TC) of SsoMCM-ssDNA-SsoDnaG or (b) the binary complex (BC) of SsoMCM-ssDNA in the presence and absence of SsoDnaG, respectively. Separate channels of a fluorescent EMSA showing signals for 100 nM 66mer 3'-tail-30 nt (5'Cy3) ssDNA, 0 or 200 nM Cy5-DnaG, and a titration of unlabeled *Sso*MCM (16-1066 nM hexamer). A Cy5-DnaG-DNA complex (lane 2) does not stably form alone at these concentrations. Only upon increasing MCM concentrations (lanes 3-10) does the Cy5 signal (from constant DnaG concentration) appear near the top of the gel. Free unbound Cy3-DNA was electrophoresed off of this particular gel to highlight BC and TC.

bacterial-like *Sso*DnaG physically interacts with the eukaryotic-like *Sso*MCM helicase, forming a ternary complex on DNA. The site of interaction at the N-terminus of *Sso*MCM is analogous to primosome complexes from bacteria and phage. Conserved acidic active-site mutants of *Sso*DnaG (D220A and D222A) show decreased binding affinity to ssDNA as well as decreased priming activity verifying functional homology to bacterial *Ec*DnaG. The unwinding activity of *Sso*MCM is inhibited in the presence of *Sso*DnaG by probable restriction of the *Sso*MCM conformation or disruption of the interaction with the DNA strands. Characterization of this unique primosome-like complex supports a role for the *Sso*DnaG primase in archaeal DNA replication as well as provides important evolutionary links between bacterial and eukaryotic DNA replication mechanisms.

The interaction between the DnaG primase and DnaB helicase constitutes the bacterial primosome and is a well-characterized subcomplex within the replisome [194, 195]. After separation of bacterial genomic DNA, the lagging strand passes through the central channel of the DnaB helicase before encountering the DnaG primase for RNA primer synthesis [84]. DNA priming on the lagging strand of eukaryotes is not as well characterized but is thought to be coupled to polymerase ε through interaction with the GINS complex [196, 197]. Although a direct primase–MCM interaction has never been detected in archaea, there is precedent from a single gene from the bacterium, *Bacillus cereus*, which encodes an MCM helicase with a putative DNA primase domain at the N-terminus [198, 199]. Interestingly, direct interactions between the eukaryotic primase (pol–prim) and simian virus 40 large T antigen (a functional MCM homolog) are known to be required for efficient primer synthesis most likely by orientating pol–prim in a proper orientation for catalysis and increased binding affinity [82, 200].

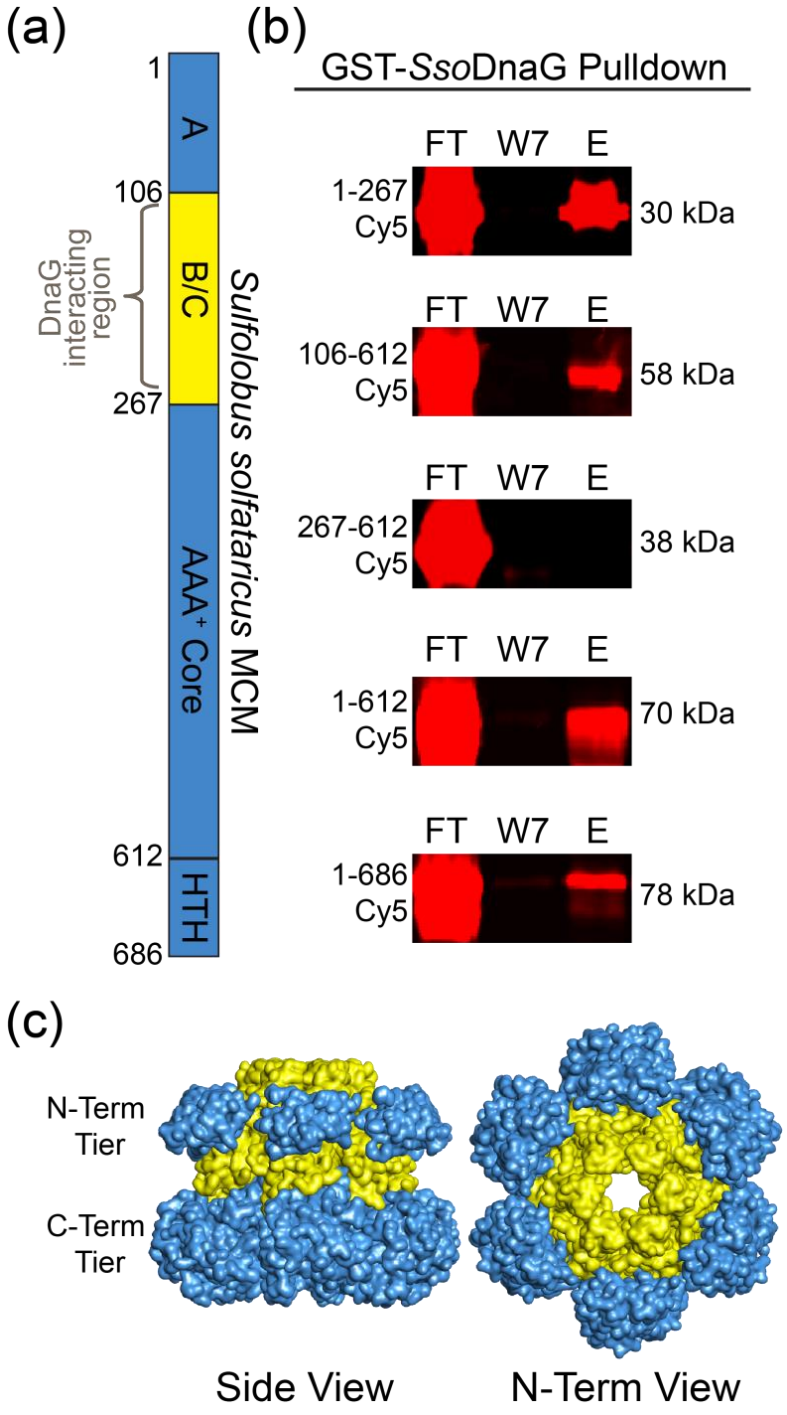


Figure 4.14: Binding of *Sso*DnaG to *Sso*MCM is localized to MCM's N-terminal B/C domains.
A) Schematic of the protein sequence of *Sso*MCM mapped for *Sso*DnaG binding identifying individual domains. **B)** Fluorescent GST-*Sso*DnaG pulldown assays showing binding of N-terminal Cy5-labeled domains of *Sso*MCM (amino acids 1-267, 106-612, 267-612, 1-612, and full length 1-686). Shown are fluorescent images of specific regions of SDS-PAGE gels of *Sso*MCM domain flow through (FT), the seventh wash (W7), and elution (E) with glutathione. **C)** *Sso*MCM hexamer model (PDB ID: 3F9V) highlighting the *Sso*DnaG interacting region (amino acids 106-267) in yellow.

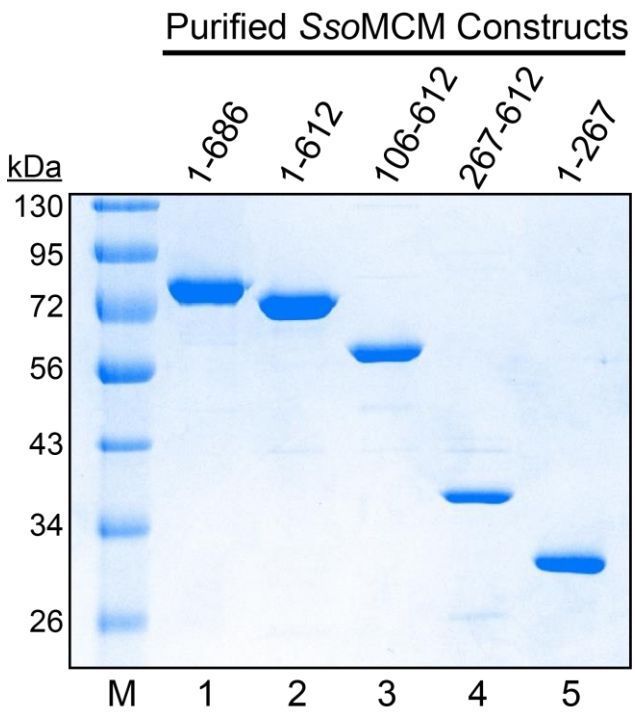


Figure 4.15: Coomasie stained SDS-PAGE gel of *Sso*MCM constructs.
Shown are purified full length untagged *Sso*MCM (1-686) and truncation (1-612, 106-612, 267-612, and 1-267) proteins.

4.5.1 Archaea DNA primase dichotomy

Archaeal organisms are the only domain of organisms that contain two purported DNA primases: a eukaryotic-like PriSL and a bacterial-like DnaG. Archaeal *Sso*PriSL is known to interact with the *Sso*GINS complex involved in helicase complex regulation, assembly, and function but has no effect on priming or unwinding activity [67]. Instead, it is hypothesized that *Sso*GINS acts as a molecular bridge between *Sso*MCM and *Sso*PriSL coupling but not increasing activities. The clamp loader, *Sso*RFC, is also found to interact with *Sso*PriSL, but its presence inhibited full-length primer formation by competing *Sso*PriSL off of the DNA template [191]. Interaction of *Sso*PriSL with either *Sso*GINS or *Sso*RFC could support a role in primer handoff on the lagging strand but may also be consistent with trailing the *Sso*MCM helicase complex on the leading strand and behaving more like a DNA polymerase than a primase under certain conditions. In addition to ribonucleotide synthesis ability, archaeal PriSL has robust deoxynucleotide synthesis activity leading to long template-dependent and -independent DNA products [58, 63, 65]. Phylogenetic analysis has revealed that archaeal PriSL may have evolved from a common ancestor of the B-family DNA polymerase, [181] but continued evolution into eukaryotes resulted in DNA polymerase X-family members known for terminal deoxytransferase activity and participation in double-strand break (DSB) repair [56, 201]. Analogously, *Sso*PriSL by itself has been shown to have promiscuous terminal deoxytransferase activity [58] that facilitates discontinuous

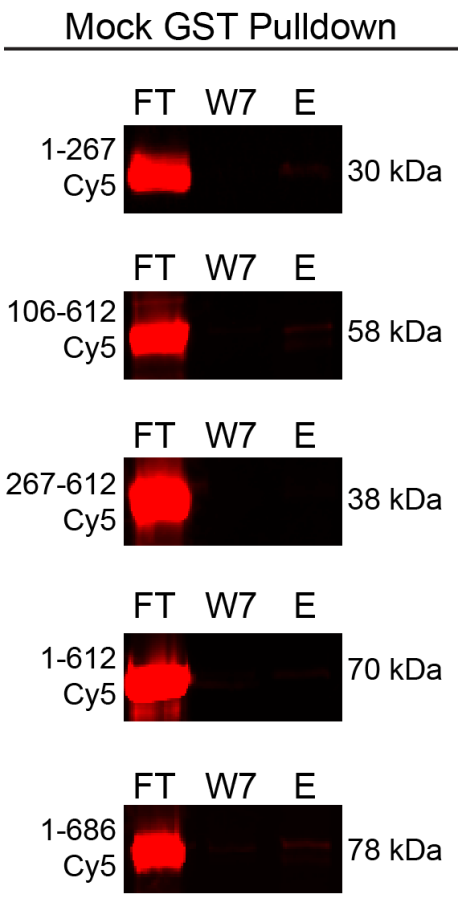


Figure 4.16: Mock fluorescent pulldown assays.
Mock fluorescent pulldown assays with GST resin and N-terminal Cy5-labeled domains of *Sso*MCM (amino acids 1-267, 106-612, 267-612, 1-612, and full length 1-686). Shown are fluorescent images of specific regions of SDS-PAGE gels of *Sso*MCM domain flow through (FT), the seventh wash (W7), and elution (E) with glutathione. No significant background binding of any Cy5-labelled MCM construct is noted.

polymerization across multiple templates mimicking DSB repair [66]. *Sso*PriSL also interacts with Rad50, which is involved in DSB repair [202] and is upregulated slightly in the presence of UV damage [203]. Although PriSL is suggested to be the DNA primase in archaea, the experimental evidence is actually more consistent with a DNA polymerase repair role [66, 204].

Unfortunately, experimental results on the role of archaeal DnaG are also conflicting. Previously, *Sso*DnaG was found to interact with the exosome, but definition of a functional role is hampered by no detectable associated enzymatic activity [98, 205]. The exosome is a large multisubunit complex responsible for degradation and turnover of nascent mRNA transcripts and has exactly the opposite activity of a DNA primase. *Sso*DnaG has been found to exist in at least two different populations in both soluble and membrane-bound states in the archaeal cell [206-208]. Interestingly, DNA replication in archaea was also recently found to be localized at the same peripheral membrane location [209]. It is possible that the oligomeric state of *Sso*DnaG or interactions with other proteins modulate its activity to more than one location and metabolic process in archaea.

On the other hand, the *in vitro* enzymatic primase ability of archaeal *Sso*DnaG is clear. *Sso*DnaG synthesizes discrete RNA primers of 13 nt in length as well as smaller dimer and tetramer products selecting against deoxyribonucleotides [96]. Moreover, the core TOPRIM active site of DnaG is conserved between bacteria and archaea. In addition to the essential glutamate involved in the chemistry of primer synthesis [96], we have now characterized additional proposed metal binding aspartates that, when mutated, mimic the phenotypes of *Ec*DnaG mutants [72, 74, 192, 193]. Interestingly, only E175Q abolished priming activity [96], while D220A and D222A

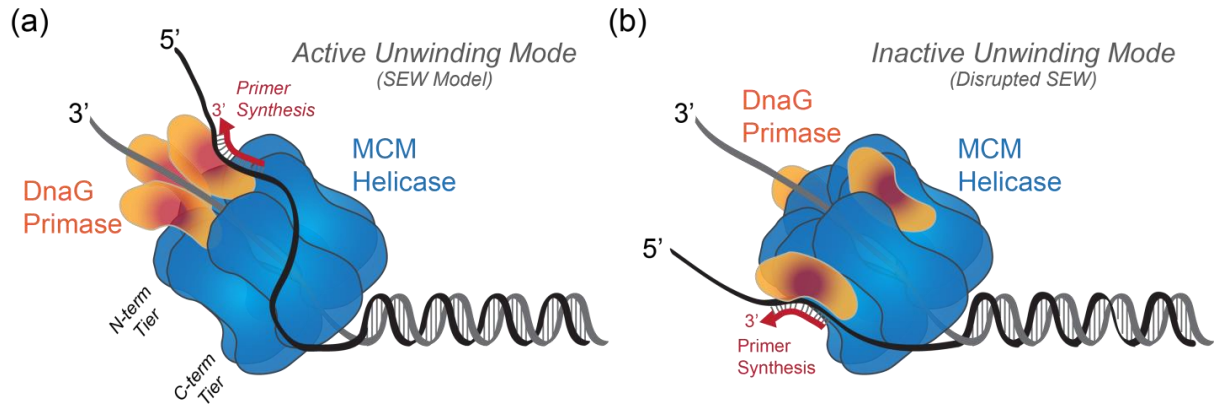


Figure 4.17: Active and inactive unwinding mode based on the SEW model.
Model of archaeal DnaG-MCM primosome complex on DNA in proposed (a) active and (b) inactive ternary complexes. MCM (blue) encircles the 3' strand placing the C-terminal tier toward the duplex region and the N-terminal tier. The excluded 5' strand interactions on the exterior surface of the MCM helicase in the steric exclusion and wrapping (SEW) model. DnaG (orange) interacts at the extreme N-terminal tier of MCM poised to synthesize RNA primers on the lagging strand. Inhibition of unwinding but not priming could be the result of an alternatively positioned 5'-tail that disrupts the SEW model and/or more loosely arranged MCM hexamer conformation on DNA.

only reduce priming. *Sso*DnaG also has a relatively high affinity to bind ssDNA even though it lacks the N-terminal zinc binding domain found in *Ec*DnaG. The metal binding aspartate dyad mutants (D220A or D222A) in particular decreased the binding affinity of DNA to immeasurable values. In *E. coli*, the homologous residues (D309/D311) have recently been verified to coordinate metals A and B, respectively, required for nucleotide binding and catalysis [74]. Mutation of *Sso*DnaG D179A reduced DNA binding to micromolar affinity but did not significantly affect priming under the conditions tested. The homologous aspartate (D269) in *Ec*DnaG seems to coordinate a third metal C only upon nucleotide binding [74]. Therefore, it seems that for *Sso*DnaG, coordination of metals A and B by E175, D220, and D222 disrupts binding of catalytic metal A, reducing DNA binding and priming activity, while coordination of metal C by D179 has much less of an inhibitory effect. The cumulative mutagenesis results for *Sso*DnaG are strikingly similar to that for *Ec*DnaG and suggest a high degree of conserved structural and functional features of priming between the archaeal and bacterial TOPRIM domains.

4.5.2 Architecture of a potential archaeal primosome complex

After validating that archaeal *Sso*DnaG interacts with *Sso*MCM using yeast two-hybrid assays, pull-down assays, and EMSAs, we mapped the interacting region to the extreme N-terminal tier

of *Sso*MCM. Interestingly, this interaction site is analogous to interactions and orientations found in bacterial and phage primase–helicase complexes [84, 194, 210]. Importantly, *Sso*DnaG stimulates the basal ATPase activity of *Sso*MCM, suggesting that their interaction promotes the arrangement of the hexamer into a more active state. By separately monitoring ssDNA, *Sso*DnaG, and *Sso*MCM using fluorescent EMSA experiments, we can now verify that a ternary primosome complex exists with reasonable affinity. The interaction of *Sso*MCM and *Sso*DnaG without DNA is much weaker and is similar to interactions between DnaG and DnaB in *E. coli* [80]. Therefore, stable ternary complex formation is dependent on interactions of *Sso*DnaG with both DNA and *Sso*MCM.

We have previously suggested that *Sso*DnaG minimally forms a dimer on DNA and has the tendency to aggregate at high concentrations [96]. Bacterial DnaG also has the propensity to self-associate but only when bound to DnaB as a mechanism to control primer length [78, 211]. Because *Sso*MCM (267–612) had a significant monomer population and showed no detectable interaction with *Sso*DnaG, it is also possible that *Sso*DnaG interacts minimally with a dimer of *Sso*MCM. Nevertheless, the interaction of *Sso*DnaG is also seen with *Sso*MCM (1–267), consistent with N-terminal tier binding. The stoichiometry of bacterial or phage helicase to primase subunits varies from 1:1 to 6:1 [76, 79, 84, 212-214]. The absolute ratio of *Sso*MCM to *Sso*DnaG will require further future experiments, but in our model (Figure 4.17), newly separated and excluded DNA will interact with the external surface of *Sso*MCM in the steric exclusion and wrapping (SEW) unwinding model [123] before being handed off to the interacting *Sso*DnaG primase. Interactions of the lagging strand template on the external surface of *Sso*MCM will provide the flexibility needed for coupled unwinding and priming activities [215]. Our unwinding experiments in the presence of *Sso*DnaG and associated mutants would suggest that *Sso*MCM is rendered in an inactive conformation for unwinding possibly by disrupting the *Sso*MCM interaction with the 5′-tail known to be important for unwinding in the SEW mechanism [123]. Alternatively, *Sso*DnaG may alter the conformation of the *Sso*MCM hexamer and its interactions with the encircled 3′ DNA strand, inhibiting unwinding and preventing further stimulation of the ATPase rate. Because the priming activity is unaffected, the binding of DNA to *Sso*DnaG is unaffected. It may be that additional factors are required for formation of an optimal primosome conformation required for efficient coupled priming and unwinding activities.

Formation of the bacterial primosome increases both the priming and unwinding activities of DnaG and DnaB, respectively [77, 79, 80]. We were unable to see stimulation in either primer synthesis or DNA unwinding when *Sso*DnaG and *Sso*MCM were included together, but the ATPase activity of *Sso*MCM was stimulated in the presence of *Sso*DnaG. Analogously, *Sso*PriSL priming inhibition was also seen upon interaction with *Sso*RFC with increased *Sso*RFC ATPase activity [191]. Unlike *Ec*DnaG, *Sso*DnaG has a relatively high DNA binding ability by itself and may not require interactions with the helicase to increase catalytic priming ability through recruitment to the template. Increased ATPase activity of *Sso*MCM suggests that *Sso*DnaG interacts specifically and causes a conformational change within the hexamer to stimulate the hydrolysis activity. This conformation is different from an *Sso*MCM–DNA complex as further simulation in ATPase activity was not seen when forked DNA is added, suggesting that *Sso*MCM is interacting with DNA in an incompetent state.

The inactive *Sso*MCM–DNA–*Sso*DnaG ternary complex can be controlled by equilibria processes. Unwinding inhibition was strongest at or above stoichiometric concentrations of *Sso*DnaG or when DNA is limiting. When the conserved acidic *Sso*DnaG active-site residues were mutated or when DNA is in excess, the inhibitory effect on DNA unwinding and ATPase activity was reduced. Because the *Sso*DnaG aspartate mutants also show reduced DNA binding affinity, we suspect that unwinding inhibition is caused through interaction with both DNA and *Sso*MCM. The EMSA experiments show that *Sso*DnaG is assembled on the *Sso*MCM–DNA template, but the ATPase assays reflect that although *Sso*MCM is arranged in a more competent ATP hydrolysis state, its unwinding activity is compromised. Larger DNA templates or interactions with other pre-replication proteins may be required to load and arrange *Sso*MCM in an active unwinding complex in the presence of *Sso*DnaG.

In addition to the novel archaeal DnaG–MCM complex described here, it is also possible that alternative primosome complexes include PriSL as the active primase using GINS to mediate interactions between the MCM helicase, but without evidence of an associated change in activity, it is speculative. Alternatively, archaea may utilize a dual-primase system to initiate DNA replication on the leading and lagging strands with either PriSL or DnaG acting specifically on one strand. Unfortunately, stable associations of DNA primases with members of the replisome are not characteristic and therefore difficult to detect. Hence, future experiments aimed at dissecting stoichiometries, interaction sites, and the specificities of each archaeal primase compared with *in*

vivo synthesized RNA primers or Okazaki fragments will be essential in differentiating the roles of these primase enzymes in archaea.

4.6 ACKNOWLEDGEMENTS

We thank Karen Arndt for the PJ69-4A yeast strain. Thanks to the University of Pittsburgh Cancer Institute at the Hillman Cancer Center for use of the Typhoon phosphorimager. The *Ec*DnaB expression plasmid was graciously provided by Dan Kaplan. This work was supported by startup funds from the University of Pittsburgh, Department of Chemistry (M.A.T.) and by a Research Scholar Grant (RSG-11-049-01-DMC) to. M.A.T. from the American Cancer Society.

5 *SSO*MCM-*SSO*SSB INTERACTION

5.1 SUMMARY

*Sso*MCM and *Sso*SSB have been previously shown to interact in the literature, but the data was qualitative and primarily inferred. We have utilized unwinding assays, ATPase assays, FRET, 2-aminopurine fluorescence, pull-down assays, and fluorescent EMSAs to characterize this interaction more quantitatively. After extensive experimentation under a variety of assay conditions, we were unable to validate positive functional or physical interactions between *Sso*MCM and *Sso*SSB.

5.2 INTRODUCTION

DNA replication, repair, and recombination are all required processes that involve the generation of single-stranded DNA intermediates. During these processes, single-stranded DNA (ssDNA) is prone to chemical and nucleolytic cleavage. One form of protection of ssDNA comes from single-stranded binding proteins (SSBs) that are able to bind, stabilize and protect ssDNA. SSBs are present in all three domains of life and viruses and share a common fold, typically located towards the N-terminus, known as a oligosaccharide/oligonucleotide binding fold (OB fold) used for DNA binding [21, 216]. The flexible C-terminal domain, which is less conserved, typically participates in protein-protein interactions [36-38]. The oligomeric states, depending on the organism, varies from monomer through tetramer. The main eukaryotic version of SSB (containing four OB folds total) is known as Replication Protein A (RPA) and is a heterotrimer of the three subunits (RPA70, RPA30, and RPA14), where the number indicates its approximate kDa [17]. Bacterial SSB (containing one OB fold per monomer) is a more simplified version that forms a homotetramer in solution; *E. coli* SSB is 20 kDa [47, 217]. Archaeal *Sulfolobus solfataricus* SSB (containing one OB-fold per monomer) has been shown to exist in a number of oligomeric states by different groups including monomer [27-31], dimer [30], and tetramer [30, 32]. Recently, human SSB1 and 2 were discovered based on homology with archaeal SSB's (containing one OB fold per monomer) and seem to play an important role in both the DNA damage response pathway and overall genomic stability [20, 21]. A comparison schematic and ClustalW2 alignment of the four previously mentioned SSB's is shown in Figure 5.1 [168, 169]. In addition to harboring one N-terminal OB fold, each SSB also contains a disordered C-terminal region and charged C-terminus residues. Both hSSB1 and 2 contain a basic C-terminus, while *E. coli* and *Sso* contain an acidic C-terminus (Figure 5.1B).

In addition to binding DNA, SSB's have also been shown to interact with a variety of proteins performing different cellular functions. *E. coli* SSB interacts with at least 14 other proteins involved in DNA metabolism [24]. These *E. coli* SSB interactions typically occur within the disordered C-terminal domain which does not participate in DNA binding [218]. Similarly for *Sso*, the C-terminal domain participates in protein-protein interactions [27, 219]. *Sso*SSB has been shown to be able to detect DNA damage [40], interact with reverse gyrase [41], *Sso*RadA during homologous recombination [42], and *Sso*MCM, the DNA replicative helicase [9, 43]. Other

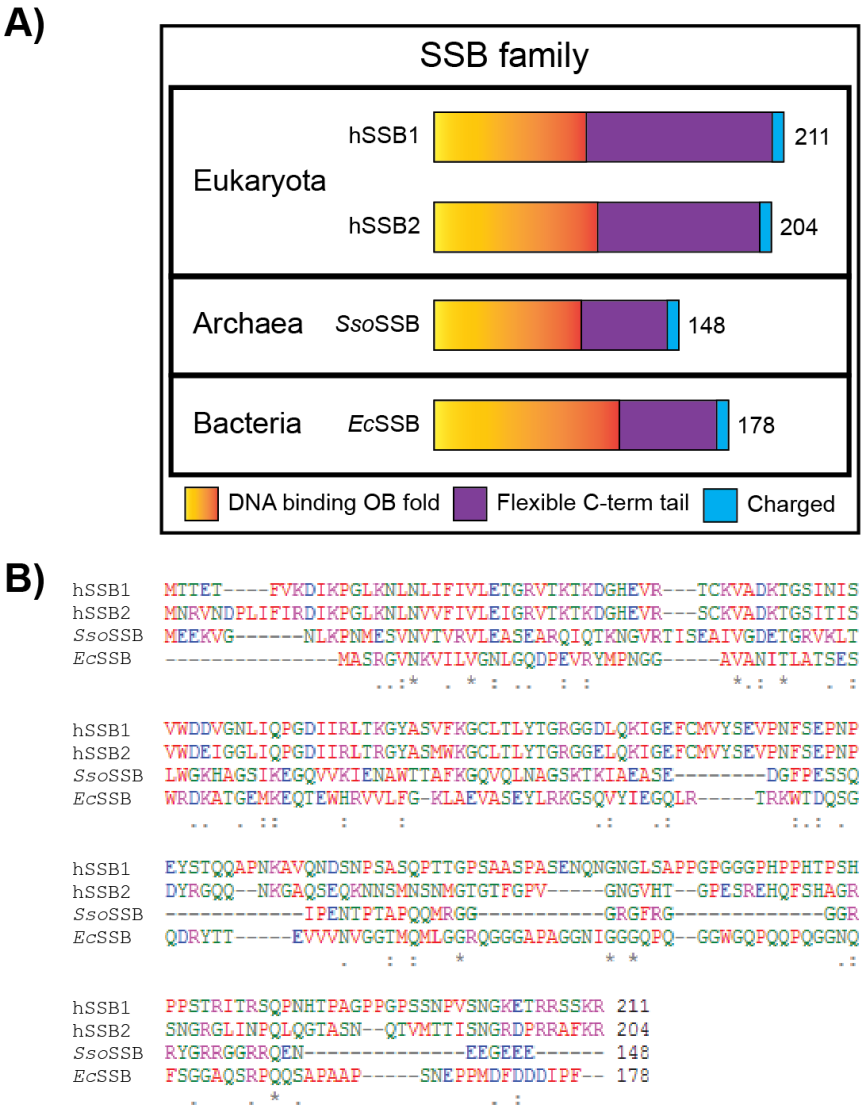


Figure 5.1: SSB comparison across the three domains
A) Single-stranded binding proteins (SSBs) domain organization and length in amino acids, where the orange region shows the DNA binding OB fold, the purple region shows the flexible C-terminal tail and the blue region represents charged residues at the C-terminus. *Sso* – *Sulfolobus solfataricus*, *Ec* – *Escherichia coli*, h – human. B) ClustalW2 alignment of hSSB1, hSSB2, *Sso*SSB, and *Ec*SSB.

analogous systems have shown interactions between DnaB (also a homohexameric DNA replicative helicase) and SSB including *Escherichia coli* [44], *Mycobacterium tuberculosis* [45], and *Helicobacter pylori* [46]. In *E. coli*, SSB stimulates DnaB unwinding but also inhibits ssDNA-dependent ATPase activity [44].

