Activation of the Mcm2-7 helicase function by ancillary replication factors GINS and Cdc45: The switch to S phase

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by

Tatjana Petojević

Berlin, 2013

1st Reviewer: Prof. Dr. Michael R. Botchan

2nd Reviewer: Prof. Dr. Petra Knaus

Defense on August 6th, 2013

Acknowledgements

First, I wish to thank my advisor Professor Michael R. Botchan at the University of California, Berkeley for accepting me to work in his lab, supporting my research for the doctorate studies and guiding me to mature as an independent scientist. I am indebted to him for giving me great scientific freedom, the possibility to explore scientific questions and many encouraging discussions.

Thanks go to my advisor Professor Petra Knaus at the Freie Universität Berlin for supervising my doctorate studies and giving me great advice throughout my stay abroad.

My stay in Berkeley for my doctorate studies was supported by a pre-doctoral fellowship from Boehringer Ingelheim Fonds and I would like to express my gratitude for both the financial and professional support during my studies. It is a great honour to be a BIF fellow and this has allowed me to meet many inspiring scientists and friends along my scientific path.

There are many people that have shaped my experiences and growth both as a scientist as well as a person and I thank everybody who has entered my life.

First and foremost, my special appreciation goes to my husband Massimiliano for his constant support and patience that gave me the strength and endurance to pursue this path. A special thank goes to his willingness to share me with science and happily take the few hours left without any complaint. Without him, none of this would have been possible. Thank you for the crazy short trips that still allowed us to discover the beauty of many national parks and our home countries, and for teaching me how to become a more patient and assertive person.

I am especially grateful to Ivar Ilves for excellent advice and mentoring in my early scientific career and for teaching me how to think and work as a true biochemist. I have greatly benefited from his professional experience and knowledge of many years in the lab, but also personally from his friendship and kindness.

My stay in the lab would have not been the same without my colleague, true friend and lastly even roommate Shirali Pandya who shared with me the ups and downs of a graduate student life. Our endless morning coffee trips to Yali's and pleasant 'Silk Road Sour' evening breaks at Revival during long nights in the lab not only diversified the lab hours,

but resulted in inspiring and fruitful scientific discussions, from which I have greatly benefitted.

I was honored to work with several very talented young scientists and help them in their early scientific steps and I would like to thank all of them. A special thank goes to Aylin Goeke for her great help in the lab, but also for being the person she is and for becoming a true friend. Her scientific enthusiasm and power was always very contagious and she truly enlightened my – sometimes - grey lab days.

I want to thank the entire Botchan lab members for great support throughout my research years. Jim Pesavento was always very generous with his advice and helpful in discussions and I deeply admire his technical inventiveness. But also I want to thank him as a person and for the priceless beer hours, which allowed us to taste his delicious home brews and become knowledgeable about the art of beer brewing.
I am very grateful to Maren Bell and her exceptional organizational skills that always kept my back free to solely dedicate my time to science. Endless thanks also for sharing the deliciousness from her garden.
Also, I want to thank Nele Tamberg for exceptional work together in the lab, funny discussions during coffee and lunch breaks and for being such a great person.
My thanks go to Daniel Blanchard and Melissa Harrison for many fruitful discussions, scientific advice and interesting conversations during social hours.

Without my family I would not be the person I am. I am deeply indebted to my parents Bosko and Jagodina without whom I could have never reached this. I thank my sister Aleksandra for being my best friend throughout life and for always giving me undivided attention, undoubtful understanding and support. An immeasurable gratitude goes to her and my brother-in-law Didi who have both never hesitated to help me in life and always cheered me up when needed. I know that it was hard to see me settle down so far away from home, and I will be eternally grateful for their willingness to support this decision.

Last, but not least, I would like to thank two of my best friends, Maria Vo and Sonja Schall, for always being there for me and having an ear open. True friendship is difficult to find and I consider myself lucky to have found you.

Abstract

The association of two auxiliary factors, Cdc45 and GINS, into the CMG (Cdc45/Mcm2-7/GINS) complex activates the pre-assembled Mcm2-7 hexameric helicase in eukaryotes and leads in the cell cycle to the switch into S phase. Protein-protein interaction studies in Psf1, a subunit of the GINS tetramer, identified four critical amino acids, which are essential for assembling the complex and, thus, for activation of the helicase.

Cdc45 and GINS proteins remain part of the CMG complex as the replication fork progresses along DNA; however, their direct role remains elusive. A scope of the work presented here is to investigate the role of the allosteric factor Cdc45 during translocation. Surprisingly, Cdc45 directly contacts the leading strand in the absence of nucleotide and displays no interaction with the lagging strand as was earlier hypothesized. Residues in Cdc45 responsible for DNA binding were identified and it was found that mutation of these residues diminished helicase activity as well as processivity of the CMG complex, demonstrating for the first time a direct effect of residues distal to the Mcm2-7 helicase motor on the processivity of the entire CMG complex.

All of the eukaryotic Mcm2-7 subunits belong to the AAA+ super-family of ATPases and the ring contains six active site pairs each between two adjacent subunits that bind and hydrolyze ATP. Elimination of the conserved ATP-hydrolysis arginine finger motifs demonstrate a functional asymmetry within the Mcm2-7 ring of the CMG complex. The nucleotide dependence for DNA binding of CMG, and significant crippling of DNA binding upon elimination of the most important active site pairs suggest a direct correlation between defects in ATPase activity and DNA binding by individual Mcm2-7 subunits.

A detailed biochemical dissection of CMG affinities for the leading and lagging strands separately also demonstrates the importance of all six Mcm2-7 subunits for the subsequent unwinding mechanism. Three candidate β-hairpin motifs of each Mcm2-7 subunit located in the central channel are implicated in DNA binding, and mutational analysis was performed for different critical residues in these hairpins. Side-by-side mutational comparison of isolated Mcm2-7 complexes to CMG helped understand differences and similarities between the inactive and active form of the helicase. Multiple interaction regions of the DmMcm2-7 proteins with the leading strand in the central channel were identified. While the N-terminal β-hairpins (NT-hp) in several Mcm2-7 subunits is crucial for DNA binding in both the free Mcm2-7 and in the CMG complex, the Pre-Sensor1 β-hairpins (PS1-hp) of other subunits are only used for DNA binding in the context of CMG. In contrast to NT-hp, primarily important for DNA binding, PS1-hp contribute directly to the translocation mechanism. The demonstrated unequal contributions with a varying degree of importance of the six Mcm2-7 subunits in the

binding and translocation of the leading strand suggest a specific order of activity around the ring, that is in agreement with a model, in which there is a sequential order of events around the Mcm2-7 ring during strand separation with a specific start site for the 'power stroke'.

Furthermore, data show that the lagging strand also binds a subset of Mcm2-7 subunits. Mutations on the external surface of the Mcm2-7 weaken CMG's affinity for this strand and decrease the helicase activity, suggesting a wrapping of the lagging strand around the external surface during strand separation.

Zusammenfassung

Die Assoziation der beiden Hilfsfaktoren, Cdc45 und GINS, in den CMG (Cdc45/Mcm2-7/GINS) Komplex aktiviert die als Hexamer vormontierte Mcm2-7 Helikase in Eukaryonten und führt im Zellzyklus zum Übergang in die S-Phase. Protein-Protein-Interaktionsstudien identifizierten in Psf1, einer Untereinheit des GINS Tetramers, vier Aminosäuren, die für den Aufbau des Komplexes und damit für die Aktivierung der Helikase kritisch sind.

Beide bleiben Bestandteil des CMG Komplexes, und ein Schwerpunkt der hier vorgestellten Arbeit ist es, die Rolle des allosterischen Faktors Cdc45 während der Translokation zu untersuchen. Cdc45 weist unerwartet einen direkten Kontakt zum Leitstrang in Abwesenheit von Nukleotid und keine Interaktion mit dem Folgestrang. Aminosäuren in Cdc45, die für das Binden von DNA verantwortlich sind, wurden identifiziert und ihre Mutation reduziert deutlich die Helikaseaktivität sowie Prozessivität des Enzyms. Erstmals wird hiermit eine direkte Wirkung von Proteinregionen distal des Mcm2-7 Motors auf den gesamten CMG Komplex gezeigt.

Alle eukaryontischen Mcm2-7 Untereinheiten gehören zur AAA+ Familie von ATPasen und der Mcm2-7 Ring enthält sechs aktive Zentren jeweils zwischen zwei benachbarten Untereinheiten, die ATP binden und hydrolysieren. Eliminierung der ‚Arginin-Finger' Motife, die für die ATP Hydrolyse verantwortlich sind, zeigen eine funktionelle Asymmetrie innerhalb des CMG-Komplexes auf.

Die Tatsache, dass die Interaktion von CMG zur DNA eine Nukleotidabhängigkeit aufweist, und dass die Mutationen der wichtigsten aktiven Zentren das Binden von DNA auflösen, deutet auf eine direkte Korrelation zwischen Defekten in ATP-Hydrolyse und Interaktion von einzelnen Mcm2-7 Untereinheiten zur DNA. Eine detaillierte biochemische Untersuchung der Affinitäten von CMG zum Leit-und Folgestrang zeigt die Bedeutung aller sechs Mcm2-7 Untereinheiten bezüglich der Strangseparation auf. Drei β-Haarnadelmotive jeder Mcm2-7 Untereinheit gelten als Kandidaten für das Binden von DNA, und eine Mutationsanalyse von kritischen Aminosäuren in diesen wurde durchgeführt. Ein paralleler Vergleich der isolierten Mcm2-7 und CMG Komplexe hat geholfen zu verstehen, welche Unterschiede und Gemeinsamkeiten zwischen der inaktiven und aktiven Form der Helikase bestehen. Mehrfache Wechselwirkungen der DmMcm2-7 Proteine zu dem Leitstrang wurden innerhalb des zentralen Kanals identifiziert. Während die N-terminalen β-Haarnadeln (NT-hp) weniger Mcm2-7 Untereinheiten von entscheidender Bedeutung für die DNA-Interaktion im isolierten Mcm2-7 sowie im CMG Komplex sind, sind die ‚Pre-Sensor1' β-Haarnadeln (PS1-hp) von anderen Mcm2-7 Untereinheiten nur für das Binden innerhalb des CMG-Komplexes wichtig. Im Gegensatz zu NT-hp sind die PS1-hp direkt für den Translokationsmechanismus verantwortlich. Die

demonstrierten ungleichen Beiträge der Mcm2-7 Proteine deuten auf eine spezifische Reihenfolge der Aktivitäten im Ring, die im Einvernehmen mit einem Modell sind, welches eine sequentielle Abfolge mit einem bestimmten Startpunkt während der Strangtrennung vorsieht.

Darüber hinaus zeigen die Daten, dass der Folgestrang ebenfalls Mcm2-7 Proteine bindet. Mutationen auf der äußeren Oberfläche des CMG Komplexes schwächen die Affinität zum Folgestrang und führen zur Verringerung der Helikaseaktivität, was darauf hindeutet, dass während der Translokation der Folgestrang sich um die äußere Oberfläche des CMG Komplexes umwickelt.

To my husband Max, my sister Aleks and my parents

Table of Contents

Table of Figures

Chapter 1

Introduction

The cell's drive to self-propagate represents one of the most fundamental principles in every growing organism. The ability to replicate its genomic information underlies this biological concept and defines each cells inherited program. While many epigenetic and environmental factors influence the cell's final destiny, the genome provides the crucial code of the building blocks for the general tools each cell is equipped with. The cell cycle consists of four main phases leading to the final division of the chromosomes to the daughter cells. During G1, S, and G2 phases, collectively called the interphase, the cells grow and ensure the copying of the genome, before the subsequent division in M (Mitosis) phase. While the correct execution of the M phase is absolutely required to provide an equal and regulated division of material onto the two daughter cells, making two identical copies of the parental cell, the pre-requisite to this arises from the duplication of the genetic information alongside the provision of the necessary tools during the growth phases. The regulation and correct execution of each step during the cell-division cycle, a process proliferating cells repeatedly undergo, are mandatory to maintain the cell's integrity. Aberration of these well-orchestrated events can result in diseases connected to dysfunctional control and progression through the cell cycle, such as uncontrolled cell growth, known as cancer, one of the most prevalent epidemics of our time. Thus, especially multi-cellular organisms must ensure the precise replication of their genomic information during S phase and limit it to only once per cell cycle as has first been shown in *Xenopus laevis* egg extracts (Blow and Laskey, 1986). Thus, each origin must fire only once per cell cycle. In addition to the accuracy of the copying event, the cell must coordinate this process with continuous transcription and chromatin assembly, thus adding to the complexity and necessity of precise regulation during cell cycle progression.

One start in each cell? – Origins and the tricky challenge of the replicon model

A cell remains at high risk each time it is required to duplicate its genome during the S phase, which represents an extremely sensitive period during the cell cycle. Not only a faithful copy of its genetic material has to be guaranteed, the DNA synthesis requires the unwinding of the double strand and increases, thus, the accessibility of mutagenic substances to nature's exposed building blocks of life that are otherwise densely packaged in chromatin (Schwob, 2004).

Central to the replication event are specific DNA sites throughout the genome, so-called origins of replication (*ori*), from which the DNA replication can initiate. Jacob and Brenner introduced 1963 the concept of a genetic unit as the *replicon* that can only replicate in its entity (Jacob and Brenner, 1963). Their hypothesis postulated that such initiation sites,

termed *replicators*, consist of specific *cis*-DNA sequences to which DNA-binding elements, so called *initiators*, can bind in *trans* and regulate.

In agreement with this 'replicon model' are the studies of various prokaryotic (e.g. *Escherichia coli*) and some eukaryotic viral (e.g. *Simian Virus* 40) DNA replication machineries that initiate their replication from a single *ori* of particular sequence specificity (Borowiec et al., 1990; Bramhill and Kornberg, 1988; Kornberg and Baker, 1992). Once initiated from the unique replication start site, the duration of the DNA synthesis phase is determined by the fork progression speed and length of the genome. Small genomes as is in *E. coli* with fast moving forks (30-60kb/min) require around 40 minutes to copy their 4.6Mb genetic information (Kornberg and Baker, 1992; Schwob, 2004). Since the process of replication represents a highly susceptible phase for internal and external DNA damage, the emphasize must lay in minimizing such exposure time of the genetic information. However, eukaryotes with their increased genome sizes and added complexity required additionally an improved accuracy of the replication process and, thus, slower moving replication forks (3kb/min) (Kornberg and Baker, 1992; Schwob, 2004). The replicon model no longer fully applied in the perception of how their chromosomes get duplicated, and nature instead evolved to the use of multiple origins of replication throughout the genome to shorten the length of S phase and still ensure error-free replica (Kornberg and Baker, 1992; Schwob, 2004).

While *Saccharomyces cerevisiae* retained the sequence specificity of the Jacob replicon model and, thus, extended the proposed hypothesis to this unicellular eukaryote, no consensus *cis*-acting sequences could be identified for higher eukaryotes. In *S. cerevisiae* such origins of replication (replicators) were identified as conserved elements of autonomously replicating sequence, or ARS, that contain the ARS consensus site (ACS) consisting of A/T-rich 11bp regions and representing the binding site for the initiator protein (Brewer and Fangman, 1987; Stinchcomb et al., 1979).

The cell's biological clock – timing and control are key

In accordance to the 'replicon' hypothesis, an initiator protein should bind to these sequence specific replicators (Jacob and Brenner, 1963). While the presence of such specific origins in eukaryotes was only identified in the budding yeast genome, the findings on the existence of a potential initiator protein are different. The role of the hexameric ORC complex as the conserved eukaryotic initiator element, consisting of six Orc polypeptides, numbered 1 through 6, was unraveled through the finding that it bound to the budding yeast ACS (Bell and Stillman, 1992; Stillman, 2005). The ORC complex as

replicator protein labels the initiation sites for the replication in the S phase in an ATP-dependent manner and provides a platform to anchor downstream proteins required for the so-called 'licensing' process by assembly of a pre-replication complex (pre-RC) that finally leads to the generation of the complex replication machinery (Bell and Stillman, 1992; Diffley et al., 1994; Klemm et al., 1997; Klemm and Bell, 2001).

Importantly, the DNA sites that become competent in the G1 phase for the subsequent copying events during the following cell cycle phase, do not initiate unwinding until the cells transition into the S phase. Several of the ORC complex subunits are ATPases that belong to the AAA+ family of proteins (ATPases Associated with various cellular Activities) (Iyer et al., 2004; Koonin, 1993; Neuwald et al., 1999). Binding to the ori leads to the inhibition of the ORC ATPase activity, but the activity of Orc1 subunit in turn undergoes activation when a second pre-RC factor and AAA+ ATPase binds, the Cdc6 protein (Bell and Dutta, 2002; Cocker et al., 1996; Coleman et al., 1996; Matsunaga et al., 2001; Stoeber et al., 1998; Wang et al., 1999). This binding to ORC is directed through its interaction with the largest of the six subunits, Orc1 protein and the interaction induces structural re-arrangements in the complex between ORC, Cdc6, and DNA that increase the specificity of ORC binding to the origin DNA (Mizushima et al., 2000; Sclafani and Holzen, 2007; Speck et al., 2005; Speck and Stillman, 2007).
The loading of the MCM helicase enzyme onto the marked ORC-Cdc6 origin sites during the G1 phase of the cell cycle requires additionally the Cdt1 protein and establishes the complete pre-RC. The requirement of this protein for the MCM loading was first reported from fission yeast and frog egg extracts (Maiorano et al., 2000; Nishitani et al., 2000). Finally, both polypeptides dissociate from the origins and the Mcm2-7 get actively loaded onto the duplex DNA as head-to-head double-hexamers in an ATP-dependent manner (Evrin et al., 2009; Gambus et al., 2011; Randell et al., 2006; Remus et al., 2009). This loading enables the origins for subsequent initiation, but the Mcm2-7 still remain inactive in G1 until additional key activators enter the scene leading to the switch into S phase and a full activation of the helicase. Thus, the strict temporal separation of the competency of the origins in G1 phase through restricting pre-RC formation to only this period, and allowing the actual initiation of replication only in S phase represent a key solution in preventing uncontrolled origin firing and guaranteeing correct replication timing.

Additionally, key to the essential ordered assembly is the proper regulation of each step and prevention of re-replication. For this purpose various instances of regulation have been introduced, involving either direct degradation induced by CDK phosphorylation as

shown for Cdc6 or ORC, ubiquitylation of Cdt1 or interactions with other replication proteins that can block the loading of the Mcm2-7 helicase complex as demonstrated for Geminin and its binding to the Cdt1 protein (Bell and Dutta, 2002; Drury et al., 2000; Elsasser et al., 1999; Havens and Walter, 2011; McGarry and Kirschner, 1998; Nguyen et al., 2001; Nishitani et al., 2006). Further, assurance that the pre-RCs are only assembled during G1 and no active Mcm2-7 loading occurs during S phase progression is achieved through the low levels of cyclin-dependent kinases (CDKs) that are necessary to recruit MCMs onto chromatin (Bell and Dutta, 2002; Botchan, 2007). In contrast to this, high CDK levels, as present in S phase, inhibit the pre-RC formation and, thus, limit the time in which the DNA can become 'competent' for replication to a specific period during the cell-division cycle and assure prevention of re-replication (Bell and Dutta, 2002).

For the correct initiation and activation of the loaded helicase motor, the sequential action of two conserved cell cycle kinases, the mentioned CDK and the Dbf4-dependent Cdc7 kinase (DDK) are crucial to ensure the switch to S phase. Main target in eukaryotes for the latter Cdc7 kinase are several of the Mcm2-7 proteins that contain long extensions at their N-termini such as Mcm2, Mcm4, and Mcm6 in budding yeast (Sclafani and Holzen, 2007). This kinase gets activated through the regulatory subunit Dbf4 and is therefore known as DDK (Jackson et al., 1993; Sclafani and Holzen, 2007). Compared to the main structured N-terminal and C-terminal domains these subunits share with the archaeal MCMs, these additional amino acid extensions are mostly unstructured, preceeding the structured body and present thus ideal access for the kinase (Randell et al., 2010; Sheu and Stillman, 2006, 2010). The binding and phosphorylation of these proteins was shown to require the origin-loaded Mcm2-7 state and is initially required to recruit Sld3 and Cdc45 to the origins before the CDK activation (Heller et al., 2011; Sheu and Stillman, 2006). Unlike the identification of DDK target proteins, the exact mechanism behind the initiation remains unanswered. Besides providing direct binding sites on Mcm2-7 for downstream replication factors, heretofore, mainly two ideas have been proposed for the DDK activation. First, the structural analysis of a mutant in *S. cerevisiae* Mcm5 (*mcm5-bob1*) that was able to bypass the DDK requirement revealed minor structural changes within the Mcm2-7 protein complex (Fletcher et al., 2003; Hardy et al., 1997). Since Mcm5 is not a direct target of the complex, the possibility was discussed that structural re-arrangements might be induced by the phosphorylation critical for the activation of the helicase (Labib, 2010). Second, a recent study from the Stillman lab revealed an inhibitory domain in the N-terminus of Mcm4 that gets alleviated upon phosphorylation by DDK (Sheu and Stillman, 2010). However, such domain could not be found in other targets of DDK like

Mcm2 and Mcm6, suggesting a more complicated mechanism, the combination of several effects or unknown other kinase targets (Sheu and Stillman, 2010).

Subsequently and different from the DDK activation, CDK's main target proteins are factors necessary during the initiation steps of DNA replication, but do not remain part of the replisome progression complex. Majorly, Sld3 and Sld2 proteins are phosphorylated by CDK, leading to their docking on N-terminal and C-terminal sites of Dpb11 (Botchan, 2007; Masumoto et al., 2002; Tanaka et al., 2007; Zegerman and Diffley, 2007). In budding yeast, these two proteins have been considered to represent the 'minimal set' of CDK target proteins during the initiation process that still can promote DNA replication since a phospho-mimetic mutation in Sld2 together with an Sld3/Dpb11-fusion protein was able to bypass the CDK phosphorylation requirement for Sld3 as well as the presence of BRCT repeats at the N-terminus of Dpb11 through which it binds (Zegerman and Diffley, 2007). The phosphorylation of these factors results in the final recruitment of additional crucial replication factors such as GINS and polymerase ε to the origins that maintain interaction with the replication fork during the following elongation phase (Gambus et al., 2006; Heller et al., 2011; Pacek et al., 2006).

Partners of the MCM helicase motor – fully equipped, ready for action, but in need for helpers

Since the discovery of the MCM proteins in the 1980's through a genetic screen in *Saccharomyces cerevisiae* that displayed defects in minichromosome maintenance, much progress has been made to decipher their discrete role in DNA replication (Forsburg, 2004; Maine et al., 1984). In archaea six copies of the MCM protein oligomerize to form a ring structure, while eukaryotes possess six different genes encoding for six homologous, but different Mcm2-7 polypetides that have been demonstrated to be required for initiation as well as elongation steps (Tye and Sawyer, 2000).

Consistent with other DNA helicase proteins, the MCM proteins belong to the AAA+ superfamily of ATPases (Koonin, 1993; Tye and Sawyer, 2000). Characteristic to this family of proteins is the presence of a 'MCM box' that contains the required motifs for their ATPase activity (Forsburg, 2004; Neuwald et al., 1999). This feature comprises an about 200 amino acids region within the C-terminus of each MCM polypeptide including the distinctive ATPase motifs such as Walker A and Walker B box for nucleotide binding, as well as the arginine finger for subsequent nucleotide hydrolysis (Forsburg, 2004; Neuwald et al., 1999). The arginine finger is provided in *trans* to the neighboring subunit containing the Walker A and Walker B motifs in *cis* to catalyze the cycles of ATP-binding

and -hydrolysis that are critical to fuel the MCM motor during translocation (Hanson and Whiteheart, 2005; Koonin, 1993).

This suggests that pairs of MCM proteins need to be in close proximity, and indeed the Mcm2-7 subunits assemble not only in a hexameric ring structure, but also with a specific order of the six proteins that distinctively identify the next neighboring subunits around the ring (Crevel et al., 2001; Davey et al., 2003; Schwacha and Bell, 2001). The identification of corresponding ATPase pairs in *S. cerevisiae* led to the proposed unique order of Mcm2-7 in '5-3-7-4-6-2' and has been confirmed in other organisms (Crevel et al., 2001; Davey et al., 2003). Structural studies from archaeal MCM homo-hexamers have demonstrated a characteristic 'dumbbell' shape of each MCM protein with an isthmus between the N- and C-termini, leading to an overall structure of the MCM ring with the N-terminal domain of each subunit forming one ring that is stacked upon another ring formed by the C-terminal domains of the six proteins (Brewster and Chen, 2010; Brewster et al., 2008; Costa et al., 2006; Fletcher et al., 2003; Pape et al., 2003). The crystallographic studies of either N-terminal fragments or full-length MCMs further revealed specific positively charged β–hairpin structures throughout the N- and C-terminal domains of each Mcm2-7 subunit with implications in DNA binding within the central channel (Brewster et al., 2008). Corresponding hairpin motifs in papillomavirus E1 and simian virus SV40 Large T antigen directly participate in pulling nucleotides through the central pore, leading to posit similar mechanistic functions for these hairpin motifs in Mcm2-7 (Enemark and Joshua-Tor, 2006; Gai et al., 2004).

The Mcm2-7 hetero-hexamer has been suggested to constitute the putative eukaryotic replicative helicase since all necessary motifs for ATPase activity and binding are present and the archaeal counterpart possesses robust unwinding activity (Forsburg, 2004). In contrast to this, purified Mcm2-7 complexes failed to display helicase activity, and for long only a sub-complex of Mcm4, Mcm6, and Mcm7 showed unwinding *in vitro* while Mcm2, Mcm3, and Mcm7 were attributed negative regulatory functions (Ishimi, 1997; You et al., 1999). However, recently, a purified Mcm2-7 complex from budding yeast demonstrated weak intrinsic helicase activity *in vitro* and this is in agreement with the *in vivo* requirements for all six subunits (Bochman and Schwacha, 2008; Tye and Sawyer, 2000).