We have attempted to fully characterize the interaction between *Sso*MCM and *Sso*SSB. *Sso*SSB is monomeric in solution and inhibits *Sso*MCM's unwinding and ATPase activity in the

presence of DNA. *Sso*SSB, however may provide stabilization of MCM binding to forked DNA. However, we were unable to show a definitive physical interaction. It is our view that *Sso*SSB plays a regulatory role for *Sso*MCM, where at low concentrations it is able to protect recently unwound DNA, but at high concentrations it removes MCM, preventing unwinding.

5.3 MATERIALS AND METHODS

5.3.1 Materials

ATP was obtained from Invitrogen. Optikinase was purchased from USB. All other materials were from commercial sources and were analytical grade or better. Helicase buffer is used in all unwinding and binding reactions and consists of 125 mM potassium acetate, 25 mM Tris acetate (pH 7.5) and 10 mM magnesium acetate.

5.3.2 DNA substrates

Oligonucleotides (Table 5.1) were purchased from IDT Corp and gel purified [144]. [γ-^{32}P]ATP was purchased from Perkin Elmer (Waltham, MA) and Optikinase was used to ^{32}P label the 5′-end of DNA. Fluorescent DNA was synthesized and HPLC purified by IDT. Complementary DNA was added in a ratio of 1.2:1. ^{32}P-labeled and fluorescently-labeled DNA substrates were heated at 95 °C for 5 min and then cooled to room temperature after turning off the heat block.

5.3.3 *Sso*SSB cloning

*Sso*SSB was PCR amplified from genomic DNA using primers *Sso*SSB FWD 5'-CACCATGGAAGAAAAAGTAGGTAATCTAAAACC and *Sso*SSB REV 5'-CTACTCCTCTTCACCTTCTTCGTTTTC and then cloned into the pENTR SD-TOPO Gateway® cloning vector (Invitrogen). After DNA sequencing verification (U. Pittsburgh Genomics Core) *Sso*SSB was recombined into pDEST14 using LR Clonase II™. Briefly, a 10 µL reaction was prepared with pENTR *Sso*SSB was added to 1 µL of pDEST14 and 2 µL Clonase II in TE buffer. The reaction was vortexed briefly twice and incubated at room temperature for 1 hr. 1 µL of Proteinase K (2 µg/µL) was then added and incubated at 37 ºC for 10 min. 5 µL of the resulting reaction was transformed in Mach I cells, selected, and the resulting pDEST14-*Sso*SSB purified. pDEST14-*Sso*SSB (G81C) was prepared by Quikchange (Jessica Meyers).

Table 5.1: DNA substrates

DNA[1]	Length	Duplex	Sequence[2,3]
3'-tail-30 nt	66	36	5'-CACCTCTCCCTACGCTTCCCACCCACCCCGACCGGC ATCTGCTATGGTACGCTGAGCGAGAGTAGC
5'-tail-30 nt	66	36	5'-CGATGAGAGCGAGTCGCATGGTATCGTCTAGCCGGT CGGGGTGGGTGGGAAGCGTAGGGAGAGGTG
5'-tail-50 nt	86	36	5'-CGATGAGAGCCGATGAGAGCCGATGAGAGCGAGTC GCATGGATCGTCTAGCCGGTCGGGGTGGGTGGGAAG CGTAGGGAGAGGTG
3'-tail-30 nt (5'Cy3)	66	36	5'-**3**CACCTCTCCCTACGCTTCCCACCCACCCCGACCGG CATCTGC TATGGTACGCTGAGCGAGAGTAGC
3'-tail-31 nt (5'-B)(3'Cy3)	49	18	5'-**B**TGGCGACGGCAGCGAGGCTTTTTTTTTTTTTTTTT TTTTTTTTTTT**3**T
5'-tail-51 nt (5'Cy5)	69	18	5'-**5**TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTGCCTCGCTGCCGTCGCCA
3'-tail-30/33 nt (Int2AP)	71	38/41	5'-CACCTCTCCCTACGCTTCCCACCCACCCCGACCGGC **2**CTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
5'-tail-30 nt (2AP-2bp)	68	38	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGCCGGTC GGGGTGGCTGGGAAGCGTAGGGAGAGGTG
5'-tail-30 nt (2AP-5bp)	71	41	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAAGTGCCG GTCGGGGTGGGTGGGAAGCGTAGGGAGAGGTG

[1]nt – nucleotides; [2]Modifications are bolded and underlined; [3]2 – 2-aminopurine, 3 – Cy3, 5 - Cy5, B – biotin

5.3.4 *Sso*SSB purification

pDEST14-*Sso*SSB wild-type or G81C was transformed into Rosetta2 Bl21DE3 and grown in ZYP media using an auto-induction protocol for ~16 hr at 37 ºC [187]. The supernatant was resuspended

in buffer A (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA pH 8.0, 5 mM β-mercaptoethanol). Cells were lysed, sonicated, and heat treated at 70º C for 30 min. The supernatant was loaded onto an SP column and eluted with a 20 mL linear gradient from buffer A to B (buffer A with 1 M NaCl). Positive fractions were pooled and dialyzed into buffer A. *Sso*SSB was loaded onto a Heparin HP column and eluted with a 60 mL linear gradient from buffer A to B. Positive fractions were loaded onto an analytical gel filtration column (Superdex 200) that was equilibrated with *Sso*SSB storage buffer (buffer A with 10% glycerol). Positive fractions were pooled and concentrated. The concentration was determined using UV-Vis ($\varepsilon = 12660\ M^{-1}\ cm^{-1}$).

5.3.5 *Ec*SSB purification

Escherichia coli SSB (*Ec*SSB) was purified as previously described [220]. Briefly, *Ec*SSB was grown in 2XYT media to $A_{600} = 0.5$ and then induced with 0.2 mM IPTG. Cells were grown for seven hours at 37 ºC. After centrifugation, cells were resuspended in lysis buffer (50 mM Tris-HCl pH 8.3, 200 mM NaCl, 1 mM EDTA pH 8.0, and 10% w/v sucrose). Cells were lysed, sonicated, and centrifuged. 10% w/v polyethyleneimine (Sigma Aldrich) was added to 0.4% w/v (final concentration) slowly with stirring at 4 ºC to precipitate *Ec*SSB. After centrifugation, the pellet was resuspended in TGE buffer (50 mM Tris-HCl pH 8.3, 20% v/v glycerol, 1 mM EDTA pH 8.0) with 400 mM NaCl with stirring. After centrifugation, ammonium sulfate is added with stirring to 150 g/L over the course of 30 min. The pellet was gently resuspended in TGE buffer with 300 mM NaCl. Ammonium sulfate was subsequently added again to ~40% w/v with stirring. Finally, the pellet was resuspended in *Ec*SSB storage buffer (20 mM Tris-HCl pH 8.3, 500 mM NaCl, 1 mM EDTA pH 8.0, 1 mM β-mercaptoethanol, 50% v/v glycerol) and dialyzed overnight. The concentration was determined using UV-Vis ($\varepsilon = 26840\ M^{-1}\ cm^{-1}$).

5.3.6 Unwinding assays

Unwinding assays were performed as previously described [123]. Briefly, *Sso*MCM was incubated with 15 nM 5'-radiolabeled DNA (3'-tail-30 nt annealed to 5'-tail-30 nt, sequences in Table 5.1) in the presence of *Sso*SSB or *Ec*SSB as specified for 5 min at 60 ºC and initiated with ATP. Reactions were quenched with an equal volume of glycerol quench (0.5% SDS w/v, 50% v/v

glycerol, 0.1% w/v bromophenol blue, 100 mM EDTA pH 8.0, and 300 nM trap ssDNA. Quenched reactions were stored on ice until the addition of 4 µL Proteinase K 20 mg/mL (Thermo Fisher). Samples were incubated at 37 ºC for 1 hr to allow for digestion. DNA was resolved on the indicated percentage acrylamide gels with 0.1% SDS to remove any remaining bound protein. Gels were then phosphorimaged.

5.3.7 ATPase assays

ATPase assays for *Sso*MCM incubated in the absence or presence of unlabeled DNA (3'-tail-30 nt annealed to 5'-tail-30 nt, sequences in Table 5.1), *Sso*SSB or *Ec*SSB as previously described [97]. Briefly, 30 µL reactions were incubated at 60 ºC for 5 min and ATP was added to initiate the reaction. Samples were quenched at 5, 10, and 15 min after initiation in equal volumes of 0.7 M formic acid. A total of 0.8 µL of quenched reaction was spotted on Millipore TLC PEI Cellulose F, allowed to dry, resolved in 0.6 potassium phosphate (pH 3.5), and then phosphorimaged.

5.3.8 Bulk FRET

FRET experiments were performed using a Fluoromax-3 fluorometer (HORIBA Scientific, Edison NJ) 50 nM forked DNA (3'-tail-31 nt (5'-B)(3'Cy3) annealed to 5'-tail-51 nt (5'Cy5), sequences in Table 5.1) was used for each experiment with 2 µM *Sso*MCM monomer, if present. *Sso*SSB was titrated and spectra collected at each point indicated. Slit widths were adjusted for optimal signal. 550 nm excitation wavelength was used and emission was collected from 560-700 nm. Apparent energy transferred (E_{app}) was calculated according to:

$$E_{app} = I_A/(I_A + I_D) \quad (1)$$

where I_A and I_D are the acceptor and donor emission values at 670 and 570 nm, respectively.

5.3.9 2-aminopurine fluorescence

Steady-state fluorescence measurements were acquired in helicase buffer. Forked templates containing 2-aminopurine (sequences in Table 5.1) were heated to 95 ºC for 5 min and cooled to room temperature. 400 µL reactions containing 300 nM DNA ± 2.0 µM *Sso*MCM. *Sso*SSB, *Ec*SSB, or BSA was titrated to five concentrations: 230 nM, 460 nM, 1.2 µM, 2.3 µM, and 4.6 µM. Spectra were collected using a Fluoromax-3. The excitation wavelength was 315 nm and fluorescence emission was collected from 340-400 nm. The entrance and exit slits were adjusted for optimal intensities. Spectra were collected for buffer and titrant alone to correct for any background fluorescence. Samples containing DNA only were corrected by subtracting of the spectrum of buffer alone; while titrant containing samples were corrected by subtracting the spectrum for the titrant alone.

5.3.10 Pull-down assay

Columns were prepared by adding 400 µL of HisPur^{TM} Cobalt Resin (Thermo Fisher) in two Spin-X columns (Sigma Aldrich). They were washed five times with 400 µL wash buffer (25 mM Tris-HCl pH 8.3, 100 mM NaCl, 10 mM imidazole, and 5 mM β-mercaptoethanol) and centrifuged 30 s at 6000 × g. 200 µL samples of *Sso*MCM and *Sso*SSB (20 µM each) or *Sso*SSB (20 µM) in wash buffer were incubated at room temperature for 45 min before adding to spin columns for 2 min. Columns were centrifuged collecting the flow through. Columns were washed with 200 µL wash buffer seven times collecting the flow through after each centrifugation. Columns were eluted with 600 µL elution buffer (wash buffer with 500 mM imidazole). Samples were analyzed by 12% SDS-PAGE stained with Coomassie Brilliant Blue.

5.3.11 Protein labeling

Purified *Sso*MCM-His was N-terminally labeled with Cy5 succinimidyl ester in N-terminal labeling buffer (50 mM Hepes, pH 6.8, 100 mM NaCl, 5 mM β-mercaptoethanol, and 10% glycerol) as previously described [97]. Briefly, proteins were reacted with 3X molar excess of dye

for 30 min at room temperature. Excess fluorophore was removed by extensive dialysis in N-terminal labeling buffer. *Sso*SSB G81C was cysteine labeled using AlexaFluor 488 maleimide in cysteine labeling buffer (25 mM HEPES pH 7.4 and 100 mM NaCl). 1.5X molar excess of dye was added and nutated for 1 hr at room temperature. Another 1.5X molar excess was added and nutated for 2 hr at room temperature. 2 mM BME was added to quench the labeling reaction and G-25 columns followed by extensive dialysis were used to remove excess fluorophore. Labeling efficiency was determined through UV-Vis spectroscopy.

5.3.12 Fluorescent electrophoretic mobility shift assays

Fluorescent EMSAs were performed as previously described [97]. Cy5 N-terminally-labeled *Sso*MCM and Alexa Fluor 488 labeled *Sso*SSB (G81C) were used with Cy3-labeled forked DNA (3'-tail-30 nt (5'Cy3) annealed to 5'-tail-50 nt, sequences in Table 5.1). Briefly, 10 µL reactions were prepared by the stepwise addition of Cy3-labeled DNA, Cy5-labeled *Sso*MCM or unlabeled BSA, if present, and Alexa Fluor 488-labeled *Sso*SSB. Reactions were incubated for 10 min at 60 ºC to promote native thermodynamic complex formation and resolved on 5% native polyacrylamide gels. The gels were imaged using a Typhoon phosphorimager with separate fluorescence channels.

5.4 RESULTS

5.4.1 *Sso*SSB is a monomer in solution

*Sso*SSB has been reported to adopt a number of oligomeric states including monomer, dimer, and tetramer. During our purification, we also probed the oligomeric state using gel filtration. The analytical gel filtration column was calibrated with protein standards including: thyroglobulin (669 kDa), ferritin (440 kDa), catalase (250 kDa), conalbumin (75 kDa), myoglobin (17.6 kDa), and vitamin B12 (1.3 kDa). The molecular weights and their corresponding elution volumes are shown at the top of Figure 5.2A. *Sso*SSB is 16.1 kDa and elutes just after myoglobin a one uniform peak

indicating the presence of monomeric species. Figure 5.2B shows that *Sso*SSB has eluted within the fraction collected in an SDS-PAGE gel. This is in agreement with a number of publications [27-29, 31]. The actual oligomeric state in solution remains a matter of debate, though a recent review favors the monomeric view [21].

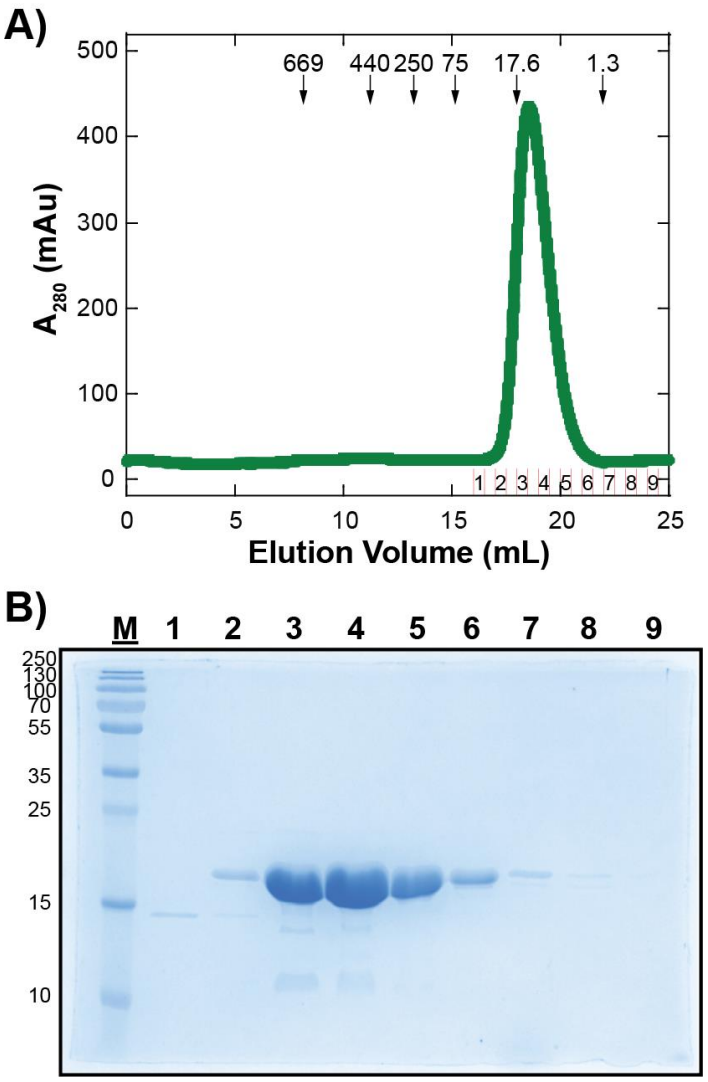


Figure 5.2: *Sso*SSB elutes as a monomer **A)** A_{280} (milli absorbance units – mAu) versus elution volume (mL). *Sso*SSB was applied to an analytical gel filtration column. Standard molecular weights (in kDa) are shown at the top with arrows pointing to their corresponding elution volumes. During the elution, 0.5 mL fractions collected are indicated with red lines separating the fractions collected (numbered 1-9). **B)** 12% SDS-PAGE showing markers (M) and the fractions (1-9) that were collected in A).

5.4.2 *Sso*SSB inhibits *Sso*MCM's unwinding

*Sso*SSB, though unable to physically unwind DNA, is able to bind ssDNA effectively separating forked DNA into two ssDNA strands thermodynamically [219, 221]. In Figure 5.3A, a representative unwinding experiment for *Sso*MCM and/or *Sso*SSB is presented. The fraction unwound was quantified and is shown in Figure 5.3B. *Sso*MCM's fraction unwound is 0.35 ± 0.02. Interestingly, and in contrast to previous studies, *Sso*SSB inhibits unwinding as its concentration is titrated. *Sso*SSB's site size of 4-6 nt per monomer has been shown by a variety of groups, when discussing the site size, 5 nt per monomer of SSB will be used [27, 29-31]. The forked DNA used (15 nM with 132 total nt) has an effective concentration of 400 nM, where 400 nM equals the total number of binding in this reaction. When equimolar to the effective DNA concentration at 400 nM *Sso*SSB, there ise little thermodynamic

separation in the absence of DNA and little unwinding inhibition in the presence of *Sso*MCM. The highest *Sso*SSB concentration tested (4 µM) decreased the fraction unwound by *Sso*MCM 2.1-fold to 0.17 ± 0.01 and is similar to *Sso*SSB alone (0.13 ± 0.04). High concentrations of *Sso*SSB may prevent *Sso*MCM loading possibly through sequestering the DNA and thermodynamically unwinding it. This result is in agreement with previous studies in which *Sso*SSB inhibited *Sso*MCM [221] and *Mth*MCM was inhibited by its single-stranded binding protein known as RPA [222].

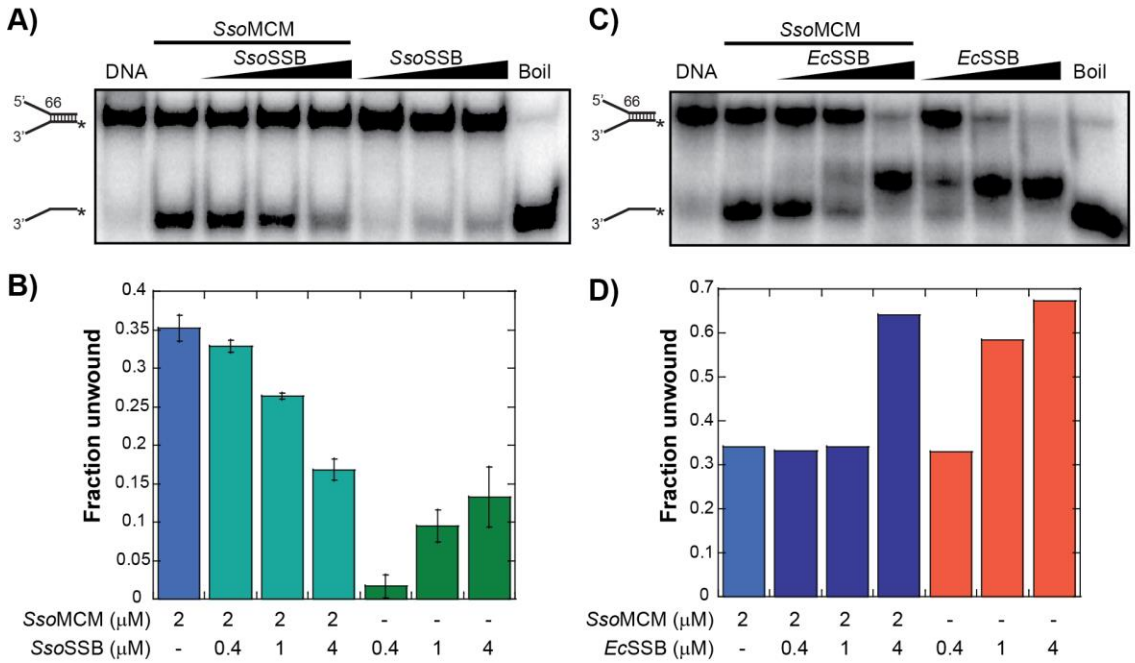


Figure 5.3: Unwinding comparison for SSB's in the absence of presence of *Sso*MCM.
A) Forked DNA unwinding by *Sso*SSB and/or *Sso*MCM. If present, *Sso*MCM was 2 µM, three *Sso*SSB concentrations were used (0.4, 1, and 4 µM) and all reactions contain 15 nM ^{32}P-forked DNA (3'-tail-30 nt annealed to 5'-tail-30 nt). Unwinding reactions were for 15 minutes at 60 ºC prior to quenching and Proteinase K treatment as specified in the 'Materials and Methods.' Bands were resolved on 14% polyacrylamide gels containing 0.1% SDS and phosphorimaged. DNA alone is shown in lane 1 and boiled DNA is shown in lane 9. **B)** Quantification of fraction unwound of A). Error bars represent the standard error from at least three independent unwinding experiments. **C)** Forked DNA unwinding by *Ec*SSB and/or *Sso*MCM. If present, *Sso*MCM was 2 µM, three *Ec*SSB concentrations were used (0.4, 1, and 4 µM) and all reactions contain 15 nM ^{32}P-forked DNA. Reactions proceeded as A). **D)** Quantification of the fraction unwound from C).

*Ec*SSB was tested in parallel experiments to probe species specificity and interactions (Figure 5.3C and D). Here, *Ec*SSB seemed to both robustly separate forked DNA into ssDNA and tightly bind the product as indicated by gel shifted bands above boiled DNA but below duplex

DNA. At the lowest concentration of *Ec*SSB (0.4 μM), a similar fraction is thermodynamically separated compared with *Sso*MCM unwinding alone. This may be due to *Ec*SSB binding in its $(SSB)_{65}$ mode in which tetrameric *Ec*SSB binds 65 nt per tetramer, the effective concentration here would be 30 nM per tetramer [29, 48, 223]. As *Ec*SSB was titrated further, apparent unwinding is merely due to *Ec*SSB's strong ssDNA binding and stabilization ability which effectively separates dsDNA.

5.4.3 ATPase confirms *Sso*SSB's DNA sequestering ability

We have previously characterized the ATPase ability of *Sso*MCM in the absence and presence of DNA [97] and our results were in agreement with previous work [111]. In the presence of DNA, *Sso*MCM's ATPase rate is stimulated approximately two-fold, allowing for unwinding activation or translocation [97, 111, 170]. Therefore, if *Sso*SSB were sequestering DNA, monitoring ATP hydrolyzed would provide insight into *Sso*MCM's ability to access DNA. In the left panel of Figure 5.3A we present the ATP hydrolyzed (pmol/min) versus *Sso*SSB concentration in the presence and absence of 0.5 μM forked DNA. The effective concentration of the DNA in these assays (500 nM with 132 total nt) is 13 μM. The dashed box region (0 - 2 μM) is shown enlarged in right panel of Figure 5.4A. *Sso*MCM's ATP hydrolyzed increases 2.1 fold from 290 ± 30 to 610 ± 10 pmol/min in the absence of *Sso*SSB. In the absence of DNA, as *Sso*SSB is titrated to 5 μM, ATP hydrolyzed remains within error. However, in the presence of DNA, as *Sso*SSB is titrated to 20 μM, ATP hydrolyzed decreases until reaching 310 ± 20 pmol/min at 20 μM *Sso*SSB; this concentration of *Sso*SSB is ~1.5-fold higher than the effective concentration. This falls within the error of *Sso*MCM's ATP hydrolysis rate in the absence of DNA. Representative ATPase plates in the absence of DNA (Figure 5.4B) and in the presence of DNA (Figure 5.4C).

Additional experiments were also performed by titrating DNA and keeping *Sso*MCM and *Sso*SSB constant at 2 μM each. The effective DNA concentrations used are 1.3, 2.6, 6.6, and 13 μM, respectively for the four DNA concentrations used (50, 100, 250, and 500 nM all with 132 total nt). *Sso*SSB is only in excess at the lowest DNA concentration. Figure 5.5A shows that as forked DNA is titrated, *Sso*MCM in the absence and presence of 2 μM *Sso*SSB, displays enhanced ATP hydrolysis rates to 0.5 μM DNA. The increase in ATP hydrolyzed by *Sso*MCM is due only

to the excess DNA present. TLC plates of ATPase experiments in the absence (Figure 5.5B) and presence (Figure 5.5C) of 2 μM *Sso*SSB are shown.

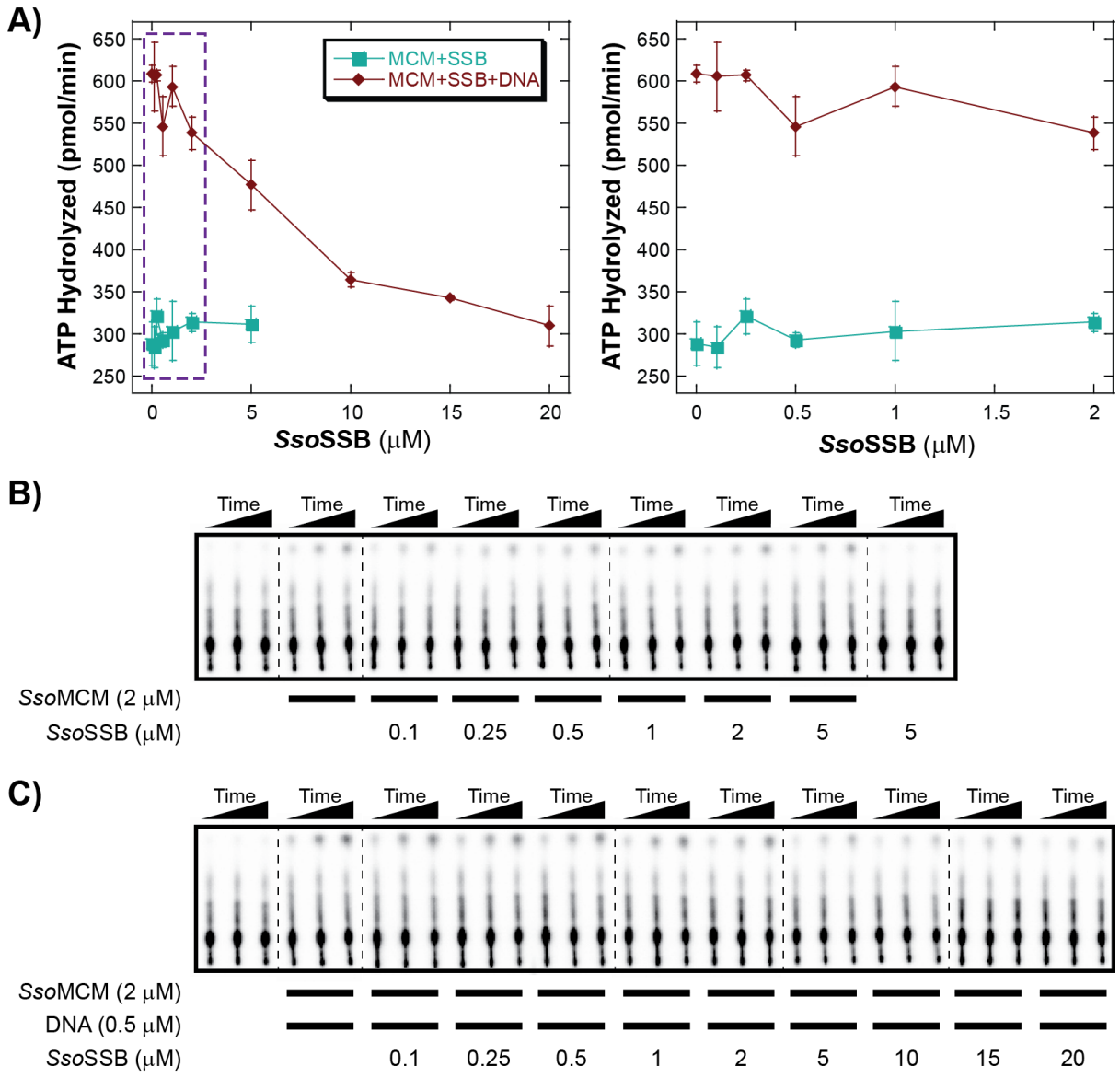


Figure 5.4: ATPase assays demonstrates *Sso*SSB sequesters DNA from *Sso*MCM.
A) ATP hydrolyzed (pmol/min) versus *Sso*SSB (μM) at 60 ºC in the absence and presence of DNA. 2 μM *Sso*MCM and/or *Sso*SSB (concentrations specified) were incubated in the absence or presence of 0.5 μM forked DNA (3'-tail-30 nt annealed to 5'-tail-30 nt). The region contained within the dashed purple box is shown enlarged at right. The ATPase rate resultant from *Sso*SSB titrated into *Sso*MCM in the absence and presence of 0.5 μM DNA is shown with turquoise squares and burgundy diamonds, respectively. Error bars represent the standard error from at least three independent ATPase experiments. **B)** Representative ATPase plate of data collected in the absence of DNA. **C)** Representative ATPase plate of data collected in the presence of equal length tail forked DNA. The ATP hydrolysis rate shown in A) was calculated as a function of time (5, 10, and 15 minutes) shown in B) and C).