In other eukaryotic organisms, isolated Mcm2-7 still represent an inactive form of the helicase complex, and robust unwinding was only observed after incorporation of the two auxiliary replication factors Cdc45 and the four GINS proteins into the so-called CMG complex (Cdc45/Mcm2-7/GINS) (Moyer et al., 2006). The biochemical analysis of isolated *Drosophila* Mcm2-7, side-by-side with the DmCMG complex, showed that various intrinsic activities of the Mcm2-7 hetero-hexamer undergo a dramatic activation upon binding of

Cdc45 and GINS. This led to propose that the CMG complex represents the true activated helicase in eukaryotes, in which the auxiliary factors induce allosteric remodeling of the Mcm2-7 core and, thus, optimize subunit organization for enhanced activities (Ilves et al., 2010).

Additionally, an electro-microscopic analysis of the DmCMG showed structural roles for Cdc45 and GINS during binding to the Mcm2-7 helicase motor. In isolation, the Mcm2-7 complex can adopt two different conformational states: either a lock-washer shaped spiral state or a planar ring form with a discontinuity between Mcm2/Mcm5 (Costa et al., 2011). Within the CMG the equilibrium of these two states is shifted towards the planar configuration and is reinforced by the association of Cdc45 and GINS as a bridge over the '2/5 gap' on the exterior of Mcm2, Mcm3 and Mcm5 (Costa et al., 2011). The gap in the Mcm2-7 ring is fully sealed upon nucleotide binding, leading to a constriction of the Mcm2-7 central channel that could suggest an enhanced grasping of the DNA through the Mcm2-7 motor (Costa et al., 2011).

Besides its essential role in initiation of DNA replication as one of the first proteins loaded onto the replication origins in the switch to S phase, the Cdc45 protein was shown to be further required for elongation during S phase progression and to travel with the replication fork (Jares and Blow, 2000; Tercero et al., 2000; Zou et al., 1997).

Similarly, the tetrameric GINS complex (named from Japanese 'Go-Ichi-Ni-San' for '5-1-2-3') consisting of Sld5, Psf1-3 proteins that constitutes one of the more recently discovered protein complexes in DNA replication was not only shown to load early onto the origins and be required for the initiation in the transition from G1 to S phase of the cell cycle, but also to additionally remain associated with the replication fork during progression through S phase (Gambus et al., 2006; Kanemaki et al., 2003; Kubota et al., 2003; Takayama et al., 2003).

The nature of both proteins will be discussed in more detail in Chapters 2 and 4, but direct physical interactions between Cdc45-Mcm2-7-GINS have been repeatedly reported through co-immunoprecipitations in *Xenopus* egg extracts or in biochemical purifications of *Drosophila* or human CMG complexes (Ilves et al., 2010; Im et al., 2009; Kang et al., 2012; Kubota et al., 2003; Moyer et al., 2006). Both accessory protein factors maintain robust association with the Mcm2-7 helicase motor throughout the S phase and must play a critical role – together with CDK and DDK - in the transition from the assembled pre-RC complex in G1 to the activated CMG helicase complex and, thus, switch to the S phase of the cell cycle. Similar to other replicative helicases, CMG couples energy from NTP hydrolysis to separate the parental duplex DNA into single-stranded DNA as substrates for

polymerases that subsequently step in. *Drosophila* CMG translocates in 3' to 5' direction, however, little is known about the molecular unwinding mechanism (Moyer et al., 2006).

Proposed models for MCM unwinding

The central question of how replisomes unwind duplex DNA has led to various proposed mechanistic models as well to the long-standing discussion if the two replisomes retain their association during strand separation or work independently of each other. Elegant single-molecule studies were able to demonstrate that, after their activation, the two sister replisomes do not retain physical contact, but move bi-directionally away from the origin of replication in a manner that allows both separated helicase hexamers to work independently (Yardimci et al., 2010). While the Mcm2-7 are loaded as double-hexamers onto DNA during G1 phase of the cell cycle, the subsequent separation of the two complexes in the transition to S phase is required to assure the independent replisomes (Evrin et al., 2009; Gambus et al., 2011; Remus et al., 2009). This is in agreement with biochemical studies of the activated *Drosophila* CMG helicase that functions as single hexamer in solution (Ilves et al., 2010). Since both *Drosophila* and human CMG helicases cannot bind to duplex DNA, but require single-stranded regions for the binding, major structural remodeling must take place after the double hexamers are loaded onto duplex, leading to the additional melting of the double-stranded DNA and segregation of the separated strand (Ilves et al., 2010; Kang et al., 2012).

For long, the Large T antigen of the simian virus (SV)40 represented a model helicase for studying eukaryotic DNA replication. Electron microscopic studies displayed 'rabbit ear' structures during strand separation and based a model of functional double-hexamers during the unwinding mechanism on these observations (Wessel et al., 1992). These structures demonstrated the exit of single-stranded DNA through side channels of each of the two physically connected hexamers (Wessel et al., 1992). In support of this double-hexameric state of this helicase, mutations in Tag that abrogated the formation of the double-hexamer also additionally inactivated the enzyme, and the helicase displayed increased activity as double hexamer compared to the single hexameric state (Smelkova and Borowiec, 1997; Weisshart et al., 1999). However, recent single molecule studies were able for the first time to prove that these double hexamers that are formed for the initiation are not required for the elongation step and can indeed work independently of each other (Yardimci et al., 2012). These findings support the idea that different replicative helicases like SV40 large Tag, papillomavirus E1 and MCMs share a common unwinding mechanism based on single hexamers that work autonomously.

With regard to the mechanism that MCMs might use for strand separation, major remodeling must take place within the central pore that allow a transition from the double hexamers loaded onto double-stranded DNA to the establishment of the fully activated CMG single hexamer encircling single stranded DNA. Several lines of evidence were provided from studies of the separate N- and C-terminal domains of archaeal MCMs and showed different binding affinities of dsDNA and ssDNA in these two regions (Liu et al., 2008; Pucci et al., 2007). The findings led to the proposed 'strand extrusion model' in which the duplex DNA is engaged in the C-terminus of the MCM hexamer and the two separated strands leave the MCM core through different pores: the leading strand remains in the central channel and is guided outside through the N-terminal domain of the center while the lagging strand emerges from the side channels to the exterior. Consistent with this model are observations in archaea that MCMs are oriented with their C-terminus towards the duplex DNA and the model requests that the central channel is large enough to accommodate two strands simultaneously (Moreau et al., 2007).

Different from this, crystallographic studies of E1 have previously demonstrated that this helicase encircles exclusively single-stranded DNA in its central channel consistent with a 'steric exclusion' mechanism (Enemark and Joshua-Tor, 2006). This mechanism entails that the leading strand remains within the central channel of the hexameric helicase after initial unwinding, while the separated lagging strand is physically segregated and excluded to the outside. Thus, at no given time duplex DNA is present during the unwinding and elongation phases of the DNA synthesis after the origin melting. In agreement with this are the observations from the DmCMG helicase complex that does not bind double-stranded DNA, in par with the structural studies that suggest that the central channel of the Mcm2-7 is not large enough to accommodate duplex DNA (Costa et al., 2011; Ilves et al., 2010). Additionally, work from archaeal MCMs reported dynamic interactions of the lagging strand with the exterior MCM surface and recently proposed a 'steric exclusion and wrapping mechanism' in which the segregated strand wraps around the MCM ring during the unwinding mode and the leading strand is guided through the central channel (Graham et al., 2011; Rothenberg et al., 2007).

Thesis overview

The thesis presented here characterizes both structural and biochemical aspects of the Mcm2-7 core and the auxiliary factors Cdc45 and GINS within the *Drosophila melanogaster* CMG complex - the eukaryotic active helicase. The structural analysis of GINS mainly focuses on discrete interactions within the high-molecular weight complex and targets the identification of residues that are absolutely crucial for the formation of the complex and, thus, the activation of the enzyme. As pre-assembled Mcm2-7 proteins remain inactive until the incorporation of GINS and Cdc45, the identification of several amino acids in one of the GINS subunits that lead a fully abrogated CMG formation when eliminated could, thus, constitute possible novel targets for anti-cancer therapies by directly preventing the activation step.
Additionally, biochemical dissection of the second auxiliary factor Cdc45 identifies residues responsible for the interaction with the leading strand and suggests a role of as 'guardian of the Mcm2/5 gate' for Cdc45. This analysis, for the first time, unravels sites outside the Mcm2-7 motor that directly affect the helicase activity and processivity of the enzyme suggesting a possible regular participation during unwinding.

As the association of Cdc45 and GINS is required for the activation of the *Drosophila* Mcm2-7 helicase motor, detailed aspects about the DNA binding within this CMG activated complex were additional focus of the thesis. Understanding how the active helicase complex binds the two separated DNA strands was sought to shed light onto the underlying mechanism as to how the enzyme separates dsDNA.

Therefore, three main aspects of DNA binding properties and helicase activities of the CMG complex were addressed in this thesis demonstrating unequal contributions and a strong functional asymmetry of six Mcm2-7 subunits:
(A) The contributions of all three β-hairpin motifs in each of the six Mcm2-7 proteins within the central channel to leading strand binding
(B) The role of the six external β-hairpins of Mcm2-7 in DNA binding
(C) The role of the six ATPase active sites and their effect on DNA binding

The biochemical analysis of intrinsic activities such as ATPase, helicase and DNA binding represents the first complete dissection of the CMG complex by targeting specific putative DNA binding motifs or ATPase active sites and understanding their inter-dependencies. The gained knowledge provides insights into why all six subunits are required and attributes them different specific roles.

The major conclusions that can be drawn from this work include:

1) Psf1-B domain represents one of the crucial regions for the CMG formation and might comprise an anchor for other proteins of the replisome.

2) Contributions of the six Mcm2-7 ATPase sites to the CMG activities are not equal and suggest a functional asymmetry.

3) Role of the N- and C-terminal hairpin motifs in the Mcm2-7 central channel of CMG is unequal displaying strong DNA binding asymmetries with preference of specific subunits for the 'power stroke'; this analysis corroborates that the Mcm2-7 pore encircles the leading strand.

4) CMG binds the lagging strand through a wrapping mechanism on the exterior Mcm2-7 surface.

5) Cdc45 directly contacts the leading strand, possibly in a regular manner, and guards the Mcm2/Mcm5 discontinuity in the Mcm2-7 ring during translocation to prevent a leading strand slippage.

Chapter 2

Psf1-B domain is critical for the formation of the replicative *Drosophila* CMG helicase

2.1 Abstract

In eukaryotes, an activation of the pre-assembled Mcm2-7 helicase requires the action of two kinases, CDK and DDK, and the association of the two auxiliary factors, Cdc45 and the tetrameric GINS complex that lead to the switch into S phase of the cell cycle. While weak pair-wise interactions could be detected between the three components of this activated CMG (Cdc45/Mcm2-7/GINS) helicase complex, stable interactions take only place in the complete eleven-membered helicase complex.

In this study, co-immunoprecipitations were used to identify important domains for the CMG stability. In contrast to the isolated GINS complex, in which the flexible C-terminus (B-domain) of Psf1 is dispensable for the complex formation, the corresponding deletion mutant in *Drosophila* Psf1 failed to yield an intact CMG complex. Homology modeling of the *Drosophila* Psf1 C-terminus was based on the human Sld5 crystal structure, and allowed secondary structure prediction that provided information through the evolutionary conservation of the GINS complex. Additional smaller deletions helped narrow down the region responsible for CMG formation to the last 18 carboxy-terminal amino acids of Psf1. Alanine scanning of a predicted α–helix domain in this region identified four amino acids that are absolutely essential for the assembly of the CMG complex. Only a recent crystallographic study provided the spatial position for the highly flexible Psf1 C-terminus within an archaeal GINS complex. Docking of the homology model for DmC-Psf1 onto the electron density map of CMG based on this latter work suggests the location of this B-domain towards Cdc45 within the CMG complex. In par with a direct interaction between Cdc45 and GINS that was biochemically confirmed, the structural data suggests that C-Psf1 most likely interacts with the N-terminus of Cdc45. In contrast to this prediction, deletion of up to 99 N-terminal amino acids from Cdc45 still yielded an intact CMG complex and further did not abolish the direct binding of the GINS in a Cdc45-GINS sub-complex. This data leads to postulate that the direct interaction of the α-helix within C-Psf1 resides somewhere else in Cdc45, and that this domain might move into close proximity upon structural re-arrangements of CMG and establish the direct interaction that together with interactions to the Mcm2-7 ring lead to a tight complex assembly.

2.2 Introduction

The loading and activation of the helicase enzyme critical for separation of the parental DNA for subsequent synthesis is of utmost importance for the DNA replication process conducted in all dividing cells once per cell division. In eukaryotic cells this requires a highly coordinated multistep mechanism. In a first step, the origin recognition complex (ORC), assisted by Cdc6 and Cdt1, loads the helicase motor complex consisting of six minichromosome maintenance subunits (Mcm2-7) onto duplex DNA. This step is accomplished by deposition of a double hexamer in a head-to-head manner, resulting in a catalytically still inactive proto-helicase (Bowers et al., 2004; Evrin et al., 2009; Gambus et al., 2011; Remus et al., 2009). In contrast to its archaeal counterpart, isolated Mcm2-7 complexes show only a weak helicase activity *in vitro* (Bochman and Schwacha, 2008; Chong et al., 2000; Kelman et al., 1999; McGeoch et al., 2005). Concomitant with the switch to S phase, additional replication proteins Cdc45 and the GINS tetramer associate with Mcm2-7 to form the eleven-membered CMG helicase (Gambus et al., 2006; Moyer et al., 2006). This assembly leads to an activation of the Mcm2-7 helicase motor, and the CMG tracks in 3' to 5' direction on the leading strand to provide single-stranded DNA (ssDNA) for the DNA polymerase to step in and duplicate the genomic material (Ilves et al., 2010; Pacek et al., 2006). The biochemical analysis of the recombinant *Drosophila* CMG reported that upon the association of Cdc45 and GINS intrinsic activities such as ATP hydrolysis, DNA binding and unwinding of the Mcm2-7 motor are elevated by orders of magnitude (Ilves et al., 2010). Despite of pairwise interactions that were detected between Mcm2-7, GINS, and Cdc45, tight assembly only takes place in the complete CMG complex (Ilves et al., 2010).

One part of this stable ternary complex consists of the GINS complex, named in Japanese for go-ichi-ni-san (5-1-2-3) and is composed of the four polypeptides Sld5 (synthetic lethal with the *dpb-11* mutant-5) and Psf1-3 (partner of Sld5) (Kanemaki et al., 2003; Takayama et al., 2003). This tetrameric complex was originally identified from *Saccharomyces cerevisiae* and its homolog from *Xenopus laevis* (Kanemaki et al., 2003; Kubota et al., 2003; Takayama et al., 2003). These studies showed that the complex played a key role in the initiation process after assembly of the pre-replicative complex and that it was crucial for regular replisome progression (Gambus et al., 2006; Pacek et al., 2006). Ever since, homologous complexes from various eukaryotes have been identified and X-ray crystallographic structures of the human and recently one archaeal GINS complex were reported (Chang et al., 2007; Choi et al., 2007; Kamada et al., 2007; Oyama et al., 2011). Besides its integration into the CMG complex and its functional role in the helicase

activation, evidence has been accumulated that GINS functionally interact with other factors of the replisome, such as polymerases, and was accredited importance from initiation to elongation steps during DNA replication (De Falco et al., 2007; Takayama et al., 2003). The crystal structures demonstrated extensive interaction surfaces required for the formation of the compact GINS complex and implicated several flexible regions for putative interactions with other replication proteins (Kubota et al., 2003). The complex shows a high degree of conservation amongst eukaryotes, and each of the four GINS members is a small protein comprising of approximately 200 amino acids (Kubota et al., 2003; Takayama et al., 2003). In general, all four polypeptides consist of a major α-helix-rich 'A'- and a minor β-sheet-rich 'B'-domain (Kamada et al., 2007; Makarova et al., 2005). The order for Psf1 and Sld5 is A at the N-terminus and B at the C-terminus (A-B), and is circular permuted in Psf2 and Psf3 to B-A. Consistent with the fact that archaea possess a simplified version of the replisome that is related to the eukaryotic one, they also have a simpler GINS complex. Similar to the six copies of only one MCM polypeptide that homo-hexamerize to form the replicative helicase, instead of the six different Mcm2-7 proteins that hetero-hexamerize, archaea encode only two forms of GINS homologues, named Gins15 and Gins23 (Makarova et al., 2005; Marinsek et al., 2006).

Information about the second auxiliary factor Cdc45 lacked for long behind. Only recent bioinformatics studies revealed an evolutionary link between this protein and bacterial RecJ exonucleases, thus opening a novel perpective for studies to gain insights into its discrete role (Krastanova et al., 2012; Makarova et al., 2012; Onesti and Macneill, 2013; Sanchez-Pulido and Ponting, 2011). The biochemical analysis of *Drosophila* Cdc45 within the CMG complex is focus of Chapter 4 and will be discussed there.

However, how exactly the GINS and Cdc45 contribute to the activation of the Mcm2-7 helicase is not known. Also how these eleven proteins exactly associate with each other when incorporated into the CMG complex remains an unanswered question. Therefore, a single-particle electron microscopic study of the isolated *Drosophila* Mcm2-7 and CMG complexes was performed and resulted not only in important insights into the overall architecture, but laid the structural fundament for the activation mechanism (Costa et al., 2011). The study demonstrated that the free Mcm2-7 complex can transition between a planar, notched-ring and a spiral, lock-washer conformation and contain a discontinuity between subunits Mcm2 and Mcm5. The planar Mcm2-7 state with the present gap is reinforced in the CMG complex in absence of nucleotide and by association of the auxiliary factors Cdc45 and GINS. Docking of the homologous human GINS crystal structure onto the EM map revealed extensive contacts between Cdc45, GINS and the Mcm2-7 ring (Costa et al., 2011).

GINS and Cdc45 share a large contact surface with each other, in particular Cdc45 interacts with the N-terminus of Psf2 and associates with the N-terminal domain of Mcm2 (Costa et al., 2011). Furthermore, GINS binds to the external surface of the N-terminus of Mcm3 and Mcm5 through interactions with mainly Psf2 and Psf3; two proteins of the GINS tetramer that are evolutionary closely related (Costa et al., 2011; Makarova et al., 2005). Archaea possess two homologous ancestral GINS subunits, of which Gins23 was shown to bind to the N-terminus of the MCM ring (Marinsek et al., 2006). Additional contacts were observed upon nucleotide binding, mainly between the GINS and the AAA+ domains of Mcm3 and Mcm5 of DmCMG (Costa et al., 2011). Taken together, the structural studies of the DmCMG complex indicated that GINS and Cdc45 bridge the Mcm2/5 gap as a handle on the longitudinal axis of one side of the Mcm2-7 hetero-hexamer and further help structurally narrow the existent '2/5' discontinuity (Costa et al., 2011).

The enhanced interactions, which could be structurally visualized for the DmCMG, and the bridging of the Mcm2/5 gap by Cdc45/GINS provide one explanation why the complex's intrinsic activities are manifold improved compared to the Mcm2-7 (Costa et al., 2011; Ilves et al., 2010). Once the Mcm2-7 are loaded as double-hexamers around duplex DNA, subsequent steps will require the exclusion of the lagging strand from the central core to ensure only one strand in the central cavity. This might be rendered possible by the 'Mcm2/5 gap' (see Chapter 4). However, for robust translocation on the leading strand, the functional closure of the gate might constitute a fundamental criterium to ensure that the DNA remains continuously encircled. Such transition from dsDNA to ssDNA is implicated by the observation that, unlike Mcm2-7, CMG can only bind to single-stranded DNA, and, thus, requires major structural remodeling of the DNA between the recruitment and the activation of the helicase complex (Evrin et al., 2009; Ilves et al., 2010; Remus et al., 2009). Observations from the direct biochemical comparison of free Mcm2-7 and CMG complexes that reported a vast activation of the MCM motor by association of Cdc45 and GINS (Ilves et al., 2010; Moyer et al., 2006) led to suggest, that Cdc45 and GINS represent allosteric factors of the Mcm2-7 proteins and assist in the proper coordination of the six subunits in the Mcm2-7 motor that subsequently then lead to more efficient DNA separation (Ilves et al., 2010).

The CMG complex comprises a complicated and extensive network of direct protein-protein interactions. Even though many domains involved in direct contacts have been implicated essential for cell viability, it was crucial to understand if one could identify residues that are absolutely critical for the formation and stability of the CMG complex without altering the formation of either the Mcm2-7 or GINS complex. Identifying such

direct protein-protein interactions that might abrogate the CMG formation could provide an important target for drug development in future cancer therapeutics by inhibiting the activation of one of the most fundamental enzymatic tasks of each dividing cell: the separation of the duplex DNA.

2.3 Results

Drosophila Psf1 B-domain can be modeled based on homology to human Sld5

Crystal structures of human GINS attempting to crystallize the full-length GINS complex remained unsuccessful for long and showed that the C-terminus of Psf1 was susceptible for protease degradation due to a long linker region reminiscent of the one found in Sld5 (Kamada et al., 2007). Hence, the removal of this flexible region resulted in stable complex formation with the same stoichiometry as wildtype GINS (Chang et al., 2007; Choi et al., 2007; Kamada et al., 2007). However, the deletion of the C-terminus in Psf1 resulted in impaired chromatin binding of GINS and led to abrogated DNA synthesis in *Xenopus* egg extracts (Kamada et al., 2007). Since this β–sheet rich (B) domain is dispensable for the formation of the GINS complex, it was suggested to likely constitute a flexible domain of the GINS core complex by a linker between N- and C-termini of the protein (Kamada et al., 2007). Further, its potential role as binding surface for other replication proteins was earlier implicated (Kamada et al., 2007). To understand whether this domain was required for the formation of the eukaryotic CMG helicase, I wished to test the CMG stability upon deletion of the C-terminus of *Drosophila* Psf1 (Chang et al., 2007; Kamada et al., 2007). No direct structural evidence existed due to its absence in other crystal structures of human full-length GINS (Chang et al., 2007). However, comparison of sequences indicated that Psf1 and Sld5 share some conservation and are closely related as ancestral paralogs (Makarova et al., 2005). Thus, DmPsf1 can be modeled based on its homology to the hSld5 (Figure 2.1A and B). Threading of the modeled DmPsf1 3D structure onto the crystal structure of full-length human Sl5 suggests a location for the C-terminal B-domain of Psf1 (Figure 2.1B and C). All four GINS polypeptides consist of an α–helix rich (A-domain) and a β–strand rich (B-domain) region. Similar to the schematic diagram for the four human GINS presented by Kamada et al., (2007) the secondary structure prediction for *Drosophila* Psf1 by PsiPred guided to generate such schematic. It shows, analogous to Sld5, that the order of these two structural domains is A-B (Figure 2.1D). The secondary structure prediction was used to first generate a deletion construct for Psf1, named Psf1ΔC ($Psf1^{1-139}$) that removes the entire predicted B-domain of the protein (Figure 2.1D). The sequence alignment of this domain between different species shows several absolutely conserved residues and a high degree of primary sequence similarity (Figure 2.1E). Previous work on the recombinant *Drosophila* CMG complex established that strong contacts between Cdc45/Mcm2-7/GINS are pivotal for the stability of the helicase, and, thus, raised an interest in understanding direct contacts between the three parts of the high-molecular weight CMG complex. Biochemical analysis during my diploma thesis

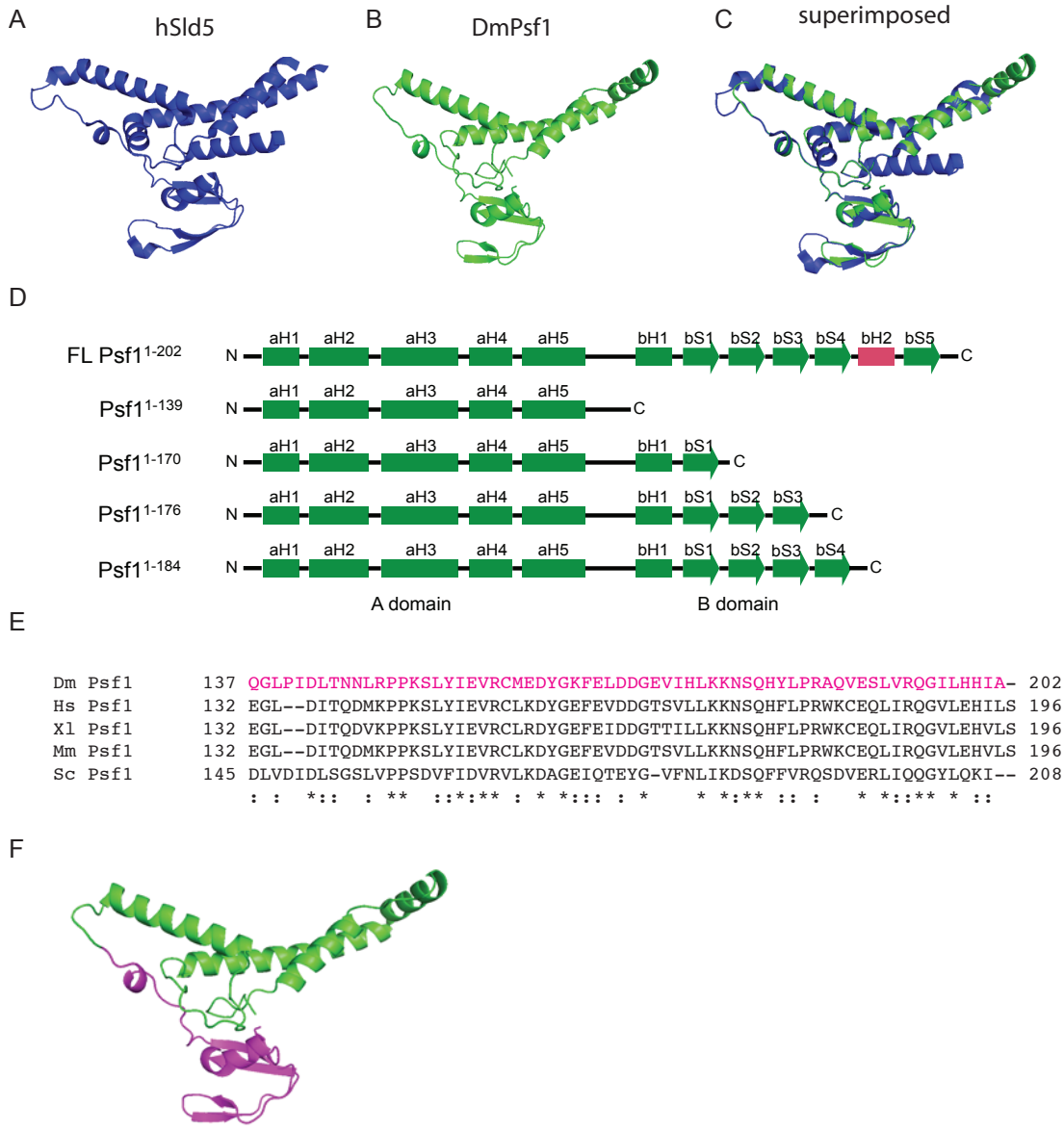

Figure 2.1. Modeling of *Drosophila* Psf1 based on homology to human Sld5.
(A) Shown is a cartoon representation of the crystal structure of the human Sld5 subunit of the GINS complex (PDB entry: 2E9X) in *blue*. The image was created by the PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.
(B) Cartoon representation of the homology model of *Drosophila* Psf1 protein (in *green*) based on the hSld5 crystal structure using SWISS-MODEL (Arnold et al., 2006; Kiefer et al., 2009) and PyMOL.
(C) The homology model of DmPsf1 (in *green*) was superimposed onto the crystal structure of hSld5 (in *blue*) using PyMOL.
(D) The secondary structure prediction for DmPsf1 obtained by PsiPred and homology modeling using SWISS-MODEL is shown schematically. The arrangement of the two domains A (α-helix rich) and B (β-sheet rich) is depicted. The schematic diagram was derived from the representation shown for human GINS (Kamada et al., 2007). Highlighted (in *magenta*) is a helix located at the very C-terminus of DmPsf1.
(E) Multiple sequence alignment of C-termini of Psf1 between different species created by ClustalW2. Shown is the B-domain of *Drosophila* (Dm, *magenta*), *Homo sapiens* (Hs), *Xenopus laevis* (Xl), *Mus musculus* (Mm), and *Saccharomyces cerevisiae* (Sc) Psf1 proteins. Conserved residues are indicated by asterisks, and residues with similar amino acids by a colon.
(F) The B-domain at the C-terminus of Psf1 is highlighted (in *magenta*) on the cartoon representation of the homology model of DmPsf1.

revealed several direct interactions between Mcm2-7, and Cdc45 or GINS, and a recent single-particle electron microscopic study of the DmCMG substantiated some of these findings. More importantly the latter report provided the basis for a more detailed knowledge about the interaction network within the eleven-membered complex (Costa et al., 2011; Ilves et al., 2010). As the Mcm2-7 helicase gets only activated upon the binding of Cdc45 and GINS, it raised the interest in identifying regions of interaction that are absolutely critical for the formation of the CMG complex, and, thus, for an activation of the pre-assembled proto-helicase (Ilves et al., 2010; Moyer et al., 2006).