5.4.4 Fluorescence spectroscopy indicate low *Sso*SSB concentrations stabilize *Sso*MCM binding

The potential interaction between *Sso*MCM and *Sso*SSB was further probed by utilizing bulk fluorescence resonance energy transfer (FRET) between the 3'-Cy3 and 5'-Cy5 ends of forked DNA. Previously, we have shown that upon *Sso*MCM binding to this FRET fork substrate, there is a large increase in FRET associated with an external interaction of ssDNA with *Sso*MCM [123]. FRET experiments were performed by titrating *Sso*SSB in the absence and presence of *Sso*MCM. The effective concentration of the DNA in these assays (initially 50 nM with 118 total nt) is 1.2 μM. As expected, when *Sso*MCM binds this DNA substrate, FRET increases as shown in Figure 5.6, where the left panel shows the entire concentration range and the right panel is limited to 1.2 μM [123]. The right panel demonstrates that at *Sso*SSB concentrations below 0.6 μM, which is also below the effective DNA concentration, *Sso*SSB increases the FRET value over *Sso*MCM alone. This increase in FRET for *Sso*MCM and *Sso*SSB is consistent with fluorophores in closer proximity. We hypothesized that *Sso*SSB may be stabilizing the 5'-tail interaction yielding an increase in FRET. A similar increase in FRET is also present for *Sso*SSB titration onto DNA alone, potentially indicative of *Sso*SSB binding both strands simultaneously drawing the fluorophores together into closer proximity. As the concentrations of *Sso*SSB increase beyond 0.6 μM, there is a FRET decrease both in the absence and presence of MCM, though never reaching the value for DNA alone. For *Sso*SSB alone, the FRET value above DNA is likely due to the duplex region not being completely thermodynamically separated, even at the highest concentrations (~8.6 μM). However, when *Sso*SSB binds in the presence of *Sso*MCM, prior to reaching concentrations high enough to remove a large fraction of *Sso*MCM, *Sso*SSB may alter *Sso*MCM's binding location by pushing it forward into the ss-dsDNA junction to facilitate binding. This may release the 5'-tail from *Sso*MCM causing a drop in FRET.

To test this hypothesis, we utilized 2-aminopurine containing DNA to monitor the degree of duplex DNA with bound *Sso*MCM and *Sso*SSB. 2-aminopurine (2AP) is a fluorescent nucleotide analog that is significantly quenched upon base pairing to either T or C. We tested three substrates: 2AP-ssDNA, 2AP forked DNA (2 bp inside ss-dsDNA junction), and 2AP forked DNA (5 bp inside ss-dsDNA junction). The effective DNA concentrations (300 nM with 71, 139, and 142 total nt) are 4.3, 8.3 and 8.5 μM, respectively. Figure 5.6A-C are 2AP-ssDNA, 2AP forked

DNA 2 bp inside ss-dsDNA junction, and 2AP forked DNA 5 bp inside ss-dsDNA junction, respectively. To each of these substrates in the absence and presence of *Sso*MCM, we added *Sso*SSB, *Ec*SSB, or bovine serum albumin (BSA). In Figure 5.6A, adding *Sso*MCM to 2AP-ssDNA increases the fluorescence from 1.0 to 1.5 demonstrative of a reduction in base pairing, likely due to *Sso*MCM loading onto the DNA which would disrupt any non- specific base pairing.

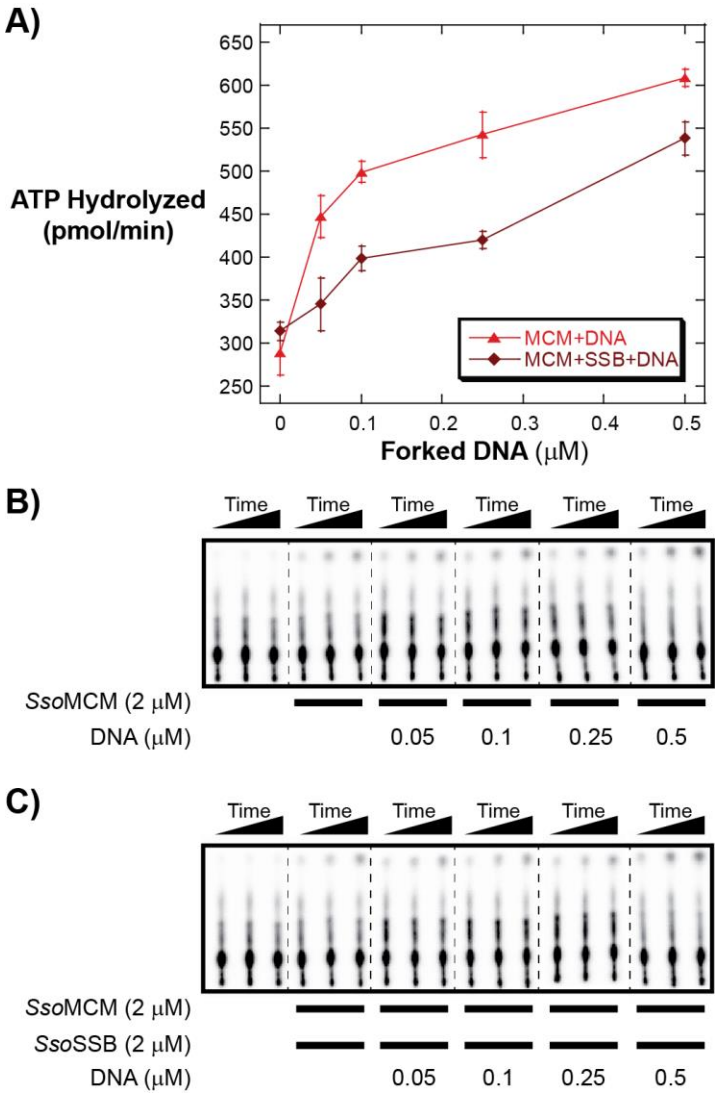


Figure 5.5: *Sso*SSB does not enhance *Sso*MCM's ATPase activity in the presence of forked DNA.
A) ATP hydrolyzed (pmol/min) versus equal length tail forked DNA (μM) at 60 ºC in the absence and presence of *Sso*SSB. 2 μM *Sso*MCM was incubated in the absence or presence of 2 μM *Sso*SSB and increasing amounts of forked DNA (3'-tail-30 nt annealed to 5'-tail-30 nt) as specified. The ATPase rate resultant from forked DNA titrated into *Sso*MCM in the absence and presence of *Sso*SSB is shown with red triangles and burgundy diamonds, respectively. Error bars represent the standard error from at least three independent ATPase experiments. **B)** Representative ATPase plate of data collected in the absence of DNA. **C)** Representative ATPase plate of data collected in the presence of equal length tail forked DNA. The ATP hydrolysis rate shown in A) was calculated as a function of time (5, 10, and 15 minutes) shown in B) and C).

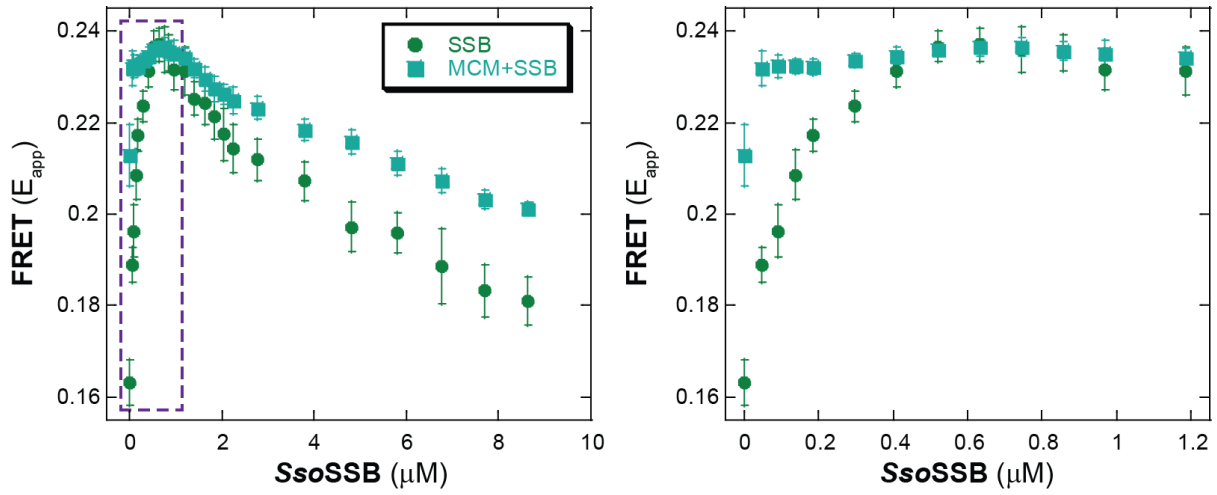


Figure 5.6: *Sso*SSB increases FRET for MCM-DNA at low concentrations.
FRET versus *Sso*SSB in the absence and presence of *Sso*MCM. *Sso*SSB was titrated in the absence and presence of 2 μM *Sso*MCM with 50 nM Cy3-Cy5 forked-biotinylated DNA (annealed 3'-tail-31 nt (5'-B)(3'Cy3) and 5'-tail-51 nt (5'Cy5)). FRET values resultant from the *Sso*SSB titration in the absence and presence of *Sso*MCM are shown in green circles and turquoise squares, respectively. The left panel contains the entire titration. The concentrations contained within the purple dashed line (0-1.2 μM) are shown in the right panel.

*Sso*SSB and *Ec*SSB titrations in the absence and presence of MCM also increase the fluorescence indicative of less base pairing, while BSA, which does not bind DNA, does not shown a similar fluorescence increase. Both SSB's bind the ssDNA and similar to *Sso*MCM disrupt non-specific base pairing. BSA, on the other hand, provides no binding and the fluorescence remains similar over the titration. In Figure 5.6B, adding MCM to DNA decreases the fluorescence for 2AP forked DNA 2 bp inside ss-dsDNA junction from 1.0 to 0.73. This is likely due to the 5'-tail stably binding on the exterior of *Sso*MCM [123]. BSA and *Ec*SSB do not appear to effect the fluorescence over their respective titrations. Interestingly, *Sso*SSB, in the absence of *Sso*MCM also appears to decrease the fluorescence, indicating more stable base pairing. At the highest *Sso*SSB concentration (4.6 μM, still well below the effective concentration of 8.3 μM), the fluorescence increases indicating less stable pairing and possibly the beginning of thermodynamic strand separation noted above (Figure 5.6). *Sso*SSB in the presence of *Sso*MCM shows an enhanced stabilization at low concentrations, similar to what was seen for in Figure 5.6. This stabilization ceases when the concentration of *Sso*SSB is greater than 2.3 μM again indicating that strand separation may be beginning to occur or *Sso*MCM may be pushed towards the duplex. At the same time, in the absence and presence of *Sso*MCM at 2.3 and 4.6 μM, the fluorescence values are

similar for *Sso*SSB indicating that *Sso*MCM may have dissociated due to *Sso*SSB binding. Figure 5.7C shows a similar pattern of fluorescence using 2AP forked DNA 5 bp inside ss-dsDNA junction. Both Figure 5.7B and C show that at higher concentrations of *Sso*SSB, *Sso*MCM's effect on the DNA base pairing is minimized, consistent with data above that *Sso*SSB is most likely removing beginning to remove *Sso*MCM from the DNA. However, at lower concentrations of *Sso*SSB, a direct interaction may provide stabilization forming an *Sso*MCM-DNA-*Sso*SSB complex.

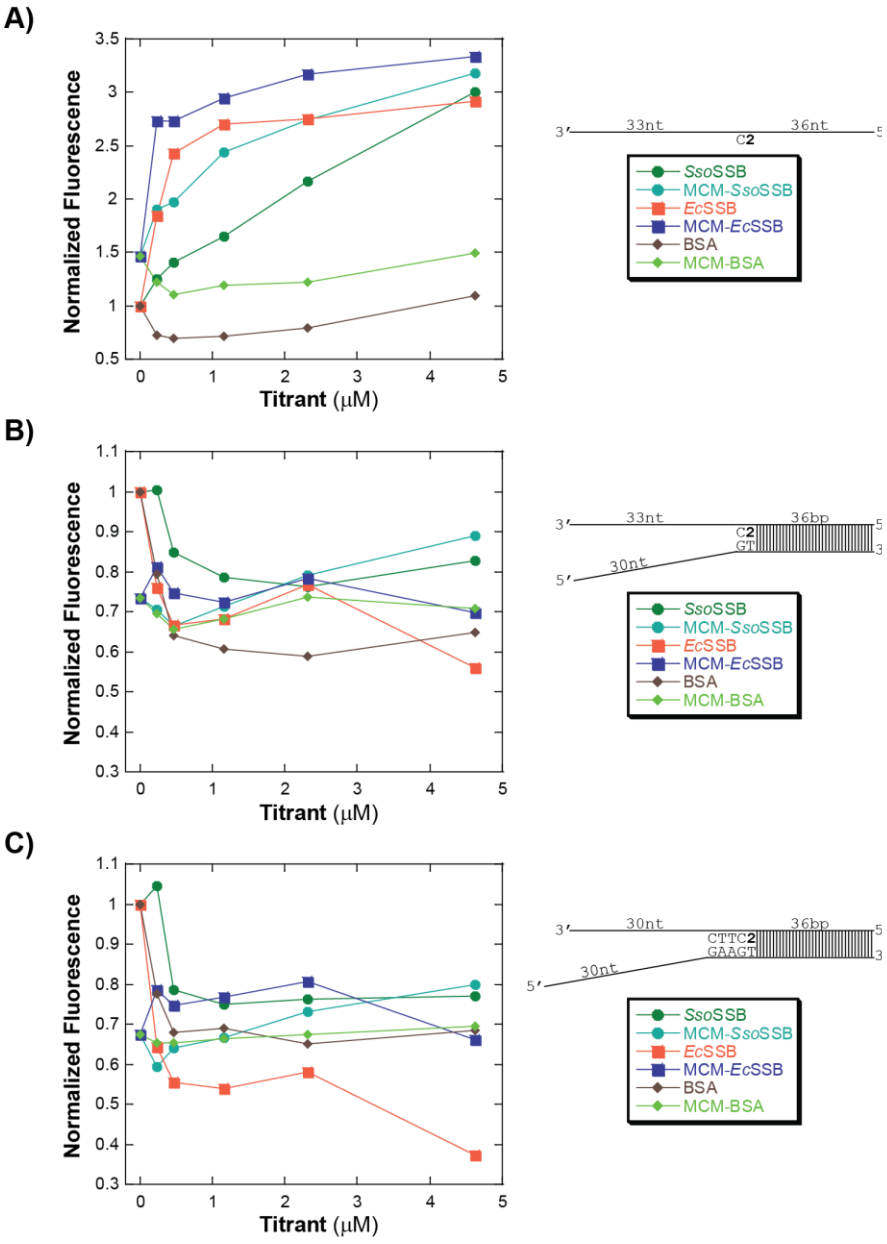


Figure 5.7: Effect of SSB's on *Sso*MCM's position at the ss-dsDNA junction. **A)** Normalized fluorescence values at the observed maxima (373 nm) for 2AP positioned for i) ssDNA alone (71 nt in length). A cartoon of the DNA substrate is shown. Increasing amounts of titrant were added to either DNA or MCM-DNA where green and turquoise circles are SsoSSB titrated into DNA and MCM-DNA, respectively; where orange and purple squares are EcSSB titrated into DNA and MCM-DNA, respectively; and where brown and light green diamonds are BSA titrated into DNA and MCM-DNA, respectively. The DNA cartoon and labels are present for B) and C). **B)** Normalized fluorescence values at the observed maxima (373 nm) for 2AP positioned for forked DNA 2AP 2 bp in from ss-dsDNA junction (3'-tail 33 nt and 5'-tail 30 nt with a 38 bp duplex). **C)** Normalized fluorescence values at the observed maxima (373 nm) for 2AP positioned for forked DNA 2AP 5 bp in from ss-dsDNA junction (3'-tail 33 nt and 5'-tail 30 nt with a 41 bp duplex).

5.4.5 Pull-down assays do not demonstrate physical interaction between *Sso*MCM and *Sso*SSB

Previous reports have demonstrated a physical interaction between *Sso*MCM and *Sso*SSB using ELISA, far Western blots [9], and bacterial two-hybrid [43]. The interaction shown by bacterial two-hybrid was not shown in the publication, however. We probed the interaction by *in vitro* solution pull-down assays. Here, we incubated *Sso*SSB in the absence (Mock – lanes 3-5) and presence (His-MCM – lanes 8-10) of His-*Sso*MCM as seen in Figure 5.8. Mock pull-downs of *Sso*SSB alone show no background binding of *Sso*SSB to the Co-column. When His-*Sso*MCM is included there is also no evidence of *Sso*SSB in the elution (lane 10). It may be possible that DNA and/or another protein is required for the interaction.

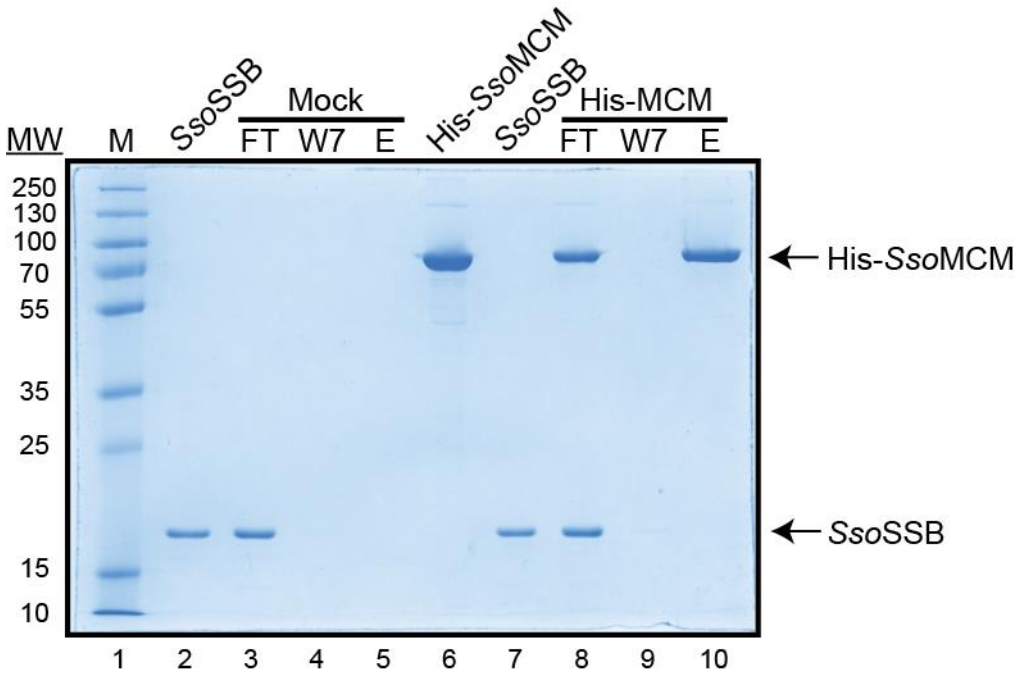


Figure 5.8: Pull-down assay does not detect physical interaction between *Sso*MCM and *Sso*SSB.
Coomassie stained 12% SDS-PAGE gel showing purified proteins *Sso*SSB (lanes 2 and 6) and His-*Sso*MCM (lane 5). His-pull down with (lanes 8-10) and without (lanes 3-5) immobilized His-*Sso*MCM. *Sso*SSB and/or His-*Sso*MCM flow through (FT, lanes 3 and 8), seventh wash (W7, lanes 4 and 9) and elution with high imidazole (E, lanes 5 and 10). The lack of *Sso*SSB co-eluting with His-*Sso*MCM in lane 10 is shown by the arrow on the lower right.

5.4.6 Fluorescent EMSAs show DNA binding locations for both *Sso*MCM and *Sso*SSB

In order to more directly visualize simultaneous binding of *Sso*MCM and *Sso*SSB to forked DNA, fluorescent EMSAs can be employed. Here, we utilized fluorescently-labeled proteins and Cy3-labeled DNA to individually track each species on a native polyacrylamide gel. The effective DNA concentration (20 nM with 152 total nt) for all fluorescent EMSAs is 610 nM. In Figure 5.9, we titrated *Sso*SSB in the absence and presence of 2 μM *Sso*MCM. *Sso*SSB is able to bind DNA as expected, but in the presence of *Sso*MCM, more *Sso*SSB is bound. However, only at the highest *Sso*SSB concentration do we see some potential overlap between the *Sso*MCM and *Sso*SSB bands. Viewing the lower left panel, the amount of DNA bound to *Sso*MCM at the upper portion of the gel decreases as [*Sso*SSB] increases in favor of *Sso*SSB/DNA complexes in the middle of the gel. In Figure 5.10, *Sso*SSB is held constant at 1 μM, and *Sso*MCM is titrated. Interestingly, as *Sso*MCM is titrated and viewing either the AF488-*Sso*SSB channel (upper right) or the Cy3-DNA channel (lower left), there is an apparent upward shift indicating complex formation. Alternatively, the presence of *Sso*MCM may increase the binding affinity of *Sso*SSB for DNA. At the two highest concentrations of *Sso*MCM (1 and 2 μM), there is a fraction of *Sso*SSB, *Sso*MCM, and DNA co-localizing at the upper portion of the gel. These results suggested that there may be a non-specific interaction that was promoting *Sso*SSB binding to DNA. To test this hypothesis, we titrated *Sso*SSB in the absence and presence of 2 μM BSA as shown in Figure 5.11. Comparing Figure 5.11 to Figure 5.9 AF-*Sso*SSB (upper right panels), there are very similar binding shifts for *Sso*SSB in the presence of either *Sso*MCM or BSA at 2 μM. To further test the hypothesis that non-specific protein interactions or crowding affect SSB binding to DNA, we added 2 μM BSA to all lanes and then titrated *Sso*SSB in the absence and presence of 2 μM *Sso*MCM (Figure 5.12). It seems that either BSA or *Sso*MCM promotes *Sso*SSB binding to DNA. Unfortunately, these methods did not produce definitive evidence of an interaction between *Sso*MCM and *Sso*SSB and exposed how spurious protein binding can be affected by molecular crowding at high total protein concentration. At the same time, higher concentrations of *Sso*SSB in the presence of *Sso*MCM showed overlap at the upper portion of the gel and we cannot rule out the possibility that a ternary complex has formed consisting of *Sso*MCM-DNA-*Sso*SSB.

The fluorescent EMSAs show that a ternary complex may form at the higher concentrations of *Sso*SSB in Figures 5.9, 10, and 12. The FRET experiment shows that at the highest *Sso*SSB

concentration tested (~8.6 μM) in the presence of *Sso*MCM the FRET is higher (~0.20) than in the absence of *Sso*MCM (~.18). This indicates that some MCM is still present as *Sso*SSB either thermodynamically separates the DNA or removes MCM; in either situation we would see this as a decreases in FRET. Since this is a bulk measurement we see the average, so we can only speculate. In both types of experiments (fluorescent EMSAs and FRET), we examine *Sso*SSB concentrations below and above the effective DNA concentration. At 20 nM DNA in Figure 5.9, 2 μM MCM completely binds the DNA. The fluorescent EMSA Figures 5.9-12 all show that molecular crowding aids in *Sso*SSB binding (compare lanes 3-6 to 7-10). In the FRET assays, MCM is initially 2 μM, but over the course of the titration MCM's concentration decreases due to the increase in volume from *Sso*SSB, lessening the crowding. This would explain why *Sso*SSB is able to more easily bind at the lower concentrations (below 0.6 μM), but not bind as well at higher concentrations, even though the ATPase results indicate that when above the effective DNA concentration, thermodynamic separation occurs (Figure 5.4A left panel).

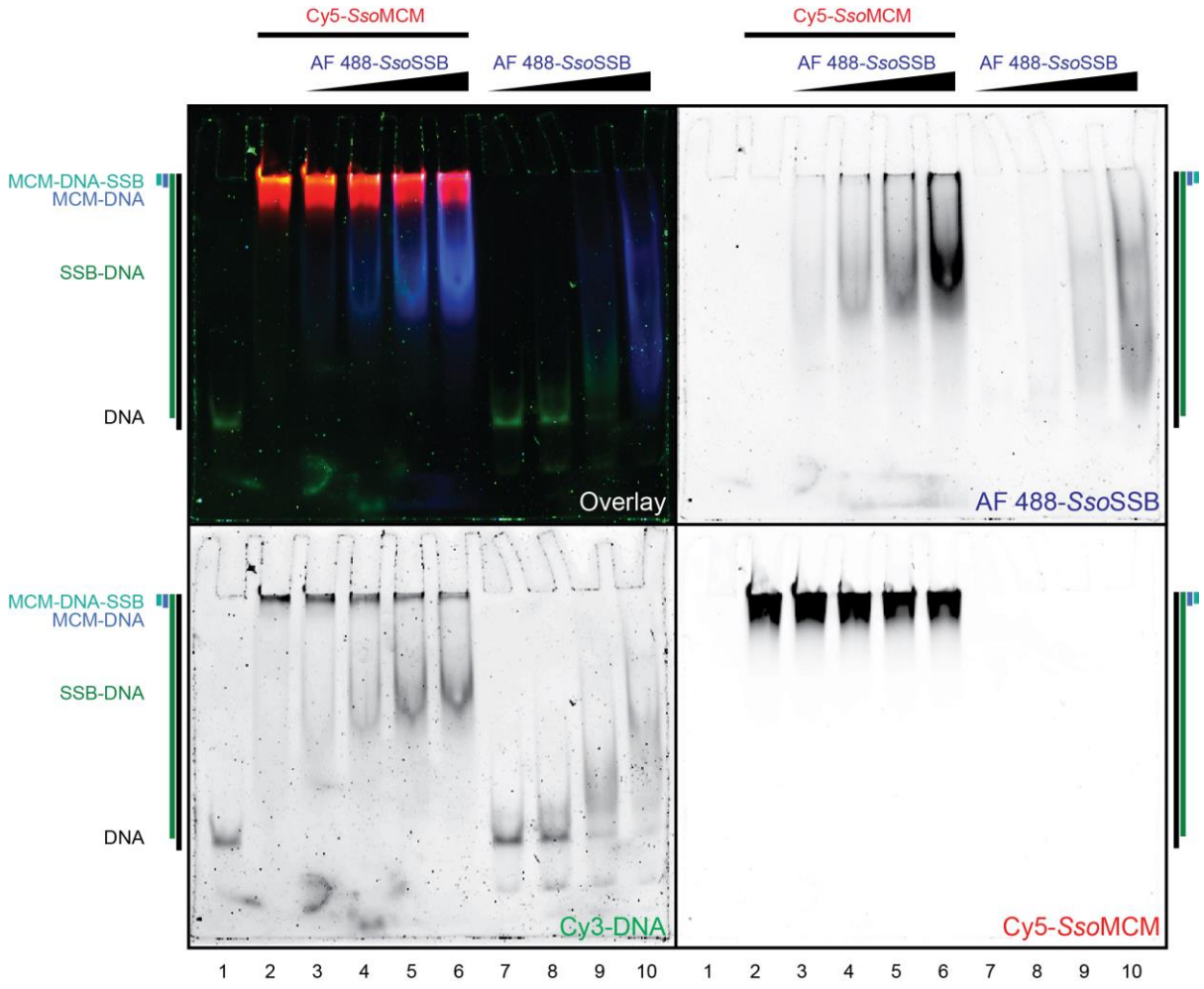


Figure 5.9: *Sso*SSB binding is enhanced by the presence of *Sso*MCM.
Fluorescent EMSA – *Sso*SSB titration in the absence and presence of *Sso*MCM. Alexa Fluor 488-*Sso*SSB (AF 488-*Sso*SSB) was titrated at four concentrations (0.4, 1, 2, and 4 μM) in the absence or presence of Cy5-*Sso*MCM (2 μM monomer) and Cy3-DNA (20 nM - 3'-tail-30 nt (5'Cy3) annealed to 5'-tail-50 nt). Samples were incubated at 60 ºC for 10 min to promote complex formation and subsequently resolved on a 5% native polyacrylamide gel. The gel was scanned using three channels to isolate each fluorescently-labeled species. The overlay of the three channels is shown in the upper left, the Cy3 channel is shown in the lower left, the AF 488 channel is shown in the upper right, and the Cy5 channel is shown in the lower right. The location of the DNA both bound and unbound is represented by a black bar on the side of each channel. SSB-DNA is shown with a green bar, MCM-DNA is shown with a blue bar, and MCM-DNA-SSB is shown with a turquoise bar.