The position of this flexible C-terminal B-domain is highlighted on the homology model of DmPsf1 in Figure 2.1F, suggesting its highly accessible location, and, thus, implication in direct interactions with other proteins of the replisome. As this domain is dispensable for the GINS complex formation, it does not represent critical interactions for the core of the GINS, but could rather serve as an anchor for other replication proteins, and was, thus, of interest for this study.

Psf1 B-domain is absolutely critical for CMG formation

To test the assumption that this domain might be important for other replication proteins, we deleted the complete flexible B-domain of *Drosophila* Psf1 and asked if this Psf1ΔC ($Psf1^{1-139}$) construct might affect the formation of the CMG complex. Co-immunoprecipitations were used to address this question. Evan expression levels of all proteins, including the truncated Psf1 protein, were confirmed and, first, the formation of the four-membered GINS complex was tested (Figure 2.2A). As expected from the findings about the human GINS, the C-terminal deletion of Psf1 was capable to form a complex with its remaining three GINS partners, with similar efficiency as wildtype GINS (Figure 2.2B, lane 2). Next, formation of the CMG complex containing Psf1ΔC was tested. All proteins were expressed to similar levels (Figure 2.2C). However, immunoprecipitation of the CMG complex showed that removal of the B-domain in Psf1 disabled its assembly into an intact CMG complex (Figure 2.2D, lane 2). Others have shown that this domain in human Psf1 was also dispensable for its association into the GINS complex, but that GINS, upon deletion of the C-terminus, could no longer bind to chromatin, and that this region was essential for DNA replication (Kamada et al., 2007). The structural study of the DmCMG had previously revealed an intricate network of interactions and identified most of the interactions between Mcm2/3/5 and Cdc45/GINS (Costa et al., 2011). Biochemical analyses of pairwise interactions suggested the necessity of a ternary complex between Mcm2-7/Cdc45/GINS for robust stability (Costa et al., 2011; Ilves et al., 2010). In agreement with this, withholding Mcm5 entirely disrupted the CMG formation, and

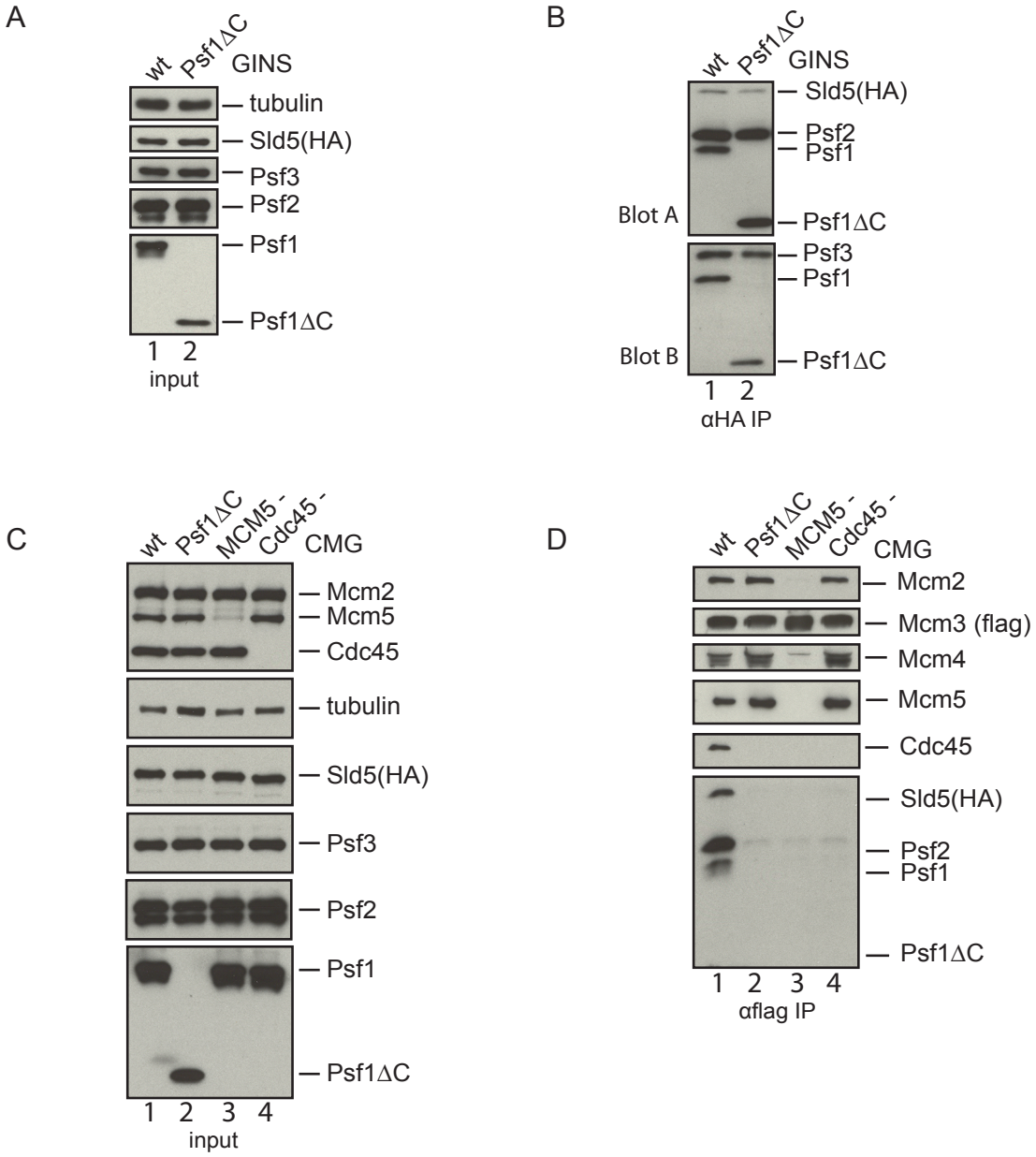

Figure 2.2. B-domain of Psf1 is critical for the formation of the CMG complex.
(A) Expression levels of individual proteins used for testing GINS formation were detected by antibodies against the four GINS subunits indicated on the right of the immunoblot. Loading was controlled by tubulin detection.
(B) IP experiments testing the stability of the GINS complex were performed by pull-downs of the HA-tagged Sld5 subunit. The formation of the four-membered complex is shown for wt GINS (lane1), and the C-terminal B-domain of Psf1 is dispensable for GINS formation (lane2). Blots A and B were probed with antibodies against indicated proteins.
(C) Expression levels of individual proteins used for testing CMG formation were detected by antibodies against proteins indicated on the right of the western blot. Loading was controlled by tubulin.
(D) Immunoprecipitation (IP) experiments analyzing the CMG stability. CMG wildtype (1); CMG with a C-terminal truncation in Psf1 (2); CMG without MCM5 (3); CMG without Cdc45 (4). Immunoprecipitation of Mcm3 subunit via a FLAG-tag yields the intact CMG complex, whereas removal of the Psf1 B-domain ($Psf1^{1-139}$, named Psf1ΔC), or absence of Mcm5 or Cdc45 proteins, disrupts the CMG formation. The truncation of Psf1 C-terminus is shown in the schematic diagram in Figure 2.1D.

neither the GINS complex nor Cdc45 co-immunoprecipitated with Mcm3 (Figure 2.2D, lane 3). Similarly, when Cdc45 was not provided in the co-expression, the GINS complex could not be detected in the pull-down reactions (Figure 2.2D, lane 4). This data suggests that for the stability of the complete CMG complex interactions between all eleven subunits require a complicated network and further explains why purifications of pairwise or ternary sub-complexes failed (Figure 2.2D, lane 3 and 4).

This biochemical finding reported here about the B-domain of Psf1 demonstrated one region within the complex network of interactions between the eleven proteins that is absolutely essential for the CMG formation.

An α-helix at the C-terminus of Psf1 is essential for CMG formation

To further narrow down which residues within the last 63 amino acids of the C-terminus in Psf1 might be crucial for the CMG formation, a series of truncations in the B-domain of Psf1 were introduced by removing the last 32 ($Psf1^{1–170}$), 26 ($Psf1^{1–176}$) or 18 ($Psf1^{1–184}$) amino acids from the C-terminus (Figure 2.1D). The expression of all three Psf1 deletion constructs was confirmed (Figure 2.3A, lanes 3-5). Consistent with the observation that the Psf1 deletion mutant removing the entire B-domain still gets incorporated into the GINS complex, all three Psf1 smaller truncations were capable to associate with the remaining GINS partners into a stable complex and this was tested by pull-down experiments (Figure 2.3B, lanes 3-5). To test if these Psf1 constructs could associate with the remaining ten proteins into the CMG complex, I co-immunoprecipitated co-expressed proteins from whole cell extracts in which all eleven proteins were co-infected using a FLAG-tag on Mcm3 (Figure 2.3C). Such pull-down yields an intact stable CMG complex, unless proteins e.g. Mcm5 or Cdc45 are withheld during the co-expression and, thus, prevent a stable complex formation (Figure 2.3D, lanes 3,4). This biochemical study revealed that all of the Psf1 deletion constructs failed to associate into the CMG complex, in par with the truncation of the complete B-domain (Figure 2.3D, lanes 5-7). These findings indicate that residues within the last 18 amino acids at the carboxy-terminus of Psf1 are absolutely critical for the formation of CMG (Figure 2.3D, lane 7). Analysis of the secondary structure prediction of DmPsf1 by PsiPred showed the presence of an α-helix and a β-sheet within this region (Figure 2.1D).

The helical region is composed of six amino acids with two positively charged, two hydrophobic and two uncharged residues (Figure 2.4A). Alanine substitutions of different amino acids in this helical region were introduced and subsequently tested for their incorporation into the CMG complex (Figures 2.4A-C). The expression level of all the Psf1

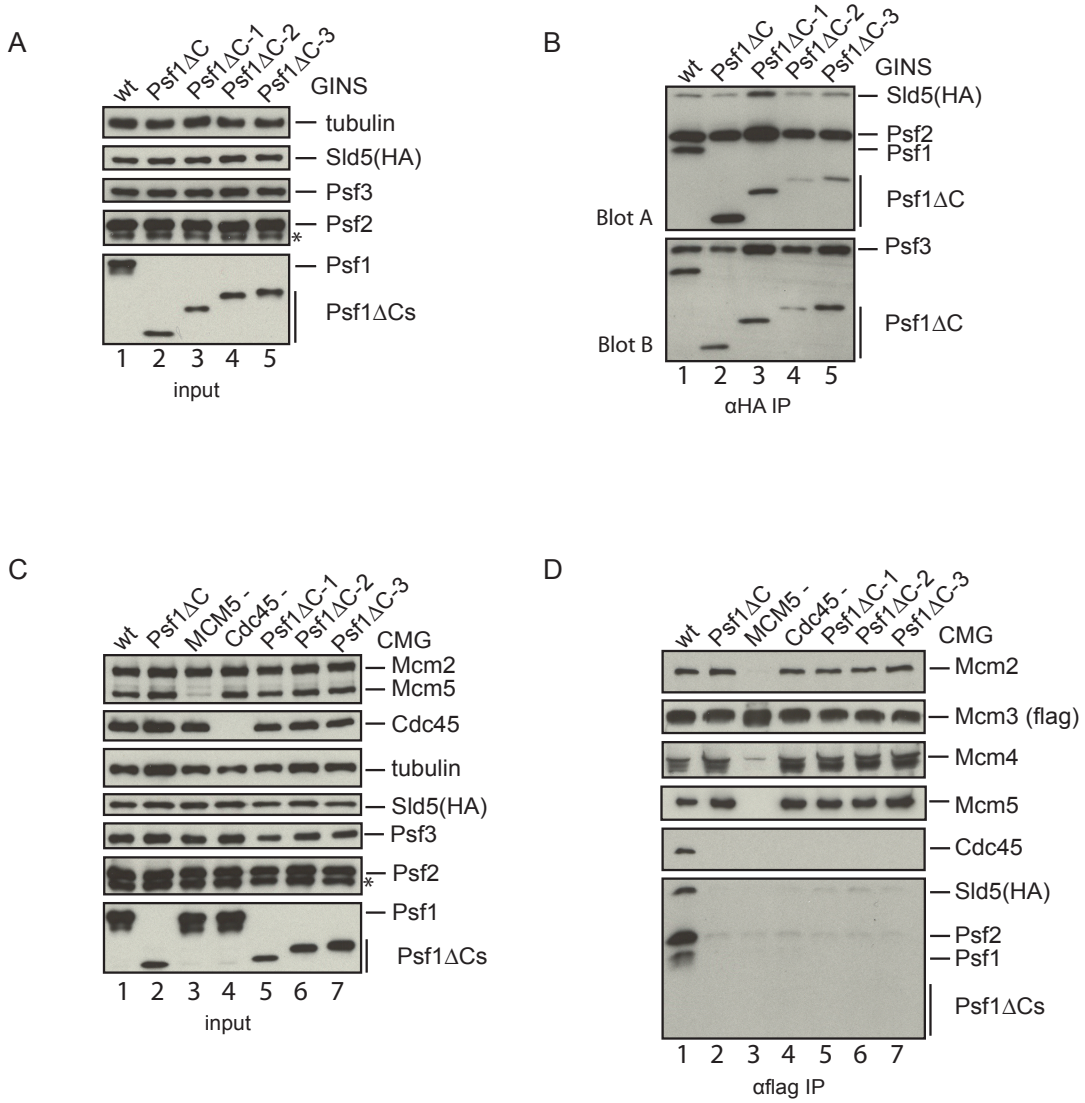

Figure 2.3. Last 18 N-terminal residues of Psf1 are essential for the CMG formation.
(A) Expression levels of individual proteins used for testing GINS formation were detected by antibodies against the four GINS subunits indicated on the right of the immunoblot (asterisk indicates a degradation product of Psf2 or an unspecific band). Loading was controlled by α-tubulin antibody.
(B) IP experiments testing the stability of the GINS complex were performed as described in Figure 2.2B. The formation of the four-membered complex is shown for wt GINS (1), and all the four different C-terminal truncations of Psf1 are incorporated into the GINS complex (2, Psf1$^{1–139}$, named Psf1ΔC; 3, Psf1$^{1–170}$, named Psf1ΔC-1, 4, Psf1$^{1–176}$, named Psf1ΔC-2; 5, Psf1$^{1–184}$, named Psf1ΔC-3).
(C) Expression levels of individual proteins used for testing CMG formation were detected by antibodies against proteins indicated on the right of the western blot (asterisk indicates a degradation product of Psf2 or an unspecific band). Loading was controlled by α-tubulin antibody.
(D) Immunoprecipitation (IP) experiments analyzing the CMG stability: CMG wildtype (1); CMG with various C-terminal truncations in Psf1 (2, 5-7); CMG without MCM5 (3); CMG without Cdc45 (4). Immunoprecipitation of FLAG-tagged Mcm3 subunit yields the intact CMG complex, whereas removal of the entire Psf1 B-domain (Psf1ΔC), smaller truncations of Psf1 B-domain (5, Psf1ΔC-1; 6, Psf1ΔC-2; 7, Psf1ΔC-3) or absence of Mcm5 or Cdc45 proteins, disrupts the CMG formation. The different truncations of Psf1 C-terminus are shown in the schematic diagram in Figure 2.1D.

proteins containing various point mutations within the α-helix was confirmed before subjecting the extracts to co-immunoprecipitations (Figure 2.4B). With the exception of one four-fold point mutant ('Psf1-2'), all remaining single and double point mutations in Psf1 yielded the formation of an intact CMG complex (Figure 2.4C, lanes 4 and 6-9). However, the alanine substitution in E190, L192, V193, and R194 ('Psf1-2') at the same time abrogated the association of Psf1 into the CMG and, thus, formation of the complex (Figure 2.4C, lane 5). Thus, this biochemical analysis revealed four amino acids within the B-domain of Psf1 that are essential for the formation of the eleven-membered complex.

Analysis of the predicted positions of these four amino acids on the homology model of DmPsf1 shows that E190 and R194 are located on the same side of the helix and in such way to point away from the helix and are, thus, likely accessible (Figure 2.4D). In contrast, L192 and V193 are positioned on the internal side of the helix. The biochemical dissection further showed that the simultaneous mutation of the two hydrophobic residues L192 and V193 ('Psf1-6') alone did not abolish the CMG complex formation (Figure 2.4C, lane 9). Also, the alanine substitutions of either R194 alone ('Psf1-3'), or E190 together with L192 ('Psf1-1') in Psf1 still allowed its association into the complex (Figure 2.4C, lanes 4 and 6). Given the position of the two charged residues, it is tempting to hypothesize that E190 and R194 might together present the discrete interactions that are absolutely required for the formation of the CMG helicase complex.

Recent structural studies from an archaeal GINS complex from *Thermococcus kodakarensis* (Tko) reported the crystal structure of the full-length TkoGINS complex for the first time in presence of the corresponding flexible B-domain of Psf1 in the archaeal Gins51 paralog (Oyama et al., 2011). The crystal structure (PDB entry: 3ANW) revealed the position of this region within the GINS complex and shows a counterclockwise shift towards the α-helix-rich (A) domain compared to the homology modeling of DmPsf1 based on Sld5 (Figure 2.4E). Similar to the residues in the helical region of Psf1, the corresponding α-helix of TkoGins51 also displays two charged amino acids (E177 and R173) that point to the outside, with two hydrophobic residues between (Figure 2.4E). The sequence alignment between different species also confirmed a high conservation of the amino acids within this α-helix (Figure 2.1E). The presence of such conserved combination of two charged residues flanking two hydrophobic amino acids further leads me to speculate, that the hydrophobic amino acids in addition to the charged residues (E and R) might be necessary for a stable complex formation. Interestingly, despite of the large interaction surface that has been demonstrated by the single-particle electron microscopy of DmCMG, four amino acids could be identified that establish strong enough contacts that are able to entirely abrogate the formation of the eleven-membered CMG complex if missing.

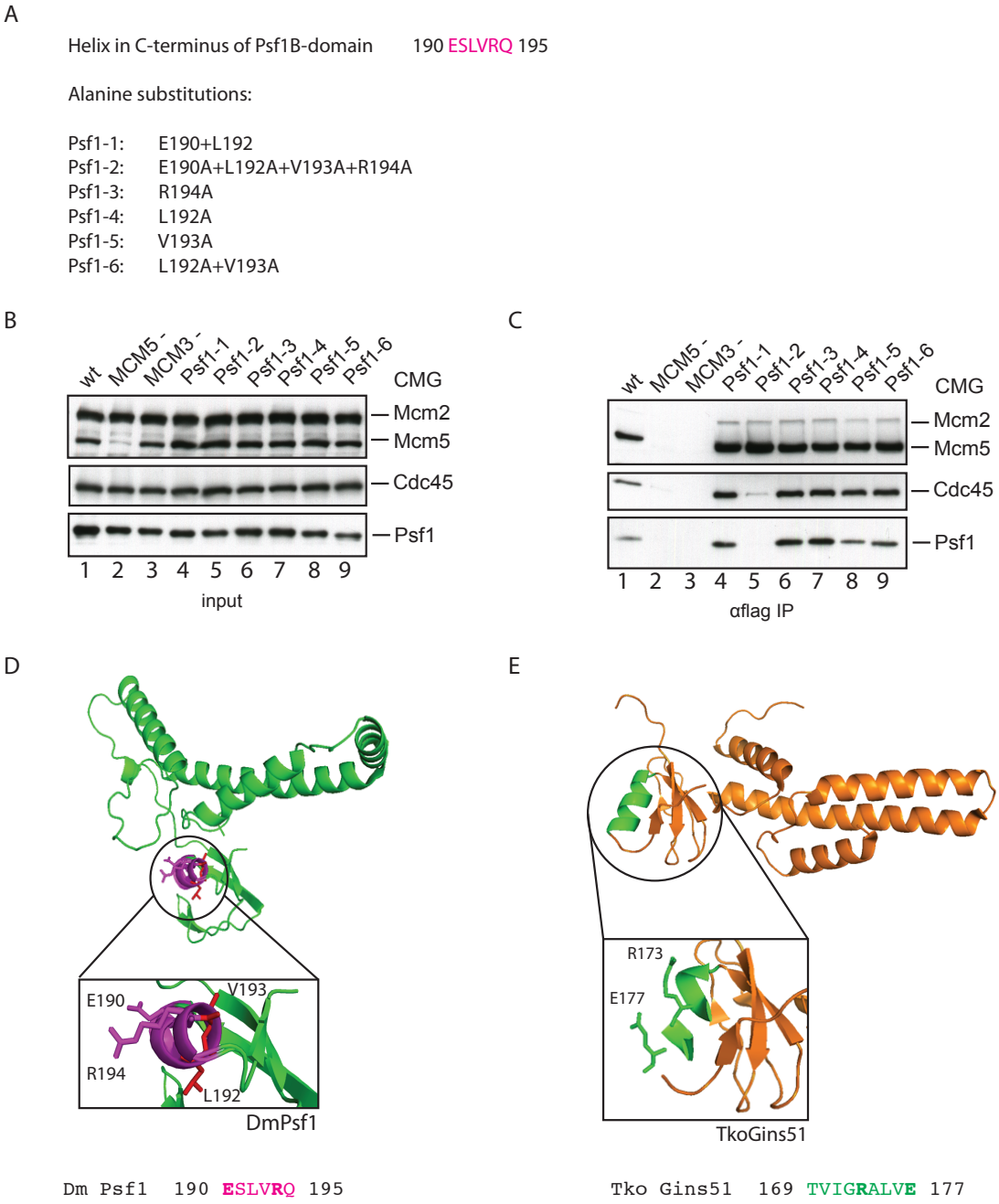

Figure 2.4. An α-helix in the C-terminus of Psf1 is crucial for CMG stability.
(A) Sequence of a predicted α-helix (residues 190-195) located in the last 18 C-terminal amino acids of DmPsf1 is shown. Various alanine substitutions have been introduced in this helical region and are listed below.
(B) Expression levels of individual proteins used for testing CMG formation were detected by antibodies against several CMG proteins indicated on the right of the western blot.
(C) Immunoprecipitation (IP) experiments analyzing the CMG stability: CMG wildtype (1); CMG with various alanine substitutions in Psf1 C-terminal helix (4-9); CMG without MCM5 (2); CMG without Mcm3 (3). Immunoprecipitation of Mcm3 via a FLAG-tag on this subunit (1), and most point mutations in the C-terminal helix of Psf1 yield the intact CMG complex (4,6-9), whereas simultaneous mutation of four helical residues (5), absence of Mcm5 or Mcm3 (bait) protein (2,3), disrupt the CMG formation. The different alanine substitutions of Psf1 are listed in Figure 2.4A.
(D) Homology model of DmPsf1 based on hSld5 as shown in Figure 2.1F with the four critical amino acids in the helix represented as sticks. Two of the residues (E190 and R194, in *magenta*) point

away from helix, while the other two residues (L192 and V193, in *red*) are positioned inwards. The sequence of the α-helix is included below.
(E) Crystal structure of *Thermococcus kodakarensis* (Tko) Gins51 (PDB entry: 3ANW, in *orange*). Encircled is the corresponding B-domain showing its significant clockwise shift compared to the position of this domain in DmPsf1 based on homology modeling after hSld5. Highlighted is the corresponding α-helix (in *green*) and the zoom-in view shows two of the residues (R173 and E177) pointing away from helix. The sequence of the α-helix is included below.