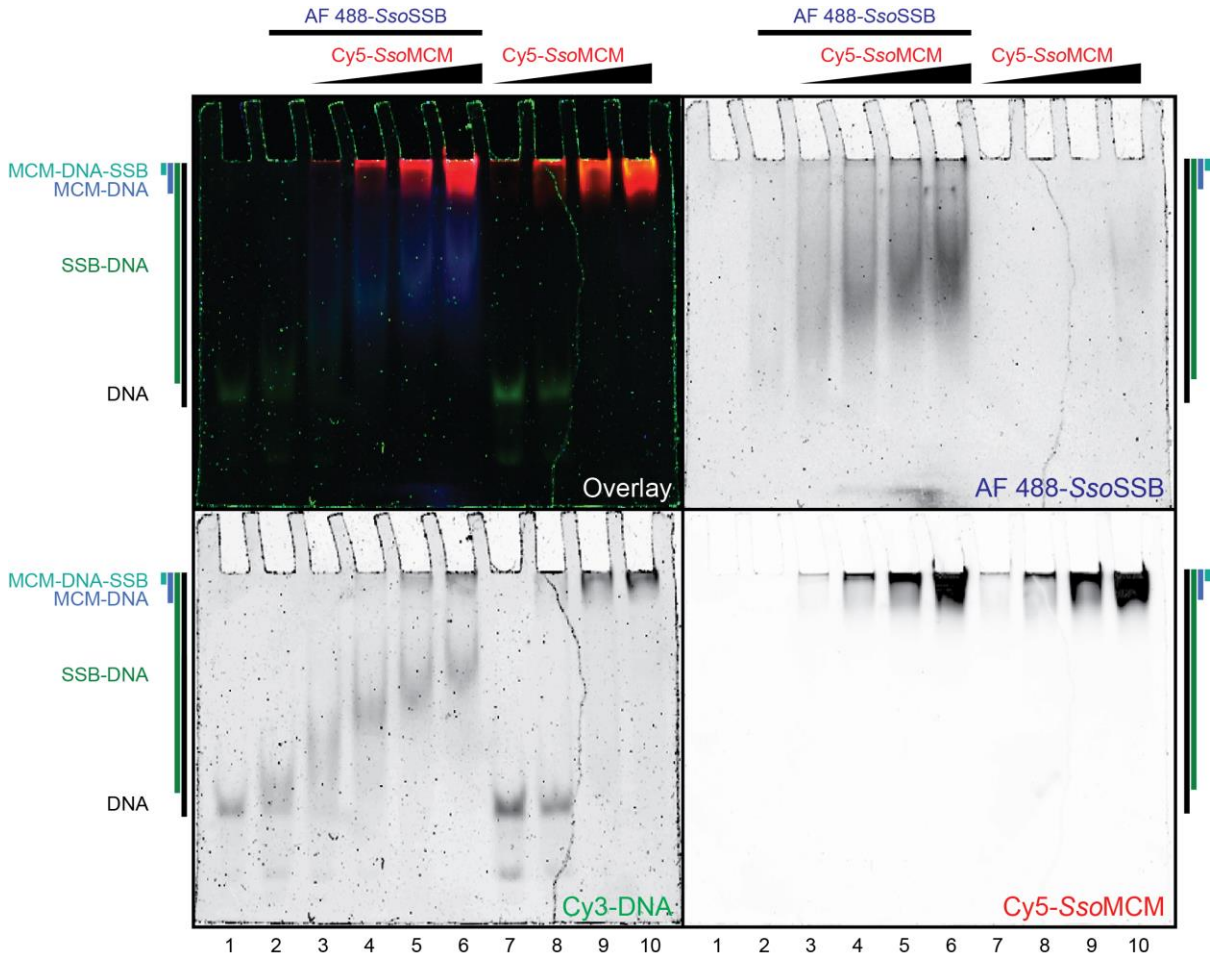


Figure 5.10: *Sso*MCM titration enhances *Sso*SSB binding.
Fluorescent EMSA – *Sso*MCM titration in the absence and presence of *Sso*SSB. Cy5-*Sso*MCM was titrated at four concentrations (0.2, 0.5, 1, and 2 µM monomer) in the absence or presence of AF 488-*Sso*SSB (1 µM) and Cy3-DNA (20 nM - 3'-tail-30 nt (5'Cy3) annealed to 5'-tail-50 nt). Samples were incubated at 60 ºC for 10 min to promote complex formation and subsequently resolved on a 5% native polyacrylamide gel. The gel was scanned using three channels to isolate each fluorescently-labeled species. The overlay of the three channels is shown in the upper left, the Cy3 channel is shown in the lower left, the AF 488 channel is shown in the upper right, and the Cy5 channel is shown in the lower right. The location of the DNA both bound and unbound is represented by a black bar on the side of each channel. SSB-DNA is shown with a green bar, MCM-DNA is shown with a blue bar, and MCM-DNA-SSB is shown with a turquoise bar.

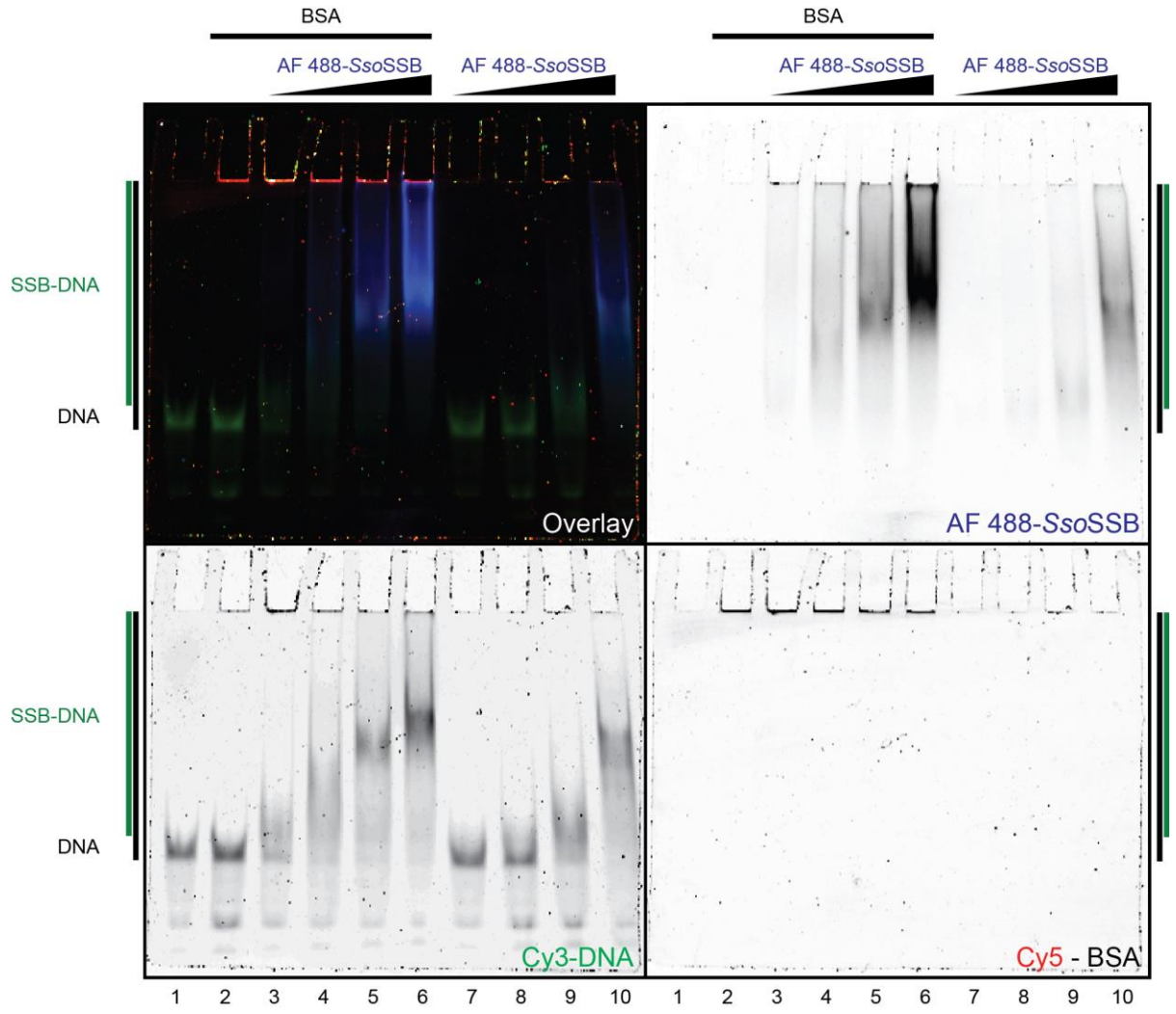


Figure 5.11: BSA enhances *Sso*SSB binding.
Fluorescent EMSA – *Sso*SSB titration in the absence and presence of BSA. AF 488-*Sso*SSB was titrated at four concentrations (0.4, 1, 2, and 4 µM) in the absence or presence of unlabeled BSA (2 µM) and Cy3-DNA (20 nM - 3'-tail-30 nt (5'Cy3) annealed to 5'-tail-50 nt). Samples were incubated at 60 ºC for 10 min to promote complex formation and subsequently resolved on a 5% native polyacrylamide gel. The gel was scanned using three channels to isolate each fluorescently-labeled species. The overlay of the three channels is shown in the upper left, the Cy3 channel is shown in the lower left, the AF 488 channel is shown in the upper right, and the Cy5 channel is shown in the lower right. The location of the DNA both bound and unbound is represented by a black bar on the side of each channel. SSB-DNA is shown with a green bar.

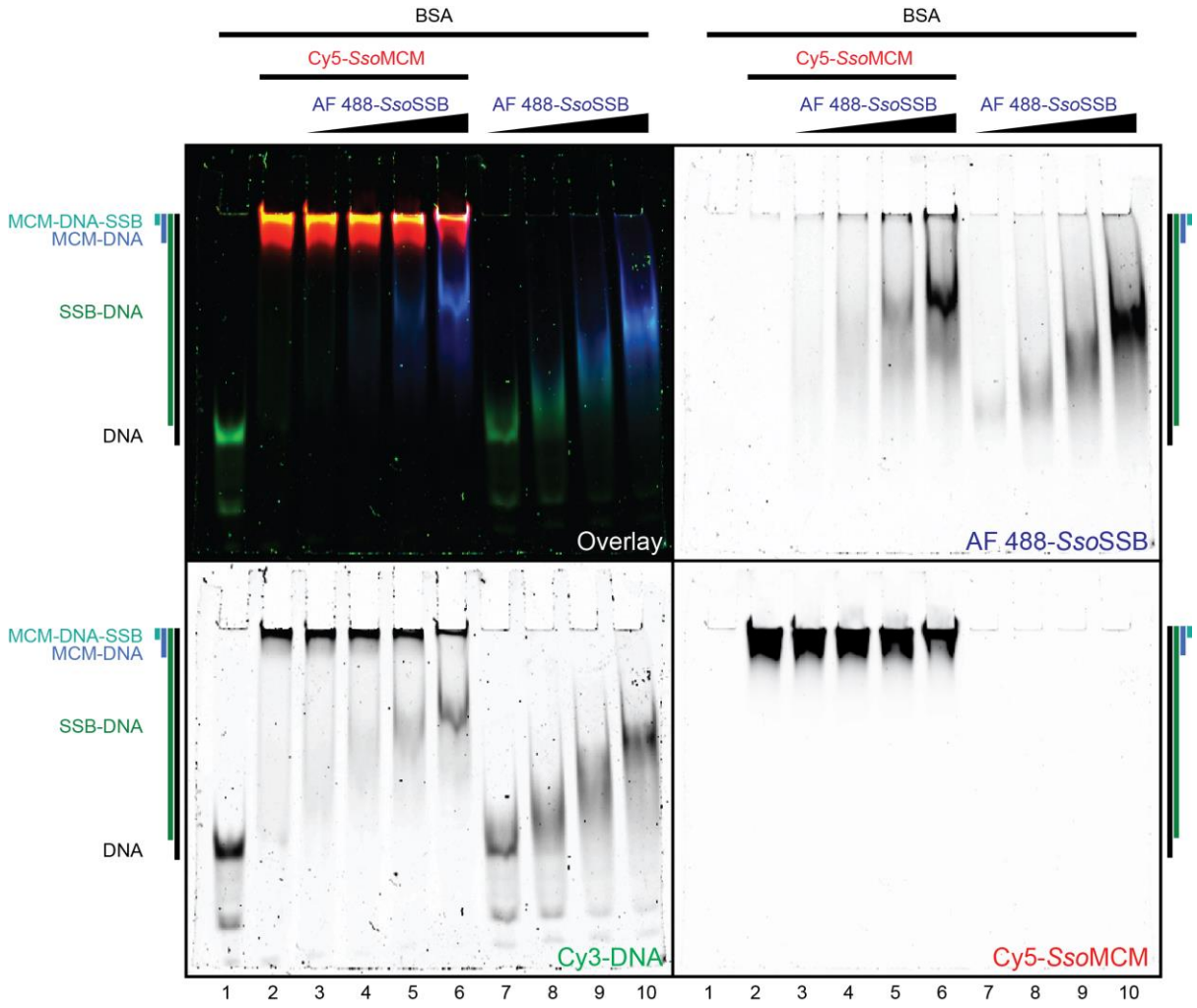


Figure 5.12: Protein enhances *Sso*SSB binding.
Fluorescent EMSA – BSA containing *Sso*SSB titration in the absence and presence of *Sso*MCM. Alexa Fluor 488-*Sso*SSB (AF 488-*Sso*SSB) was titrated at four concentrations (0.4, 1, 2, and 4 µM) in the absence or presence of Cy5-*Sso*MCM (2 µM monomer), Cy3-DNA (20 nM - 3'-tail-30 nt (5'Cy3) annealed to 5'-tail-50 nt) and unlabeled BSA (2 µM). Samples were incubated at 60 ºC for 10 min to promote complex formation and subsequently resolved on a 5% native polyacrylamide gel. The gel was scanned using three channels to isolate each fluorescently-labeled species. The overlay of the three channels is shown in the upper left, the Cy3 channel is shown in the lower left, the AF 488 channel is shown in the upper right, and the Cy5 channel is shown in the lower right. The location of the DNA both bound and unbound is represented by a black bar on the side of each channel. SSB-DNA is shown with a green bar, MCM-DNA is shown with a blue bar, and MCM-DNA-SSB is shown with a turquoise bar.

5.5 DISCUSSION

We were unable to verify a physical or positive functional interaction between *Sso*MCM and *Sso*SSB. Other groups have shown a variety of oligomeric states including monomer [27-31], dimer [30], and tetramer [30, 32]. Three groups have shown that *Sso*SSB exists as a monomer in solution [27, 28, 31], while one group has shown the existence of monomeric, dimeric, and tetrameric species which persists after an 80 minute digestion at room temperature in the presence of 20 ug/mL DNase I [42]. DNase I, however is most active at 37 ºC and its specificity for ssDNA is 500 fold less than for dsDNA [224]. It is possible that ssDNA was still present which may have yielded the higher ordered species. The isolated tetrameric *Sso*SSB complex both bound DNA with a higher affinity and a larger site size compared to monomer [42]. Another group showed *Sso*SSB tetramer formation of by native PAGE [32]. It is probable, that inconsistences in protein purification protocols, buffer conditions, or incomplete removal of genomic DNA is responsible for the apparent differences in oligomeric states that need further study. In our preparations, the A_{260}/A_{280} ratio was 0.67 indicative that our prep contained at least 99.2% protein [225].

Both unwinding and ATPase assays have shown that *Sso*MCM is inhibited by *Sso*SSB either through *Sso*SSB sequestering the DNA or by inducing a conformational change in *Sso*MCM binding to DNA that yields an inactive state. Our cumulative evidence favors the former interpretation where direct competition of a higher affinity binding of *Sso*SSB for DNA dislodges any prebound *Sso*MCM at a critical concentration. Cellular concentrations of *Sso*MCM and *Sso*SSB are ~1.8-3.6 µM monomer (equivalent to ~1800-3600 molecules per cell) [127] and ~4.5 µM (equivalent to ~4500 molecules per cell) [226], respectively [227]. DNA replication is similar across the three domains of life with a helicase unwinding DNA into two strands; a leading strand which is continuously replicated and a lagging strand that is discontinuous requiring RNA primers and Okazaki fragments which must be ligated. Okazaki fragments in eukarya and archaea are ~100 bp, while in bacteria they are much longer at 2000 bp [228]. Since *Sso* has three replication origins and replicates bidirectionally [131, 132], there could be ~600 nt of ssDNA (assuming looping out occurs on the lagging strand prior to replication). As previously established, *Sso*SSB also functions in other cellular processes beyond DNA replication, so assuming *Sso*SSB binds 5 nt, only 120 SSB molecules would be necessary, which is likely not inhibitory for *Sso*MCM. This correlates well, with our FRET results (Figure 5.6), which show that lower *Sso*SSB concentrations stabilize

*Sso*MCM binding on the 3'-tail. Since *Sso*SSB generally binds non-cooperatively [27, 29, 40], it will initially bind on any free 3'- or 5'-tail ends and stabilize *Sso*MCM towards the ss-dsDNA junction. *Sso*SSB protects ssDNA from nucleolytic cleavage, likely aids in the unwinding process *in vivo*. However, *in vitro*, this inhibition and likely regulatory role may not be uncommon as inhibition and competition by *Sso*SSB was shown for interactions with *Sso*XPF (a nucleotide excision repair nuclease) [229] and *Sso*RadA (strand-exchange protein involved in homologous recombination) [42].

We were unable to capture a definitive physical interaction of *Sso*MCM and *Sso*SSB through either pull-down or fluorescent EMSA. *Sso*SSB may just capture ssDNA after unwinding without a direct interaction with the helicase (*Sso*MCM). Another possibility is that another protein mediates the interaction between *Sso*SSB and *Sso*MCM. A recent report showed that the helicase binding domain of *Ec*DnaG interacts with the acidic C-terminal tail of *Ec*SSB [39]. In addition to *Ec*DnaG, a large number of other replication and repair proteins interact with the acidic C-terminal tail of *Ec*SSB as a binding motif [24]. *Sso*SSB also contains an acidic C-terminal tail that may facilitate similar interactions in this species and would be interesting to test. As shown in Figure 1B, there is homology between *Sso* and *E. coli* in the C-terminal region. However, differences within the archaeal domain also exist in which crenarchaea (*Sso*), thaumarchaea, and korarchaea contains SSB, while euryarchaea generally harbor an RPA homolog and do not contain SSB [33]. In *Sso*, we have shown that *Sso*MCM and *Sso*DnaG (a bacterial-like primase) interact [97]. *Sso*DnaG may bridge the gap between *Sso*MCM and *Sso*SSB. Functionally, *Sso*MCM unwinds duplex DNA into ssDNA while *Sso*DnaG makes primers from that separated ssDNA and *Sso*SSB is able to protect ssDNA from degradation and/or reannealing.

5.6 CONCLUSIONS

*Sso*SSB forms a monomer in solution and harbors DNA thermodynamic melting capabilities. At high concentrations, it inhibits *Sso*MCM's unwinding and ATPase abilities likely by sequestering DNA. *Sso*SSB stabilizes MCM binding to DNA at low concentrations prior to thermodynamic melting. No physical interaction between *Sso*SSB and *Sso*MCM was detected.

5.7 ACKNOWLEDGEMENTS

We would like to thank James Keck for providing the *Ec*SSB plasmid.

APPENDIX A: CROSSLINKING TO MAP *SSO*MCM 5'-TAIL DNA INTERACTION SITES

A.1 INTRODUCTION

DNA helicases unwind duplex DNA to yield two complementary strands prior to replication. Minichromosome maintenance protein MCM 2-7 functions as the eukaryotic DNA replicative helicase as a heterohexamer. *Sulfolobus solfataricus* MCM, an ancestrally related archaeal species, functions similarly as the DNA replicative helicase, but as a simplified homohexameric. *Sso*MCM unwinds DNA with 3'-5' directionality [111], and during unwinding it has been shown that the 5'-tail of a forked substrates dynamically interacts the exterior surface [110]. We have recently demonstrated that the helicase utilizes a steric exclusion and wrapping (SEW) model of unwinding where the 5'-tail wraps around the exterior surface of MCM [123]. Utilizing the recent near full-length crystal structure [113], we identified two basic residues (K323 and R440) that when mutated to alanine disrupt this interaction [123]. In order to more thoroughly characterize the binding path, a crosslinking protocol was designed to map the external binding sites of the 5'-tail. Briefly, biotinylated-DNA was crosslinked to MCM through either chemical agents or ultraviolet (UV) light. Trypsin digestion of the crosslinked DNA-MCM followed by biotin/streptavidin bead selection will be used to isolate the crosslinked peptide fragment. Afterwards, the crosslinks will be reversed, and the tryptic fragments will be subjected to electrospray tandem mass spectrometry (ESI MS-MS) for identification. This basic protocol has been used recently to characterized RNA-protein and protein-protein interactions utilizing glutaraldehyde, trypsin digestion, and MS [230]. We hoped to identify DNA-protein interactions with the 5'-tail with this approach. Unfortunately, we were unable to complete this mapping due to low abundant products after isolation. However, we were able to show crosslinking of DNA to MCM, trypsin digestion, and reversal of crosslinks.

A.2 MATERIALS AND METHODS

A.2.1 Materials

Optikinase was purchased from USB (Cleveland, OH). Glutaraldehyde was purchased from Amresco (Solon, OH). TPCK-treated Trypsin and 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF, a trypsin protease inhibitor) were acquired from Sigma (St. Louis, MO). *Dynabeads® M-280 Streptavidin* were acquired from (GE Healtcare, Piscataway, NJ). All other materials were from commercial sources and were analytical grade or better. Helicase buffer is used in all binding reactions and consists of 125 mM potassium acetate, 25 mM HEPES acetate (pH 7.5) and 10 mM magnesium acetate.

A.2.2 Cloning and protein purification

Full-length His-tagged (WT) *Sso*MCM was purified as previously described using 70 °C heat treatment as well as MonoQ, heparin and gel filtration columns to isolate the hexameric species [111].

A.2.3 DNA substrates

Oligonucleotides (Table A.1) were purchased from IDT Corporation and gel purified [144]. [γ-^{32}P]ATP was purchased from Perkin Elmer (Boston, MA) and used with Optikinase to ^{32}P label the 5′-end. Complementary DNA was added in a ratio of 1.2:1. ^{32}P-labeled DNA substrates were heated at 95 °C for 5 min and then cooled to room temperature after turning off the heat block.

Table A.1: DNA substrates

DNA	Sequence[1, 2]
3'-tail-30 nt	5'-CACCTCTCCCTACGCTTCCCACCCACCCCGACCGGCATCTGCTATGGTAC GCTGAGCGAGAGTAGC
5'-tail-50 nt (5'-B)	5'-CGATGAGAGCCGATGAGAGCCGATGAGAGCGAGTCGCATGGTATCGTCTA GCCGGTCGGGGTGGGTGGGAAGCGTAGGGAGAGGTG**B**
5'-tail-50 nt (BrdU)	5'-CGATGAGAGCCGATGAGAGCCGATGAGAGC(**BrdU**)GAGTCGCATGGTATCG TCTAGCCGGTCGGGGTGGGTGGGAAGCGTAGGGAGAGGTG

[1]Modifications bolded and underlined [2]B – biotin, BrdU - bromodeoxyuridine

A.2.4 Glutaraldehyde crosslinking

20 µL reactions with helicase buffer, DNA (radiolabeled), MCM and glutaraldehyde are incubated at room temperature for the specified amounts of time. Reactions were quenched with Laemmli Sample Buffer (LSB), consisting of 60 mM Tris pH 6.8, 10% glycerol, 5% BME, 2% SDS, and 0.01% bromophenol blue w/v supplemented with 100 mM Tris pH 8.0. 5% SDS-PAGE was used to resolve the bands. If applicable the gels were exposed to phosphor screens, imaged using a Storm 820 Phosphorimager (GE Healthcare), and the fraction crosslinked was calculated. Gels were also stained with Coomassie Brilliant Blue.

A.2.5 Trypsin digestion

Trypsin was dissolved to 1 mg/mL in PBS. 5 mM DTT in 333 mM ammonium bicarbonate (DTT buffer) has been shown to promote trypsin digestion [231]. 20 µL reactions in the absence of crosslinker were assembled using helicase buffer, DTT in NH_4HCO_3, MCM, and trypsin. The reactions were incubated at the temperatures and times specified, until quenching with 2 µL of 50 mg/mL AEBSF and storage on ice until loading. Crosslinked trypsin digestions omitted DTT in NH_4HCO_3 and additionally has 10 mM $CaCl_2$, radiolabeled DNA and the amounts of urea specified. In both instances 5% or 15% SDS-PAGE were used to resolve the bands, with either staining or phosphorimaging as necessary.

A.2.6 BrdU UV crosslinking

The basic protocol for UV crosslinking DNA to protein was followed [232]. Briefly, 10 µL reactions consisting of helicase buffer, BrdU-containing radiolabeled DNA, and MCM were incubated in covered open tubes at a distance of either 5 or 10 cm from a UVP Transilluminator for the specified amounts of time (Upland, CA). Two different wavelengths were used: 254 and 302 nm. Reactions were quenched with LSB prior to loading. 6% SDS-PAGE was used to resolve the bands with either staining or phosphorimaging as necessary.

A.3 RESULTS

A.3.1 Crosslinking to reveal surface residues of *Sso*MCM responsible for binding the 5'-tail

We have previously confirmed the ability for *Sso*MCM to stably bind DNA on its external surface [123]. As seen in Figure A.1, a multistep protocol was developed to more specifically identify the surface residues interacting with the DNA 5'-tail. Briefly, MCM is loaded onto biotinylated-DNA and then chemically crosslinked. Crosslinked DNA-MCM is then trypsin digested and enriched by pulling down biotin via a biotin-streptavidin interaction. After reversing the crosslinks, the tryptic fragments are subjected to electrospray ionization mass spectrometry-mass spectrometry (ESI MS-MS) whereby the residues interacting with the 5'-tail of the DNA can be identified.

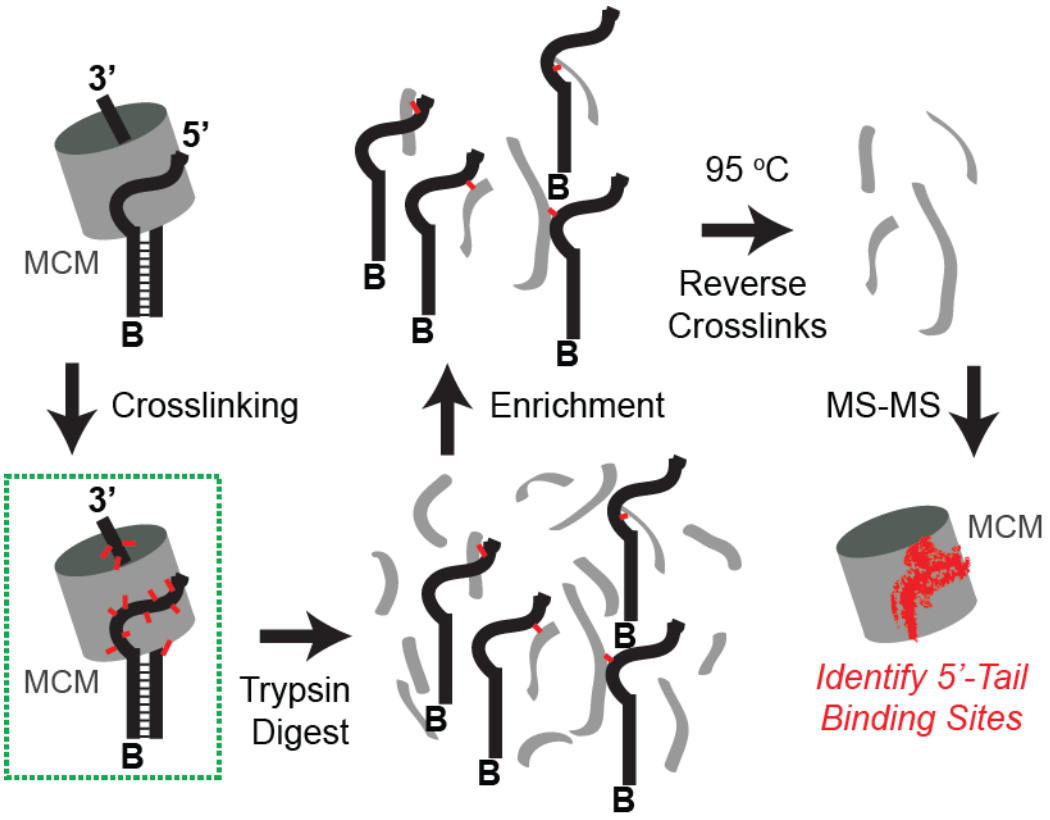


Figure A.1: Crosslinking protocol
Crosslinking protocol to determine external binding sites for *Sso*MCM. First MCM is loaded onto biotinylated-DNA. Then crosslinker is added which covalently bonds DNA to the protein surface. Trypsin is subsequently added to digest the protein. Enrichment is achieved via a biotin-streptavidin interaction through the use of *Dynabeads® M-280 Streptavidin*. After the removal of non-biotinylated material, the crosslinks are reversed through boiling at 95º C to release the tryptic fragments that had once bound the exterior surface of MCM. Finally, the fragments will be identified through MS-MS allowing the identification of residues on the 5'-tail binding site.