C-terminus of Psf1 is predicted to interact with the N-terminus of Cdc45 within the CMG

These findings raised the interest about the discrete partner in this direct protein-protein interaction between the four amino acids in the α-helix of Psf1 C-terminus and the remaining CMG subunits.

The location for the C-terminus of DmPsf1 could now be more precisely suggested based on its position reported for the Tko Gins51 crystal structure (Oyama et al., 2011). The modeling was adjusted and accordingly docked into the electron density of an 18Å EM structure of the CMG complex (Costa, 2013, manuscript in preparation). The location of the B-domain of Psf1 now indicates its spatial proximity to the N-terminus of Cdc45 (Figure 2.5A) (Costa, 2013, manuscript in preparation). Heretofore, no structural knowledge existed for Cdc45. Only very recent bioinformatics studies unraveled an evolutionary link between Cdc45 and the RecJ protein, a 5'-3' exonuclease, involved in DNA repair mechanisms (described in detail in Chapter 4). The RecJ crystal structure (PDB entry: 1IR6) could be docked into the electron density of Cdc45 in the EM of the *Drosophila* CMG complex. The N-terminal and C-terminal domains of *Drosophila* Cdc45 could be independently modeled; the N-terminal lobe based on its direct sequence homology to RecJ and the C-terminus through its structural homology found between bacterial RecJ and a related exopolyphosphatase (data provided by Alessandro Costa). The homology model of DmCdc45 can be superimposed onto the crystal structure of RecJ in the CMG-EM, and confirmed the spatial vicinity of the N-Cdc45 to C-Psf1 (Figure 2.5A). The structural knowledge about the Cdc45 N-terminus reveals the proximity of the α-helix in C-Psf1 to a specific predicted α-helix in N-Cdc45 (Figure 2.5A, amino acid sequence 'KRAYGEHQGD').

A direct interaction between Cdc45 and GINS was tested in co-immunoprecipitations of both proteins when all four members of the GINS complex were co-expressed with Cdc45 (Figure 2.5B and C). Subsequent reciprocal pull-downs either by GINS (HA-tagged Sld5) or by Cdc45 (α-Cdc45 antibody) independently confirmed the direct contacts between the two members of the CMG complex (Figure 2.5B and C).

The secondary structure of the N-terminus of Dm Cdc45 (amino acids 1-120) was predicted by PsiPred and used as a guideline to introduce a series of N-terminal deletions of Cdc45 (Figure 2.5D). The different truncations are depicted in a schematic diagram and include removal of the first 15 ($Cdc45^{\Delta 1-15}$), 52 ($Cdc45^{\Delta 1-52}$), 65 ($Cdc45^{\Delta 1-65}$), or 99 ($Cdc45^{\Delta 1-99}$) amino acids of the protein (Figure 2.5E).

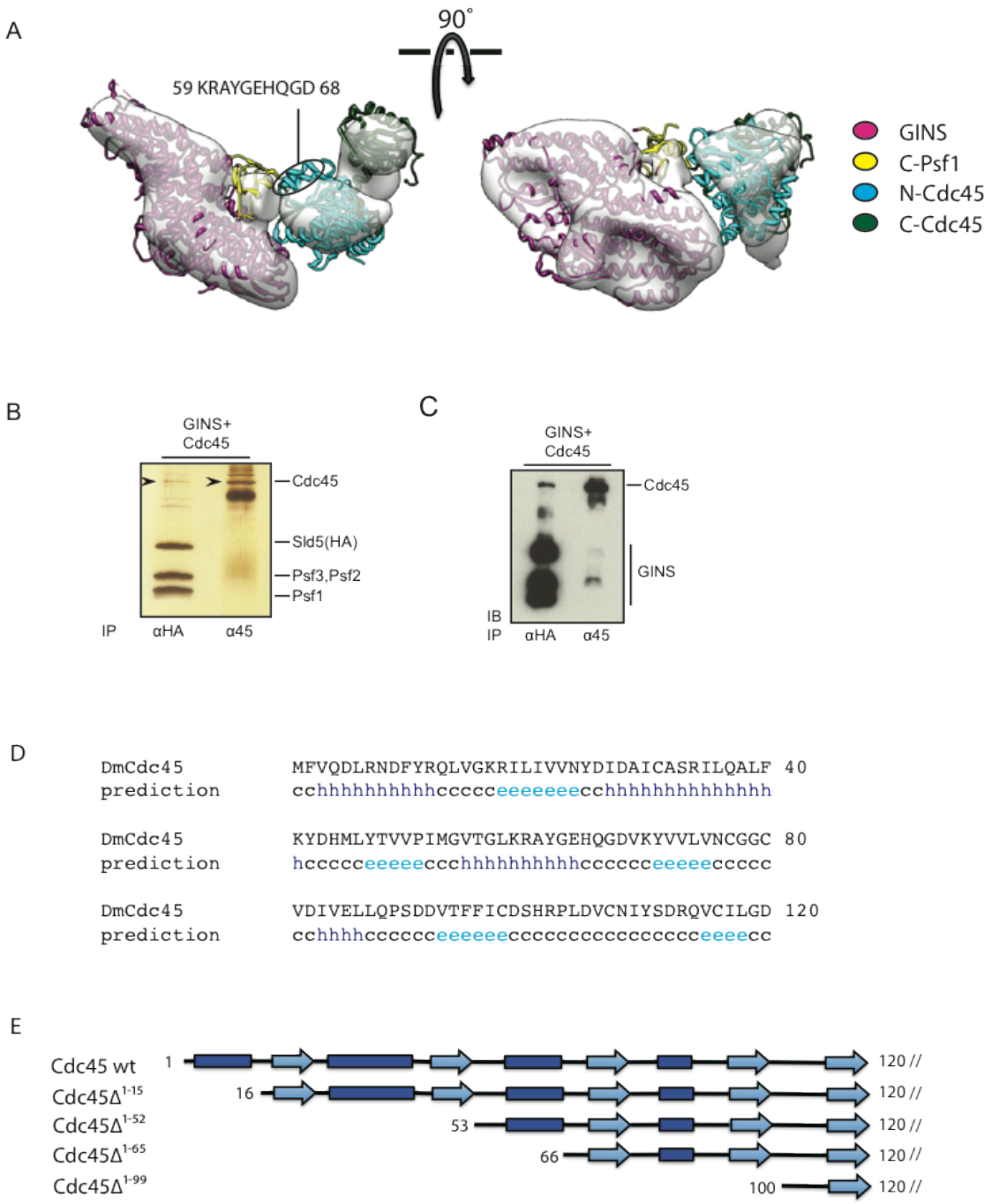

Figure 2.5. GINS interact directly with Cdc45.
(A) 3D reconstruction of the Cdc45-GINS side handle within the CMG complex. The crystal structures of the RecJ horseshoe (N in *cyan*, C in *green,* PDB entry: 1IR6) and of human GINS (in *magenta*, PDB entry: 2Q9Q) are fitted into the single-particle EM reconstruction (18Å) viewed from the AAA+ (*left panel*), and the side view on Cdc45-GINS in front of Mcm5 and Mcm2 of the Mcm2-7 ring (*right panel*). The position of the C-terminus of Psf1 (B-domain, in *yellow*) proximal to N-Cdc45 (*cyan*) is highlighted by overlapping *Thermus kodakarensis* GINS 51 (PDB entry: 1IR6) onto the human Psf1 subunit (Costa, 2013, manuscript in preparation).
(B) Silver-stained SDS-polyacrylamide gel showing the immunoprecipitation (IP) of GINS and Cdc45 proteins. Reciprocal pull-downs through GINS (HA-tagged Sld5) or Cdc45 (α-Cdc45 peptide B antibody) confirm the direct interaction between these two CMG members. Arrowheads indicate the position of Cdc45.

(C) Western blot of the same samples as shown in (A). The immunoblot (IB) was probed with antibodies detecting Cdc45 or one of the four GINS proteins.
(D) The secondary structure prediction obtained by PsiPred of the N-terminal region of DmCdc45 (residues 1-120). Structured regions are indicated in blue (h, α-helix; e, β-strand), and unstructured in black (c, random coil).
(E) Schematic diagram depicting the secondary structure predictions of the first 120 N-terminal amino acids of DmCdc45. The schematic illustrates various N-terminal truncations that have been introduced in Cdc45.

Expression of all Cdc45 deletion constructs was tested, and subsequently its assembly into the CMG complex assayed (Figure 2.6A and B). All Cdc45 mutants were equally well incorporated into the CMG, suggesting that the first 99 amino-terminal amino acids of Cdc45 are dispensable for the complex formation (Figure 2.6B, lanes 2-5). If the removal of the N-terminus in Cdc45 only mildly weakened the interaction to Psf1, the large interaction surface within the eleven-membered complex could promote interactions through other strong protein-protein contacts. Therefore, I wanted to test the direct binding of Cdc45 to the GINS complex. The different Cdc45 deletion mutants were co-expressed with the four GINS proteins and subsequently subjected to pull-downs through the HA-tagged Sld5 (Figure 2.6C and D). The data presented here demonstrate that Cdc45 still could bind to the GINS complex when the first 99 residues were removed (Figure 2.6D, lane 5). Thus, the interaction of C-Psf1 to Cdc45 must lie in a region different from residues 1-99, but still likely within the N-terminus of the protein based on the structural prediction together with the electron microscopic study. Further, it is possible that DmCdc45 undergoes structural re-arrangements when incorporated into the stable CMG complex that cause movements and alignments of specific motifs required for the direct interaction with C-Psf1. Additionally, the stability of the CMG complex containing $Cdc45^{\Delta 1\text{-}99}$ will still have to be tested through the stringency of the biochemical multistep purification (outlined in Chapter 3) to exclude that the co-immunoprecipitation of the intact CMG complex was majorly based on weak interactions. This approach will test the extent to which the interactions might have been altered or weakened upon deletion of the N-terminus of Cdc45. Since the IP reactions were performed in presence of ethidium bromide, I exclude that the protein-protein interactions were promoted by DNA. Taken together, however, the findings suggest, that the amino acids 1-99 of Cdc45 are not absolutely required for the CMG formation and that the direct Psf1-binding region resides further downstream in the protein's amino acids sequence.

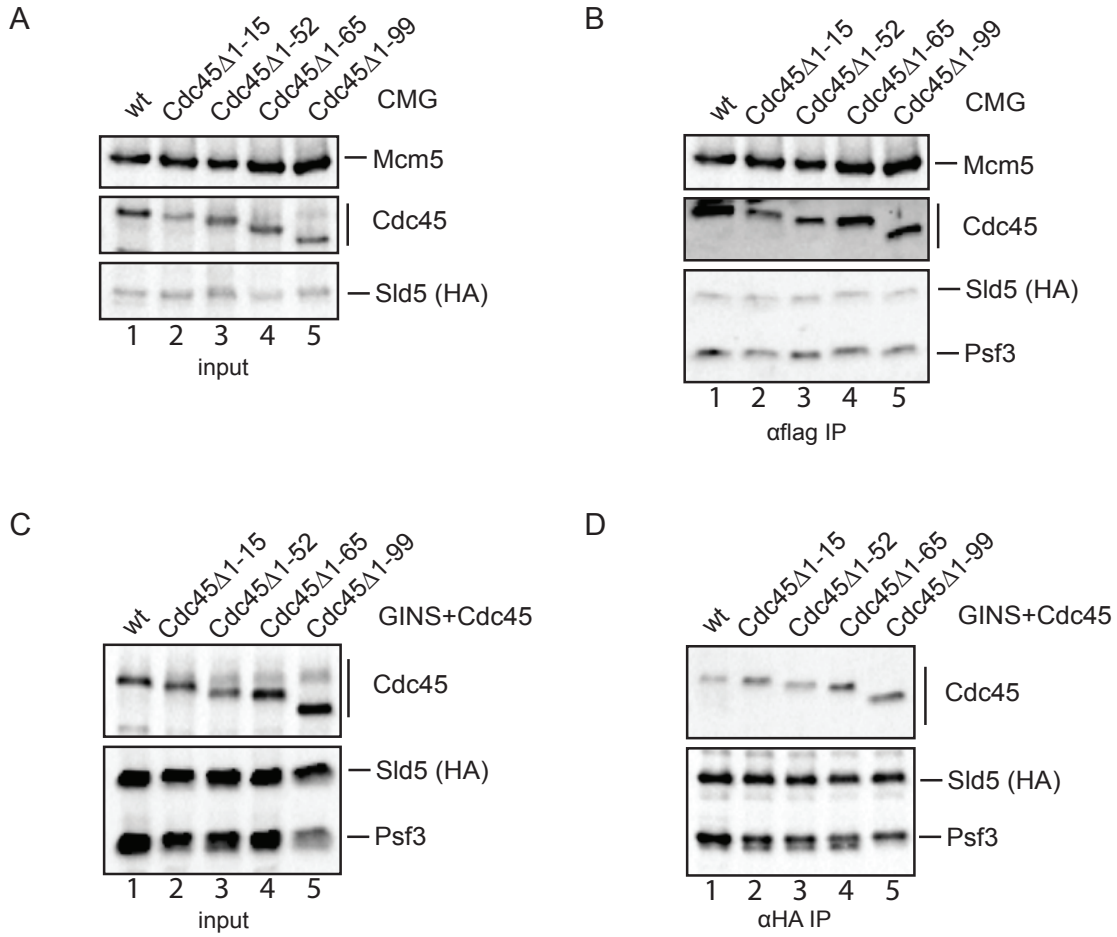

Figure 2.6. The deletion of N-terminus in Cdc45 still allows CMG formation.
(A) Expression levels of individual proteins used for testing CMG formation were detected by antibodies against proteins indicated on the right of the western blot.
(B) Immunoprecipitation (IP) experiments analyzing the CMG stability: CMG wildtype (1); and CMG with various N-terminal deletions of Cdc45 (2-5); Pull-down via FLAG-tagged Mcm3 yields the intact CMG complex for all tested protein constructs (1-5). Proteins were detected with antibodies against several of the CMG members and are indicated on the side.
(C) Expression levels of individual proteins used for testing GINS-Cdc45 interactions were detected by antibodies against proteins indicated on the right of the western blot.
(D) IP experiments analyzing the direct interaction between Cdc45-GINS. IP via HA-tagged Sld5 resulted in intact Cdc45-GINS interaction for wt and all tested truncated Cdc45 protein constructs (1-5). Proteins were detected with antibodies against Cdc45 and several of the GINS members and are indicated on the side.

2.4 Discussion

The focus of the study presented here was to examine residues responsible protein-protein interactions within the eleven-membered CMG complex and, thus, directly to the stability of the complex. The central question of interest was to identify specific contacts that abrogate the formation of the CMG helicase complex upon elimination and that thus lead to an inactivation of the unwinding process in cells.

In eukaryotes, it has been shown that the activation of the pre-assembled Mcm2-7 within the pre-replication complex (pre-RC) requires a complex link of well-controlled events, including the action of two cell-cycle kinases, CDK and DDK, the association of other replication factors, and subsequently the melting of the duplex leading to a significant structural remodeling of the DNA at the origins of replication (ori) (Evrin et al., 2009; Fu et al., 2011; Heller et al., 2011; Ilves et al., 2010; Remus et al., 2009). For *Drosophila*, it was demonstrated that critical to this activation step at the S phase transition is the binding of two auxiliary proteins, Cdc45 and the GINS tetramer, to the Mcm2-7 hetero-hexamer (Ilves et al., 2010). An electron microscopic study of the DmCMG complex reported an intricate network of protein contacts with an extensive interaction surface (Costa et al., 2011). The six Mcm2-7 polypeptides assemble into a toroidal structure with a specific order around the ring, leaving a gap in the center to encompass DNA (Brewster et al., 2008; Crevel et al., 2001; Davey et al., 2003). The EM showed that GINS and Cdc45 associate as a handle to the longitudinal side of the exterior of Mcm2, Mcm3, and Mcm5 (Costa et al., 2011). The finding that Cdc45 interacts with the N-terminus of Mcm2, while GINS bind to N-and C-termini of Mcm3/5 was partially previously observed in biochemical analysis of direct interactions (Ilves et al., 2010; Im et al., 2009).

Here, the stability of the complex was tested by co-immunoprecipitation experiments of the complete CMG complex. For this purpose, all eleven proteins were co-expressed and pulled-down by specific protein tags. In contrast to all biochemical interaction studies so far, in which only pair-wise interactions were tested, this approach for the first time took into account the analysis of interactions within the intact CMG complex. While no robust interactions can be observed between sub-complexes, a stable complex was only found when all eleven proteins were co-expressed simultaneously, suggesting that the most stable contacts are only formed within the CMG. In agreement with this, if individual proteins such as Mcm5 or Cdc45 were not provided during the co-infections with the remaining ten members, neither GINS nor Cdc45 co-immunoprecipitated with the tagged Mcm3 polypeptide.

The data described here demonstrated the importance of the C-terminus (B-domain) of Psf1 for the *Drosophila* CMG stability. Several X-ray crystallographic studies of the human GINS complex solved the structure only in absence of this B-domain in Psf1 and demonstrated a high flexibility of this region, which led to functional implications as potential interactions platform for other factors of the replisome (Chang et al., 2007; Choi et al., 2007; Kamada et al., 2007). In agreement with this, the removal of this region in *Drosophila* Psf1 still yielded an intact tetrameric GINS complex. In contrast to this, the deletion of the C-terminus in Psf1 prevented the formation of the CMG complex, suggesting that some residues in this region are essential for the helicase stability.
Smaller truncations of this domain helped subsequently pinpoint the interaction domain to the last 18 amino acids at the very carboxy-terminus of the Psf1 protein.
Secondary structure predictions identified an α-helical structure at the C-terminus of Psf1 consisting of six amino acids. Alanine scanning of this helix showed that simultaneous mutations of four individual amino acids - two charged and two hydrophobic (ELVR) - accounted for the hindrance of the CMG stability and, thus, complex formation. However, CMG that contained mutations of the two hydrophobic residues alone (L192 and V193) could still be formed. This observation suggests, that E190 and R194 might represent the direct interactions that are essential for the CMG complex. Interestingly, alignment of the B-domain between different species shows the presence of these charged residues (E and R) in the corresponding helical regions, as well as the presence of such a hydrophobic patch between them.

A recent crystallography report for the archaeal *T. kodakarensis* GINS homologue solved the structure of the full-length GINS complex in presence of the flexible C-terminus. My analysis of the sequence of this region showed that such corresponding α–helix in Tko Gins51 subunit also contains the two charged residues, separated by three amino acids, out of which two are hydrophobic (Oyama et al., 2011). The homology model of DmPsf1 that was initially based on the homology to the human Sld5 was now revised and correctly oriented within the electron density map (18Å) of the DmCMG complex according to the published Tko GINS crystal structure (Costa et al., 2013, manuscript in preparation).
Hence, it became evident from the EM study that the C-Psf1 is positioned towards the very N-Cdc45 within the CMG complex and motivated the construction of deletions of various lengths within the very N-terminus of Cdc45 (Figure 2.5A, kindly provided by Alessandro Costa). However, Cdc45 in which the first 99 amino-terminal amino acids were removed was still able to associate into the complete CMG complex. Additionally, studies of pair-

wise protein-protein interactions demonstrated earlier direct contacts between Cdc45 and GINS. Since CMG displays a complicated and extensive network of interactions, the direct interaction of Cdc45/GINS was tested with these N-terminal deletion mutants of Cdc45. Consistent with the observation from the CMG co-immunoprecipitation experiments, the removal of the first 99 residues of Cdc45 did not alter its binding to GINS complex, substantiating that this region is dispensable for the direct Cdc45-GINS interaction. Thus, it is likely that the direct contact surface resides elsewhere in Cdc45 and outside of amino acids 1-99. Given the significant structural re-arrangements within the eleven-membered CMG complex, it cannot be ruled out that structural remodeling during complex assembly might cause movement of more distal residues closer to C-Psf1.

Taken together, this study identified four amino-acids within a α-helix in the C-terminus of Psf1 that are absolutely essential for the formation of the CMG complex. This B-domain was shown to contain a high degree of flexibility an its removal resulted in impairment of DNA replication and could not contribute to a replication complex (Kamada et al., 2007). Functional interactions of GINS and polymerases have been multiply reported and lead to speculation whether particularly the accessibility of B-domain in Psf1, with its here unraveled important role in the CMG stability, could serve as a platform for association of replication proteins further downstream of the unwinding initiation (De Falco et al., 2007; Takayama et al., 2003). One can speculate on residues in close proximity to the ones identified here that ensure a physical link between the helicase and polymerase enzymes necessary to faithfully fulfill the entire replication cycle.
Lastly, the identification of a direct interaction surface that abrogates the formation and activation of the entire eukaryotic CMG helicase complex could represent a powerful novel target in development of future cancer therapeutics by targeting and disrupting those.

2.5 Materials and Methods

Cloning and construction of baculovirus

Viruses used in the studies were constructed following the manufacturer's manual for the Bac-to-Bac expression system from Invitrogen. The pFastBac1 recombination vectors containing the cDNAs for all 11 wildtype CMG subunits were provided by Ivar Ilves and Stephen Moyer. MCM3 and SLD5 genes were tagged with FLAG or HA epitopes at the amino termini through 5'PCR oligonucleotides that inserted the epitopes in frame. Mcm2, Mcm3, Mcm4, Mcm5, Psf1, Psf2, and Psf3 cDNAs were inserted between EcoRI and SpeI restriction sites of the pFastBac1 vector. The cDNA of Mcm6 and Mcm7 was inserted between BamHI and SpeI sites, and Cdc45 and HA-Sld5 cDNA between EcoRI and Xba restriction sites. These vector templates were in general use for generation of CMG mutants through PCR based mutagenesis. Sequencing was used to verify the entire protein coding regions of all generated pFastBac1 constructs.

In detail following specific deletion mutants of either Psf1 C-terminus or Cdc45 N-terminus were designed:

Psf1: $Psf1^{1\text{-}139}$, $Psf1^{1\text{–}170}$, $Psf1^{1\text{–}176}$, $Psf1^{1\text{–}184}$

Cdc45: $Cdc45^{\Delta 1\text{-}15}$, $Cdc45^{\Delta 1\text{-}52}$, $Cdc45^{\Delta 1\text{-}65}$, $Cdc45^{\Delta 1\text{-}99}$

The C-terminal deletions were constructed by introducing STOP codons after the desired residue and removing the remaining C-terminal cDNA. The N-terminal deletions were constructed by removal of the appropriate cDNA and introduction of a START codon in front of the desired residue. The same restriction enzyme sites were used to sub-clone the constructs into the pFastBac1 vector as previously described.

Alanine substitutions in the C-terminus of Psf1 were introduced into following residues (individually or in parallel) by PCR based site-directed mutagenesis:

E190, L192, V193, R194

Preparation of cell extracts of baculovirus infected insect cells

Sf9 or ES-Sf9 cells used for baculovirus amplification were grown in TNM-FH media (AppliChem GmbH) supplemented with 10% fetal bovine serum (HyClone1 Laboratories Inc., Logan, UT) or ESF 921 Insect Cell Culture Medium, Protein-Free (Expression System). Hi5 cells used for protein expression were grown either adherently in serum-free ExCell media (SAFC Biosciences, Lenexa KS) or in ESF 921 Insect Cell Culture Medium. For each desired protein co-expression, 1ml of each virus was added per 150mm tissue culture dish of Hi5 cells (~$2.0 * 10^7$ cells per dish). Proteins were expressed for 72hours and cells were collected, washed once with PBS buffer, and resuspended in hypotonic buffer H (15mM

KCl, 25mM Hepes pH 7.6, 0.02% Tween20, 10% glycerol, 1mM EGTA, 0.4mM PMSF) supplemented with complete protease inhibitor cocktail from Roche Diagnostics. The cell suspension was kept at -80°C until used.

Antibodies and immunoprecipitation experiments

The antibodies against individual CMG subunits have been previously described (Moyer et al., 2006). Monoclonal anti-FLAG (M2) and anti-HA (F-7) antibodies were purchased from Sigma-Aldrich and Santa Cruz Biotechnology, respectively. Mouse monoclonal anti-alpha tubulin antibody (DM1A) was purchase from Abcam for loading controls of total cell lysate. For immunoprecipitation experiments from the cell extract, the infected cell suspension was thawed and cells were broken in a dounce homogenizer. The final concentration of KCl was adjusted to 100mM and the extract was cleared by centrifugation. A total volume of 500µl of protein extract (from 1 x 10^7 infected Sf9 cells) was added to 10µl of antibody beads (either anti-Cdc45-peptide B as described in Moyer et al., 2006, cross-linked to protein A beads; or monoclonal anti-HA agarose (Monoclonal Anti-HA Agarose Conjugate Clone HA-7) or anti-FLAG agarose beads (ANTI-FLAG M2 Affinity Gel) from Sigma-Aldrich) and the binding reactions were incubated rotating for 3 h at 4°C in presence of 50µg/ml ethidium bromide. The resin was washed three times each with 1 ml of wash buffer (1M KCl, 25mM Hepes pH 7.6, 0.02% Tween-20, 10% glycerol, 1mM EDTA, 1mM EGTA, 0.4mM PMSF, supplemented with 1mg/ml BSA, and 1mM β- mercaptoethanol), followed by two washes with 1 ml of buffer A (25mM Hepes pH 7.6, 10% glycerol, 50mM sodium acetate, 10mM magnesium acetate, 0.2mM PMSF, 1mM DTT, 250µg/ml insulin). Proteins bound to the resin were eluted in 25µl glycine buffer pH 2.5 (50mM glycine, 150mM NaCl) and the eluate was immediately neutralized with 2.5µl of 1M Tris HCl, pH 8.0. Unless otherwise specified, 30% of the total immunoprecipitated material was loaded to each lane of the SDS-PAGE gel for analysis by silver-staining or Western blotting. A total of 2µl (out of 500µl) cell-extract was loaded to test the protein expression.

Homology modeling and secondary structure predictions

The Sld5 protein within the crystal structure of the human GINS complex (PDB entry: 2E9X) was used as template to generate a homology model for the *Drosophila* Psf1 protein via SWISS-MODEL (Arnold et al., 2006; Kiefer et al., 2009). The obtained model was used to identify specific residues. Images were created by the PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC. Predictions about secondary structures of Psf1

C-terminus or Cdc45 N-terminus were obtained by using PsiPred. Other models and EM figures were generously provided by Alessandro Costa.

Acknowledgements

I want to thank Alessandro Costa for generously providing Figure 5A showing the docking of human GINS with modeled Psf1 B-domain based on (Oyama et al., 2011), as well as TkoRecJ onto the electron density of CMG.

Chapter 3

ATPase active sites of the Mcm2-7 ring illustrate the functional asymmetry within the CMG complex

3.1 Abstract

Precisely regulated steps in eukaryotic DNA replication require the interaction of different protein complexes at replication origins. Helicase activity, which serves to unwind duplex DNA and create single strands for polymerases, is essential for dividing cells; this activity has long been attributed to the MCM (minichromosome maintenance) proteins. While in archaea six copies of one MCM protein hexamerize to form an active circular helicase, in eukaryotes this activity is attributed to a Mcm2-7 hetero-hexamer. Each of these six Mcm2-7 proteins is essential *in vivo*, suggesting a strict requirement for each in potentially specialized tasks. The subunits of the Mcm2-7 complex each belong to the AAA+ super-family of ATPases, which contain a highly conserved arginine in a short sequence defining the arginine finger element and an invariant lysine in the Walker A consensus sequence. Together, these two motifs, with the lysine from one and the arginine from the adjacent subunit, form active site pairs at the interface of two MCM subunits. While the Walker A box was ascribed the critical function of nucleotide binding, the arginine finger was shown to be important for ATP hydrolysis. As opposed to the archaeal MCMs and yeast Mcm2-7, Mcm2-7 isolated from higher eukaryotes lacks helicase activity *in vitro*. In recent studies, however, the involvement of the tetrameric GINS complex and Cdc45 have both emerged as important factors in the initiation and elongation phases of replication. Their recruitment to the origins of replication promotes assembly with the hetero-hexamer Mcm2-7 and forms an eleven-membered complex known as CMG (Cdc45/Mcm2-7/GINS). Previous work from the Botchan lab has shown that the CMG complex serves as the activated eukaryotic helicase. By introducing alanine substitutions into the conserved arginine finger motif of various Mcm2-7 subunits, and thus eliminating the hydrolysis activity of individual active sites, I was able to demonstrate a functional asymmetry of the Mcm2-7 ring within the CMG complex. Mutation of arginine in active sites of one-half of the Mcm2-7 ring entirely abolished the helicase activity whereas other mutations show only mild defects. Elimination of both nucleotide binding as well as subsequent ATP hydrolysis further suggest an unequal contribution of the Mcm2-7 subunits to DNA binding of CMG.

3.2 Introduction

During each S phase of the cell cycle, a single copy of the entire genome needs to be synthesized before the genetic material is segregated to daughter cells during mitosis. This requirement is universal for all proliferating cells and must be precisely executed and tightly controlled. In eukaryotes, DNA replication is initiated when a multimeric complex, called the pre-replication (pre-RC) complex, is formed on chromatin. Many studies have contributed to our current model of replication initiation and our understanding of the order of events at a molecular level. The core of this multi-protein complex is comprised of the Mcm2-7 proteins, a hetero-hexameric helicase motor that provides the unwinding function at the replication fork. Once unwinding of the duplex DNA is achieved, the cell is committed to DNA replication. Thus, the loading and activation of a helicase represents an early step in the generation of a replication fork (Kornberg and Baker, 1992). The highly regulated chain of assembly and activation events is initiated by the recognition and subsequent binding of the six subunit ATPase Origin Recognition Complex (ORC, ORC1-6) to origins of replication (ori). Together with Cdc6 and the loading factor Cdt1, ORC then helps recruit a dimer of Mcm2-7 to replication origins, resulting in the formation of the inactive pre-RC (Aparicio et al., 1997; Bell and Dutta, 2002; Bell and Stillman, 1992; Tanaka et al., 1997). These events take place during G1 phase of the cell cycle and lead to the deposition of a head-to-head double Mcm2-7 hexamer on double-stranded DNA (Evrin et al., 2009; Remus et al., 2009). The ordered actions of the two key cell-cycle dependent kinases DDK (Dbf4-dependent kinase) and CDK (cyclin-dependent kinase) have been recently shown to lead to the association of further replisome factors and eventually to helicase activation (Heller et al., 2011; Sclafani and Holzen, 2007; Stillman, 2005). While the full replisome consists of many proteins, only a subset of these are attributed the function of an active eukaryotic DNA helicase. This active helicase, known as the CMG complex, is comprised of one copy each of Cdc45, Mcm2-7, and GINS proteins (Aparicio et al., 2006; Ilves et al., 2010; Moyer et al., 2006; Pacek et al., 2006) with the MCMs representing the motor of the enzyme.

Unlike in eukaryotes, where the hexameric Mcm2-7 ring displays no helicase activity before its activation by the allosteric factors Cdc45/GINS and their incorporation into a higher-order multi-protein complex, helicase activity in archaea derives from the MCM motor alone. There, we find a simpler version of the DNA helicase, where six copies of one MCM protein homo-hexamerize to form the enzymatically active complex.

The MCM protein family was originally identified through a genetic screen in *Saccharomyces cerevisiae* for defects in minichromosome maintenance, and early characterization of mutant phenotypes suggested roles in DNA replication (Maine et al.,

1984). Unlike archaea, where six copies of one MCM protein oligomerize, the eukaryotic MCM complex is composed of six homologous, but different proteins (Mcm2-7), each containing putative nucleotide binding and hydrolysis motifs (Koonin, 1993). The Mcm2-7 subunits represent a subgroup of the AAA+ family of proteins, which, as their name indicates, are adenosine triphosphatases (ATPases) with various cellular activities (Neuwald et al., 1999). The MCM proteins couple energy from the nucleotide hydrolysis with conformational changes, to initiate translocation on DNA. A common feature of the AAA+ proteins is a stretch of 200 amino acids that represent the ATP binding domain (Hanson and Whiteheart, 2005). The Mcm2-7 proteins possess a version of this characteristic domain, and this so-called MCM box includes the ATPase consensus motifs, known as the Walker A and Walker B boxes and 'arginine finger' motifs. The P-loop of the Walker A active site interacts with the phosphates of ATP, more precisely through an invariant lysine residue within the MCM-specific consensus sequence GDPxx(S/A)KS (Forsburg, 2004). Mutation of this critical lysine leads to the abolishment of ATP binding and, thus, to inactivation of the protein. This is consistent with observations made in *Drosophila* Mcm2-7 ring within the CMG complex, as well as in yeast Mcm2-7, upon substitution of this conserved residue with alanine (Ilves et al., 2010; Schwacha and Bell, 2001). The Walker B motif D(D/E) resides in MCM proteins within the conserved sequence IDEFDKM, where the aspartate is implicated in coordination of a catalytically important Mg^{2+}-ion and the glutamate in activation of the water molecule for the ATP hydrolysis reaction (Iyer et al., 2004). The consensus motif SRFD defines the arginine finger. When the critical arginine residue in this motif is mutated, ATP hydrolysis activity is impaired (Hanson and Whiteheart, 2005; Ogura et al., 2004).

Even though each Mcm2-7 subunit possesses all motifs required for binding and hydrolysis of ATP, no helicase activity can be observed with isolated Mcm2-7 proteins (Davey et al., 2003). Another striking hallmark of AAA+ proteins is that the location of the active sites resides at the interface of two subunits and that the direct participation of motifs in both proteins is necessary for ATP hydrolysis. In Mcm2-7, the ATPase active sites are formed by the Walker A and B motifs from one subunit and the arginine finger provided in *trans* by another adjacent subunit.

Unlike the homo-multimeric archael MCM helicase, the eukaryotic MCM motor is composed of six essential homologous proteins that need to assemble in a unique way to form the active ATPase sites. Genetic analyses of yeast Mcm2-7 have suggested that the correct assembly of the six proteins into a complex is essential for its function *in vivo* (Lei et al., 2002; Sherman et al., 1998). This requirement makes this complex a unique model system for studying the mutational consequences when the ATPase motifs are disrupted.

As opposed to homo-hexameric helicases, where each of the six active sites has an equal contribution, the ATPase active sites of the Mcm2-7 may be functionally nonequivalent. To understand the asymmetry of the eukaryotic helicase, interaction studies of human MCMs as well as extensive two-hybrid screens with *Drosophila* MCMs have helped identify the nearest neighbors within the Mcm2-7 ring (Crevel et al., 2001; Su et al., 1996; Yu et al., 2004). Further, reconstitution experiments with the *S. cerevisiae* Mcm2-7 complex have unraveled certain pairs of Mcm2-7 subunits that have ATPase activity; such pairs were Mcm3/7, Mcm4/7 and Mcm2/6, and these approaches together resulted in the proposal of a unique order of subunits in the intact Mcm2-7 complex (Figure 3.1)(Davey et al., 2003; Schwacha and Bell, 2001).

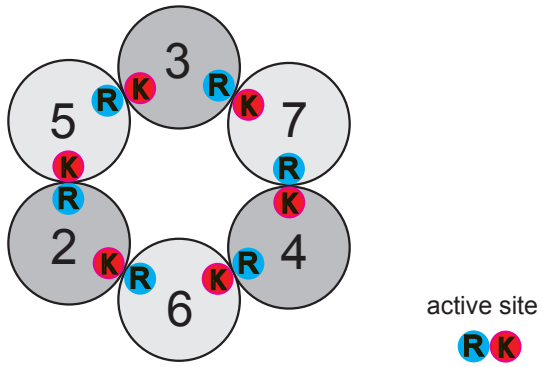

Figure 3.1 Model of the Mcm2-7 hetero-hexamer.
This figure shows the model adapted and changed from Davey et al., (2003) and depicts the unique order of the six Mcm2-7 subunits when circularized. The circularization is necessary for activity, and the figure shows the six ATPase active sites between the interfaces of two neighboring subunits. Each active site consists of the Walker A motif with its crucial lysine (K), which is necessary for ATP binding, and the arginine finger with its critical arginine (R), protruding from the next subunit, needed for nucleotide hydrolysis.

Subsequent *in vitro* studies of ATPase activities with mutants in ATP binding and hydrolysis elements in the different MCM subunits helped provide significant insights into the mechanistic link between the harvest of energy from ATP hydrolysis and subsequent induction of conformational changes (Schwacha and Bell, 2001).

Reports of the large tumor antigen (LTag) of simian virus 40 (SV40) have suggested a concerted model for its nucleotide occupancy mode and for simultaneous binding and hydrolysis of ATP by the six active sites (Gai et al., 2004). In contrast, in another Superfamily III (SF-III) helicase - E1 of papillomavirus - each individual subunit undergoes cycles of binding, hydrolysis and release sequentially, while the DNA-binding loops guide the ssDNA through the central channel (Enemark and Joshua-Tor, 2006). However, studies of MCM complexes and MCM-associated complexes (e.g. CMG) in archaea as well as in

eukaryotes have not yielded clear evidence for the possible mechanistic translocation steps of these Superfamily VI (SF-VI) helicases.

Although the precise link between ATP hydrolysis and translation into generation of movement is not understood at a mechanistic level, studies with budding yeast Mcm2-7 ATPase activities have proven that not all subunit pairs contribute equally to ATP hydrolysis (Bochman et al., 2008; Davey et al., 2003; Schwacha and Bell, 2001).

While archael MCM complexes were demonstrated to possess helicase activity, a similar *in vitro* unwinding activity seemed to be lacking from the eukaryotic Mcm2-7 complexes (Chong et al., 2000; Davey et al., 2003; Ishimi, 1997; Kelman et al., 1999; Lee and Hurwitz, 2001; Schwacha and Bell, 2001; Shechter et al., 2000).

In contrast, a sub-complex consisting of Mcm4,6,7 subunits possesses weak helicase activity *in vitro* and for a long time it was believed that this half of the helicase ring constitutes the active center of the MCM motor while Mcm2,3,5 were required for regulatory roles (Bochman et al., 2008; Ishimi, 1997; You et al., 2003; You et al., 1999). The paradox between these *in vitro* observations and *in vivo* studies implicating that the full Mcm2-7 hetero-hexamer is the true helicase in cells was finally resolved by a recent report of studies done with yeast Mcm2-7 that demonstrated *in vitro* helicase activity of the purified complex (Bochman and Schwacha, 2008). This biochemical study concluded that a discontinuity between subunits Mcm2 and Mcm5 accounted for the lack of helicase activity in previous experiments. This ATP-dependent 'gate' showed strong anion-dependence and only specific reaction conditions induced a closure of the gate, which was required for its unwinding activity. These findings support the view that a complete circularization of the Mcm2-7 toroid is necessary for its helicase activities. Such a gap is not present in the Mcm4,6,7 sub-complex and could explain why the dimerization of this (4-6-7)-trimer displays weak helicase activity. Helicase activity of the Mcm2-7 in higher eukaryotes was still absent, suggesting a possible requirement for additional factors or re-arrangement of the Mcm2-7 ring.

Both the intrinsic ATPase as well as helicase activity of *Drosophila* Mcm2-7 was strongly activated by the association of Cdc45 and GINS into the Cdc45/Mcm2-7/GINS (CMG) complex (Ilves et al., 2010). These cofactors themselves lack any canonical ATPase motifs and have been suggested to induce the activation of the Mcm2-7 motor within CMG through its allosteric remodeling and optimization of subunit interactions that then enhance the rates of ATP hydrolysis. Introduction of mutations in the ATP binding element Walker A of single Mcm2-7 subunits allowed the dissection of the contributions of each subunit to the overall ATPase and helicase activities of the CMG complex. In agreement

with what was seen in yeast Mcm2-7, the *Drosophila* Mcm2-7 within the CMG complex also displays a functional asymmetry with non-equal contributions of all six active sites (Ilves et al., 2010). So far, nothing is known about the DNA binding contributions of each Mcm2-7 subunit within the activated eukaryotic CMG helicase, and it remains an important unanswered question in the field. Introduction of mutations in the canonical ATPase consensus motifs also represents a powerful tool to further analyze the participation of the different subunits with respect to DNA binding.

3.3 Results

Eukaryotic helicases possess six different MCM subunits

Eukaryotic Mcm2-7 rings are formed by circularization of six different homologous MCM proteins (Mcm2-7) in a unique order (Figure 3.1) (Davey et al., 2003). These proteins represent the motor of the activated eukaryotic CMG (Cdc45/Mcm2-7/GINS) helicase. Alignment of all *Drosophila* MCM proteins shows a strong sequence homology between the different subunits. These proteins belong to the AAA+ superfamily of ATPases with a N-terminal domain, a AAA+ domain and a C-terminal region (Figure 3.2).

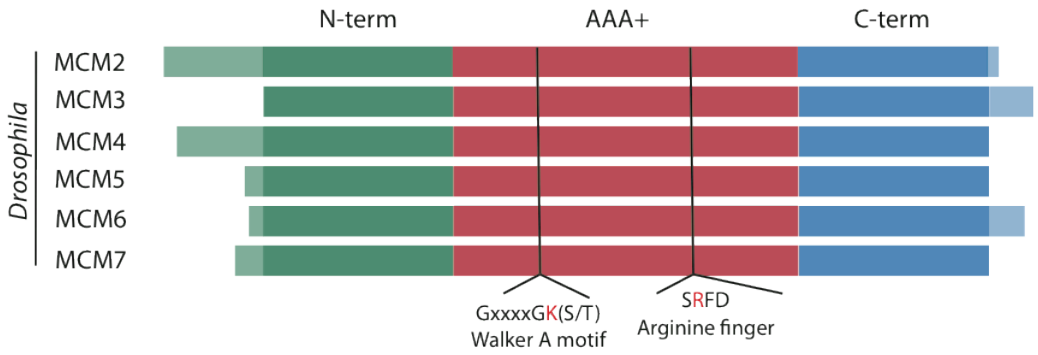

Figure 3.2. Schematic of *Drosophila* Mcm2-7 alignment.
Schematic representation of a multiple primary sequence alignment of all six *Drosophila* Mcm2-7 subunits using ClustalW2. The N-terminal (*green*), AAA+ (red) and C-terminal (*green*) domains are color-coded. The relative locations and consensus sequences of the ATPase active site motifs Walker A and arginine finger are indicated. Compared to archael MCM proteins, any N-terminal (*light green*) and C-terminal (*light blue*) extensions in each of the *Drosophila* Mcm2-7 proteins are depicted in the schematic.

The AAA+ domain contains all motifs required for ATPase activity, amongst which an invariant lysine in the Walker A box was shown to be essential for nucleotide binding while a conserved arginine from the arginine finger motif of the neighboring subunit protrudes to hydrolyze the bound ATP (Hanson and Whiteheart, 2005; Neuwald et al., 1999; Ogura et al., 2004). The central question of interest then is why eukaryotes have evolved to posses six different genes encoding six different, but related, MCM proteins, instead of only one MCM protein of which six copies hexamerize to form the active MCM helicase as in archaeal organisms.

Unlike archaeal MCMs in which the six ATPase active sites in the six identical MCM proteins presumably contribute equally to the biochemical activities of the helicase complex, the eukaryotic Mcm2-7 represents an ideal system to dissect the precise

contributions of each ATPase pair to the activity of the entire complex. By manipulating specific residues of interest, we can identify individual contributions and understand if the six proteins have developed specialized primary roles.

Previous work has shown unequal roles for the different active site pairs with regard to intrinsic ATPase as well as helicase activity for *Saccharomyces cerevisiae* Mcm2-7 and the *Drosophila* CMG helicase. While mutational defects in various ATPase elements of yeast Mcm2-7 have been assayed, the only mutational effect reported for CMG is the introduction of an alanine substitution of a conserved lysine in the Walker A motif (Bochman and Schwacha, 2007, 2008; Ilves et al., 2010; Schwacha and Bell, 2001). These mutations were termed 'KA' for the substitution of the invariant lysine to alanine.

The elimination of the conserved lysine required for the Walker A box to bind ATP in each Mcm2-7 subunit in CMG leads to the significant ablation of ATPase and helicase activities (Ilves et al., 2010). The extent to which these activities were affected varied strongly among the six active sites, suggesting that some sites were more important for the overall activities than others, at least under the chosen experimental conditions *in vitro*. These observations are consistent with the reported ATP-dependent DNA binding of CMG. Unlike Mcm2-7 alone, do subunits in the Mcm2-7 ring show functional asymmetry that is congruous with our previous findings when we allow nucleotide to bind, but prevent their subsequent hydrolysis?

Purification of recombinant Drosophila CMG (Cdc45/Mcm2-7/GINS) and Mcm2-7 complexes containing a mutation in the arginine finger motifs

The data reported in Ilves et al., (2010) suggested that the intrinsic ATPase and helicase activities of the Mcm2-7 proteins were significantly enhanced upon binding of the cofactors Cdc45/GINS, factors that lack any canonical ATP binding motifs themselves. In a study analogous to that with the CMG KA mutants, I wished to test the direct contribution of each Mcm2-7 subunit arginine finger on ATPase activity. To analyze if certain ATPase active sites display a more important role in the overall activity of the complete complex, I sought to understand the effects on the CMG helicase if I eliminated the critical arginine residue in the arginine finger motif, known to be essential for ATP hydrolysis. Each of these arginine fingers, together with the Walker A motif from the adjacent MCM subunit, form an active ATPase site pair. Mutations of the arginine finger should not affect ATP binding, with the nucleotide trapped in the binding pocket, while its hydrolysis should be abolished. Based on findings of the critical ATPase pairs from the Walker A motif studies in CMG shown by Ilves et al., (2010). I sought to understand if there existed a similar

pattern of defects around the ring. During the experimental studies for my diploma thesis, I had introduced point mutations in the six Mcm2-7 subunits that substitute the invariant arginine with alanine and named them 'RA'. The helicase activity of two CMG complexes with mutations in the arginine finger motifs of Mcm2 and Mcm7were assayed. To obtain the full picture of the level of defects, caused by mutation of each arginine finger, I wished to purify and assay six recombinant CMG complexes that each contained one RA mutant Mcm2-7 subunit in presence of the five remaining wildtype Mcm2-7 proteins.

Ilves et al., (2010) have successfully worked out and described the protocol to purify recombinant *Drosophila* CMG and Mcm2-7 complexes. I followed the established chromatographic steps to reconstitute CMG and Mcm2-7 proteins for my subsequent mutant studies. An overview of these steps is outlined in Figure 3.3A, and most steps corresponded to the published CMG purification protocol. Acetate-containing buffers were generally used in place of chloride anions during the chromatographic steps of Mcm2-7 reconstitution. This change was found to stabilize the Mcm2-7 ring and fully separate incomplete MCM sub-complexes (e.g. Mcm4/6/7) from the full hetero-hexamer.

To confirm that my purified complexes behaved as described by Ilves et al., (2010), I ran the wildtype CMG complex through a 15-35% glycerol gradient centrifugation. Using protein standards of known molecular weight, I was able to confirm that my protein preparation also consisted of a single homogenous population (Figure 3.3B and C). No higher-order complexes that could indicate formation of aggregates were present in the fractions. Figure 3.3C shows that the banding pattern of the silver-stained CMG proteins from the glycerol gradient is consistent with the eleven members within the CMG complex as previously reported (Ilves et al., 2010; Moyer et al., 2006).

The ATPase active sites of Mcm2-7 illustrate a functional asymmetry

Following the modified protocol, I purified all six CMG complexes containing mutations of the conserved arginine in the arginine finger element ('RA') of each of the six Mcm2-7 subunits at a time (Figure 3.3E and D). Each of the CMG complexes had identical properties during the purification steps and displayed the same 1:1:1 Cdc45:Mcm2-7:GINS stoichiometry as described for wildtype CMG (Ilves et al., 2010). Additionally, I wished to compare the ATPase defects of the RA mutations in the *Drosophila* Mcm2-7 complex in isolation to those caused by the same mutations within the CMG complex. The single particle electron-microscopic study from Costa et al., (2011) compared the structures of free *Drosophila* Mcm2-7 to those within the CMG complex. The report concluded that upon incorporation into the CMG complex there are striking structural re-arrangements within

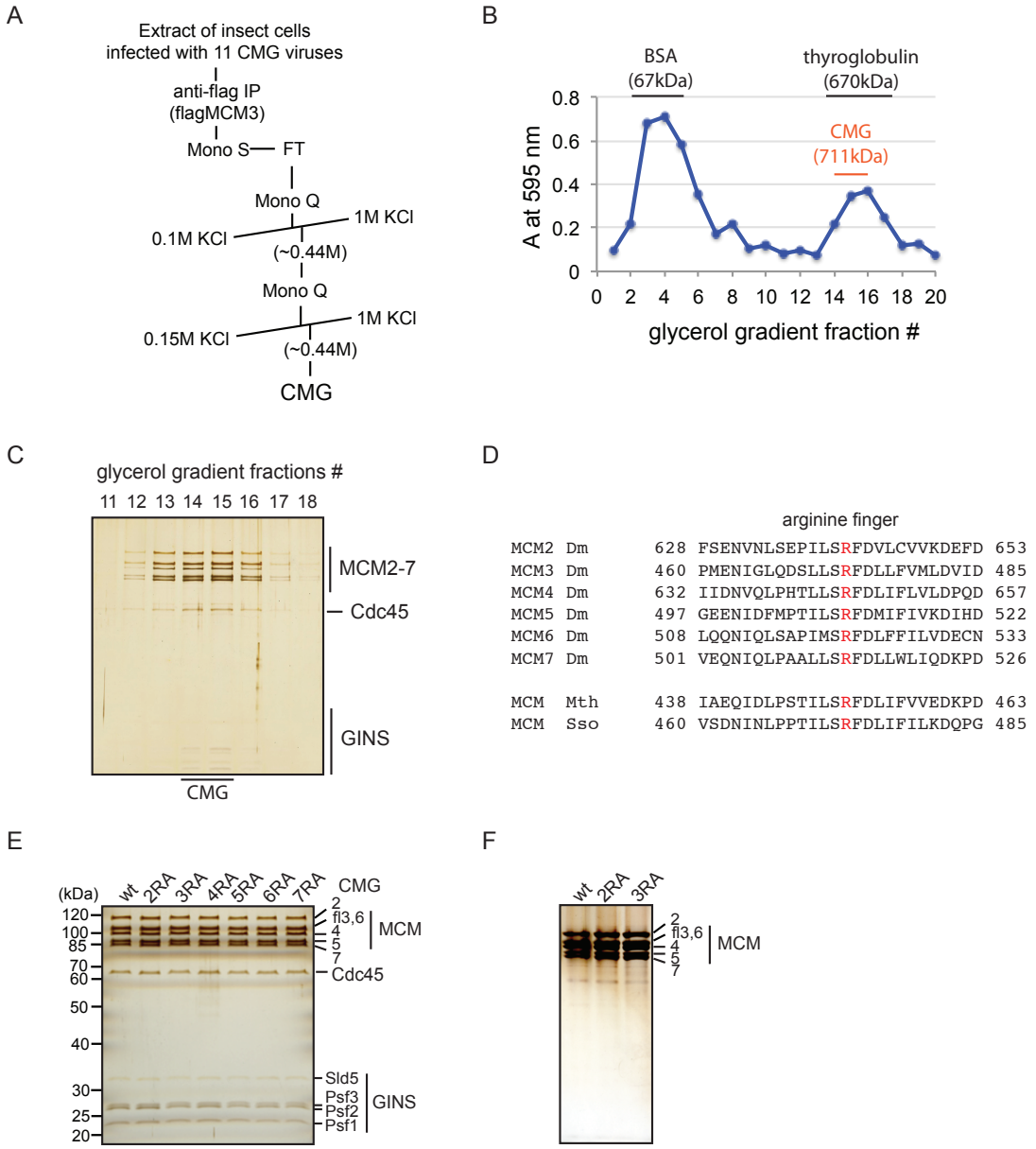

Figure 3.3. Purification of recombinant *Drosophila* CMG complexes.
(A) Schematic of established purification protocol for CMG complexes according to Ilves et al, (2010). The purification includes sequential immunoaffinity and ion-exchange chromatography steps starting from cell extracts of insect cells that have been infected with eleven viruses for CMG.
(B) Fractions of glycerol gradient centrifugation (15-35%) of the two standard proteins BSA (67kDa) and thyroglobulin (670kDa). Plotted is the Absorption (A) value at 595nm measured in the Bradford assay of the 20 collected fractions. Fractions containing CMG (711kDa) are indicated in red. The purification of Mcm2-7 complexes following this general scheme, but contains different buffers (see Material and Methods).
(C) CMG proteins from glycerol gradient fractions were separated by SDS-PAGE (10%) and silver-stained.
(D) Shown is the primary sequence alignment of the arginine finger motifs of all six *Drosophila melanogaster* (Dm) Mcm2-7 proteins and MCM of *Methanobacter thermoautotrophicus* (Mth) and *Sulfolobus solfataricus* (Sso). The arginine (R in *red*) has been substituted by an alanine and the complexes containing these mutations were termed 'RA'.
(E) Purified recombinant CMG wildtype and arginine finger 'RA' mutants were separated on a 10% SDS polyacrylamide gel and silver-stained.
(F) Purified Mcm2-7 wt, '2RA' and '3RA' complexes were separated by SDS-PAGE (10%) and proteins were silver-stained for visualization.

the Mcm2-7 ring that then favor a tighter engagement of the DNA strand through the central channel.