A.3.2 Glutaraldehyde chemically crosslinks DNA to MCM

Glutaraldehyde has been used for crosslinking due its ease of use and lack of specificity in binding any two amino groups in close proximity [233, 234]. The five-carbon dialdehyde exists in at least 11 different forms depending on the pH of the solution, where free monomeric glutaraldehyde is present at lower pH values [235]. Glutaraldehyde has been shown to be capable of crosslinking DNA to protein [236, 237], although the mechanism remained under debate until recently [235, 238]. Briefly, in alkaline solution, a Schiff base forms with an amino group from one protein molecule and a C-N bond forms by Michael addition to β-carbon of its adjunct double bond from

an amino group of a neighboring protein group [238]. As seen in Figure A.2A, lane 2 of the Coomassie Blue gel shows primarily monomeric and some dimeric MCM after 10 minutes of crosslinking. As glutaraldehyde is titrated, there is more complete crosslinking in lanes 5-10. The minimum concentration for the 10 minute time point in lanes 5 and 6 were 0.013% w/v glutaraldehyde. The phosphorimage gel in Figure 2 A.2A shows similar maximum crosslinking for lanes 5-10. A quantification of the fraction crosslinked in A.2B shows that additional glutaraldehyde does not enhance crosslinking. The maximal fraction crosslinked was approximately 0.08. Having selected the optimal glutaraldehyde concentration, we performed a time course to optimize the incubation time. Figure A.3A shows that ten minutes (lane 5) provides full crosslinking on the Coomassie Blue gel, while phosphorimage quantification in Figure A.3B shows that 20 minutes provides maximal crosslinking with a maximal fractional value of approximately 0.14. The difference in values represents detection of MCM hexamer crosslinking compared to MCM-DNA crosslinking, respectively.

Prior to enrichment, crosslink reversal was optimized. Glutaraldehyde crosslinks are very stable, and if lysine is crosslinked it does not get regenerated after 24 hours at 110 ºC in 6 N HCl [239]. Since we did not know what residues were being crosslinked, we heat treated at 95 ºC to reverse the crosslinks. Figure A.4, interestingly, shows the monomeric through hexameric MCM species, labeled at left. Optimized crosslinking conditions (0.013% glutaraldehyde for 20 minutes at room temperature) were used for lanes 3-6. DNA was omitted from these reactions to assess the ability of crosslinking reversal for MCM alone. This was necessary due to MCM-MCM crosslinks additionally formed in the course of glutaraldehyde crosslinking. Lanes 4-6 were boiled at 95 ºC for 15, 30, and 60 minutes, respectively. Boiling occurred within quench that contained excess Tris pH 8.0, which provided an abundant source of amino groups to cease crosslinking. Somewhat surprisingly, qualitatively, lanes 4-6 all show some monomer and smaller portion of dimer while the majority of MCM is still crosslinked as a hexamer. Critically, only a small amount of protein was released from the crosslinks.

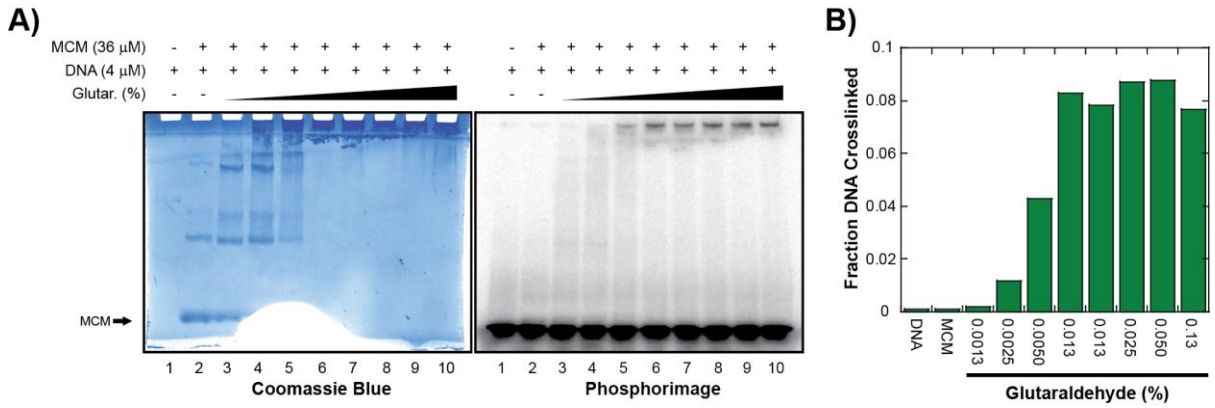


Figure A.2: Glutaraldehyde titration
A) Glutaraldehyde titration for *Sso*MCM bound to ^{32}P-biotinylated-forked DNA. 36 μM monomer MCM was bound to 4 μM DNA. Glutaraldehyde was titrated into MCM bound to DNA in lanes 3-10 at 0.0013, 0.0025, 0.005, 0.013, 0.013, 0.025, 0.05, and 0.13% w/v, respectively and incubated at room temperature for 10 min. Quenched reactions, as specified in the 'Materials and Methods,' were loaded onto 5% SDS-PAGE and subsequently Coomassie Blue stained to identify crosslinked protein and phosphorimaged to show crosslinked DNA. **B)** Quantification of the fraction DNA crosslinked for each lane of the phosphorimaged gel.

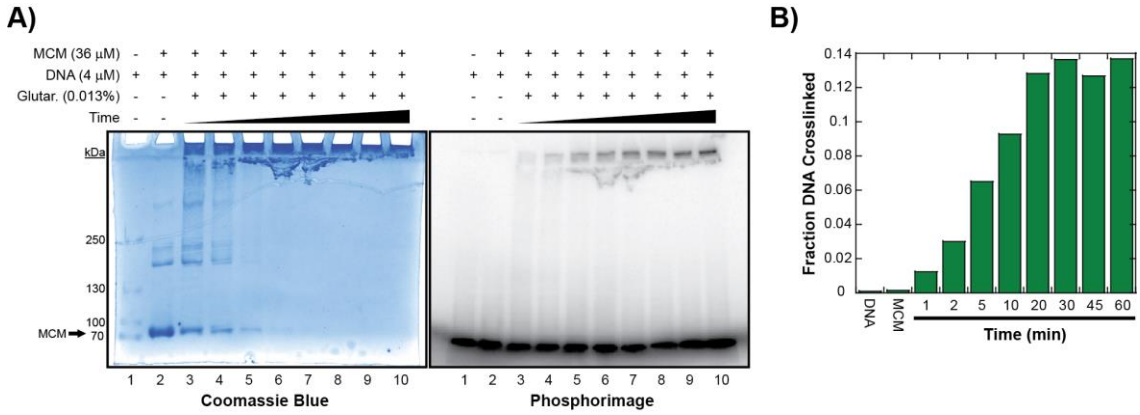


Figure A.3: Glutaraldehyde time course
A) Glutaraldehyde time course for *Sso*MCM bound to ^{32}P-biotinylated-forked DNA. 36 μM monomer MCM was bound to 4 μM DNA and then incubated at room temperature with 0.013% w/v glutaraldehyde. Reactions in lanes 3-10 were quenched at 1, 2, 5, 10, 20, 30, 45, and 60 minute(s), respectively. Quenched reactions, as specified in the 'Materials and Methods,' were ran on 5% SDS-PAGE and subsequently Coomassie Blue stained to identify crosslinked protein and phosphorimaged to show crosslinked DNA. **B)** Quantification of the fraction DNA crosslinked for each lane of the phosphorimaged gel.

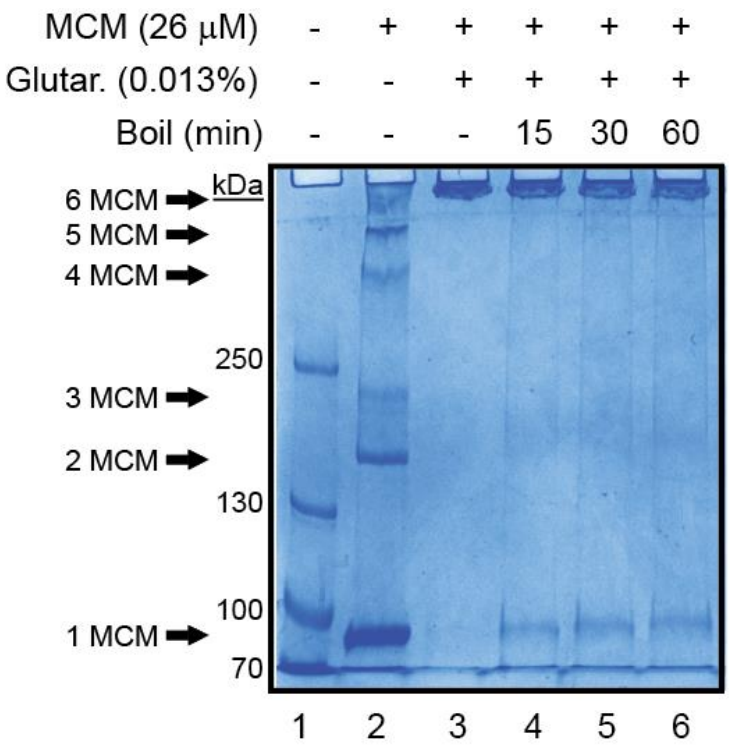


Figure A.4: Glutaraldehyde crosslinking reversal time course
Glutaraldehyde reversal time course for *Sso*MCM in the absence of DNA. 26 μM monomer MCM was crosslinked to itself with 0.013% glutaraldehyde for 20 minutes at room temperature and quenched. Lanes 4-6 were then boiled at 95º C for 15, 30, and 60 minutes, respectively. Quenched reactions, as specified in the 'Materials and Methods,' were ran on 5% SDS-PAGE and and subsequently Coomassie Blue stained to identify crosslinked protein.

A.3.3 Tryptic digests yield conflicting results

Tryptic digestions have been shown to be enhanced in the presence of urea [240]. Protein denaturation promotes trypsin cleavage by unfolding and thus permitting trypsin enhanced access to buried protein regions. In Figure A.5, we show trypsin digestion in the absence and presence of glutaraldehyde crosslinker. Urea was titrated to optimize cleavage products in the presence of glutaraldehyde crosslinks. Surprisingly, in the absence of crosslinker (lanes 3-5), the trypsin digestion products compared to MCM alone in lane 2 are either too small to be resolved on the gel or perhaps precipitated prior to loading. Lanes 6-9 show that the majority of MCM remains crosslinked and undigested. As a control, we trypsin digested MCM in the absence of crosslinker and DNA, as seen in Figure A.6. We utilized three different temperatures to verify that the trypsin was cleaving and that MCM can be trypsin digested. After the 14 hour digestion time, lanes 3-5 show a disappearance of full length MCM and only various fragments <25 kDa are present. Crosslinking likely impedes trypsin cleavage.

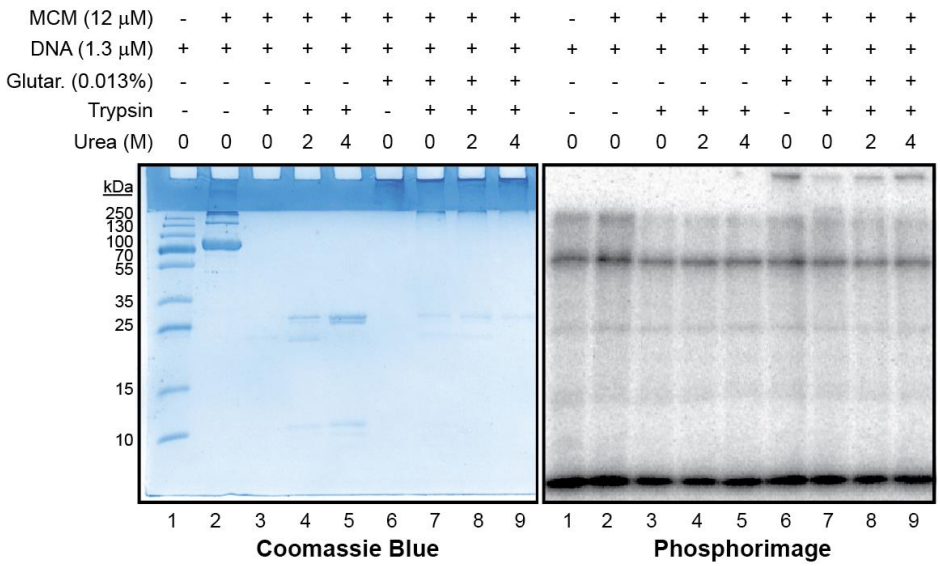


Figure A.5: Trypsin digestion of *Sso*MCM in the absence and presence of glutaraldehyde crosslinker
Trypsin digestion with urea titration in the absence and presence of glutaraldehyde crosslinker. *Sso*MCM (12 μM monomer) was bound to ^{32}P-biotinylated-forked DNA (1.3 μM). Lanes 6-9 were crosslinked with 0.013% w/v glutaraldehyde for 20 minutes at room temperature. Lanes 3-5 and 7-9 were trypsin digested for 14 hours at 37 ºC, where either 0, 2, or 4 M urea was present as specified. Quenched reactions, as specified in the 'Materials and Methods,' were ran on 15% SDS-PAGE and subsequently Coomassie Blue stained to identify crosslinked protein and phosphorimaged to show crosslinked DNA.

A.3.4 BrdU crosslinking provides specific DNA-MCM crosslinks

Because, the efficiency of both crosslinking and release was low, a different tactic was taken where DNA containing internal bromodeoxyuridine (BrdU), a halogenated analog of thymidine that can be crosslinked to protein via UV light, was utilized [232]. A cartoon showing
the forked DNA substrate and BrdU location is shown in Figure A.7A. The 5'-tail is 50 nt in length and the BrdU is 30 nt from the 5'-end. Based on our previous models, this location should be bound to the exterior of MCM [123]. The experimental setup is shown in Figure A.7B. Briefly, reactions are prepared in 1.5 mL Eppendorff tubes and are covered with parafilm. The racked tubes are then placed under the UV transilluminator and exposed to UV light to allow for crosslinking.

An example experiment is shown in Figure A.8A, here we used 10 cm exposure distance and 302 nm wavelength light. Inspection of the Coomassie Blue gel does not readily show crosslinking, but when examining the phosphorimage as time progresses, more DNA is gel-shifted and therefore crosslinked. A quantification of the phosphorimage is shown in Figure A.8B.

Compared to glutaraldehyde crosslinking, it seems that an extended UV exposure time is beneficial. As with glutaraldehyde crosslinking, we wanted to be able to reverse the crosslinks. In Figure A.9A, we compared the effect of boiling samples that were exposed to UV light in a time course. Boiling for five minutes showed a 1.6-fold reduction in the fraction UV crosslinked for the 60 minute time point (.075 to 0.048), while the shorter 15 minute time point showed a 4.8-fold reduction when boiled (.043 to .0091). Therefore, the UV exposure distance was shortened to 5 cm and two wavelengths (302 and 254 nm both for 60 minutes) were used in subsequent experiments. Figure A.10A shows that the shorter wavelength 254 nm, provides more crosslinking as seen the phosphorimage. Quantifications show that there is a 2.5-fold increase compared to 302 nm light.

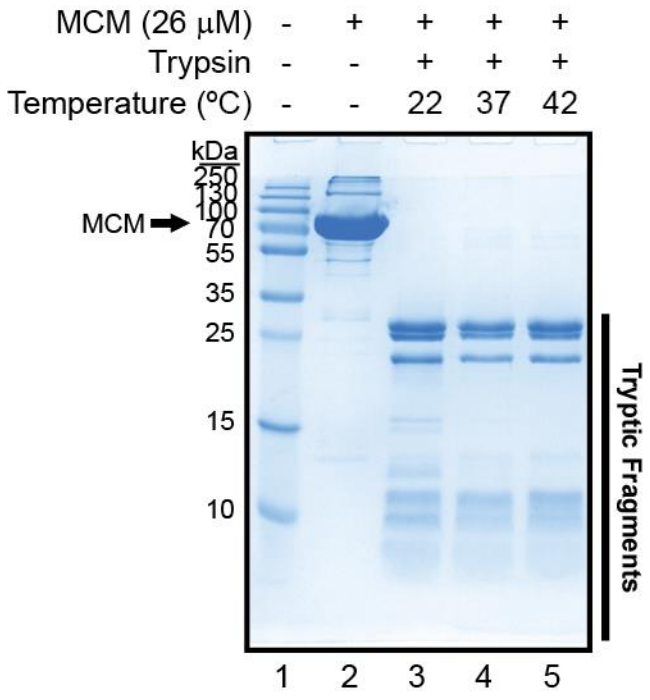


Figure A.6: Trypsin digestion at different temperatures
Trypsin digestion of *Sso*MCM in the absence of crosslinker at different temperatures. 26 µM monomer MCM was digested for 14 hours at 22, 37 and 42 ºC. The quenched reactions, as specified in the 'Materials and Methods,' were ran on 15% SDS-PAGE and subsequently Coomassie Blue stained.

A.3.5 Comparison of chemical to UV crosslinking

Optimized glutaraldehyde crosslinking shown in Figure A.3A and B demonstrates that the maximal fraction chemically crosslinked was 0.13. Similar results were seen with UV crosslinking in Figure A.10 A and B, where the maximal fraction UV crosslinked was 0.14. Perhaps owed to its lack of specificity, glutaraldehyde forms many MCM-MCM crosslinks forming the hexamer, as shown in Figure A.4. UV crosslinking, due to only one crosslink possible per hexamer loaded onto BrdU-DNA, does not show higher ordered MCM species (Figure A.10A Coomassie Blue). Crosslink reversal for both chemical and UV crosslinks were effective to varying degrees.

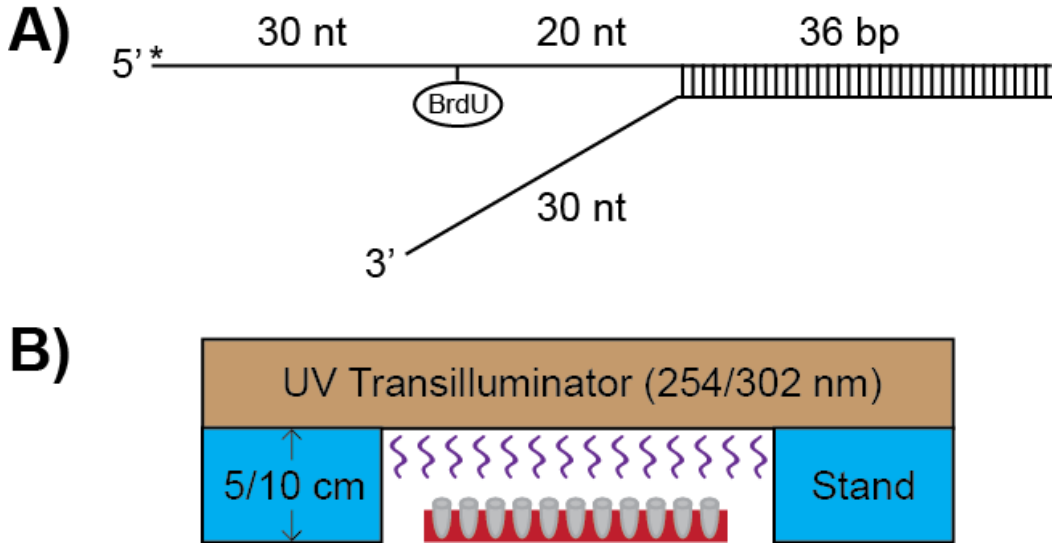


Figure A.7: Internally-labeled BrdU DNA substrate and experimental setup
A) Schematic of the ^{32}P-BrdU-forked DNA. The duplex is 36 bp and the 5'-tail (^{32}P denoted by an asterisk) is 50 nt while the 3'-tail is 30 nt. BrdU is on located 30 nt from the 5'-tail and 20 nt from the ss-dsDNA junction. **B)** Experimental setup for UV crosslinking. The UV Transilluminator was placed on a stand yielding a height or either 5 or 10 cm to the bottom of the tubes held by a rack. The tubes were covered with parafilm and exposed to UV light at either 254 or 302 nm for varying amounts of time.

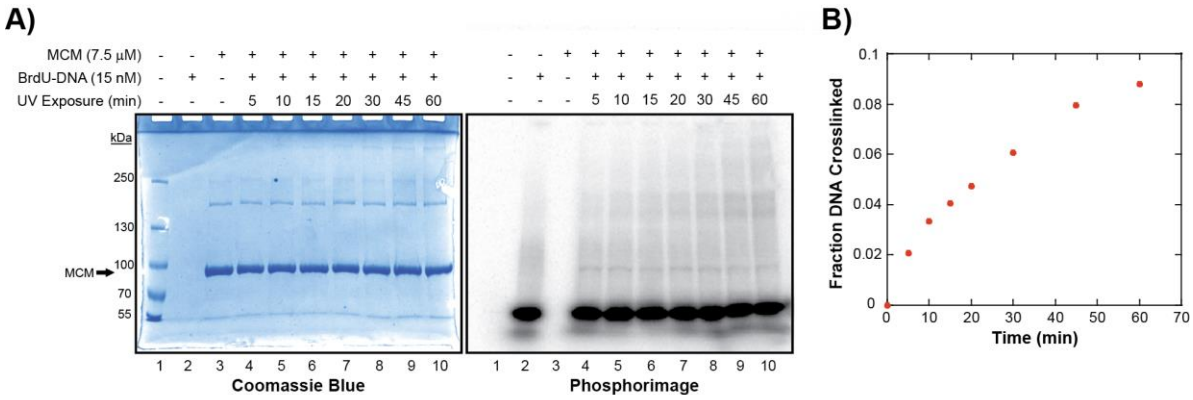


Figure A.8: BrdU crosslinking time course
A) UV crosslinking time course *Sso*MCM bound to ^{32}P-BrdU-forked DNA. 7.5 μM monomer MCM was bound to 15 nM DNA. Reactions in lanes 4-10 were exposed to 302 nm light at a distance of 10 cm for 5, 10, 15, 20, 30, 45, and 60 minutes, respectively. Quenched reactions, as specified in the 'Materials and Methods,' were ran on 6% SDS-PAGE and subsequently Coomassie Blue stained to identify crosslinked protein and phosphorimaged to show crosslinked DNA. **B)** Quantification of the fraction DNA crosslinked versus time for the phosphorimaged gel.

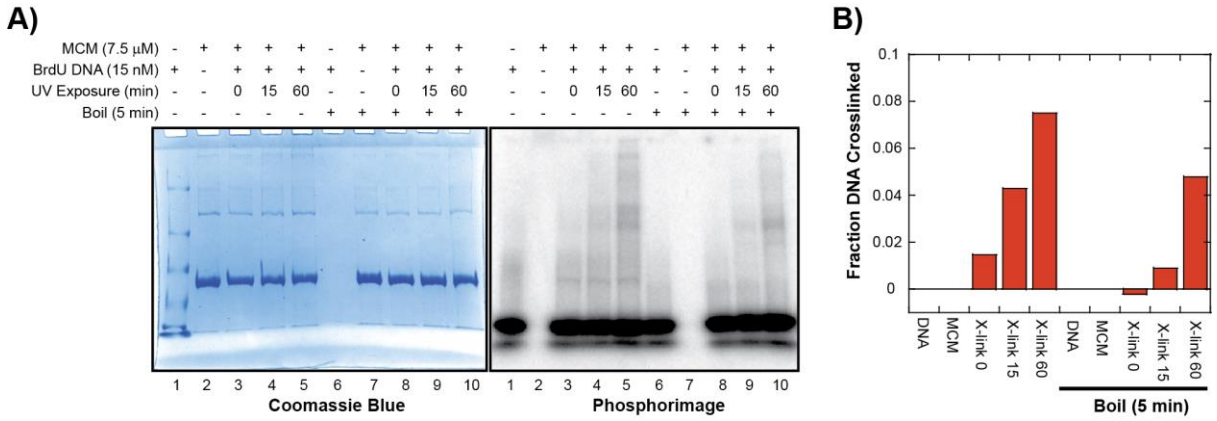


Figure A.9: BrdU crosslinking reversal time course
A) UV crosslinking time course *Sso*MCM bound to ^{32}P-BrdU-forked DNA versus crosslinking reversal by boiling. 7.5 µM monomer MCM was bound to 15 nM DNA fix the figure. Reactions in lanes 3-5 and 8-10 were exposed to 302 nm light at a distance of 10 cm for 0, 15, or 60 minutes, respectively for each set. Lanes 6-10 were boiled at 95º C for 5 minutes to reverse the crosslinks Quenched reactions, as specified in the 'Materials and Methods,' were ran on 6% SDS-PAGE and subsequently Coomassie Blue stained to identify crosslinked protein and phosphorimaged to show crosslinked DNA. **B)** Quantification of the fraction DNA crosslinked normalized to DNA alone for either lanes 1-5 (non-boiled) or 6-10 (boiled) of the phosphorimaged gel.

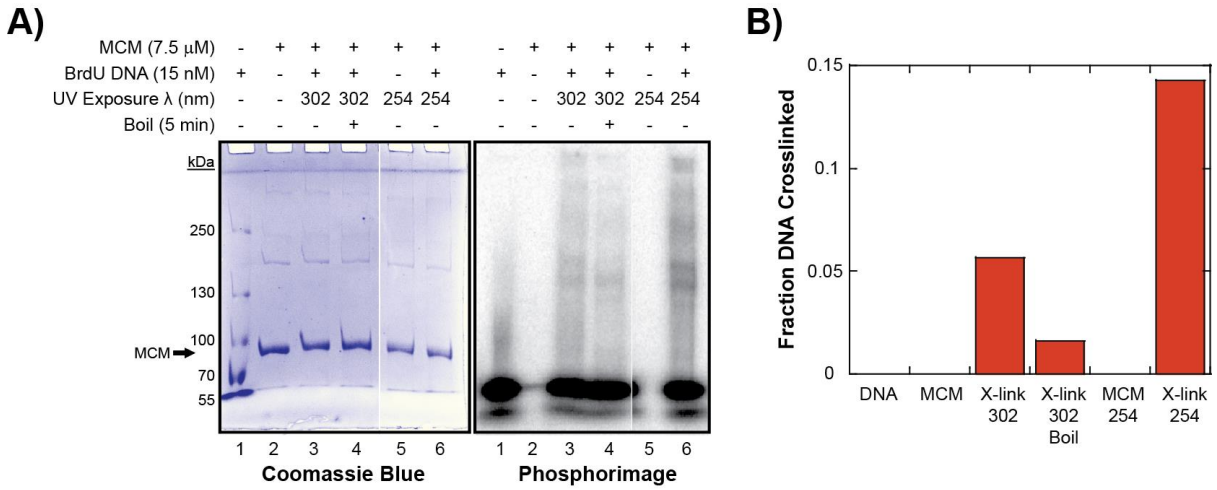


Figure A.10: BrdU crosslinking at two wavelengths
A) UV crosslinking time course *Sso*MCM bound to ^{32}P-BrdU-forked DNA versus crosslinking reversal by boiling. 7.5 µM monomer MCM was bound to 15 nM DNA. Lane one shows markers in the Coomassie gel and DNA in the phosphorimaged gel. Lane two shows MCM alone. Reactions in lanes 3-6 were exposed to 302 or 254 nm light at a distance of 5 cm for 60 minutes. Lane 4 was boiled for 5 minutes at 95º C to reverse the crosslinks. Quenched reactions, as specified in the 'Materials and Methods,' were ran on 6% SDS-PAGE and subsequently Coomassie Blue stained to identify crosslinked protein and phosphorimaged to show crosslinked DNA. **B)** Quantification of the fraction DNA crosslinked normalized to DNA alone of the phosphorimaged gel.

A.4 CONCLUSIONS AND FUTURE DIRECTIONS

Ultimately, this project did not yield the results that were expected in Figure A.1. Crosslinking DNA to MCM was shown, but at best 14% was crosslinked. This number is slightly better than what has been reported for common crosslinking yields, likely due to our optimization efforts [232]. Trypsin digestion for glutaraldehyde crosslinked MCM to DNA (Figure A.5) was unsuccessful when compared trypsin digestion of MCM alone (A.6). The reversal process was also incomplete. In Figure A.9B, the 60 minute time point (with the highest crosslinked percentage that was boiled) shows only 36% released after boiling. In summary, too much protein is either unused (not crosslinked), undigested (by trypsin), or remains crosslinked still in spite of our efforts to reverse them for effective mass spectrometry analysis. ESI MS-MS also requires fairly high concentrations of protein for analysis. According to our facility, at least 20 µL of 1 mg/mL is required for strong signal. For full-length *Sso*MCM, this equates to approximately 13 µM monomer (2.6×10^{-10} moles). Calculating concentrations after the enrichment step would also prove difficult due to the molar absorptivity based on full-length protein, while we would only have fragments. Assuming other steps in the protocol are also inefficient, this would require initial concentrations above where *Sso*MCM is stable (~500 µM). UV crosslinking, due to its specificity may prove more useful, yet would require using multiple BrdU locations on the 5'-tail to allow for complete mapping. Moreover, the amount of protein required for this protocol is outside the range of useful experimental conditions unless a parallel scale up procedure is employed.

APPENDIX B: 2-AMINOPURINE STUDIES OF TRANSCRIPTION BUBBLE SIZE FOR *SACCHAROMYCES CEREVISIAE* RNA POLYMERASE II[4]

B.1 INTRODUCTION

Transcription of DNA to RNA results in the expression of genes and is required prior to translation for protein production in the cell. In *Saccharomyces cerevisiae*, this process occurs through a transcription complex consisting of RNA polymerase II (Pol II), its general transcription factors TFIIB, TFIID, TFIIF, TFIIE, and TFIIH, and a DNA template. The transcription cycle occurs in three phases, initiation, elongation, and termination. During initiation, one of the first steps is DNA promoter melting by TFIIH helicase forming a transcription bubble (TB) that is 7-9 bases. Elongation is marked by continued TB unwinding to 18-25 bases followed by the synthesis of a short RNA strand complementary to the template strand. If the RNA primer reaches 10 nt in length, productive transcription occurs, otherwise transcription is aborted and the RNA is released. Finally, productive transcription ends with termination and release of the RNA transcript [241-244].

Even after more than 50 years of study, some of the mechanistic details of transcription are still not known, including the size of the transcription bubble, which has been shown to vary between 8-14 nucleotides among the three domains of life [241, 245]. It is also possible that the TB is not a static size but instead may depend on the stage of transcription by Pol II. One possibility for initial transcription is the "DNA scrunching" model, in which small hairpins structures form on both the template and non-template strands [246, 247]. Additionally, backtracking or interactions with other previously mentioned transcription factors may have an effect on the size(s) of the TB.

---

[4] Barnes, C.O., Calero, M., **Graham B.W.**, Malik M., Cohen, A., Lin, G., Brown, I.S., Zhang, Q., Pullarra, F., Trakselis, M.A., Kaplan, C., and Calero, G. *Structural Basis of Transcription: Crystal structure of a RNA Polymerase II transcribing complex at 4.0 Å* Science, 2014. **Submitted.** Graham performed and analyzed the 2-aminopurine fluorescence assays.

Work from the Calero laboratory has determined the crystal structure for *Saccharomyces cerevisiae* RNA Pol II bound to a complete nucleic acid scaffold (NAS) at a resolution of 4.0 Å (Figure B.1). This is the first structure that simultaneously includes diffraction data for the RNA Pol II bound to upstream double helix and the non-template strand, showing the full transcription bubble. In the downstream region of the structure, scrunching of the DNA occurred to such an extent that it was possible that some of the base pairs had become open. We set out to more closely examine the downstream base pairing through changes in 2-aminopurine florescence.

B.2 MATERIALS & METHODS

B.2.1 2-aminopurine fluorescence

Steady-state fluorescence measurements were acquired in buffer containing 50 mM Hepes (pH 7.0), 120 mM NaCl, 5 mM βME, 2 mM $MgCl_2$, 5 mM Tris (pH 7.5), 0.1 mM EDTA and 1 μM $ZnCl_2$. DNA-RNA primer templates containing 2-aminopurine (see Table B.1) were heated to 95 ºC for 5 minutes, cooled to room temperature, purified by gel filtration, and quantified using a NanoDrop to ensure stoichiometric annealed concentrations. If present, polymerase was then added to the primer template with non-template strand, i1s. Alternatively, polymerase was added to a preannealed primer template followed by addition of the fully complementary or bubble forming non-template strand. The ratio of polymerase to DNA was 1.1:1 (330 nM:300 nM). Spectra were collected using a Fluoromax-3 (HORIBA Scientific, Edison NJ). Excitation wavelengths included both 280 and 315 nm and fluorescence emission was collected from 340-400 nm. The entrance and exit slits were adjusted for optimal intensities. Spectra were collected for buffer and polymerase alone to correct for any background fluorescence. Samples containing DNA only were corrected via the subtraction of the spectrum of buffer alone; while polymerase bound samples were corrected according to (Equation 1):

$$Corrected\ (Pol-DNA)\ spectrum = (Pol-DNA)_{315} - (Pol_{315} \times \frac{(Pol-DNA)_{280}}{Pol_{280}}) \quad (1)$$

where Pol-DNA is polymerase bound to DNA, Pol is polymerase alone, and the subscripts are the excitation wavelengths [248]. The ratio of $(Pol\text{-}DNA)_{280}/Pol_{280}$ ranged from 0.68-0.98 and corrected for any drop in protein fluorescence upon binding.

Table B.1: DNA and RNA substrates

Oligo	Sequence
9mer RNA	5'- UCGAGAGGA
i2TS*	5'- CGTACCGATAAGCAAA**2**ATCCTCTCGAAGCACGGATCTTTATGGC
i3TS*	5'- CGTACCGATAAGCAA**2**AATCCTCTCGAAGCACGGATCTTTATGGC
i5TS*	5'- CGTACCGATAAGC**2**AAAATCCTCTCGAAGCACGGATCTTTATGGC
i8TS*	5'- CGTACCGATA**2**GCAAAAATCCTCTCGAAGCACGGATCTTTATGGC
compNTS#	5'- GCCATAAAGATCCGTGCTTCGAGAGGATTTTTGCTTATCGGTACG
i1compNTS#	5'- TTTTGCTTATCGGTACG
12NTS#	5'- GCCATAAAGATCCGTGGAAGCTCTCCTATTTTGCTTATCGGTACG
13NTS#	5'- GCCATAAAGATCCGTGGAAGCTCTCCTAATTTGCTTATCGGTACG
15NTS#	5'- GCCATAAAGATCCGTGGAAGCTCTCCTAAAATGCTTATCGGTACG

*Template strand DNA where **2** represents 2 – aminopurine, A designates the i+1 site and underlined residues denote where RNA binds.
#Non-template DNA strands used for 2 – aminopurine experiments, where compNTS represents the fully complementary strand, and i1compNTS represents downstream complementary DNA starting at i+1 site. Bubble substrates 12, 13, and 15bubbleNTS with the bubble regions shown in blue.

B.3 RESULTS

B.3.1 Crystal structure acquired by Calero lab

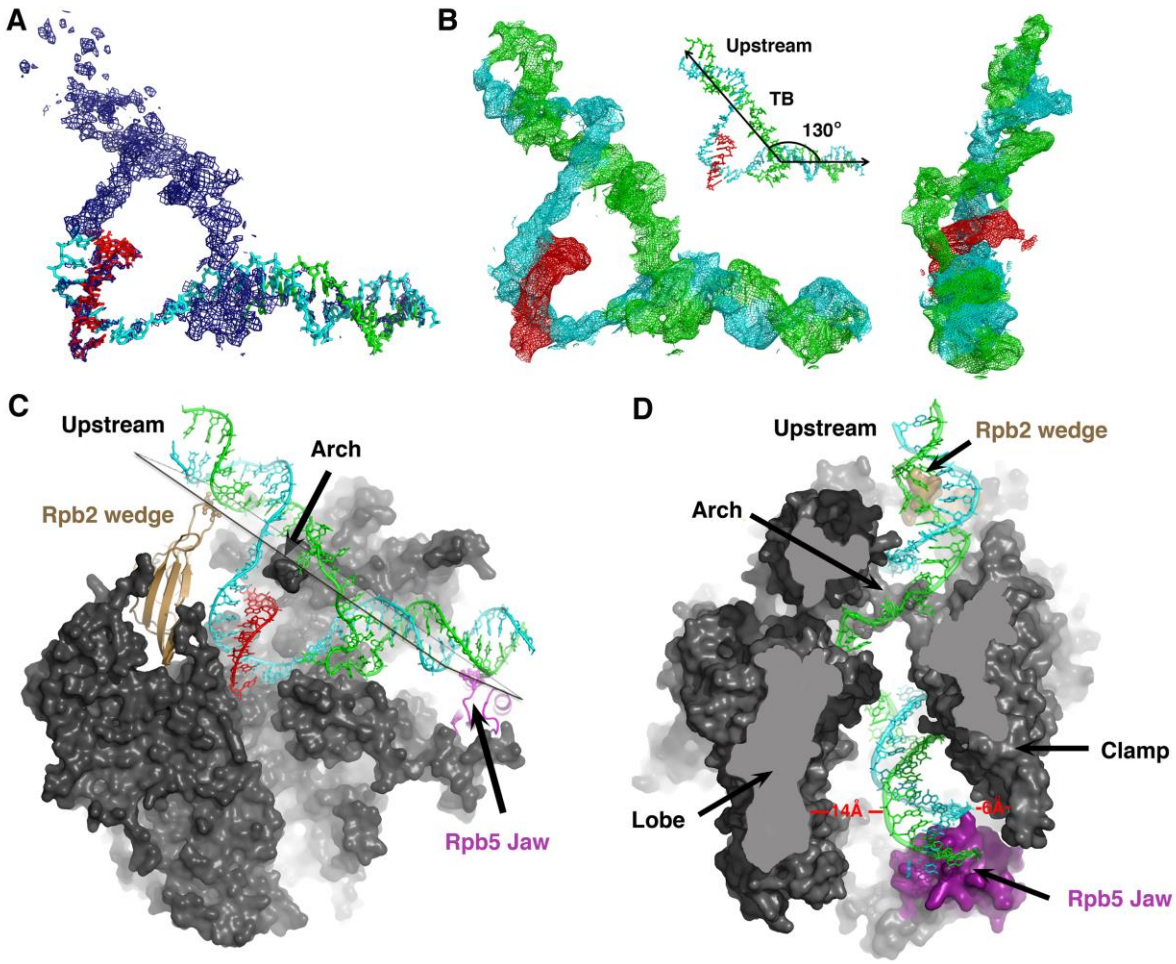


Figure B.1: Overall structure and binding of the NAS to Pol II
A) Difference F_{obs}-F_{calc} electron density map contoured at 2σ. B) F_{obs}-F_{calc} map of the final refined map contoured at 0.8σ. The DNA-RNA hybrid was easily identified and was used as reference to place the rest of the DNA. The refined nucleic acid scaffold (NAS) is 40 nucleotides-long; downstream and upstream duplexes form an angle of approximately 130º degrees. The following color scheme for the NAS will be used throughout: cyan, template strand (TS); green, non-template strand (NTS); red, RNA transcript. C) Surface representation (side view, Rpb2 removed) illustrating the position of the NAS inside Pol II and its stabilization by Rpb2 wedge-residues, Rpb5 jaw-residues and arch residues comprising Rpb1 rudder and Rpb2 FL2 residues. The NAS is supported by three Pol II structures – wedge, jaw and arch residues– lying on almost a perfect plane, minimizing strain during elongation. D) Cut-away surface representation (top view) illustrating wedge-, arch- and jaw-NAS interactions. The downstream duplex binds asymmetrically inside Pol II's cleft, cf. clamp-DNA distance of ≈6 Å vs. lobe DNA distance of ≈14 Å due to interactions with clamp and jaw residues.

B.3.2 2-aminopurine substrate

2-aminopurine (2AP) is a fluorescent nucleotide analog that is significantly quenched upon base pairing to either T or C. In these experiments, 2-aminopurine is placed in the template strand (TS) 2, 3, 5, or 8 residues downstream from the RNA primer (Table B.1). Five separate non-template strands (NTS) were utilized (Table B.1). Fully complementary NTS spans the length of the TS. i1s is complementary only to the region downstream from the primer. Additionally three bubble forming NTS were used with bubbles spanning 12, 13, or 15 base pairs. Bubble formation begins two base pairs prior to the RNA primer and terminates at the i+2 position for 12, i+3 position for 13 and i+5 for 15, corresponding i+2, i+3, and i+5 TS substrates were used. Upon addition of polymerase, if there are disruptions of the base stacking interactions we will see an increase in fluorescence, whereas fluorescence values that remain constant are indicative of no disruption.

B.3.3 2-aminopurine fluorescence shows RNA Pol II opens fully complementary and i1s non-template strand to the i+3 position

A cartoon showing the results for binding to primer-template bound to fully complementary NTS is shown in Figure B.2A. As expected, when the primer-template is annealed to complementary non-template strand, full quenching is observed consistent with stable base pairing of 2AP (Figure B.2B). When polymerase is added to fully complementary non-template strand substrates we see a 4.5-fold increase in fluorescence for the i+2 substrate and a 2.8-fold increase in fluorescence for the i+3 (Figure B.2B). This increase in fluorescence indicates a disruption in 2AP base pairing at both locations. There is essentially no fluorescence enhancement for i+5 and i+8 suggesting 2AP is more stably base paired at those positions. Similarly when comparing polymerase bound to primer-template and primer-template annealed to fully complementary NTS, there are significant decreases in fluorescence for i1s at the i+5 and i+8, indicating more stable base pairing at 2AP (Figure B.2C). However, for i+2 and i+3 substrates, the fluorescence decreases to 0.68 and 0.39, respectively, relative to ssDNA alone, again suggesting 2AP base pairing disruption.

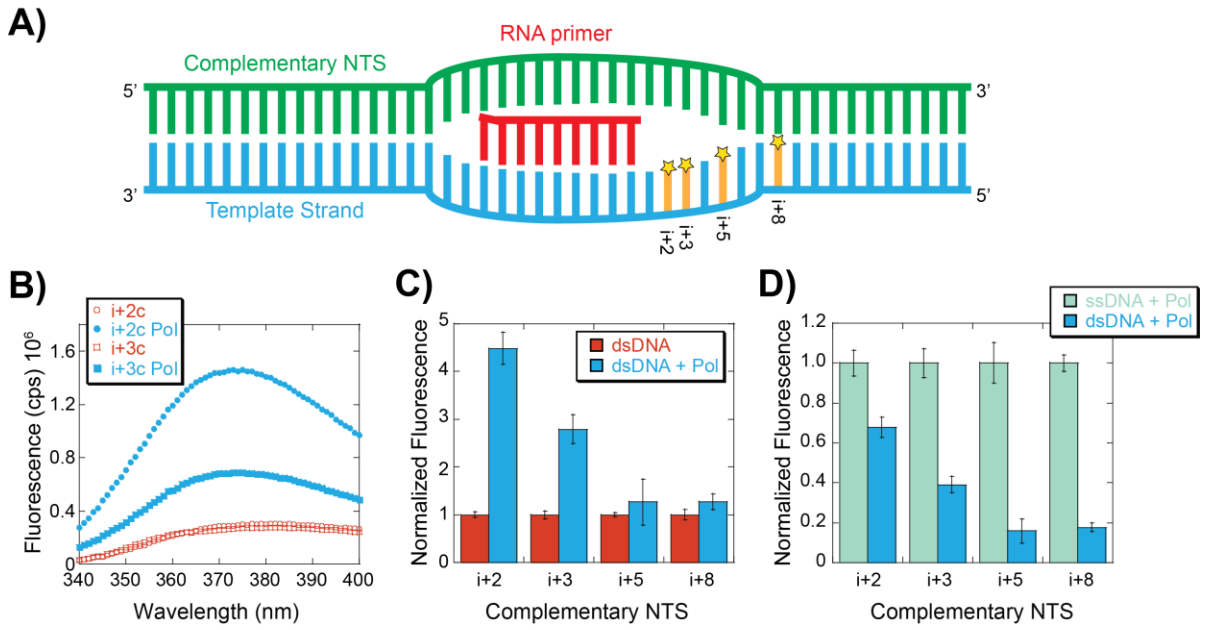


Figure B.2: 2-aminopurine fluorescence and spectra for fully complementary NTS
A) Cartoon of DNA-RNA hybrid bound to fully complementary NTS where DNA TS (blue), while RNA primer is shown in red, the fully complementary NTS (green), and the 2AP locations are shown in orange with stars. B) Representative spectra of i+2 and i+3, bound to fully complementary non-template strand (c) in the absence (open orange circles and squares, respectively) and presence (blue circles and squares, respectively) of polymerase (Pol). The excitation wavelength was 315 nm and the fluorescence emission (shown in counts per second $\times 10^6$) was collected from 340-400 nm. C) Normalized fluorescence values at the observed maxima (373 nm) for 2AP positioned at i+2c, i+3c, i+5c, and i+8c, where c indicates the presence of fully complementary non-template strand (NTS) for dsDNA n the absence (orange) and presence (blue) of polymerase. D) Normalized fluorescence values for polymerase bound to ssDNA (primer-template) where 2AP is at the i+2, i+3, i+5 or i+8 position (green) and polymerase bound to dsDNA (primer-template annealed to i1s non-template strand) (blue). The average fluorescence from at least three experiments is plotted with standard error.

Similar results are seen for primer-template bound to downstream i1s NTS (Figure B.3A). Comparable increases for i+2 (4.1-fold) and i+3 (2.3-fold) are noted upon the addition of polymerase (Figure B.3B and C). Again in the case of i+5 and i+8, there are no significant increases in fluorescence. When comparing polymerase bound to primer-template to primer-template annealed to i1s, there are significant decreases for i+5 and i+8 positions indicating stable binding. Decreases for i+2 and i+3 2AP positions only show decreases to 0.73 and 0.45, respectively. These results indicate that for the i1s NTS base pair opening occurs at both the i+2 and i+3 positions.

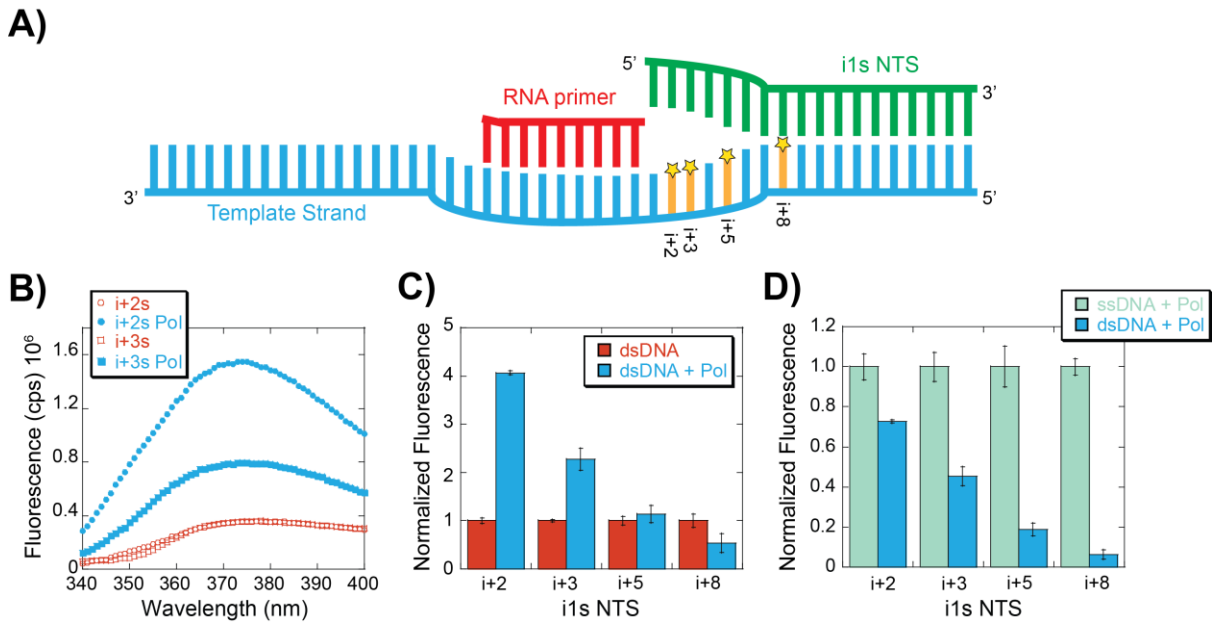


Figure B.3: 2-aminopurine fluorescence and spectra for i1s NTS
A) Cartoon of DNA-RNA hybrid bound to i1s complementary NTS; where DNA TS (blue), the RNA primer (red), i1s NTS (green), and the 2AP locations are shown in orange with stars. B) Representative spectra of i+2 and i+3, bound to i1s complementary NTS (s) in the absence (open orange circles and squares, respectively) and presence (blue circles and squares, respectively) of polymerase (Pol). The excitation wavelength was 315 nm and the fluorescence emission (shown in counts per second $\times 10^6$) was collected from 340-400 nm. C) Normalized fluorescence values at the observed maxima (373 nm) for 2AP positioned at i+2, i+3, i+5, and i+8, bound to i1s NTS) for dsDNA in the absence (orange) and presence (blue) of polymerase. D) Normalized fluorescence values for polymerase bound to ssDNA (primer-template) where 2AP is at the i+2, i+3, i+5 or i+8 position (green) and polymerase bound to dsDNA (primer-template annealed to i1s non-template strand) (blue). The average fluorescence from at least three experiments is plotted with standard error.

B.3.4 Terminal 2-aminopurine bubble substrate shows opening at i+5

Three bubble substrates were also tested. A cartoon example showing i+5 primer-template bound to 15 bubble-forming non-template strand is shown in Figure B.4A. i+2 bound to 12 and i+3 bound to 13 retract the bubble size to the terminal 2AP-labeled base pair. Three bubble sizes result where i+5 bound to 15 yields a 15 nt bubble, i+3 bound to 13 yields a 13 nt bubble and i+2 bound to 12 yields a 12 nt bubble. Interestingly for the bubble substrates, there is a slightly different trend than expected based on the results from i1s and fully complementary NTS. Increases in fluorescence for 12 and 15 in the presence of polymerase are observed, but not for 13 (Figure B.4B and C). The fluorescence increase for 12 (1.4-fold) and for the first time, we see an increase at i+5 position for 15 (1.8-fold) (Figure B.4B and C). When comparing Pol II bound to primer-template to Pol II

bound to the bubble substrates, we see a slight decrease for bubble substrates to values 12 (0.98), 13 (0.72), and 15 (0.82). This is indicative of an open base pairing at the 2AP position at i+2 for 12, i+3 for 13, and i+5 for 15. We attribute this apparent opening to the results in Figure B.4B for the bubble substrates alone. For 12 and 13 in the absence of Pol II, we see maxima values in the range of 1.5 and 1.3 × 10^6 cps, while for 15 the value is closer to 0.5 × 10^6 cps. This would indicate that 15 is base paired while 12 and 13 are likely not base paired or fractionally base paired even though the terminal residue, the 2AP site, should be base paired for all three bubble substrates. When comparing Figure B.3B to Figures B.1B and 2B at the i+2 and i+3 positions we see much more quenching (0.3 × 10^6 cps). However it is likely that the 15 bubble substrate is base paired and the increase shown in Figure B.3B and C in the presence of Pol II likely demonstrates base pair opening.

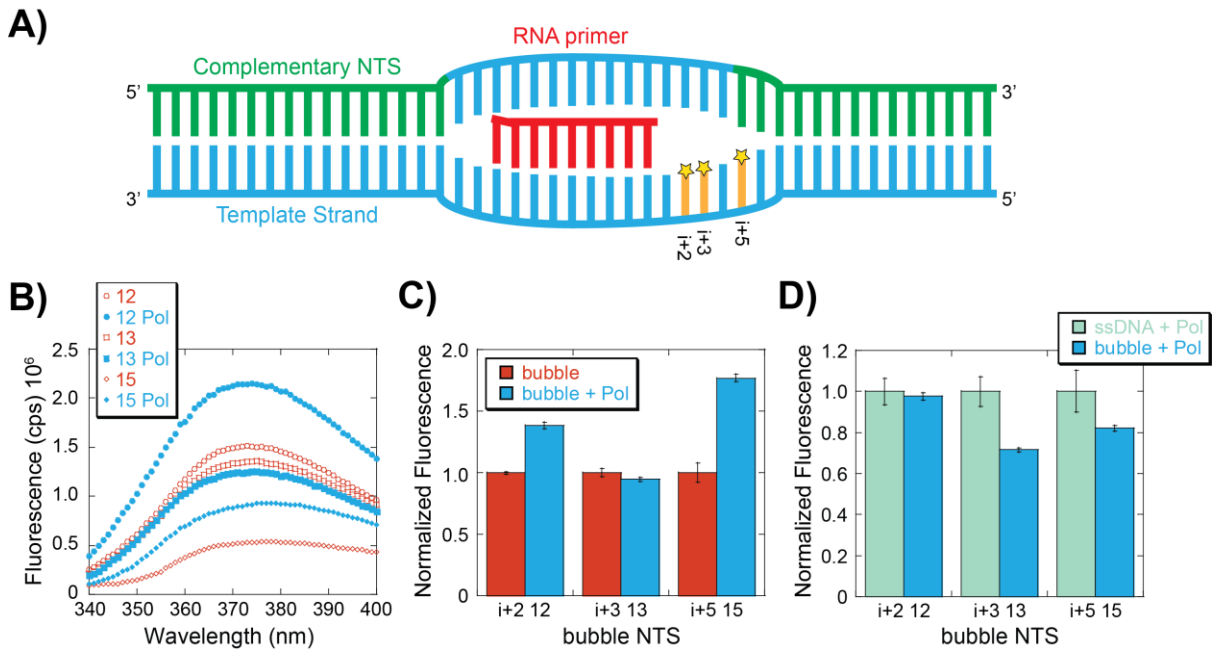


Figure B.4: 2-aminopurine fluorescence and spectra for bubble substrates
A) Example cartoon of i+5 DNA-RNA hybrid bound to 15 bubble NTS; where DNA TS (blue), the RNA primer (red), i1s NTS (green), and the 2AP locations are shown in orange with stars. i+2 and i+3 are bound to 12 and 13 bubble NTS, respectively. B) Representative spectra of bubble substrates 12, 13, and 15 in the absence (open orange circles, squares, and diamonds respectively) and presence (blue circles, squares, and diamonds respectively) of polymerase (Pol). The excitation wavelength was 315 nm and the fluorescence emission (shown in counts per second ×10^6) was collected from 340-400 nm. C) Normalized fluorescence values at the observed maxima (373 nm) for bubble substrates i+2 12, i+3 13, and i+5 15 for dsDNA in the absence (orange) and presence (blue) of polymerase. D) Normalized fluorescence values for polymerase bound to ssDNA (primer-template) where 2AP is at the i+2, i+3, or i+5 position (green) and polymerase bound to dsDNA (primer-template annealed to the bubble substrate 12, 13, and 15, respectively) (blue). The average fluorescence from at least three experiments is plotted with standard error.

B.4 DISCUSSION

2AP fluorescence monitoring allows for the study of base pairing in the downstream region. *E. coli* RNA Pol II unwinds or disrupts base pairing prior to transcribing the DNA into RNA. This process would allow for the proper orientation of DNA into the active site of Pol II. The crystal structure (Figure B.1) demonstrates that base pairs are opened in the presence of Pol II. This result has been confirmed through the use of 2AP in various positions. The results indicate that the i+2 are both fully complementary downstream DNA oligos. The 15 bubble substrate showed that it is likely that base pair opening continues beyond i+3 to i+5. As Pol II is actively transcribing, base pairs are destabilized and unwound ahead of the complex. Since this assay did not allow for active monitoring of the elongation of the RNA primer, reannealing was likely occurring in the downstream region which may have not allowed us to capture i+5 in its open state. Previously, 2AP had shown that only one downstream base pair is in an open conformation [249]. However, these experiments indicate that both i+2 and i+3 are open in the presence of complementary NTS and RNA Pol II, and that i+5 bound to 15 bubble forming NTS suggests that base opening could extend to that position during active transcription.

By combining crystallographic information with the fluorescent 2AP experiments, we can correlate our results to provide additional support for destabilization of the downstream DNA that occurs during transcription by RNA Pol II. A "DNA scrunching" model has been previously shown in which the polymerase remains stationary and pulls downstream DNA towards itself while synthesizing RNA [242, 246, 247]. Similar scrunching of DNA has been shown in the current crystal structure, suggesting that scrunching is conserved in all domains of life. Unwinding in the downstream region may serve other functional purposes for RNA transcription, which remain open until further study.