More importantly, however, the association of Cdc45/GINS and subsequent nucleotide binding close a discontinuity between subunits Mcm2 and Mcm5 within the free Mcm2-7. The existence of this 'Mcm2/5 gate' was initially proposed based on biochemical studies of yeast Mcm2-7 (Bochman et al., 2008; Bochman and Schwacha, 2008). I was first interested in understanding if the observed functional asymmetry within the Mcm2-7 ring of the CMG complex with respect to ATPase activity was altered in the Mcm2-7 complex alone. The presence of the '2/5 gate' in the isolated Mcm2-7 ring would suggest that the defects caused by mutation of the ATPase active sites around the gap would be different from the effects of such mutations when the gap is closed.

To test this hypothesis, I chose to purify two Mcm2-7 complexes with mutations in the ATP hydrolysis motifs (arginine finger) in the two active sites at different locations on the Mcm2-7 ring. All mutants in the arginine finger motifs were named 'RA'. The two active sites chosen for my analysis showed different defects in helicase activity when the corresponding 'KA' mutations were assayed. As mutants of interest, I first chose to mutate the arginine to alanine in Mcm2 ('RA'). This arginine finger motif forms the second half of the active site with the P-loop in the Walker A element of Mcm5, thus forming the '2/5 active gate'. The elimination of the critical lysine in the Walker A motif (KA) of Mcm5 entirely abrogates DNA unwinding (Ilves et al., 2010). The second mutant I wished to assay was Mcm3RA, which together with Mcm7, forms another active site. Elimination of the lysine in the Walker A motif of Mcm7 resulted in a moderate reduction of CMG helicase activity (Ilves et al., 2010). Both Mcm2-7 RA complexes formed with identical biochemical properties to wildtype Mcm2-7 and showed equal stoichiometry (Figure 3.3F).

The introduction of RA mutations in Mcm2 and Mcm3 resulted in defects in the ATPase activity of both CMG as well as the free Mcm2-7 complexes, as indicated by the reduced V_{max} values (Figures 3.4A and B). The kinetic measurements yielded V_{max} values for the CMG RA mutants (± standard error) expressed as percentages of wt CMG: 31.2 ± 6.3% for CMG 2RA and 35.5 ±7.3% for CMG 3RA. If all active sites contributed equally to the activity, one would expect that elimination of one site would result in $1/6^{th}$, or 17% of the overall activity of the complex. However, the alanine substitution in the Mcm2 and Mcm3 subunits each resulted in a decrease in activity of more than 17% of the wt CMG, suggesting that the Mcm2-7 motor does not consist of six independent and equally contributing ATPase motifs. Compared to the CMG complexes containing Walker A mutations in these two corresponding active site pairs (Mcm5KA and Mcm7KA), the extent of defects in ATPase activity was roughly equivalent. Surprisingly, the comparison

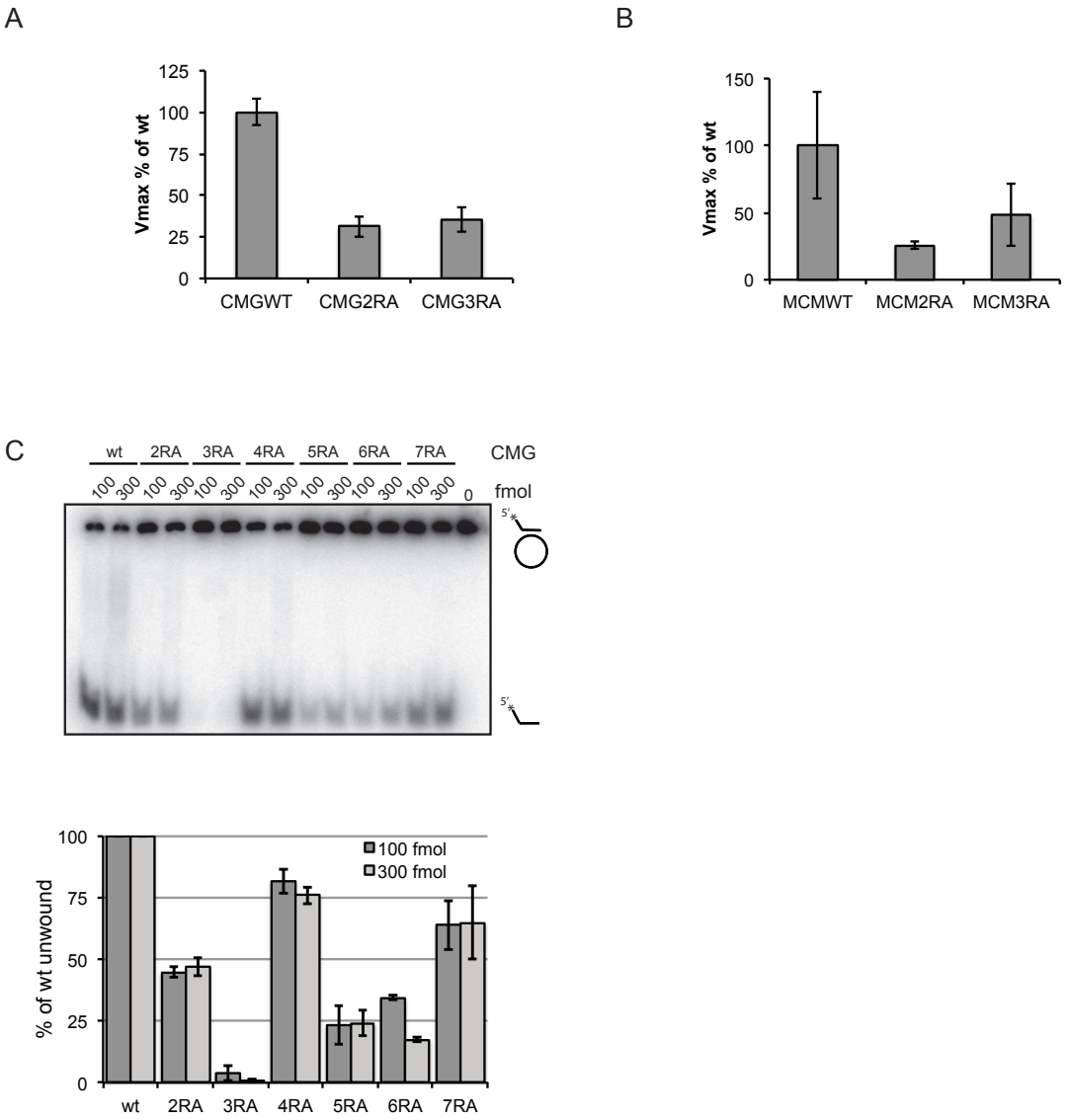

Figure 3.4. Functional asymmetry of arginine finger mutant CMG complexes.
(A) Quantification of ATPase activity assays of CMG wt, 2RA and 3RA. The x-axis indicates the protein complex measured and the y-axis the V_{max} (1/min) values expressed as percentage of CMG wt. Standard deviations of two independent series are included.
(B) Quantification of ATPase assays of Mcm2-7 wt, 2RA and 3RA complexes. The protein complexes are plotted against V_{max} (1/min) values as percentages of CMG wt with standard deviations of two independent series.
(C) Helicase activity assay of all CMG RA complexes on a circular substrate (*top panel*). The average values of different CMG complexes are plotted against the activity measurements indicated as percentages of wt CMG (*lower panel*) and include the standard deviations of two independent series. The unwinding activity was measured with two protein concentrations (100 and 300fmol).

of Mcm2RA and Mcm3RA in free Mcm2-7 versus in the CMG complexes resulted in similar levels of abolishment of ATPase activity.

Kinetic measurements yielded V_{max} values for the Mcm2-7 RA mutants (± standard error) expressed as percentages of wtMcm2-7: 25.6 ± 2.6% for Mcm2RA and 48.7 ±23.4% for Mcm3RA. These data suggest, that the structural closure of the gate in the Mcm2-7 ring within the CMG does not significantly alter the contributions of the Mcm2-7 subunits. Despite the physical disconnection between the Mcm2/5 ATPase pair in the free Mcm2-7 complex, inter-subunit communication might stay in place and determine each subunit's contribution to the overall activity.

Helicase activity assays illustrate a functional asymmetry of the Mcm2-7 complex

To analyze the precise contributions of the MCM subunits on CMG helicase activity, I next tested unwinding of each of the six purified CMG RA complexes at different protein concentrations on a circular substrate (Figure 3.4C). The analyses revealed a reduced helicase activity for all mutant complexes, but the extent of abrogation varied significantly between the complexes (Figures 3.4C). The arginine finger mutation in Mcm3 (RA) showed the most severe defect, with activities of only 1-4% of the wildtype CMG. Strongly reduced activity was found for CMG complexes with Mcm5RA and Mcm6RA with activities of 24% and 18-34% of wt, respectively. In contrast to this, CMG complexes harboring RA mutations in Mcm4 and Mcm7 were the least affected, retaining 65-82% of the wt helicase activity. CMG 2RA had an intermediate effect with 45-47% of wt activity. Given that the Mcm2-7 subunits assemble into a unique order around the ring, one can also order the quantified effects of the RA mutations on the CMG helicase activity and schematically attribute differing levels of ATPase activity contribution to the active site pairs around the Mcm2-7 toroid. Previously, the effects from the CMG KA mutant studies have been thus displayed and it was found that the weakest activity lies around the Mcm4 subunit, while the strongest effect on the CMG helicase is around the Mcm5 subunit with these ATP binding mutants (Ilves et al., 2010). I also noted with the CMG RA mutants an intriguing pattern around the Mcm2-7 ring. The most striking ablation of helicase activity was observed with the RA substitutions in Mcm3 and Mcm5, while the weakest defects were seen around the Mcm4 and Mcm7 subunits. This is consistent with the previous work with the CMG KA mutant from Ilves et al., (2010) showing that the most crippling mutations are located exactly opposite to the side with the least deleterious mutations. In summary, ATPase active sites in one-half of the Mcm2-7 ring (2/3/5) appear to contribute to a larger extent on helicase activity of the complex than in the second-half of the ring (4/6/7) (Figure 3.7A and B). These findings undeniably oppose the long held belief that the

Mcm4/6/7 subunits are the true activity subunits, while Mcm2/3/5 are only important for coordination and possibly for negative regulation of the CMG. What is striking is that the active site pair between subunits Mcm4 and Mcm7 display very weak defects when either of the two active sites is eliminated individually (Walker A in Mcm4 or arginine finger in Mcm7). It is possible that the defects are minor because the other half of the ring's activity can substitute for the eliminated one. Given the results of the remaining five ATPase pairs, however, this hypothesis appears less likely. The results more likely suggest that the Mcm4/7 ATPase site is not needed for the helicase activity of the complete CMG complex. To test this view, I purified during research for my diploma thesis a CMG that introduces an arginine to alanine substitution in the corresponding Mcm7 arginine finger motif along with the lysine to alanine substitution in the Walker A motif of Mcm4. In agreement with my hypothesis, the effect of this mutation on the overall CMG helicase activity was very comparable to that of each of the mutations alone, showing that the two mutations are not additive and that this active site is not crucial for the overall activity.

DNA binding of the CMG helicase is strongly ATP-dependent

The introduction of mutations in either the ATP binding or –hydrolysis motifs has unraveled asymmetric defects around the Mcm2-7 ring of the CMG. As helicases are enzymes that utilize energy from nucleotide hydrolysis and translate it into mechanical work, namely to translocation on DNA, I wished to understand what importance these active sites have with regard to the DNA binding of the complex.

Estimation of half-saturation points from DNA binding assays yielded K_d values in the range of 10-20nM for the CMG (Ilves et al., 2010). Not only does this represent a 10-fold increase in binding affinity compared to Mcm2-7 with K_d values in the range of 200-300nM, the same study also showed that the binding of CMG to DNA was strongly dependent on ATP, unlike free Mcm2-7. Because binding requires ATP, and elimination of the ATP binding site in different Mcm2-7 subunits displays unequal effects with respect to ATP hydrolysis, I wished to test if there was also an unequal contribution by each subunit to the DNA binding. I and others have shown that active pairs on one-half of the Mcm2-7 ring are particularly important for ATPase and helicase activities (Figure 3.7A) (Bochman et al., 2008; Bochman and Schwacha, 2008; Ilves et al., 2010; Schwacha and Bell, 2001). Is there then a direct correlation between these subunits and DNA binding by the CMG complex? For another AAA+ hexameric helicase, E1 from papillomavirus, the molecular translocation mechanism has been suggested. In E1, each of the six subunits progresses through sequential cycles of ATP binding, ATP hydrolysis (ADP) and nucleotide release (apo) (Enemark and Joshua-Tor, 2006). During progression through these cycles, each

subunit “escorts” one nucleotide through discrete DNA binding hairpins (Enemark and Joshua-Tor, 2006)(see Chapter 5). The precise mechanism of how CMG couples ATP hydrolysis to DNA translocation remains a central unanswered question in the DNA replication field. The data presented above indicate that not all six nucleotide-binding sites are required in a catalytically active form for the helicase activity of the entire CMG complex. If some of the ATPase active sites are dispensable for the activity of the unwinding complex, the question arises how nucleotide-binding overall affects the complex if all subunits are forced into an ATP-bound state. More specifically, I sought to understand what discrete effect increasing concentrations of ATP might show on CMG’s DNA binding as well as on helicase activity. In support of the nonequivalence of the six different ATPase sites, when I titrated ATP into the helicase reactions, I observed a very dramatic departure from what one would anticipate in case of simple enzymatic kinetics with a maximal rate at a maximal substrate concentration as described by Michaelis and Menten (Figure 3.5). With regard to the unwinding, I initially observed significant increase in helicase activity upon ATP titration to 1-2mM, but when further ATP was added to the reaction, it resulted in a prominent decrease in helicase activity (Figure 3.5A). Similar enzymatic behavior was reported for the *S. sulfolobus* MCM helicase activity (Moreau et al., 2007). For *Drosophila* CMG, ideal helicase activity was observed in the ATP concentration range of 100-300μM ATP, after which the complex most likely shows a negative cooperativity for further binding. CMG binding to DNA was shown to strongly depend on nucleotide presence, with no essential requirement for its hydrolysis, as strong stimulation was seen with incubation with non-hydrolyzable ATPγS. In accordance to this and the summarized findings that not all the ATP sites are catalytically necessary, I observed also with the DNA binding of CMG a drastic ATP dependence (Figure 3.5B). Initial addition of nucleotide vastly stimulated DNA binding, but with an ATP concentration of ≥ 300μM ATPγS a marked decrease in binding was noted. Optimal DNA binding by CMG is found to occur at 10μM ATPγS, while helicase activity is most stimulated at around 300μM ATP. Taken together, it appears that a low ATP concentration is favorable for DNA binding, while higher ATP concentrations are required
for helicase activity. The latter concentration is well within the physiological range of the total ATP concentration in eukaryotic cells, which was estimated to be in the low millimolar range (Traut, 1994).

Having established the optimal conditions for testing DNA binding, I first tested the arginine finger mutant CMG complexes for ATP binding at a concentration of 10μM ATPγS (Figure 3.6A). Given that the hydrolysis of ATP seems dispensable for binding, I expected mostly mild defects, if any. Indeed, the DNA binding of the six arginine finger mutant

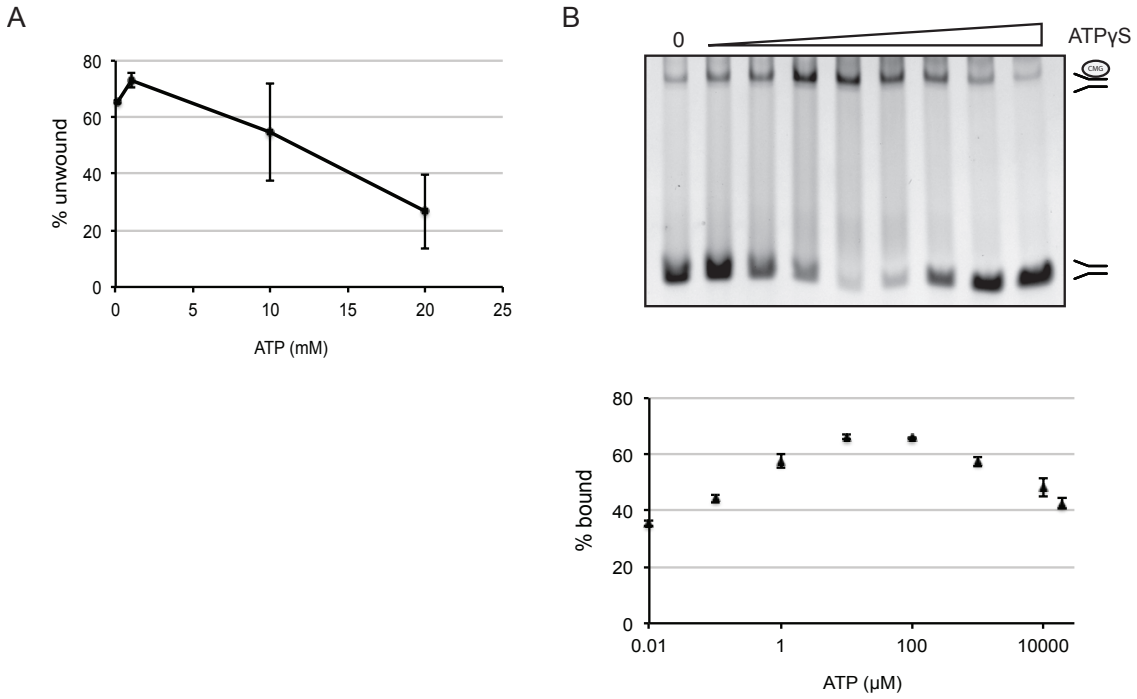

Figure 3.5. CMG activities are nucleotide dependent.
(A) Quantification of CMG wt helicase activity in ATP-concentration dependent manner on circular DNA substrate. ATP amounts were increased from 0.1mM, 1mM, 10mM, to 20mM ATP, average values of the displaced oligonucleotide were measured and the standard deviations of two independent series included.
(B) Electron-mobility shift assay (EMSA) showing DNA binding in an ATP-concentration dependent manner for wt CMG. Per reaction 100fmol of purified protein was added to 100fmol of FITC-labeled 'fork' substrate (3'+5' polyT tails). ATP amounts were increased from 10nM, 100nM, 1μM, 10μM, 100μM, 1mM, 10mM, to 20mM ATPγS. Positions of free 'fork', and CMG-bound substrate are indicated on the side. Quantification of gel-shifts are shown below. Average values of CMG-bound substrate at increasing ATPγS concentrations were plotted against percent of bound fork, including standard deviations of two independent series.

complexes was measured to be in the K_d range of 10nM on par with that of wtCMG. The strongest decrease in DNA binding was found with CMG 5RA and CMG 7RA with about 27% reduction of wt binding. Moderate decreases in the range of 16-20% of wt were displayed by CMG 3RA, 4RA, and 6RA, while CMG 2RA showed wt DNA binding, together confirming that the defects for this mutant were indeed rather mild. Thus, we can rule out that nucleotide hydrolysis is required for DNA binding. Instead, I would like to propose the possibility that the ATP hydrolysis motif might contribute in a mild way to the ATP binding of the complex. This could either be a direct effect on the nucleotide binding site or an indirect effect on the DNA binding motifs (e.g. hairpins). However, I did not assess this precise question in this study and therefore no data is present about the direct effect of mutations in the arginine finger motif on the ATP binding status in the active site.

In parallel, I wished to understand how the elimination of the discrete ATP binding motifs (Walker A) would contribute to the overall DNA binding. As anticipated by the ATP binding requirement for CMG's affinity for DNA, all CMG complexes containing a Walker A box mutation in one of the Mcm2-7 proteins (lysine to alanine, 'KA') displayed reduced DNA binding (Figure 3.6B). The weakest defects were found with CMG 2KA and CMG 4KA at about 85-86% of wildtype binding. Strong reduction in binding to 32-40% of wt was observed with CMG 3KA, 5KA and 7KA, whereas CMG 6KA showed an intermediate decrease. This data not only support the earlier finding that proper nucleotide binding is necessary for DNA binding by CMG. Further, it shows that the Mcm2-7 subunits within the CMG complex are non-equivalent with regard to the DNA binding. The ability to mutate individual ATP binding sites allowed me to dissect the different contributions of the six Mcm2-7 proteins and unravel the most important subunits that must be occupied with ATP to allow the helicase to bind DNA. To confirm that elimination of all six Walker A motifs indeed can entirely abolish ATP binding, which is required for DNA binding and subsequent unwinding, I purified a complex that simultaneously contains alanine substitutions in the lysines of all six Walker A elements (named '6xKA'). As expected, based on the previous work reporting that Walker A mutations (KA) in Mcm3 or Mcm5 individually already abrogate the helicase activity of CMG, I confirmed that the CMG 6xKA mutant had no unwinding activity (Figure 3.6C). Additionally, analysis of the same mutant CMG complex for DNA binding also showed that it had entirely lost its affinity for DNA (Figure 3.6D). It is important to note that even though this result is consistent with the hypothesis that mutation of the ATP binding sites have eliminated nucleotide binding by the complex and therefore abolished DNA binding, we cannot directly test that indeed no nucleotide is bound by the complex in these assays. However, given previous analysis of

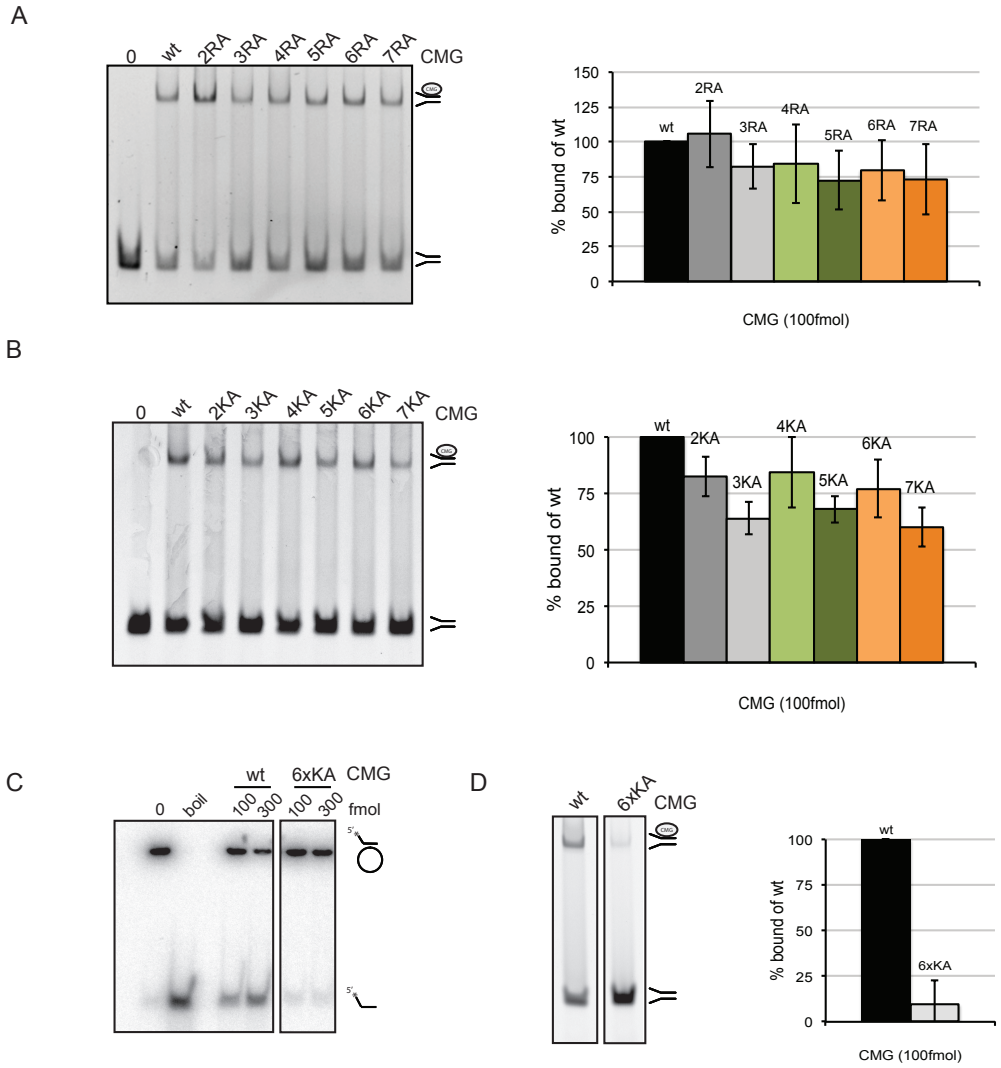

Figure 3.6. Defects in ATPase active sites correlate to DNA binding of CMG.
(A) EMSA of wildtype CMG and CMG complexes with alanine substitutions of the arginine finger residue in Mcm2-7 proteins ('RA'). Each protein complex contains an arginine mutation in one MCM subunit in the context of five wt MCM subunits. '0' indicates the control lane without protein. All lanes contain 100fmol of FITC- labeled forked DNA substrate (3'+5'polyT, 'fork') and, where indicated, 100fmol of purified CMG protein in presence of 10μM ATPγS. Reactions were separated on a native TBE polyacrylamide gel (4%). Positions of free 'fork', and substrate bound by CMG are indicated on side. EMSA assays were quantified in the right panel. The average values of all 'RA' CMG complexes are plotted, expressed as percentage of wt CMG gel-shift, and include standard deviations from six independent series.
(B) EMSA of wt CMG and CMG complexes with mutations in the Walker A box residue in Mcm2-7 proteins ('KA') in which a conserved lysine was substituted with an alanine. Experimental setup as described in (A). Quantification of EMSA of CMG 'KA' series with standard deviations from four independent series on the right panel.
(C) Autoradiograph of helicase activity assay on a circular substrate with wt CMG and CMG with mutations in the Walker A box of all six MCM subunits simultaneously ('6xKA'). Position of circular substrate with annealed oligonucleotide and displaced oligonucleotide alone are indicated on side. Two amounts of purified CMG proteins (100, 300fmol) were added to 1fmol of DNA substrate and 300μM ATP. Unwound products were separated by TBE-PAGE (8%). '0' and 'boil' are substrate controls without protein or boiled substrate, respectively.
(D) Gel-shift assay of CMG wt and '6xKA'. Per reaction, 100fmol of purified protein were added to 100fmol of FITC-labeled 'fork' substrate (3'+5' polyT tails). The results were quantified (*right panel*) and expressed as percent of wt CMG shifts with standard deviations from two independent series.

consensus motifs in other AAA+ proteins, it is reasonable to conclude that the essential lysine within the P-loop of the Walker A box is responsible for ATP binding.

In Figure 3.7, I have schematically summarized the defects in both helicase activity as well as DNA binding of the complete CMG complex when individual mutations in each Mcm2-7 subunit were introduced one at a time. Abrogation of helicase activity by mutations of the ATP binding and –hydrolysis motifs in Mcm3 and Mcm5 point to the particularly critical roles of these subunits with regard to the unwinding activity (Figures 3.7A and B). A similar schematic overview on the effects of these active site mutations on the DNA binding of CMG reveals that here as well mutations in the same half of the Mcm2-7 ring cause a drastic reduction in the binding affinity (Figure 3.7C and D). A marked decrease in binding affinity is observed with introduction of mutations in the Walker A and arginine finger motifs of the Mcm5-3-7 side of the hexameric ring. Further analysis of the discrete DNA binding motifs (hairpins) within the Mcm2-7 ring will be required to test each subunit's direct interaction with DNA and perhaps unravel an order of events around the ring to understand if a specific site is involved in the power stroke that determine a discrete start site for unwinding (see Chapter 5). For now, I propose that the ATP dependence on the DNA binding by CMG most likely suggests that nucleotide occupation of Mcm5-3-7 side of the Mcm2-7 ring is particularly important in establishing the proper DNA binding. Based on the molecular findings from studies with other helicases and my initial mutational data, it is tempting to hypothesize that a specific combination of ATP, ADP and free states around the Mcm2-7 ring is necessary to establish the correct environment for the CMG to bind DNA and subsequently provide the 'power stroke' required for helicase activity.

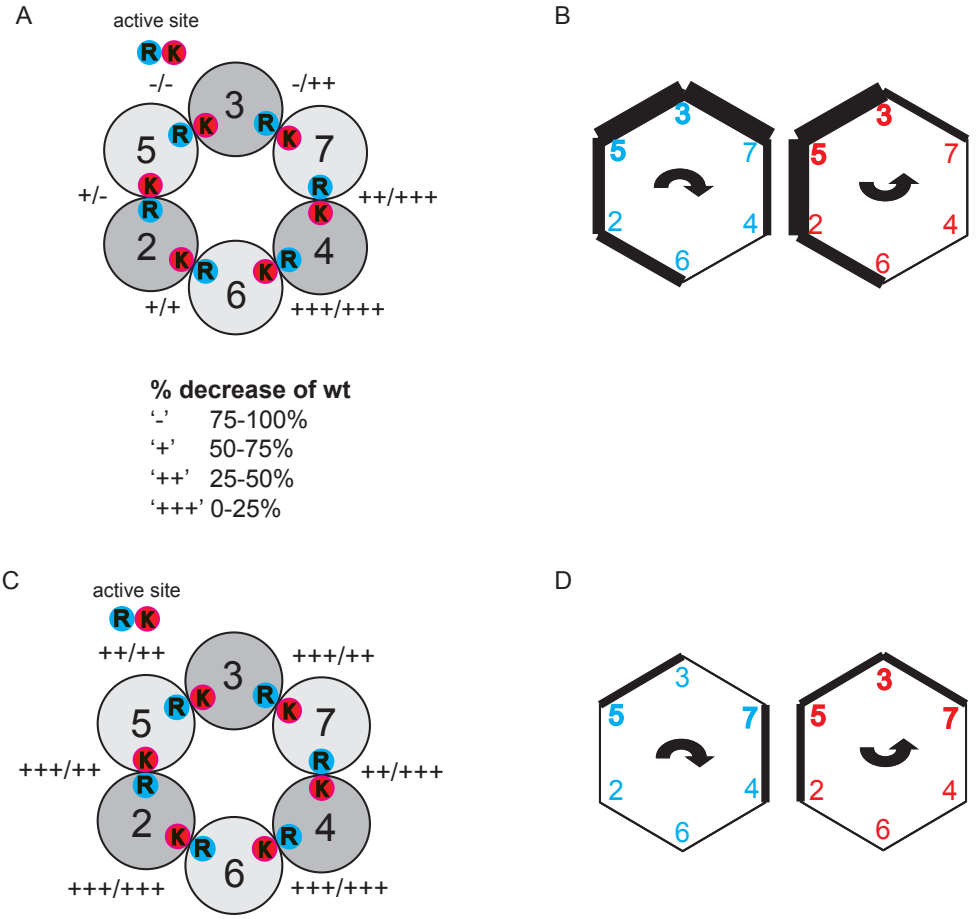

Figure 3.7. Schematic of effects on CMG activity when active sites are mutated.
(A) Schematic of the Mcm2-7 ring showing the order of all six MCM subunits as shown in Figure 3.1. The bipartite ATPase active sites created by the Walker A box ('K' in *red*) and arginine finger ('R' in *blue*) of the neighboring subunit are highlighted for all six MCM pairs.
The schematic illustrates the relative effects of arginine finger mutations side-by-side with effects of previously published Walker A box mutations on the helicase activity of the CMG complexes. The percentages of decreased activity after introducing the alanine substitutions are shown for all arginine finger/Walker A box pairs next to the ATPase active site pairs around the MCM ring. Hereby '-' indicates 75-100%, '+' 50-75% , '++' 25-50%, and '+++' 0-25% diminishment of helicase activity.
(B) Schematic summarizing the effects listed in (A). The thickness of the lines indicates the severity of effects. Arginine finger mutants (*blue*) are to be read clockwise, and the Walker A box mutants (*red*) counterclockwise. E.g. the thick line between Mcm5 and 3 for the arginine finger mutants in *blue* indicated the effect that the Mcm5RA mutation has on the entire CMG. On the other hand, the line between Mcm5 and 3 for the Walker A mutants in *red* would indicate the effect that the corresponding Mcm3KA mutation has on the entire CMG. The strongest affected MCM subunits are highlighted with thicker lines.
(C) Same schematic as (A) but now illustrating DNA binding by CMG complexes with alanine substitutions in the arginine finger or Walker A box motifs.
(D) Schematic summarizing (C) in the same way as described in (B).

3.4 Discussion

The data presented in this study demonstrate the importance of the nucleotide-bound state for the CMG helicase function. Some functional parallels to the homologous *S. cerevisiae* Mcm2-7 helicase complex could be drawn from this biochemical analysis of the ATP binding and –hydrolysis sites in the *Drosophila* CMG. Analogous to reports from budding yeast Mcm2-7, mutations in Walker A motifs in the six MCM subunits have earlier been shown to significantly eliminate ATP binding and cripple both the ATPase and helicase activities of the CMG complex (Ilves et al., 2010). Previously, Ilves et al., (2010) discussed the activation of the DmMcm2-7 intrinsic activities through association with the allosteric factors Cdc45 and GINS, and together with an electron microscopic study these present evidence for striking structural re-arrangements of the Mcm2-7 ring within the CMG (Costa et al., 2011). Consistent with the demonstration of the requirement for bound ATP reported in this biochemical analysis, the structural studies elegantly showed how nucleotide binding is necessary for an initial closure of an existent discontinuity in the Mcm2-7 ring and subsequent constriction of the ring's central cavity for better engagement of the DNA (Costa et al., 2011).

The work presented here on the nucleotide hydrolysis 'arginine finger' motif not only confirms the findings from the initial ATP binding Walker A site studies in the CMG, but further substantiates the importance of certain active site pairs around the Mcm2-7 ring. Previous work in budding yeast Mcm2-7 demonstrated how single point mutations in one of the ATPase active site consensus motifs could drastically decrease the ATPase and helicase activities of the motor and established its cooperative nature (Bochman and Schwacha, 2008; Schwacha and Bell, 2001). Similar to this interdependence, mutations in the residues critical for ATP binding in Mcm2, 3, 5, or 7 decreased ATPase rates of the entire CMG complex to levels far below $1/6^{th}$ of wildtype, which we would expect in a stochastic model, in which all six subunits contribute equally and independently to the overall activity of the complex (Ilves et al., 2010). Mutations in the corresponding nucleotide hydrolysis part of the bipartite active site confirmed a reduction of ATPase rates below 17% of wildtype. If the CMG also follows a sequential model for hydrolysis as has been suggested for the T7 and E1 DNA helicases, one might speculate that a certain order of nucleotide occupancy and 'free' state (apo) around the Mcm2-7 ring is essential for the entire activity. If one subunit is crippled, the specific order around the ring could be changed, perhaps leading to the arrest of ATP hydrolysis for the entire complex (Crampton et al., 2006; Enemark and Joshua-Tor, 2006). Despite the serious structural impact that Cdc45/GINS have on the Mcm2-7 ring, leading to major spatial re-arrangements, the work presented here revealed that the mutations led to similar effects both in free Mcm2-7 and

in the higher-molecular weight CMG complex. The extent to which the ATPase activity is reduced after elimination of the arginine finger in single Mcm2-7 subunits within the CMG is comparable to that in the free Mcm2-7 ring. The gap in the DmMcm2-7 ring is located between Mcm2 and Mcm5 with the bipartite active site formed by the Walker A box from Mcm5 and the arginine finger from the adjacent Mcm2 subunit. This discontinuity is significantly narrowed upon incorporation of the free Mcm2-7 ring into the CMG complex and closed upon nucleotide binding. In spite of the discontinuity in the Mcm2-7 complex, the introduction of an arginine finger mutation in Mcm2 caused a decrease in ATPase activity of either the isolated Mcm2-7 or the complete CMG complex to a similar range. This result strongly supports the view that the six subunits and active sites do not work as the simple sum of six independent motors, but rather display an enormous inter-subunit dependence and communication. Although the allosteric factors Cdc45/GINS vastly enhance the intrinsic activities of the Mcm2-7 within the CMG, certain active sites in the complex appear to have the same level of contribution and importance in either of the two forms of Mcm2-7. One might posit that the exact order of events is determined within the Mcm2-7 ring, and the association of the accessory factors help further optimize the helicase activity through four main functions: 1) to activate the helicase in a timely manner at the G1 to S phase transition, 2) to structurally close the discontinuity and align the active site motifs in an optimal way to enhance the intrinsic activities, 3) to possibly guide one of the unwound strands during translocation, 4) and to bridge the interactions between the MCM motor and other replication proteins required downstream of the unwinding action (see also Chapter 4).

Helicases, as AAA+ ATPases, utilize the energy obtained through cycles of ATP binding and -hydrolysis to fuel the MCM motor for translocation on DNA. The functional asymmetry of ATPase active sites raised the question if this non-equivalent sensitivity of ATP binding and –hydrolysis sites might result from a sequential movement during helicase activity that requires a unique starting position within the Mcm2-7 ring. Consistent with the reported asymmetric nature of the active sites, unwinding studies on CMG complexes containing an alanine substitution in the critical arginine finger motif demonstrated that the extent of helicase reduction drastically differed between the six ATPase active site mutants. These findings correlated well with the effects of the Walker A box mutations on the helicase activity of CMG, together pointing to Mcm2/5, Mcm5/3 and Mcm3/7 as the most critical pairs for the unwinding activity. Interestingly, the strongest effects localize around the subunits Mcm2, 3, 5, which establish strong interactions with GINS/Cdc45, while the weakest effects were located around Mcm4, at the opposite side of the ring. The closure of the discontinuity between Mcm2 and Mcm5 after association of the ancillary

factors, brought along by subsequent nucleotide binding, might represent a structural requirement for the initial melting. Due to the insufficiently robust helicase activity of the purified recombinant DmMcm2-7 complexes, it was unfortunately never possible to analyze if a mutational sensitivity might exist in the MCMs alone, and if this sensitivity was altered upon associations in the CMG, or if it followed the same pattern. However, comparison of ATPase rates of several active sites in Mcm2-7 versus CMG strongly favor the latter possibility, suggesting that the mutational sensitivity of the hexameric Mcm2-7 ring likely remains unaltered after its association into the CMG complex with regard to the active sites.

Lastly, simultaneous mutation of both parts of an active site revealed that at least one pair neither required ATP binding nor its hydrolysis for the overall helicase activity of the CMG. The active ATPase sites in the Mcm2-7 ring are formed by the interface of two adjacent MCM subunits where the critical 'arginine finger' is provided in *trans* to initiate the hydrolysis of ATP that is bound by the conserved lysine residue in the Walker A motif (reviewed in Forsburg, 2004; Hanson and Whiteheart, 2005). The composition of observed *Drosophila* sub-complexes formed during protein purification confirmed the nearest neighbor analysis of Mcm2-7 in earlier reports (Crevel et al., 2001; Davey et al., 2003). Mutational analysis of the Walker A box sites revealed that the elimination of the invariant lysine in the ATP binding site of Mcm4 barely crippled the helicase activity of the entire complex. In agreement with this, the mutation of the corresponding arginine within the ATP-hydrolyzing motif displayed the weakest defect in CMG's unwinding. The double substitution (4KA/7RA) showed similar effects as the individual mutations alone, suggesting the importance of the Mcm4/7 active site without the requirement of being in a nucleotide-occupied state (Ilves et al., 2010). In contrast, certain pairs such as Mcm5/3 already eliminated the helicase activity when only one-half of the two-part active site was abolished, strengthening the view that a unique occupation of nucleotide-bound and –free states might be essential for the complex's optimal activity.

In vivo studies of the Mcm2-7 complex have shown similarities with other replicative helicases exhibiting unwinding activities. All six subunits were found to be essential for the proliferating yeast cell (Forsburg et al., 1997; Gibson et al., 1990). However, as opposed to archaeal helicases, all biochemical studies of the eukaryotic Mcm2-7 complexes failed to demonstrate its *in vitro* unwinding activity. In contrast, studies of a trimeric sub-complex, consisting of Mcm4, 6, and 7 that dimerizes, could demonstrate weak helicase activity and has therefore often been considered the 'true catalytic' core of the Mcm2-7 ring, whereas Mcm2/3/5 were attributed a regulatory function (Ishimi, 1997).

The work reported in Ilves et al., (2010) is consistent with the findings in this study, with both studies indicating that, within the CMG, the ATP binding and –hydrolysis motifs between Mcm7/4 and Mcm4/6 are the least important to the ATPase or helicase activity of the entire complex. In contrast to these findings, reports from budding yeast have shown that Mcm2, 4, 6 are essential for ATPase activity (Schwacha and Bell, 2001). In contrast, my studies and previous studies from the Botchan lab found that both Walker A and arginine finger motifs in *Drosophila* Mcm2/5, 5/3, or 3/7 active site pairs are essential for the CMG's intrinsic ATPase and helicase activities *in vitro* (Ilves et al., 2010). The question arises whether these findings suggest an evolutionary divergence in higher eukaryotes. To date, no CMG homologous complex has been described in yeast, and the Mcm2-7 complex is still discussed as the active replicative helicase.

Crystallographic and electron microscopic studies of the archaeal MCM complexes provided pivotal insights for understanding how the eukaryotic Mcm2-7 might bind DNA and is in detail discussed in Chapter 5 (reviewed in Brewster and Chen, 2010; Brewster et al., 2008; Chong et al., 2000; Costa et al., 2006; Fletcher et al., 2003). Various eukaryotic and archaeal helicase studies have concluded that the leading strand passes through the central channel of the Mcm2-7, while the lagging strand is excluded (Fu et al., 2011; Graham et al., 2011).

While nucleotide is not required for CMG complex assembly, ATP binding, but not -hydrolysis, is essential for the formation of CMG-DNA complexes.

Interestingly, optimal DNA binding occurs in the range of lower micro-molar ATP concentration, whereas nucleotide addition beyond resulted in a marked reduction in DNA affinity. Similarly, the helicase activity undergoes robust activation upon initial ATP addition up to 300 μM, and decreases upon further ATP addition. Measurements of ATP hydrolysis rates of CMG have shown that the kinetics do not follow a simple Michaelis-Menten behavior, but surprisingly have different modes with increasing ATP concentration, suggesting nucleotide-dependent cooperative switching in the MCM ring (Ilves et al., 2010). This study showed that, at sub-micromolar ATP, the *Drosophila* CMG showed a high affinity for DNA, with its most effective binding to DNA at ~10μM. At higher nucleotide concentration, the complex showed faster hydrolysis rates, and, in parallel, showed a dramatic stimulation in helicase activity. These results lead to the hypothesis that, for optimal DNA binding only a subset of nucleotide-binding sites needs to be occupied, and occupation of the adjacent subunit might trigger hydrolysis of the first and thus lead to the right combination of ATP, ADP and free states around the Mcm2-7 ring. Thus, an excess of non-hydrolyzable ATP could occupy all ATP binding sites and worsen the optimal conformation of the DNA binding beta-hairpins within the MCM ring. The data

here support the view that ATP sites are non-equivalent around the MCM ring, strengthening the proposal that not all nucleotide-binding sites need to be catalytically active for the overall helicase activity of the complex. Similar behavior was observed for the *S.sulfolobus* MCM and was suggested to be indicative of negative cooperativity between ATP binding sites of differing strengths (Moreau et al., 2007).

From reports on AAA+ proteins, the Walker A and Walker B motifs are the only anticipated ATP binding consensus sites within MCM proteins. The Walker B motif has also been implicated in contributing nucleotide binding to some extent. However, even though no direct abrogation of nucleotide binding was tested here, the elimination of ATPase activity of the entire CMG complex upon mutation of the critical lysines in the Walker A box to alanines strongly favors the idea that these substitutions have disrupted ATP binding. Unfortunately, the biochemical methods applied here did not allow me to track the direct consequence of such inhibition within a single cycle of ATP binding, -hydrolysis, and –release during DNA unwinding within the Mcm2-7. Since the DNA binding assays with CMG require only bound nucleotide without hydrolysis – shown by binding of non-hydrolyzable ATPγS -it becomes of interest if elimination of the ATP hydrolysis sites directly correlates with a decreased binding step. Or if these sites are rather implicated in subsequent unwinding steps that require correct positioning of DNA binding motifs, which depend on the occupation state in the nucleotide binding pockets.

In agreement with the ATP-dependent DNA binding of CMG, the work here presents an interesting correlation between the elimination of the Walker A motif in the single Mcm2-7 subunits and the decrease in DNA binding. The Mcm3 and Mcm5 subunits, shown to be critical for ATPase and helicase activities, displayed, together with Mcm7, the strongest defects in DNA binding upon elimination of the critical lysine in their Walker A motifs. It remains possible that nucleotide binding affects the alignment of beta-hairpins within the central channel of the Mcm2-7 and thus leads to a reduced DNA binding through abrogation of the correct β–hairpin positioning. Besides CMG 2RA, a pattern of decreased DNA binding similar to the one revealed in the CMG KA series could be observed upon mutation of the arginine finger motifs in the MCM members. The overall extent of severity, however, was less prominent in the latter series, but pointed to the same order within the ring. Even though the arginine finger is attributed the function of hydrolysis of the bound ATP and it was established that CMG only requires ATP binding in its binding to DNA, the possibility cannot be ruled out that this motif causes a secondary effect on positioning of DNA binding elements in the nucleotide bound state. Alternatively this residue could communicate directly with the Walker A motif. It should be also

emphasized that even though these defects are mild *in vitro*, consequences would most likely be quite deleterious for the organism. Taken together, the work presented here established for the first time a direct correlation of a mutant nucleotide-binding site in an eukaryotic MCM protein with the subsequent diminishment of DNA binding in a manner that distinguishes individual subunits and assigns them an order of importance.

3.5 Materials and Methods

Cloning and construction of baculovirus

The vector templates were described in 'Material and Methods' of Chapter 2 and are in general use for generation of CMG mutants through PCR based mutagenesis. Sequencing was used to verify the entire protein coding regions of all generated pFastBac constructs.
In detail following specific residues were targeted and alanine substitutions were introduced:
Arginine finger mutants:
The R to A mutations in the arginine fingers of the six MCM subunits were introduced by targeting the following residues: R641 in MCM2, R473 in MCM3, R645 in MCM4, R510 in MCM5, R521 in MCM6, and R514 in MCM7.
Walker A motif mutants:
The K to A mutations in Walker A box of various MCM subunits targeted the following residues: K514 in MCM2, K346 in MCM3, K518 in MCM4, K384 in MCM5, K394 in MCM6, K387 and R514 in MCM7.

Baculoviruses expressing the individual 11 CMG proteins were generated according to the Bac-to-Bac system manual from Invitrogen.

Preparation of cell extracts of baculovirus infected insect cells

Sf9 or ES-Sf9 cells used for baculovirus amplification were grown in TNM-FH media (AppliChem GmbH) supplemented with 10% fetal bovine serum (HyClone1 Laboratories Inc., Logan, UT) or ESF 921 Insect Cell Culture Medium, Protein-Free (Expression System). All baculoviruses were freshly amplified for 96 hours in Sf9 or ES-Sf9cells in 150mm tissue culture dishes. Hi5 cells used for protein expression were grown either adherently in serum-free ExCell media (SAFC Biosciences, Lenexa KS) or in suspension culture in ESF 921 Insect Cell Culture Medium. 0.25ml of freshly collected FLAG-Mcm3 virus stock, 0.5 ml of Mcm2,4,5,6,7 and the four GINS virus stocks as well as 1ml of Cdc45 virus with an approximate MOI of 5 for each virus were added per 150mm tissue culture dish of Hi5 cells (~2.0 * 10^7 cells per dish) or per 20ml of Hi5 suspension culture (1.2*10^6 cells/ml). Proteins were expressed for 72-96 hours and cells were collected in centrifuge bottles, washed once with PBS buffer, and resuspended in hypotonic buffer H (15mM KCl, 25mM Hepes pH 7.6, 0.02% Tween20, 10% glycerol, 1mM EGTA, 0.4mM PMSF) supplemented with complete protease inhibitor cocktail from Roche Diagnostics. The cell suspension was kept at -80°C until protein purification was performed.

Protein purification

All subsequent steps were performed at 4°C or on ice unless otherwise indicated.