BIBLIOGRAPHY

1. Watson, J.D. and Crick, F.H.C., Molecular structure of nucleic acids. *Nature*, 1953. **171**: p. 737-738.
2. Ellis, N.A., DNA helicases in inherited human disorders. *Current Opinion in Genetics & Development*, 1997. **7**(3): p. 354-363.
3. Opresko, P.L., Cheng, W.-H., and Bohr, V.A., Junction of RecQ helicase biochemistry and human disease. *Journal of Biological Chemistry*, 2004. **279**(18): p. 18099-18102.
4. Williams, G.H., Romanowski, P., Morris, L., Madine, M., Mills, A.D., Stoeber, K., Marr, J., Laskey, R.A., and Coleman, N., Improved cervical smear assessment using antibodies against proteins that regulate DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, 1998. **95**(25): p. 14932-14937.
5. Freeman, A., Morris, L.S., Mills, A.D., Stoeber, K., Laskey, R.A., Williams, G.H., and Coleman, N., Minichromosome maintenance proteins as biological markers of dysplasia and malignancy. *Clinical Cancer Research*, 1999. **5**(8): p. 2121-2132.
6. Coleman, N. and Laskey, R.A., Minichromosome maintenance proteins in cancer screening. *European Journal of Cancer*, 2009. **45**(Supplement 1): p. 416-417.
7. Brosh Jr., R.M., Opresko, P.L., and Bohr, V.A., Enzymatic mechanism of the WRN helicase/nuclease, in Methods in Enzymology. 2006, Academic Press. p. 52-85.
8. Kelman, L.M. and Kelman, Z., Archaea: an archetype for replication initiation studies? *Molecular Microbiology*, 2003. **48**(3): p. 605-615.
9. Carpentieri, F., De Felice, M., De Falco, M., Rossi, M., and Pisani, F.M., Physical and functional interaction between the mini-chromosome maintenance-like DNA helicase and the single-stranded DNA binding protein from the crenarchaeon *Sulfolobus solfataricus*. *Journal of Biological Chemistry*, 2002. **277**(14): p. 12118-12127.
10. Barry, E.R. and Bell, S.D., DNA replication in the archaea. *Microbiology and Molecular Biology Reviews*, 2006. **70**(4): p. 876-887.
11. Woese, C.R., Kandler, O., and Wheelis, M.L., Towards a natural system of organisms: proposal for the domain archaea, bacteria, and eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, 1990. **87**: p. 4576-4579.
12. She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J., Chan-Weiher, C.C.Y., Clausen, I.G., Curtis, B.A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P.M.K., Heikamp-de Jong, I., Jeffries, A.C., Kozera, C.J., Medina, N., Peng, X., Thi-Ngoc, H.P., Redder, P., Schenk, M.E., Theriault, C., Tolstrup, N., Charlebois, R.L., Doolittle, W.F., Duguet, M., Gaasterland, T., Garrett, R.A., Ragan, M.A., Sensen, C.W., and Van der Oost, J., The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proceedings of the National Academy of Sciences of the United States of America*, 2001. **98**(14): p. 7835-7840.
13. Zillig, W., Stetter, K.O., Simon Wunderl, S., Schulz, W., Priess, H., and Scholz, I., The *Sulfolobus*-"Caldariella" group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Archives of Microbiology*, 1980. **125**: p. 259-269.

14. Brock, T.D., Brock, K.M., Belly, R.T., and Weiss, R.L., *Sulfolobus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Archives of Microbiology*, 1972. **84**: p. 54-68.
15. Sarmiento, F., Long, F., Cann, I., and Whitman, W.B., Diversity of the DNA replication system in the Archaea domain. *Archaea*, 2014. **2014**: p. 15.
16. Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., Stahl, D.A., Wagner, M., and Schleper, C., Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum *Thaumarchaeota*. *Trends in Microbiology*, 2010. **18**(8): p. 331-340.
17. Wold, M.S., Replication protein A: A heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annual Review of Biochemistry*, 1997. **66**(1): p. 61-92.
18. Ashton, N., Bolderson, E., Cubeddu, L., O'Byrne, K., and Richard, D., Human single-stranded DNA binding proteins are essential for maintaining genomic stability. *BMC Molecular Biology*, 2013. **14**(1): p. 9.
19. Alberts, B.M. and Frey, L., T4 bacteriophage gene 32: A structural protein in the replication and recombination of DNA. *Nature*, 1970. **227**(5265): p. 1313-1318.
20. Richard, D.J., Bolderson, E., Cubeddu, L., Wadsworth, R.I.M., Savage, K., Sharma, G.G., Nicolette, M.L., Tsvetanov, S., McIlwraith, M.J., Pandita, R.K., Takeda, S., Hay, R.T., Gautier, J., West, S.C., Paull, T.T., Pandita, T.K., White, M.F., and Khanna, K.K., Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature*, 2008. **453**(7195): p. 677-681.
21. Richard, D.J., Bolderson, E., and Khanna, K.K., Multiple human single-stranded DNA binding proteins function in genome maintenance: Structural, biochemical and functional analysis. *Critical Reviews in Biochemistry and Molecular Biology*, 2009. **44**(2-3): p. 98-116.
22. Murzin, A.G., OB(oligonucleotide/oligosaccharide binding)-fold: Common structural and functional solution for non-homologous sequences. *The EMBO Journal*, 1993. **12**(3): p. 861-867.
23. Chase, J.W., L'Italien, J.J., Murphy, J.B., Spicer, E.K., and Williams, K.R., Characterization of the *Escherichia coli* SSB-113 mutant single-stranded DNA-binding protein. Cloning of the gene, DNA and protein sequence analysis, high pressure liquid chromatography peptide mapping, and DNA-binding studies. *Journal of Biological Chemistry*, 1984. **259**(2): p. 805-814.
24. Shereda, R.D., Kozlov, A.G., Lohman, T.M., Cox, M.M., and Keck, J.L., SSB as an organizer/mobilizer of genome maintenance complexes. *Critical Reviews in Biochemistry and Molecular Biology*, 2008. **43**(5): p. 289-318.
25. Shishmarev, D., Wang, Y., Mason, C.E., Su, X.-C., Oakley, A.J., Graham, B., Huber, T., Dixon, N.E., and Otting, G., Intramolecular binding mode of the C-terminus of *Escherichia coli* single-stranded DNA binding protein determined by nuclear magnetic resonance spectroscopy. *Nucleic Acids Research*, 2014. **42**(4): p. 2750-2757.
26. Bochkarev, A., Pfuetzner, R.A., Edwards, A.M., and Frappier, L., Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature*, 1997. **385**(6612): p. 176-181.

27. Wadsworth, R.I.M. and White, M.F., Identification and properties of the crenarchaeal single-stranded DNA binding protein from *Sulfolobus solfataricus*. *Nucleic Acids Research*, 2001. **29**(4): p. 914-920.
28. Kernchen, U. and Lipps, G., Thermodynamic analysis of the single-stranded DNA binding activity of the archaeal replication protein A (RPA) from *Sulfolobus solfataricus*. *Biochemistry*, 2005. **45**(2): p. 594-603.
29. Kerr, I.D., Wadsworth, R.I.M., Cubeddu, L., Blankenfeldt, W., Naismith, J.H., and White, M.F., Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein. *The EMBO Journal*, 2003. **22**(11): p. 2561-2570.
30. Haseltine, C.A. and Kowalczykowski, S.C., A distinctive single-stranded DNA-binding protein from the archaeon *Sulfolobus solfataricus*. *Molecular Microbiology*, 2002. **43**(6): p. 1505-1515.
31. Han, D., Characterization of *Sso*SSB, *Sso*1450, *Sso*2001 proteins and analysis of CRISPR and cas genes from *Sulfolobus solfataricus* P2, in Department of Biochemistry. 2008, University of Bayreuth: Bayreuth, Germany. p. 141.
32. Shi, H., Zhang, Y., Zhang, G., Guo, J., Zhang, X., Song, H., Lv, J., Gao, J., Wang, Y., Chen, L., and Wang, Y., Systematic functional comparative analysis of four single-stranded DNA-binding proteins and their affection on viral RNA metabolism. *PLoS ONE*, 2013. **8**(1): p. e55076.
33. Raymann, K., Forterre, P., Brochier-Armanet, C., and Gribaldo, S., Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. *Genome Biology and Evolution*, 2014. **6**(1): p. 192-212.
34. Ghalei, H., Moeller, H.v., Eppers, D., Sohmen, D., Wilson, D.N., Loll, B., and Wahl, M.C., Entrapment of DNA in an intersubunit tunnel system of a single-stranded DNA-binding protein. *Nucleic Acids Research*, 2014. **42**(10): p. 6698-6708.
35. Nakaya, R., Takaya, J., Onuki, T., Moritani, M., Nozaki, N., and Ishimi, Y., Identification of proteins that may directly interact with human RPA. *Journal of Biochemistry*, 2010. **148**(5): p. 539-547.
36. Kozlov, A.G., Cox, M.M., and Lohman, T.M., Regulation of sngle-stranded DNA binding by the C termini of *Escherichia coli* single-stranded DNA-binding (SSB) protein. *Journal of Biological Chemistry*, 2010. **285**(22): p. 17246-17252.
37. Antony, E., Weiland, E., Yuan, Q., Manhart, C.M., Nguyen, B., Kozlov, A.G., McHenry, C.S., and Lohman, T.M., Multiple C-terminal tails within a single *E. coli* SSB homotetramer coordinate DNA replication and repair. *Journal of Molecular Biology*, 2013. **425**(23): p. 4802-4819.
38. Shereda, R.D., Reiter, N.J., Butcher, S.E., and Keck, J.L., Identification of the SSB binding site on *E. coli* RecQ reveals a conserved surface for binding SSB's C terminus. *Journal of Molecular Biology*, 2009. **386**(3): p. 612-625.
39. Naue, N., Beerbaum, M., Bogutzki, A., Schmieder, P., and Curth, U., The helicase-binding domain of *Escherichia coli* DnaG primase interacts with the highly conserved C-terminal region of single-stranded DNA-binding protein. *Nucleic Acids Research*, 2013. **41**(8): p. 4507-4517.
40. Cubeddu, L. and White, M.F., DNA damage detection by an archaeal single-stranded DNA-binding protein. *Journal of Molecular Biology*, 2005. **353**(3): p. 507-516.

41. Napoli, A., Valenti, A., Salerno, V., Nadal, M., Garnier, F., Rossi, M., and Ciaramella, M., Functional interaction of reverse gyrase with single-strand binding protein of the archaeon *Sulfolobus*. *Nucleic Acids Research*, 2005. **33**(2): p. 564-576.
42. Rolfsmeier, M.L. and Haseltine, C.A., The single-stranded DNA binding protein of *Sulfolobus solfataricus* acts in the presynaptic step of homologous recombination. *Journal of Molecular Biology*, 2010. **397**(1): p. 31-45.
43. Wang, J., Jiang, P.-X., Feng, H., Feng, Y., and He, Z.-G., Three eukaryote-like Orc1/Cdc6 proteins functionally interact and mutually regulate their activities of binding to the replication origin in the hyperthermophilic archaeon *Sulfolobus solfataricus* P2. *Biochemical and Biophysical Research Communications*, 2007. **363**(1): p. 63-70.
44. Biswas, E.E., Chen, P.-H., and Biswas, S.B., Modulation of enzymatic activities of *Escherichia coli* DnaB helicase by single-stranded DNA-binding proteins. *Nucleic Acids Research*, 2002. **30**(13): p. 2809-2816.
45. Zhang, H., Zhang, Z., Yang, J., and He, Z.-G., Functional characterization of DnaB helicase and its modulation by single-stranded DNA binding protein in *Mycobacterium tuberculosis*. *FEBS Journal*, 2014. **281**(4): p. 1256-1266.
46. Sharma, A., Nitharwal, R.G., Singh, B., Dar, A., Dasgupta, S., and Dhar, S.K., *Helicobacter pylori* single-stranded DNA binding protein – functional characterization and modulation of *H. pylori* DnaB helicase activity. *FEBS Journal*, 2009. **276**(2): p. 519-531.
47. Raghunathan, S., Kozlov, A.G., Lohman, T.M., and Waksman, G., Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nature Structural & Molecular Biology*, 2000. **7**(8): p. 648-652.
48. Bujalowski, W. and Lohman, T.M., *Escherichia coli* single-strand binding protein forms multiple, distinct complexes with single-stranded DNA. *Biochemistry*, 1986. **25**(24): p. 7799-7802.
49. Ferrari, M.E., Bujalowski, W., and Lohman, T.M., Co-operative binding of *Escherichia coli* SSB tetramers to single-stranded DNA in the $(SSB)_{35}$ binding mode. *Journal of Molecular Biology*, 1994. **236**(1): p. 106-123.
50. Roy, R., Kozlov, A.G., Lohman, T.M., and Ha, T., Dynamic structural rearrangements between DNA binding modes of *E. coli* SSB protein. *Journal of Molecular Biology*, 2007. **369**(5): p. 1244-1257.
51. Brill, S.J. and Bastin-Shanower, S., Identification and characterization of the fourth single-stranded-DNA binding domain of replication protein A. *Molecular and Cellular Biology*, 1998. **18**(12): p. 7225-7234.
52. Gamsjaeger, R., Kariawasam, R., Touma, C., Kwan, A., White, M., and Cubeddu, L., Backbone and side-chain ^{1}H, ^{13}C and ^{15}N resonance assignments of the OB domain of the single stranded DNA binding protein from *Sulfolobus solfataricus* and chemical shift mapping of the DNA-binding interface. *Biomolecular NMR Assignments*, 2013: p. 1-4.
53. Ogawa, T. and Okazaki, T., Discontinuous DNA replication. *Annual Review of Biochemistry*, 1980. **49**(1): p. 421-457.
54. Frick, D.N. and Richardson, C.C., DNA primases. *Annual Review of Biochemistry*, 2001. **70**(1): p. 39-80.
55. Lao-Sirieix, S.-H., Nookala, R.K., Roversi, P., Bell, S.D., and Pellegrini, L., Structure of the heterodimeric core primase. *Nature Structural & Molecular Biology*, 2005. **12**(12): p. 1137-1144.

56. Augustin, M.A., Huber, R., and Kaiser, J.T., Crystal structure of a DNA-dependent RNA polymerase (DNA primase). *Nature Structural & Molecular Biology*, 2001. **8**(1): p. 57-61.
57. Ito, N., Nureki, O., Shirouzu, M., Yokoyama, S., and Hanaoka, F., Crystal structure of the *Pyrococcus horikoshii* DNA primase-UTP complex: Implications for the mechanism of primer synthesis. *Genes to Cells*, 2003. **8**(12): p. 913-923.
58. Lao-Sirieix, S.-H. and Bell, S.D., The heterodimeric primase of the hyperthermophilic archaeon *Sulfolobus solfataricus* possesses DNA and RNA primase, polymerase and 3′-terminal nucleotidyl transferase activities. *Journal of Molecular Biology*, 2004. **344**(5): p. 1251-1263.
59. Zerbe, L.K. and Kuchta, R.D., The p58 subunit of human DNA primase is important for primer initiation, elongation, and counting. *Biochemistry*, 2002. **41**(15): p. 4891-4900.
60. Davey, S.K. and Faust, E.A., Murine DNA polymerase.alpha-primase initiates RNA-primed DNA synthesis preferentially upstream of a 3'-CC(C/A)-5' motif. *Journal of Biological Chemistry*, 1990. **265**(7): p. 3611-4.
61. Pellegrini, L., The pol α-primase complex, in The Eukaryotic Replisome: A Guide to Protein Structure and Function, S. MacNeill, Editor. 2012, Springer Netherlands. p. 157-169.
62. Stillman, B., DNA polymerases at the replication fork in eukaryotes. *Molecular Cell*, 2008. **30**(3): p. 259-260.
63. Le Breton, M., Henneke, G., Norais, C., Flament, D., Myllykallio, H., Querellou, J., and Raffin, J.-P., The heterodimeric primase from the euryarchaeon *Pyrococcus abyssi*: A multifunctional enzyme for initiation and repair? *Journal of Molecular Biology*, 2007. **374**(5): p. 1172-1185.
64. Ito, N., Matsui, I., and Matsui, E., Molecular basis for the subunit assembly of the primase from an archaeon *Pyrococcus horikoshii*. *FEBS Journal*, 2007. **274**(5): p. 1340-1351.
65. Chemnitz Galal, W., Pan, M., Kelman, Z., and Hurwitz, J., Characterization of DNA primase complex isolated from the archaeon, *Thermococcus kodakaraensis*. *Journal of Biological Chemistry*, 2012. **287**(20): p. 16209-16219.
66. Hu, J., Guo, L., Wu, K., Liu, B., Lang, S., and Huang, L., Template-dependent polymerization across discontinuous templates by the heterodimeric primase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Nucleic Acids Research*, 2012. **40**(8): p. 3470-3483.
67. Marinsek, N., Barry, E.R., Makarova, K.S., Dionne, I., Koonin, E.V., and Bell, S.D., GINS, a central nexus in the archaeal DNA replication fork. *EMBO Reports*, 2006. **7**(5): p. 539-545.
68. Beattie, T.R. and Bell, S.D., Molecular machines in archaeal DNA replication. *Current Opinion in Chemical Biology*, 2011. **15**(5): p. 614-619.
69. Kelman, Z. and White, M.F., Archaeal DNA replication and repair. *Current Opinion in Microbiology*, 2005. **8**(6): p. 669-676.
70. Scherzinger, E., Lanka, E., Morelli, G., Seiffert, D., and Yuki, A., Bacteriophage-T7-induced DNA-priming protein. *European Journal of Biochemistry*, 1977. **72**(3): p. 543-558.
71. Pan, H. and Wigley, D.B., Structure of the zinc-binding domain of *Bacillus stearothermophilus* DNA primase. *Structure*, 2000. **8**(3): p. 231-239.
72. Keck, J.L., Roche, D.D., Lynch, A.S., and Berger, J.M., Structure of the RNA polymerase domain of *E. coli* primase. *Science*, 2000. **287**(5462): p. 2482-2486.

73. Aravind, L., Leipe, D.D., and Koonin, E.V., Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Research*, 1998. **26**(18): p. 4205-4213.
74. Rymer, R.U., Solorio, F.A., Tehranchi, A.K., Chu, C., Corn, J.E., Keck, J.L., Wang, J.D., and Berger, J.M., Binding mechanism of metal·NTP substrates and stringent-response alarmones to bacterial DnaG-type primases. *Structure*, 2012. **20**(9): p. 1478-1489.
75. Corn, J.E., Pelton, J.G., and Berger, J.M., Identification of a DNA primase template tracking site redefines the geometry of primer synthesis. *Nature Structural & Molecular Biology*, 2008. **15**(2): p. 163-169.
76. Valentine, A.M., Ishmael, F.T., Shier, V.K., and Benkovic, S.J., A zinc ribbon protein in DNA replication: primer synthesis and macromolecular interactions by the bacteriophage T4 primase. *Biochemistry*, 2001. **40**(50): p. 15074-15085.
77. Johnson, S.K., Bhattacharyya, S., and Griep, M.A., DnaB helicase stimulates primer synthesis activity on short oligonucleotide templates. *Biochemistry*, 2000. **39**(4): p. 736-744.
78. Corn, J.E., Pease, P.J., Hura, G.L., and Berger, J.M., Crosstalk between primase subunits can act to regulate primer synthesis in *trans*. *Molecular Cell*, 2005. **20**(3): p. 391-401.
79. Bird, L.E., Pan, H., Soultanas, P., and Wigley, D.B., Mapping protein−protein interactions within a stable complex of DNA primase and DnaB helicase from *Bacillus stearothermophilus*. *Biochemistry*, 2000. **39**(1): p. 171-182.
80. Lu, Y.-B., Ratnakar, P.V.A.L., Mohanty, B.K., and Bastia, D., Direct physical interaction between DnaG primase and DnaB helicase of *Escherichia coli* is necessary for optimal synthesis of primer RNA. *Proceedings of the National Academy of Sciences*, 1996. **93**(23): p. 12902-12907.
81. VanLoock, M.S., Chen, Y.-J., Yu, X., Patel, S.S., and Egelman, E.H., The primase active site is on the outside of the hexameric bacteriophage T7 gene 4 helicase-primase ring. *Journal of Molecular Biology*, 2001. **311**(5): p. 951-956.
82. Huang, H., Zhao, K., Arnett, D.R., and Fanning, E., A specific docking site for DNA polymerase α-primase on the SV40 helicase is required for viral primosome activity, but helicase activity is dispensable. *Journal of Biological Chemistry*, 2010. **285**(43): p. 33475-33484.
83. Wang, G., Klein, M.G., Tokonzaba, E., Zhang, Y., Holden, L.G., and Chen, X.S., The structure of a DnaB-family replicative helicase and its interactions with primase. *Nature Structural & Molecular Biology*, 2008. **15**(1): p. 94-100.
84. Bailey, S., Eliason, W.K., and Steitz, T.A., Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science*, 2007. **318**(5849): p. 459-463.
85. Zechner, E.L., Wu, C.A., and Marians, K.J., Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. III. A polymerase-primase interaction governs primer size. *Journal of Biological Chemistry*, 1992. **267**(6): p. 4054-4063.
86. Arai, K. and Kornberg, A., Mechanism of dnaB protein action. IV. General priming of DNA replication by dnaB protein and primase compared with RNA polymerase. *Journal of Biological Chemistry*, 1981. **256**(10): p. 5267-5272.
87. Kuchta, R.D., Reid, B., and Chang, L.M., DNA primase. Processivity and the primase to polymerase alpha activity switch. *Journal of Biological Chemistry*, 1990. **265**(27): p. 16158-65.

88. Zechner, E.L., Wu, C.A., and Marians, K.J., Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. II. Frequency of primer synthesis and efficiency of primer utilization control Okazaki fragment size. *Journal of Biological Chemistry*, 1992. **267**(6): p. 4045-4053.
89. Lee, J.-B., Hite, R.K., Hamdan, S.M., Sunney Xie, X., Richardson, C.C., and van Oijen, A.M., DNA primase acts as a molecular brake in DNA replication. *Nature*, 2006. **439**(7076): p. 621-624.
90. Kato, M., Frick, D.N., Lee, J., Tabor, S., Richardson, C.C., and Ellenberger, T., A complex of the bacteriophage T7 primase-helicase and DNA polymerase directs primer utilization. *Journal of Biological Chemistry*, 2001. **276**(24): p. 21809-21820.
91. Podobnik, M., McInerney, P., O'Donnell, M., and Kuriyan, J., A TOPRIM domain in the crystal structure of the catalytic core of *Escherichia coli* primase confirms a structural link to DNA topoisomerases. *Journal of Molecular Biology*, 2000. **300**(2): p. 353-362.
92. Kusakabe, T., Baradaran, K., Lee, J., and Richardson, C.C., Roles of the helicase and primase domain of the gene 4 protein of bacteriophage T7 in accessing the primase recognition site. *The EMBO Journal*, 1998. **17**(5): p. 1542-1552.
93. Tougu, K. and Marians, K.J., The interaction between helicase and primase sets the replication fork clock. *Journal of Biological Chemistry*, 1996. **271**(35): p. 21398-21405.
94. Dracheva, S., Koonin, E.V., and Crute, J.J., Identification of the primase active site of the herpes simplex virus type 1 helicase-primase. *Journal of Biological Chemistry*, 1995. **270**(23): p. 14148-14153.
95. Tougu, K., Peng, H., and Marians, K.J., Identification of a domain of *Escherichia coli* primase required for functional interaction with the DnaB helicase at the replication fork. *Journal of Biological Chemistry*, 1994. **269**(6): p. 4675-82.
96. Zuo, Z., Rodgers, C.J., Mikheikin, A.L., and Trakselis, M.A., Characterization of a functional DnaG-type primase in archaea: implications for a dual-primase system. *Journal of Molecular Biology*, 2010. **397**(3): p. 664-676.
97. Bauer, R.J., Graham, B.W., and Trakselis, M.A., Novel interaction of the bacterial-like DnaG primase with the MCM helicase in Archaea. *Journal of Molecular Biology*, 2013. **425**(8): p. 1259-1273.
98. Walter, P., Klein, F., Lorentzen, E., Ilchmann, A., Klug, G., and Evguenieva-Hackenberg, E., Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Molecular Microbiology*, 2006. **62**(4): p. 1076-1089.
99. Hou, L., Klug, G., and Evguenieva-Hackenberg, E., The archaeal DnaG protein needs Csl4 for binding to the exosome and enhances its interaction with adenine-rich RNAs. *RNA Biology*, 2013. **10**(3): p. 415-424.
100. Labib, K., Tercero, J.A., and Diffley, J.F.X., Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science*, 2000. **288**(5471): p. 1643-1647.
101. Abdel-Monem, M. and Hoffmann-Berling, H., Enzymic Unwinding of DNA. *European Journal of Biochemistry*, 1976. **65**(2): p. 431-440.
102. Maine, G.T., Sinha, P., and Tye, B.-K., Mutants of *S. cerevisiae* defective in the maintenance of minichromosomes. *Genetics*, 1984. **106**(3): p. 365-385.
103. Barry, E.R., McGeoch, A.T., Kelman, Z., and Bell, S.D., Archaeal MCM has separable processivity, substrate choice and helicase domains. *Nucleic Acids Research*, 2007. **35**(3): p. 988-998.

104. Singleton, M.R., Dillingham, M.S., and Wigley, D.B., Structure and mechanism of helicases and nucleic acid translocases. *Annual Review of Biochemistry*, 2007. **76**(1): p. 23-50.
105. Ye, J., Osborne, A.R., Groll, M., and Rapoport, T.A., RecA-like motor ATPases—lessons from structures. *Biochimica et Biophysica Acta*, 2004. **1659**(1): p. 1-18.
106. Walker, J.E., Saraste, M., Runswick, M., J., and Gay, N.J., Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO Journal*, 1982. **1**(8): p. 945-951.
107. Scheffzek, K., Ahmadian, M.R., Kabsch, W., Wiesmüller, L., Lautwein, A., Schmitz, F., and Wittinghofer, A., The Ras-RasGAP complex: Structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science*, 1997. **277**(5324): p. 333-339.
108. Gorbalenya, A.E. and Koonin, E.V., Helicases: Amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology*, 1993. **3**(3): p. 419-429.
109. Hickman, A.B. and Dyda, F., Binding and unwinding: SF3 viral helicases. *Current Opinion in Structural Biology*, 2005. **15**(1): p. 77-85.
110. Rothenberg, E., Trakselis, M.A., Bell, S.D., and Ha, T., MCM forked substrate specificity involves dynamic interaction with the 5'-tail. *Journal of Biological Chemistry*, 2007. **282**(47): p. 34229-34234.
111. McGeoch, A.T., Trakselis, M.A., Laskey, R.A., and Bell, S.D., Organization of the archaeal MCM complex on DNA and implication for the helicase mechanism. *Nature Structural & Molecular Biology*, 2005. **12**(9): p. 756-762.
112. Moreau, M.J., McGeoch, A.T., Lowe, A.R., Itzhaki, L.S., and Bell, S.D., ATPase site architecture and helicase mechanism of an archaeal MCM. *Molecular Cell*, 2007. **28**(2): p. 304-314.
113. Brewster, A.S., Wang, G., Yu, X., Greenleaf, W.B., Carazo, J.M., Tjajadi, M., Klein, M.G., and Chen, X.S., Crystal structure of a near-full-length archaeal MCM: Functional insights for an AAA+ hexameric helicase. *Proceedings of the National Academy of Sciences*, 2008. **105**(51): p. 20191-20196.
114. Pucci, B., De Felice, M., Rocco, M., Esposito, F., De Falco, M., Esposito, L., Rossi, M., and Pisani, F.M., Modular organization of the *Sulfolobus solfataricus* mini-chromosome maintenance protein. *Journal of Biological Chemistry*, 2007. **282**(17): p. 12574-12582.
115. Pape, T., Meka, H., Chen, S., Vicentini, G., van Heel, M., and Onesti, S., Hexameric ring structure of the full-length archaeal MCM protein complex. *EMBO Reports*, 2003. **4**(11): p. 1079-1083.
116. Liu, W., Pucci, B., Rossi, M., Pisani, F.M., and Ladenstein, R., Structural analysis of the *Sulfolobus solfataricus* MCM protein N-terminal domain. *Nucleic Acids Research*, 2008. **36**(10): p. 3235-3243.
117. Poplawski, A., Grabowski, B., Long, S.E., and Kelman, Z., The zinc finger domain of the archaeal minichromosome maintenance protein is required for helicase activity. *Journal of Biological Chemistry*, 2001. **276**(52): p. 49371-49377.
118. Barry, E.R., Lovett, J.E., Costa, A., Lea, S.B., and Bell, S.D., Intersubunit allosteric communication mediated by a conserved loop in the MCM helicase. *Proceedings of the National Academy of Sciences*, 2009. **106**(4): p. 1051-1056.

119. Chong, J.P.J., Learning to unwind. *Nature Structural & Molecular Biology*, 2005. **12**(9): p. 734-736.
120. Patel, S.S. and Picha, K.M., Structure and function of hexameric helicases. *Annual Review of Biochemistry*, 2000. **69**(1): p. 651-697.
121. Fletcher, R.J., Bishop, B.E., Leon, R.P., Sclafani, R.A., Ogata, C.M., and Chen, X.S., The structure and function of MCM from archaeal *M. thermoautotrophicum*. *Nature Structural & Molecular Biology*, 2003. **10**(3): p. 160-167.
122. Takahashi, T.S., Wigley, D.B., and Walter, J.C., Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. *Trends in Biochemical Sciences*, 2005. **30**(8): p. 437-444.
123. Graham, B.W., Schauer, G.D., Leuba, S.H., and Trakselis, M.A., Steric exclusion and wrapping of the excluded DNA strand occurs along discrete external binding paths during MCM helicase unwinding. *Nucleic Acids Research*, 2011. **39**(15): p. 6585-6595.
124. Bochman, M.L. and Schwacha, A., The Mcm2-7 complex has *in vitro* helicase activity. *Molecular Cell*, 2008. **31**(2): p. 287-293.
125. Allen, G.C. and Kornberg, A., Assembly of the primosome of DNA replication in *Escherichia coli*. *Journal of Biological Chemistry*, 1993. **268**(26): p. 19204-19209.
126. Bell, S.P. and Dutta, A., DNA replication in eukaryotic cells. *Annual Review of Biochemistry*, 2002. **71**(1): p. 333-374.
127. De Felice, M., Esposito, L., Pucci, B., Carpentieri, F., De Falco, M., Rossi, M., and Pisani, F.M., Biochemical characterization of a CDC6-like protein from the crenarchaeon *Sulfolobus solfataricus*. *Journal of Biological Chemistry*, 2003. **278**(47): p. 46424-46431.
128. De Felice, M., Esposito, L., Pucci, B., De Falco, M., Manco, G., Rossi, M., and Pisani, F.M., Modular organization of a Cdc6-like protein from the crenarchaeon *Sulfolobus solfataricus*. *Biochemical Journal*, 2004. **381**(3): p. 645-653.
129. De Felice, M., Esposito, L., Pucci, B., De Falco, M., Rossi, M., and Pisani, F.M., A CDC6-like factor from the archaea *Sulfolobus solfataricus* promotes binding of the mini-chromosome maintenance complex to DNA. *Journal of Biological Chemistry*, 2004. **279**(41): p. 43008-43012.
130. De Felice, M., Esposito, L., Rossi, M., and Pisani, F.M., Biochemical characterization of two Cdc6/Orc1-like proteins from the crenarchaeon *Sulfolobus solfataricus*. *Extremophiles*, 2006. **10**: p. 61-70.
131. Robinson, N.P., Dionne, I., Lundgren, M., Marsh, V.L., Bernander, R., and Bell, S.D., Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell*, 2004. **116**(1): p. 25-38.
132. Lundgren, M., Andersson, A., Chen, L., Nilsson, P., and Bernander, R., Three replication origins in *Sulfolobus* species: Synchronous initiation of chromosome replication and asynchronous termination. *Proceedings of the National Academy of Sciences of the United States of America*, 2004. **101**(18): p. 7046-7051.
133. Forsburg, S.L., Eukaryotic MCM proteins: Beyond replication initiation. *Microbiology and Molecular Biology Reviews*, 2004. **68**(1): p. 109-131.
134. Madine, M.A., Khoo, C.-Y., Mills, A.D., Musahl, C., and Laskey, R.A., The nuclear envelope prevents reinitiation of replication by regulating the binding of MCM3 to chromatin in *Xenopus* egg extracts. *Current Biology*, 1995. **5**(11): p. 1270-1279.

135. Ilves, I., Petojevic, T., Pesavento, J.J., and Botchan, M.R., Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Molecular Cell*, 2010. **37**(2): p. 247-258.
136. Iyer, L.M., Leipe, D.D., Koonin, E.V., and Aravind, L., Evolutionary history and higher order classification of AAA+ ATPases. *Journal of Structural Biology*, 2004. **146**(1-2): p. 11-31.
137. Duderstadt, K.E. and Berger, J.M., AAA+ ATPases in the initiation of DNA replication. *Critical Reviews in Biochemistry and Molecular Biology*, 2008. **43**(3): p. 163-187.
138. Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E.V., AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Research*, 1999. **9**(1): p. 27-43.
139. Costa, A., Pape, T., van Heel, M., Brick, P., Patwardhan, A., and Onesti, S., Structural basis of the *Methanothermobacter thermautotrophicus* MCM helicase activity. *Nucleic Acids Research*, 2006. **34**(20): p. 5829-5838.
140. Sakakibara, N., Kasiviswanathan, R., Melamud, E., Han, M., Schwarz, F.P., and Kelman, Z., Coupling of DNA binding and helicase activity is mediated by a conserved loop in the MCM protein. *Nucleic Acids Research*, 2008. **36**(4): p. 1309-1320.
141. Li, D., Zhao, R., Lilyestrom, W., Gai, D., Zhang, R., DeCaprio, J.A., Fanning, E., Jochimiak, A., Szakonyi, G., and Chen, X.S., Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature*, 2003. **423**(6939): p. 512 - 518.
142. Gomez-Llorente, Y., Fletcher, R.J., Chen, X.S., Carazo, J.M., and Martin, C.S., Polymorphism and double hexamer structure in the archaeal minichromosome maintenance (MCM) helicase from *Methanobacterium thermoautotrophicum*. *Journal of Biological Chemistry*, 2005. **280**(49): p. 40909-40915.
143. Shin, J.-H., Heo, G.-Y., and Kelman, Z., The *Methanothermobacter thermautotrophicus* MCM helicase is active as a hexameric ring. *Journal of Biological Chemistry*, 2009. **284**(1): p. 540-546.
144. Sambrook, J. and Russell, D.W., Molecular Cloning: A Laboratory Manual. 3rd ed. 2001, Cold Spring Harbor, NY: Cold Spring Harbor Press.
145. Morris, P.D. and Raney, K.D., DNA helicases displace streptavidin from biotin-labeled oligonucleotides. *Biochemistry*, 1999. **38**(16): p. 5164-5171.
146. Mikheikin, A.L., Lin, H.-K., Mehta, P., Jen-Jacobson, L., and Trakselis, M.A., A trimeric DNA polymerase complex increases the native replication processivity. *Nucleic Acids Research*, 2009. **37**(21): p. 7194-7205.
147. Roy, R., Hohng, S., and Ha, T., A practical guide to single-molecule FRET. *Nature Methods*, 2008. **5**(6): p. 507-516.
148. Axelrod, D., Chapter 7 total internal reflection fluorescence microscopy, in Methods in Cell Biology, J.J. Correia and H.W. Detrich III, Editors. 2008, Academic Press: Burlington, MA. p. 169-221.
149. Fagerburg, M.V. and Leuba, S.H., Optimal practices for surface-tethered single molecule total internal reflection fluorescenced resonance energy transfer analysis. 1st ed. Methods in Molecular Biology, ed. G. Zuccheri and B. Samori. Vol. 749. 2011, New York, NY: Humana Press. 320.
150. Li, K., The image stabilizer plugin for Image J. 2008.
151. Rasband, W.S., ImageJ. 1997-2009, U. S. National Institutes of Health: Bethesda, MD.

152. Kaplan, D.L., Davey, M.J., and O'Donnell, M., Mcm4,6,7 uses a "pump in ring" mechanism to unwind DNA by steric exclusion and actively translocate along a duplex. *Journal of Biological Chemistry*, 2003. **278**(49): p. 49171-49182.
153. Brewster, A.S., Slaymaker, I.M., Afif, S.A., and Chen, X.S., Mutational analysis of an archaeal minichromosome maintenance protein exterior hairpin reveals critical residues for helicase activity and DNA binding. *BMC Molecular Biology*, 2010. **11**(1): p. 62.
154. Laskey, R.A. and Madine, M.A., A rotary pumping model for helicase function of MCM proteins at a distance from replication forks. *EMBO Reports*, 2003. **4**(1): p. 26-30.
155. Kaplan, D.L. and O'Donnell, M., Twin DNA pumps of a hexameric helicase provide power to simultaneously melt two duplexes. *Molecular Cell*, 2004. **15**(3): p. 453-465.
156. Gai, D., Zhao, R., Li, D., Finkielstein, C.V., and Chen, X.S., Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell*, 2004. **119**(1): p. 47 - 60.
157. Wessel, R., Schweizer, J., and Stahl, H., Simian virus 40 T-antigen DNA helicase is a hexamer which forms a binary complex during bidirectional unwinding from the viral origin of DNA replication. *Journal of Virology*, 1992. **66**(2): p. 804-815.
158. Brewster, A.S. and Chen, X.S., Insights into the MCM functional mechanism: Lessons learned from the archaeal MCM complex. *Critical Reviews in Biochemistry and Molecular Biology*, 2010. **45**(3): p. 243-256.
159. Matlock, D.L., Yeruva, L., Byrd, A.K., Mackintosh, S.G., Langston, C., Brown, C., Cameron, C.E., Fischer, C.J., and Raney, K.D., Investigation of translocation, DNA unwinding, and protein displacement by NS3h, the helicase domain from the hepatitis C virus helicase. *Biochemistry*, 2010. **49**(10): p. 2097-2109.
160. Byrd, A.K. and Raney, K.D., Displacement of a DNA binding protein by Dda helicase. *Nucleic Acids Research*, 2006. **34**(10): p. 3020-3029.
161. Enemark, E.J. and Joshua-Tor, L., Mechanism of DNA translocation in a replicative hexameric helicase. *Nature*, 2006. **442**(7100): p. 270-275.
162. Costa, A., van Duinen, G., Medagli, B., Chong, J., Sakakibara, N., Kelman, Z., Nair, S.K., Patwardhan, A., and Onesti, S., Cryo-electron microscopy reveals a novel DNA-binding site on the MCM helicase. *The EMBO Journal*, 2008. **27**(16): p. 2250-2258.
163. Bauer, R.J., Wolff, I.D., Zuo, X., Lin, H.-K., and Trakselis, M.A., Assembly and distributive action of an archaeal DNA polymerase holoenzyme. *Journal of Molecular Biology*, 2013. **425**(23): p. 4820-4836.
164. Bauer, R.J., Begley, M.T., and Trakselis, M.A., Kinetics and fidelity of polymerization by DNA polymerase III from *Sulfolobus solfataricus*. *Biochemistry*, 2012. **51**(9): p. 1996-2007.
165. Zhang, Q., Noble, K., Mao, Y., Young, N., Sathe, S., Roux, K., and Marshall, A., Rapid screening for potential epitopes reactive with a polycolonal antibody by solution-phase H/D exchange monitored by FT-ICR mass spectrometry. *Journal of The American Society for Mass Spectrometry*, 2013. **24**(7): p. 1016-1025.
166. Blakney, G.T., Hendrickson, C.L., and Marshall, A.G., Predator data station: A fast data acquisition system for advanced FT-ICR MS experiments. *International Journal of Mass Spectrometry*, 2011. **306**(2–3): p. 246-252.
167. Zhang, Q., Willison, L.N., Tripathi, P., Sathe, S.K., Roux, K.H., Emmett, M.R., Blakney, G.T., Zhang, H.-M., and Marshall, A.G., Epitope mapping of a 95 kDa antigen in complex with antibody by solution-phase amide backbone hydrogen/deuterium exchange monitored

by Fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry*, 2011. **83**(18): p. 7129-7136.
168. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G., Clustal W and Clustal X version 2.0. *Bioinformatics*, 2007. **23**(21): p. 2947-2948.
169. Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R., A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Research*, 2010. **38**(suppl 2): p. W695-W699.
170. Chong, J.P.J., Hayashi, M.K., Simon, M.N., Xu, R.-M., and Stillman, B., A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. *Proceedings of the National Academy of Sciences*, 2000. **97**(4): p. 1530 - 1535.
171. Liew, L.P. and Bell, S.D., The interplay of DNA binding, ATP hydrolysis and helicase activities of the archaeal MCM helicase. *Biochemical Journal*, 2011. **436**(2): p. 409-414.
172. Pace, C.N. and Scholtz, J.M., A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical Journal*, 1998. **75**(1): p. 422-427.
173. Hennessey, J.P., Johnson, W.C., Bahler, C., and Wood, H.G., Subunit interactions of transcarboxylase as studied by circular dichroism. *Biochemistry*, 1982. **21**(4): p. 642-646.
174. Johnson, W.C., Protein secondary structure and circular dichroism: A practical guide. *Proteins: Structure, Function, and Bioinformatics*, 1990. **7**(3): p. 205-214.
175. Lisal, J., Lam, T.T., Kainov, D.E., Emmett, M.R., Marshall, A.G., and Tuma, R., Functional visualization of viral molecular motor by hydrogen-deuterium exchange reveals transient states. *Nature Structural & Molecular Biology*, 2005. **12**(5): p. 460-466.
176. Lísal, J., Kainov, D.E., Lam, T.T., Emmett, M.R., Wei, H., Gottlieb, P., Marshall, A.G., and Tuma, R., Interaction of packaging motor with the polymerase complex of dsRNA bacteriophage. *Virology*, 2006. **351**(1): p. 73-79.
177. Houde, D., Berkowitz, S.A., and Engen, J.R., The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies. *Journal of Pharmaceutical Sciences*, 2011. **100**(6): p. 2071-2086.
178. Schroeder, L.A., Gries, T.J., Saecker, R.M., Record Jr., M.T., Harris, M.E., and deHaseth, P.L., Evidence for a tyrosine–adenine stacking interaction and for a short-lived open intermediate subsequent to initial binding of *Escherichia coli* RNA polymerase to promoter DNA. *Journal of Molecular Biology*, 2009. **385**(2): p. 339-349.
179. Froelich, C.A., Kang, S., Epling, L.B., Bell, S.P., and Enemark, E.J., A conserved MCM single-stranded DNA binding element is essential for replication initiation. eLife, ed. M.R. Botchan. Vol. 3. 2014.
180. Sakakibara, N., Kasiviswanathan, R., and Kelman, Z., Different residues on the surface of the *Methanothermobacter thermautotrophicus* MCM helicase interact with single- and double-stranded DNA. *Archaea*, 2010. **2010**(505693): p. 8.
181. Iyer, L.M., Koonin, E.V., Leipe, D.D., and Aravind, L., Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: Structural insights and new members. *Nucleic Acids Research*, 2005. **33**(12): p. 3875-3896.
182. Kuchta, R.D. and Stengel, G., Mechanism and evolution of DNA primases. *Biochimica et Biophysica Acta*, 2010. **1804**(5): p. 1180-1189.

183. Sheaff, R.J. and Kuchta, R.D., Mechanism of calf thymus DNA primase: Slow initiation, rapid polymerization, and intelligent termination. *Biochemistry*, 1993. **32**(12): p. 3027-3037.
184. Copeland, W.C. and Wang, T.S.-F., Enzymatic characterization of the individual mammalian primase subunits reveals a biphasic mechanism for initiation of DNA replication. *Journal of Biological Chemistry*, 1993. **268**(35): p. 26179-26189.
185. Maga, G., Stucki, M., Spadari, S., and Hübscher, U., DNA polymerase switching: I. Replication factor C displaces DNA polymerase α prior to PCNA loading. *Journal of Molecular Biology*, 2000. **295**(4): p. 791-801.
186. Mossi, R., Keller, R.C., Ferrari, E., and Hübscher, U., DNA polymerase switching: II. Replication factor C abrogates primer synthesis by DNA polymerase α at a critical length. *Journal of Molecular Biology*, 2000. **295**(4): p. 803-814.
187. Studier, F.W., Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 2005. **41**(1): p. 207-234.
188. Yuzhakov, A., Turner, J., and O'Donnell, M., Replisome assembly reveals the basis for asymmetric function in leading and lagging strand replication. *Cell*, 1996. **86**(6): p. 877-886.
189. Kaplan, D.L. and Steitz, T.A., DnaB from *Thermus aquaticus* unwinds forked duplex DNA with an asymmetric tail length dependence. *Journal of Biological Chemistry*, 1999. **274**(11): p. 6889-6897.
190. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T., The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 2006. **22**(2): p. 195-201.
191. Wu, K., Lai, X., Guo, X., Hu, J., Xiang, X., and Huang, L., Interplay between primase and replication factor C in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Molecular Microbiology*, 2007. **63**(3): p. 826-837.
192. Godson, G.N., Schoenich, J., Sun, W., and Mustaev, A.A., Identification of the magnesium ion binding site in the catalytic center of *Escherichia coli* primase by iron cleavage. *Biochemistry*, 2000. **39**(2): p. 332-339.
193. Rodina, A. and Godson, G.N., Role of conserved amino acids in the catalytic activity of *Escherichia coli* primase. *Journal of Bacteriology*, 2006. **188**(10): p. 3614-3621.
194. Chang, P. and Marians, K.J., Identification of a region of *Escherichia coli* DnaB required for functional interaction with DnaG at the replication fork. *Journal of Biological Chemistry*, 2000. **275**(34): p. 26187-26195.
195. Chintakayala, K., Larson, M.A., Griep, M.A., Hinrichs, S.H., and Soultanas, P., Conserved residues of the C-terminal p16 domain of primase are involved in modulating the activity of the bacterial primosome. *Molecular Microbiology*, 2008. **68**(2): p. 360-371.
196. De Falco, M., Ferrari, E., De Felice, M., Rossi, M., Hubscher, U., and Pisani, F.M., The human GINS complex binds to and specifically stimulates human DNA polymerase α-primase. *EMBO Reports*, 2007. **8**(1): p. 99-103.
197. Bermudez, V.P., Farina, A., Raghavan, V., Tappin, I., and Hurwitz, J., Studies on human DNA polymerase ϵ and GINS complex and their role in DNA replication. *Journal of Biological Chemistry*, 2011. **286**(33): p. 28963-28977.
198. McGeoch, A.T. and Bell, S.D., Eukaryotic/archaeal primase and MCM proteins encoded in a bacteriophage genome. *Cell*, 2005. **120**(2): p. 167-168.

199. Samuels, M., Gulati, G., Shin, J.-H., Opara, R., McSweeney, E., Sekedat, M., Long, S., Kelman, Z., and Jeruzalmi, D., A biochemically active MCM-like helicase in *Bacillus cereus*. *Nucleic Acids Research*, 2009: p. gkp376.
200. Zhou, B., Arnett, D.R., Yu, X., Brewster, A., Sowd, G.A., Xie, C.L., Vila, S., Gai, D., Fanning, E., and Chen, X.S., Structural basis for the interaction of a hexameric replicative helicase with the regulatory subunit of human DNA polymerase α-primase. *Journal of Biological Chemistry*, 2012. **287**(32): p. 26854-26866.
201. Kirk, B.W. and Kuchta, R.D., Arg304 of human DNA primase is a key contributor to catalysis and NTP binding: Primase and the family X polymerases share significant sequence homology. *Biochemistry*, 1999. **38**(24): p. 7727-7736.
202. Li, Z., Santangelo, T.J., Čuboňová, Ľ., Reeve, J.N., and Kelman, Z., Affinity purification of an archaeal DNA replication protein network. *mBio*, 2010. **1**(5).
203. Gotz, D., Paytubi, S., Munro, S., Lundgren, M., Bernander, R., and White, M.F., Responses of hyperthermophilic crenarchaea to UV irradiation. *Genome Biology*, 2007. **8**(10): p. R220.
204. Lao-Sirieix, S.-H., Pellegrini, L., and Bell, S.D., The promiscuous primase. *Trends in Genetics*, 2005. **21**(10): p. 568-572.
205. Evguenieva-Hackenberg, E., Walter, P., Hochleitner, E., Lottspeich, F., and Klug, G., An exosome-like complex in *Sulfolobus solfataricus*. *EMBO Reports*, 2003. **4**(9): p. 889-893.
206. Witharana, C., Roppelt, V., Lochnit, G., Klug, G., and Evguenieva-Hackenberg, E., Heterogeneous complexes of the RNA exosome in *Sulfolobus solfataricus*. *Biochimie*, 2012. **94**(7): p. 1578-1587.
207. Evguenieva-Hackenberg, E., Roppelt, V., Lassek, C., and Klug, G., Subcellular localization of RNA degrading proteins and protein complexes in prokaryotes. *RNA Biology*, 2011. **8**(1): p. 49-54.
208. Roppelt, V., Hobel, C.F.V., Albers, S.V., Lassek, C., Schwarz, H., Klug, G., and Evguenieva-Hackenberg, E., The archaeal exosome localizes to the membrane. *FEBS Letters*, 2010. **584**(13): p. 2791-2795.
209. Gristwood, T., Duggin, I.G., Wagner, M., Albers, S.V., and Bell, S.D., The sub-cellular localization of *Sulfolobus* DNA replication. *Nucleic Acids Research*, 2012. **40**(12): p. 5487-5496.
210. Bernstein, J.A. and Richardson, C.C., Characterization of the helicase and primase activities of the 63-kDa component of the bacteriophage T7 gene 4 protein. *Journal of Biological Chemistry*, 1989. **264**(22): p. 13066-13073.
211. Bhattacharyya, S. and Griep, M.A., DnaB helicase affects the initiation specifity of *Escherichia coli* primase on single-stranded DNA templates. *Biochemistry*, 2000. **39**(4): p. 745-752.
212. Mitkova, A.V., Khopde, S.M., and Biswas, S.B., Mechanism and stoichiometry of interaction of DnaG primase with DnaB helicase of *Escherichia coli* in RNA primer synthesis. *Journal of Biological Chemistry*, 2003. **278**(52): p. 52253-52261.
213. Norcum, M.T., Warrington, J.A., Spiering, M.M., Ishmael, F.T., Trakselis, M.A., and Benkovic, S.J., Architecture of the bacteriophage T4 primosome: Electron microscopy studies of helicase (gp41) and primase (gp61). *Proceedings of the National Academy of Sciences of the United States of America*, 2005. **102**(10): p. 3623-3626.

214. Jose, D., Weitzel, S.E., Jing, D., and von Hippel, P.H., Assembly and subunit stoichiometry of the functional helicase-primase (primosome) complex of bacteriophage T4. *Proceedings of the National Academy of Sciences*, 2012. **109**(34): p. 13596-13601.
215. Pandey, M., Syed, S., Donmez, I., Patel, G., Ha, T., and Patel, S.S., Coordinating DNA replication by means of priming loop and differential synthesis rate. *Nature*, 2009. **462**(7275): p. 940-943.
216. Ha, T., Kozlov, A.G., and Lohman, T.M., Single-molecule views of protein movement on single-stranded DNA. *Annual Review of Biophysics*, 2012. **41**(1): p. 295-319.
217. Raghunathan, S., Ricard, C.S., Lohman, T.M., and Waksman, G., Crystal structure of the homo-tetrameric DNA binding domain of *Escherichia coli* single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-Å resolution. *Proceedings of the National Academy of Sciences*, 1997. **94**(13): p. 6652-6657.
218. Savvides, S.N., Raghunathan, S., Fütterer, K., Kozlov, A.G., Lohman, T.M., and Waksman, G., The C-terminal domain of full-length *E. coli* SSB is disordered even when bound to DNA. *Protein Science*, 2004. **13**(7): p. 1942-1947.
219. Richard, D.J., Bell, S.D., and White, M.F., Physical and functional interaction of the archaeal single-stranded DNA-binding protein SSB with RNA polymerase. *Nucleic Acids Research*, 2004. **32**(3): p. 1065-1074.
220. Lohman, T.M., Green, J.M., and Beyer, R.S., Large-scale overproduction and rapid purification of the *Escherichia coli* ssb gene product. Expression of the ssb gene under λ P_L control. *Biochemistry*, 1986. **25**(1): p. 21-25.
221. Marsh, V.L., McGeoch, A.T., and Bell, S.D., Influence of chromatin and single strand binding proteins on the activity of an archaeal MCM. *Journal of Molecular Biology*, 2006. **357**(5): p. 1345-1350.
222. Kelman, Z., Lee, J.K., and Hurwitz, J., The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* ΔH contains DNA helicase activity. *Proceedings of the National Academy of Sciences*, 1999. **96**(26): p. 14783 - 14788.
223. Lohman, T.M. and Overman, L.B., Two binding modes in *Escherichia coli* single strand binding protein-single stranded DNA complexes. Modulation by NaCl concentration. *Journal of Biological Chemistry*, 1985. **260**(6): p. 3594-3603.
224. Latham, G.J. DNase I demystified. *Ambion's TechNotes Newsletter*, 2001. **8**, 3.
225. Glasel, J.A., Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *BioTechniques*, 1995. **18**(1): p. 62-63.
226. Choi, J.-Y., Eoff, R.L., Pence, M.G., Wang, J., Martin, M.V., Kim, E.-J., Folkmann, L.M., and Guengerich, F.P., Roles of the four DNA polymerases of the crenarchaeon *Sulfolobus solfataricus* and accessory proteins in DNA replication. *Journal of Biological Chemistry*, 2011. **286**(36): p. 31180-31193.
227. Kubitschek, H.E. and Friske, J.A., Determination of bacterial cell volume with the Coulter Counter. *Journal of Bacteriology*, 1986. **168**(3): p. 1466-1467.
228. Matsunaga, F., Norais, C., Forterre, P., and Myllykallio, H., Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Reports*, 2003. **4**(2): p. 154-158.
229. Roberts, J.A., Bell, S.D., and White, M.F., An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. *Molecular Microbiology*, 2003. **48**(2): p. 361-371.

230. Santiveri, C.M., Mirassou, Y., Rico-Lastres, P., Martínez-Lumbreras, S., and Pérez-Cañadillas, J.M., Pub1p C-terminal RRM domain interacts with Tif4631p through a conserved region neighbouring the Pab1p binding site. *PLoS ONE*, 2011. **6**(9): p. e24481.
231. Turapov, O.A., Mukamolova, G.V., Bottrill, A.R., and Pangburn, M.K., Digestion of native proteins for proteomics using a thermocycler. *Analytical Chemistry*, 2008. **80**(15): p. 6093-6099.
232. Chodosh, L.A., UV crosslinking of proteins to nucleic acids, in Current Protocols in Molecular Biology. 2001, John Wiley & Sons, Inc.
233. Fadouloglou, V.E., Kokkinidis, M., and Glykos, N.M., Determination of protein oligomerization state: Two approaches based on glutaraldehyde crosslinking. *Analytical Biochemistry*, 2008. **373**(2): p. 404-406.
234. Slusarewicz, P., Zhu, K., and Hedman, T., Kinetic characterization and comparison of various protein crosslinking reagents for matrix modification. *Journal of Materials Science: Materials in Medicine*, 2010. **21**(4): p. 1175-1181.
235. Migneault, I., Dartiguenave, C., Bertrand, M.J., and Waldron, K.C., Glutaraldehyde: Behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking *BioTechniques*, 2004. **37**(5): p. 790-802.
236. Kuykendall, J.R. and Bogdanffy, M.S., Efficiency of DNA-histone crosslinking induced by saturated and unsaturated aldehydes *in vitro*. *Mutation Research Letters*, 1992. **283**(2): p. 131-136.
237. Adami, R.C. and Rice, K.G., Metabolic stability of glutaraldehyde cross-linked peptide DNA condensates. *Journal of Pharmaceutical Sciences*, 1999. **88**(8): p. 739-746.
238. Wine, Y., Cohen-Hadar, N., Freeman, A., and Frolow, F., Elucidation of the mechanism and end products of glutaraldehyde crosslinking reaction by X-ray structure analysis. *Biotechnology and Bioengineering*, 2007. **98**(3): p. 711-718.
239. Peters, K. and Richards, F.M., Chemical cross-linking: reagents and problems in studies of membrane structure. *Annual Review of Biochemistry*, 1977. **46**(1): p. 523-551.
240. Link, A.J. and LaBaer, J., Solution protein digest. *Cold Spring Harbor Protocols*, 2011. **2011**(2): p. pdb.prot5569.
241. Pal, M., Ponticelli, A.S., and Luse, D.S., The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Molecular Cell*, 2005. **19**(1): p. 101-110.
242. Luse, D.S., Promoter clearance by RNA polymerase II. *Biochimica et Biophysica Acta*, 2013. **1829**(1): p. 63-68.
243. Nechaev, S. and Adelman, K., Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochimica et Biophysica Acta*, 2011. **1809**(1): p. 34-45.
244. Liu, X., Bushnell, D.A., Silva, D.-A., Huang, X., and Kornberg, R.D., Initiation complex structure and promoter proofreading. *Science*, 2011. **333**(6042): p. 633-637.
245. Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V., and Ebright, R.H., Structural organization of the RNA polymerase-promoter open complex. *Cell*, 2000. **101**(6): p. 601-611.
246. Kapanidis, A.N., Margeat, E., Ho, S.O., Kortkhonjia, E., Weiss, S., and Ebright, R.H., Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 2006. **314**(5802): p. 1144-1147.

247. Revyakin, A., Liu, C., Ebright, R.H., and Strick, T.R., Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*, 2006. **314**(5802): p. 1139-1143.
248. Richardson, T.T., Wu, X., Keith, B.J., Heslop, P., Jones, A.C., and Connolly, B.A., Unwinding of primer-templates by archaeal family-B DNA polymerases in response to template-strand uracil. *Nucleic Acids Research*, 2013. **41**(4): p. 2466-2478.
249. Kashkina, E., Anikin, M., Brueckner, F., Lehmann, E., Kochetkov, S.N., McAllister, W.T., Cramer, P., and Temiakov, D., Multisubunit RNA polymerases melt only a single DNA base pair downstream of the active site. *Journal of Biological Chemistry*, 2007. **282**(30): p. 21578-21582.