For purification of CMG complexes, the cells were thawed and then lysed by a dounce homogenizer, pestle B. The salt concentration of the resulting lysate was adjusted to 100mM KCl and then clarified by centrifugation at 15,000 rpm in Sorvall RC-5B for 15min. The FLAG-Mcm3 containing complexes were affinity-purified by incubation of anti-FLAG agarose beads (ANTI-FLAG M2 affinity gel, Sigma-Aldrich) with the cleared extracts for 3-4 hours. The agarose beads were pelleted, washed once in batch with 20ml of buffer A-100 (buffer A contains 25mM Hepes pH 7.6, 15-35% glycerol, 50mM sodium acetate, 10mM magnesium acetate, 0.2mM PMSF, 1mM DTT; here supplemented with 100mM KCl and 2mM 2-mercaptoethanol), and then transferred to disposable chromatography columns (Poly-Prep, BioRad) and washed three times with each 5ml of buffer A-100 on the column. The immunoprecipitated complexes were competitively eluted with buffer E (200µg/ml flag peptide in buffer A-100 supplemented with 250µg/ml insulin and complete protease inhibitor cocktail) in two consecutive steps. For the first elution step 3ml of buffer E was added to the agarose bead, incubated at RT for 15min, and the elution step was repeated one more time by addition of 2ml of buffer E and 15 minute incubation (end-over-end mixing) at RT. Both eluate fractions were combined and consecutively passed through Mono S HR 5/5 and Mono Q HR 5/5 ion exchange columns connected to an ÄKTA Purifier FPLC system (GE Healthcare Life Sciences) to separate the stoichiometric CMG or Mcm2-7 complexes from incomplete sub-complexes. Immunoprecipitated complexes were eluted with a 100-550mM KCl gradient over a total volume of 30ml (in buffer A supplemented with DTT). Further separation and final concentration of the MonoQ 1ml peak fractions of the stoichiometric complexes (~410-440mM KCl for CMG, ~ 360-390mM KCl for Mcm2-7) was achieved by a subsequent run over a Mono Q PC 1.6/5 column on a Pharmacia SMART micro-purification system. Bound complexes were gradient-eluted with 150-550mM KCl. The 50µl peak fractions of the stoichiometric complexes were combined and dialyzed for 3 hours into buffer H (25mM HEPES (pH 7.6), 50mM sodium acetate, 10mM magnesium acetate, 10% glycerol). Dialysis buffer was once exchanged after 1.5 hours. Proteins were aliquoted, frozen with liquid nitrogen and kept at -80°C until used.

This protocol was modified as follows for the purification of stoichiometric Mcm2-7 complexes. All used buffers, starting from the washes of anti-FLAG beads, contained potassium acetate instead of potassium chloride at 200mM in washing and FLAG elution buffers. The bound complexes were competitively eluted with FLAG-peptide in two steps at 4°C for first 30 and then 20 minutes. The subsequent chromatographic steps contained the following potassium acetate gradients: 200 -1100mM for the the Mono Q HR 5/5 step

on the ÄKTA Purifier FPLC system, and 300-1100 mM for the SMART Mono Q PC 1.6/5 step. The stoichiometric Mcm2-7 fractions eluted from the Mono Q column at 800-850mM potassium acetate.

All purified stoichiometric CMG and Mcm2-7 complexes were separated by SDS-PAGE (10%) and stained with SYPRO®Red Protein Gel Stain (Sigma-Aldrich) for concentration measurement by laser densitometry and confirmation that all proteins are present due to the poor staining of Sld5 protein by silver-staining. Mcm2-7 protein complexes of known concentrations were used as concentration standards.

Glycerol gradient sedimentation

Purified recombinant CMG was loaded onto a 4ml 15-35% glycerol gradient in buffer A. The samples were run for 12hrs on Beckman L7 ultracentrifuge in a Sw60 rotor at 4°C. After the run, fractions of 200µl total volume from top of the gradient were collected analyzed for the protein content by SDS-PAGE (10%) and subsequent silver-stain. BSA and thyroglobulin were used as protein standards of known molecular weight and run in parallel. The protein content in the fractions was analyzed by a Bradford assay (BioRad) on a Beckman DU® 530 Spectrophotometer.

ATPase activity assay

The measurements of the ATPase rates were performed with 300fmol of wt or mutant CMG in 20µl of buffer A (25mM Hepes pH 7.6, 10% glycerol, 50mM sodium acetate, 10mM magnesium acetate, 0.2mM PMSF, 1mM DTT, 250µg/ml insulin) in presence of the desired concentration of cold ATP (GE Healthcare) spiked with γ-^{32}P ATP (MP Biomedicals or Perkin Elmer, 1 µCi in the reactions with highest ATP concentration). The reactions were incubated for 20 minutes at 30°C and stopped by addition of 2µl of 0.5 M EDTA. 1.1µl samples were spotted on PEI-cellulose thin layer chromatography (TLC) plates (Sigma-Aldrich) and developed in 0.4M LiCl / 1M formic acid. For the Mcm2-7 assays, the reactions were performed in 10µl volume with 3pmol of protein. The analysis of the data was performed with GraphPad Prism software.

Labeling of DNA substrates

A 70-mer oligonucleotide was designed such that 40nt would anneal to M13mp18ssDNA plasmid (New England Biolabs), leaving a 30-mer polyT extension at the 5' end.

The 5' end was previsouly radioactively labeled with γ- ^{32}P ATP (MP Biomedicals or Perkin Elmer) and T4 polynucleotide kinase (New England Biolabs), subsequently purified through an illustra MicroSpin G-50 column (GE Healthcare) and mixed with the

M13mp18 ssDNA plasmid. The reactions were heat-denatured for 1 minute, and annealing was achieved by gradually cooling to room temperature. Free oligonucleotide was separated by purification through four consecutive MicroSpin S-400 HR columns (GE Healthcare).

Helicase activity assay

The helicase assay was carried out in buffer A (25mM Hepes pH 7.6, 10% glycerol, 50mM sodium acetate, 10mM magnesium acetate, 0.2mM PMSF, 1mM DTT, 250μg/ml insulin). Desired protein concentrations were mixed with 1-2fmol of a circular M13 based DNA substrate and unwinding was allowed to proceed for 30 minutes at 30°C in the presence of 0.3mM ATP in a total reaction volume of 10μl. The displacement reactions were stopped by addition of 0.1% SDS and 20mM EDTA and reaction products were immediately electrophoretically separated on TBE-acrylamide gel (8% TBE, with 0.1%SDS).

EMSA assays

The 'fork' DNA substrate used for EMSA assays, if not otherwise noted in text, is based on the 233+235 substrate described in (Bochman and Schwacha, 2007) and has been modified for a fluorescent assay in agreement to modifications described in Ilves et al., (2010).

Oligonucleotide '235' with the nucleotide sequence

(5' – CACTCGGGCTCGTTTTACAACGTCGTGACTGGGCACTTGATCGGCCAA-

$CC(T)_{40}$ - 3')

was annealed to the complementary 5'FITC (Fluorescein isothiocyanate) labeled oligonucleotide '233long' with the nucleotide sequence

(5' - (Fl)$T(T)_{39}$GGTTGGCCGATCAGTGCCCAGTCACGACGTTGTAAAACGAGCC-CGAGTG - 3')

that generates a 50bp double-stranded region with 40nt poly-thymidine extensions on both ends with a FITC labeled 5'end.

The forked substrate was PAGE-purified (8% TBE). The gel-shift assays were performed with 100fmol of fluorescently labeled DNA substrate in the presence of 10μM ATPγS and indicated amounts of purified CMG protein complexes in a total volume of 10μl. The protein binding to DNA was allowed to proceed for 30 min at 30°C and the reactions were electrophoretically separated through a native 4% acrylamide (60:1 acrylamide:bis-acrylamide)/0.5xTBE gel, supplemented with 6mM magnesium acetate and 5% glycerol. The gel was run at 4°C in 0.5xTBE buffer, supplemented with 6mM magnesium acetate for 1hour at 150V. Signals were detected with Typhoon™9400 (GE Healthcare) or ChemiDoc

MP System (BioRad), and the Image Quant or Image J software was used for analysis of the results.

Acknowledgements

I want to thank Ivar Ilves for kindly providing the baculoviruses encoding the alanine substitutions in the Walker A box motifs of *Drosophila* Mcm2-7 proteins. Also, I want to thank Nele Tamberg for help with protein purifications and assays in the initial stage of the work.

Chapter 4

Cdc45 guards the gate in the CMG helicase assuring leading strand engagement

4.1 Abstract

Separation of the parental duplex DNA is critical for replication as it allows polymerases to subsequently step in and synthesize the nascent daughter strands. This action requires an active helicase. In *Drosophila*, such activation is achieved by the association of the auxiliary factors Cdc45 and GINS through a nucleotide-induced conformational remodeling of the Mcm2-7 ring at the entry to S-phase. Structural studies show that these factors help close a discontinuity present between subunits Mcm2 and Mcm5 in the Mcm2-7 ring ('2/5 gate'). Additionally, the central channel constricts upon nucleotide binding, ensuring better engagement of DNA. Cdc45 and GINS are key parts of the multi-subunit CMG complex, which represents the active eukaryotic helicase; however, their direct roles remain elusive. Here, we report that *Drosophila* Cdc45 is an orthologue of the bacterial RecJ protein. In contradiction to previous hypotheses, Cdc45 directly contacts the leading strand in the absence of nucleotide and displays no interaction with the lagging strand. This study identifies residues in Cdc45 that are responsible for the DNA contacts and that diminish the helicase activity and processivity of CMG when mutated. On the other hand, no DNA binding by the GINS proteins in the context of CMG to either the leading or lagging strand could be detected. The majority of the DNA-protein contacts to the leading strand are established by Mcm2-7. Data from substrates that allow CMG to only bind in its 3' to 5' tracking direction suggest that the Mcm2-7 ring engages the leading strand. Furthermore, our data show that the lagging strand only binds a subset of Mcm2-7 subunits. Mutations in the external β-hairpins of Mcm2-7 reduce the direct interactions with the lagging strand and decrease helicase activity, suggesting that the lagging strand wraps around the outer surface of the Mcm2-7 ring. Thus, I propose that the *Drosophila* CMG behaves analogously to the 'steric exclusion and wrapping model' suggested for archaea. We hypothesize that Cdc45 acts as a 'guardian of the gate', ensuring that the leading strand remains engaged within the central channel of the Mcm2-7 ring during melting, progression or stalling of the replisome.

4.2 Introduction

The copying of chromosomal DNA is a fundamental biological process in all dividing cells. It is absolutely pivotal for all organisms that the duplication of its genetic information is precise, complete and limited to once per cell cycle. Hence, this process must be initiated in a strictly coordinated manner, and is comprised of a well-orchestrated multi-step pathway with checkpoints that ensure its timely and accurate execution. Loading and activation of the helicase is critical to DNA replication. This enzyme acts as a key engine to separate parental duplex DNA into single strands, thus making them accessible for DNA polymerases to step in and start replicating the genetic information. In *Drosophila*, the activation of this enzyme requires that initiator proteins associate. The subsequent binding of auxiliary factors helps create the correct molecular environment for the replication forks to form and progress.

Previous studies reported that the binding of Cdc45 and GINS proteins convert the otherwise inactive *Drosophila* Mcm2-7 hetero-hexameric helicase motor into a stable and active CMG (Cdc45/Mcm2-7/GINS) helicase (Ilves et al., 2010). These accessory factors remain part of the replisome, as it translocates in a 3' to 5' direction, as essential elements of the active eukaryotic CMG helicase complex (Aparicio et al., 1997; Moyer et al., 2006; Pacek et al., 2006).

The Botchan group previously addressed the question of how Cdc45 and GINS contribute to helicase activation and demonstrated in a single-particle electron microscopic study of *Drosophila* CMG that these co-factors play a structural role. The free *Drosophila* Mcm2-7 ring exists in either an open, lock-washer or a notched, planar configuration with a gap between the interface of subunits Mcm2 and Mcm5 (Costa et al., 2011). This discontinuity in the Mcm2-7 ring was first predicted by mutational analyses of yeast Mcm2-7 and was termed the 'Mcm2/5 gate' (Bochman et al., 2008). The Schwacha group hypothesized in their biochemical reconstitution studies of the yeast Mcm2-7 helicase activity that the active site, formed by the interface of the Mcm2/5 dimer, represents an ATP-dependent 'gate'. They suggested that this gap is strongly anion-dependent and further introduced the model that yeast Mcm2-7 displays *in vitro* helicase activity under distinct reaction conditions that induce a closed conformation of the gate (Bochman and Schwacha, 2008). The presence of this discontinuity was supported by interaction studies of the Mcm2-7 subunits (Crevel et al., 2001; Davey et al., 2003; Forsburg, 2004). Pairwise immunoprecipitation experiments of Mcm2-7 proteins led to the proposal of a specific unique order of assembly in the hetero-hexameric ring and identified adjacent MCM pairs around the ring (Mcm2/5, Mcm5/3, Mcm3/7, Mcm7/4,

Mcm4/6, and Mcm6/2). However, no direct interaction between Mcm2 and 5 could be detected, suggesting extremely weak or no direct contacts (Crevel et al., 2001; Davey et al., 2003; Forsburg, 2004).

These previous suggestions of a putative '2/5 gate' were confirmed and visualized in the recent structural study of the *Drosophila* CMG complex by Costa et al., (2011). The planar Mcm2-7 structure is reinforced by the association of Cdc45 and GINS to the exterior surface along the longitudinal axis of one-half of the Mcm2-7 ring (Mcm2/3/5). The '2/5 gate' closes upon nucleotide binding and results in constriction of the central channel of the Mcm2-7 ring, forming a structure that most likely has better engagement with DNA. This is accompanied by another structural alteration that forms a second external channel, spatially separated from the central Mcm2-7 pore (Costa et al., 2011).
Previous biochemical studies of the recombinant CMG helicase established that this activated complex only binds to single-stranded DNA, whereas the Mcm2-7 ring has been shown to load as a double hexamer in a head-to-head manner onto duplex DNA (Evrin et al., 2009; Gambus et al., 2011; Ilves et al., 2010; Remus et al., 2009).

These observations indicate that structural rearrangement of the Mcm2-7 helicase core cause the melting and subsequent segregation of the two strands of DNA. However, the molecular mechanism by which this replicative eukaryotic helicase unwinds double-stranded DNA still remains a central unanswered question in the field. Two main models have been proposed for the simpler archaeal MCM helicases, in which the enzyme was depicted to either translocate along duplex DNA, or to exclude one of the two separated strands and translocate on ssDNA (Brewster et al., 2008; Graham et al., 2011). Distinction between the two scenarios was finally rendered possible by a recent study from the Walter group that included bulky roadblocks on either leading, lagging or on both DNA strands simultaneously. By assaying the translocation of the helicase as it encountered such strand-specific obstruction this study indisputably established that the CMG can bypass a roadblock on the lagging strand but leads to helicase stalling when facing bulky groups on the leading strand (Fu et al., 2011). If the CMG were a dsDNA translocase, it would be expected to stall at roadblocks on either strand. Instead these data suggested that the leading strand passes through the central channel of the CMG helicase while the lagging strand must be excluded. These findings result in our understanding that the CMG must comprise a ssDNA translocase.

What role the auxiliary factors Cdc45 and GINS play during loading and translocation remains another central question of interest. While homologues of the GINS and MCM proteins have been described and extensively studied, no apparent homologue of Cdc45,

and hence the CMG complex, has been found in archaea (Aparicio et al., 1997; Gambus et al., 2006; Onesti and Macneill, 2013).

Cdc45 is an essential protein, important for both initiation and elongation during replication. It shows high sequence conservation amongst eukaryotes (Mimura and Takisawa, 1998; Owens et al., 1997; Tercero et al., 2000; Zou et al., 1997). It has been posited that Cdc45 might represent the major point of divergence between archaeal and eukaryotic replication machinery (Sanchez-Pulido and Ponting, 2011).

Only recently, several bioinformatics studies have independently revealed an evolutionary link for Cdc45. These analyses have uncovered a significant relationship between Cdc45 and the 'DHH family' of phosphodiesterases, with the closest relationship being with the bacterial RecJ nuclease; a protein that has been implicated in bacterial DNA repair (Krastanova et al., 2012; Makarova et al., 2012; Sanchez-Pulido and Ponting, 2011). One of the motifs required for these proteins' nuclease activity contains the amino acid sequence 'DHH' (Asp/His/His) and gives rise to the name of this protein family. The finding of Cdc45 as a potential RecJ orthologue led to an analysis of the Cdc45 primary sequence. Compared with RecJ, Cdc45 has mutations in several conserved motifs required for the coordination of metal ions in RecJ; these motifs carry out a 5'-3' exonuclease activity (Beese and Steitz, 1991; Sutera et al., 1999). The absence of these motifs raised many questions about the role of Cdc45 in DNA replication, suggesting an evolutionary divergence in the functions of these two proteins (Krastanova et al., 2012; Sanchez-Pulido and Ponting, 2011). Studies with purified recombinant human Cdc45 corroborated that it lacks any nuclease activity and shows a very weak affinity for single-stranded oligonucleotides, as expected by its similarity to RecJ (Krastanova et al., 2012).

The discovery of the homology between Cdc45 and RecJ might suggest common molecular roles for these proteins, but biochemical work so far has shown a functional divergence with respect, at least, to nuclease activity. Many studies have attributed RecJ the role of a DNA repair protein and suggested contributions in methyl-directed mismatch repair, base excision repair, homologous recombination and rescue of stalled replication forks (Chow and Courcelle, 2007; Cooper et al., 1993; Courcelle et al., 2003; Courcelle and Hanawalt, 1999; Lovett and Kolodner, 1989; Sutera et al., 1999; Wakamatsu et al., 2010; Yamagata et al., 2002). Studies show that the eukaryotic Cdc45 accumulates at stalled replication forks and, similar to its prokaryotic counterpart, RecJ, Cdc45 has been suggested to potentially be involved in the rescue of these stalled replication forks (Courcelle et al., 2003; Courcelle and Hanawalt, 1999; Pacek et al., 2006; Sanchez-Pulido and Ponting, 2011). Alternatively, Cdc45 may have adapted its ssDNA binding activity for

novel biological activities in initiation and elongation of the eukaryotic DNA replication fork (Sanchez-Pulido and Ponting, 2011). To date, most studies that have unraveled the evolutionary relationship between Cdc45 and RecJ are mainly bioinformatics analyses, and molecular properties of Cdc45 have only been analyzed with isolated recombinant human Cdc45 (Krastanova et al., 2012). Cdc45, together with GINS, have been shown to activate the pre-assembled Mcm2-7 helicase motor by significantly elevating the DNA binding and helicase activities by orders of magnitude in the CMG complex (Ilves et al., 2010). The striking conformational rearrangements of Mcm2-7 by itself versus the CMG complex have implicated roles for Cdc45 and GINS in helping the Mcm2-7 motor more efficiently grasp DNA as well as forming a second channel (Costa et al., 2011). The presence of this alternate channel led to the attractive hypothesis that the interior of the Mcm2-7 ring tracks on the separated leading strand while the lagging strand is guided through the side channel by the auxiliary factors (Costa et al., 2011).

The fact that these allosteric factors remain associated with the Mcm2-7 ring and part of the progressing replisome suggests that they might have additional roles beyond the initiation step. The Cdc45 protein has been shown to be essential for chromatin loading of replication factors like replication protein A (RPA), Polymerases alpha (Polα), epsilon (Polε) and delta (Polδ), as well as the proliferating cell nuclear antigen (PCNA) required for the establishment of the replisome (Bauerschmidt et al., 2007; Mimura et al., 2000; Nakaya et al., 2010; Pospiech et al., 2010; Tercero et al., 2000; Zou et al., 1997). Besides its direct interaction with the Mcm2-7 and GINS proteins and its role in replication initiation, the finding of *Drosophila* Cdc45 as a RecJ orthologue opened new questions regarding its role within the CMG helicase complex (Pesavento, 2013, manuscript in preparation). BLAST searches performed by Alessandro Costa have identified a strong sequence conservation between the catalytic amino-terminal domain of the bacterial RecJ and the N-terminus of Cdc45. Models of N-Cdc45 were threaded onto the N-terminal lobe of the RecJ crystal structure; a characteristic horseshoe structure, in which the N- and C-terminal lobes are connected through a long α-helix (Pesavento, 2013, manuscript in preparation). However, there is no sequence conservation between the C-termini of RecJ and Cdc45. Sequence conservation was found through a third related protein (an exopolyphosphatase) that is structurally conserved with the C-terminal lobe of RecJ. Modeling of Cdc45 based on this protein's structure allowed threading of the C-Cdc45 lobe onto the C-RecJ crystal structure (Figure 4.1A, kindly provided by Alessandro Costa). These findings allowed the independent generation of homology models for the N- and C-terminal lobes of Cdc45 and threading on *Thermus thermophilus* (Tth) RecJ crystal

structure. These modeling nicely fit into the electron density of Cdc45 from the CMG electron microscopic study (Figure 4.1B) (Costa et al., 2013, manuscript in preparation). This novel discovery raises the question in the DNA replication field whether Cdc45 displays any roles reminiscent of RecJ in DNA repair and motivate a more detailed analysis to gain more insight in its function as part of the CMG complex.

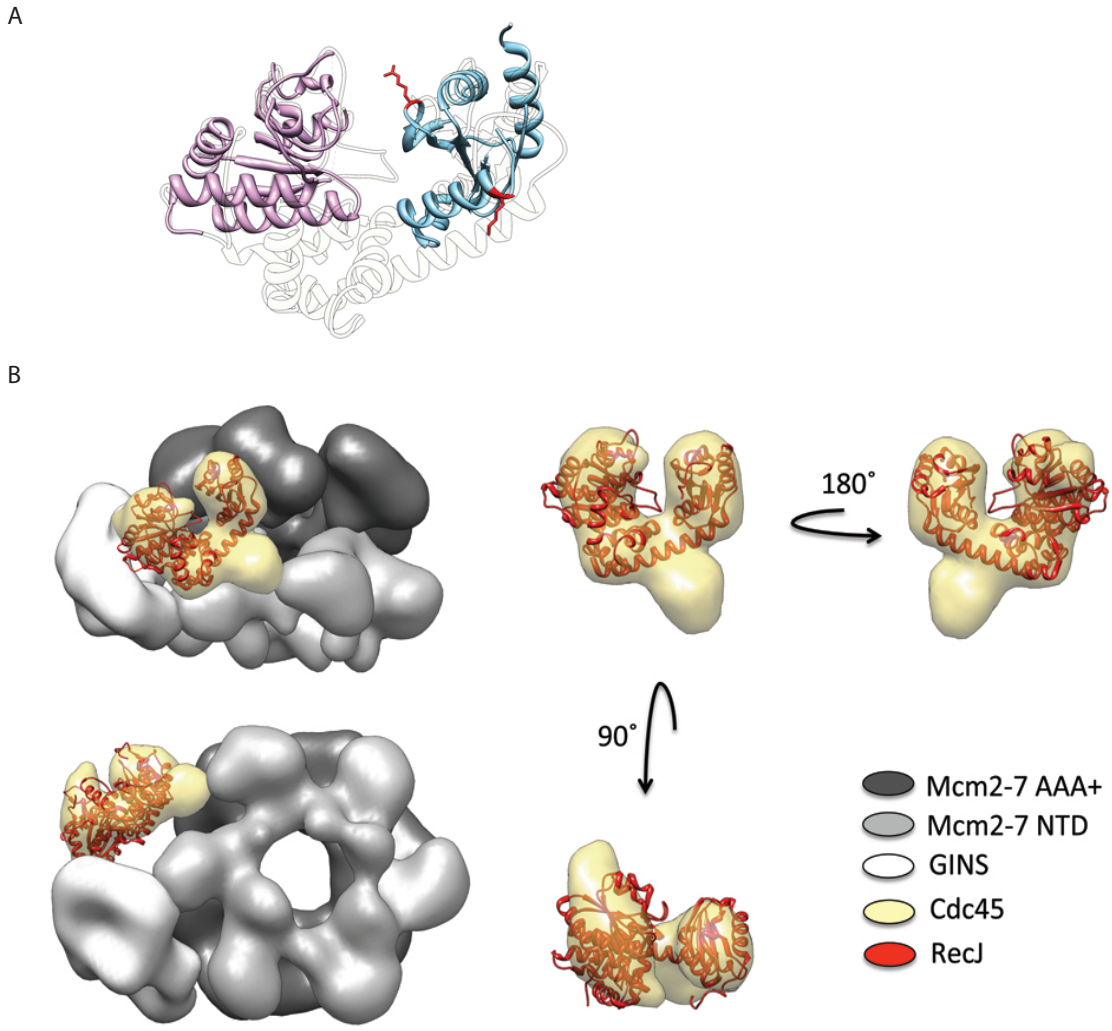

Figure 4.1. *Drosophila* Cdc45 is related to RecJ.
(A) BLAST searches have identified a strong sequence conservation between the catalytic amino-terminal domain of the bacterial RecJ and the N-terminus of Cdc45. Models of N-Cdc45 were threaded onto the N-terminal lobe of the RecJ crystal structure; a characteristic horseshoe structure, in which the N- and C-terminal lobes are connected through a long α-helix. Sequence conservation of the C-terminus between C-Cdc45 and C-RecJ was not found. Modeling was enabled through sequence conservation to a third related protein (an exopolyphosphatase) that is structurally conserved with the C-terminal lobe of RecJ. Modeling of Cdc45 based on this protein's structure allowed threading of the C-Cdc45 lobe onto the C-RecJ crystal structure.
These findings allowed the generation of a homology model for the N- and C- lobes of Cdc45 based on RecJ (PDB entry: 2zxo) and the RecJ-related exopolyphosphatase (PDB entry: 2qb6), respectively, and the threading on *Thermus thermophilus* (Tth) RecJ crystal structure
(B) AAA+ domain and N-terminal domain of Mcm2-7 proteins are shown in dark grey and light grey, respectively. Electron densities of GINS (*white*) and Cdc45 (*yellow*) are color-coded and the RecJ crystal structure is docked into electron density of Cdc45 protein within the CMG (PDB entry: 1IR6) (Yamagata et al., 2002). (Both Figures are generously provided by Alessandro Costa).

4.3 Results

CMG loads onto thymidines

Previous studies have shown that Cdc45, together with GINS, serve as necessary auxiliary factors that convert the Mcm2-7 ring into an active and stable helicase (Costa et al., 2011; Ilves et al., 2010). Both factors travel with the Mcm2-7 motor at the fork and recruit further factors (e.g. RPA, polymerases, Mcm10) required for establishing a fully functional replisome (Bauerschmidt et al., 2007; Mimura and Takisawa, 1998; Pospiech et al., 2010). The focus of this study is to understand in greater detail the role of Cdc45 during loading and translocation.

To this end, I first wished to confirm in my experimental set-up if CMG displays a preference with regard to geometry and sequence composition of various DNA substrates as has been found by James Pesavento in the Botchan lab (Figure 4.2A) (Pesavento et al., 2013, manuscript in preparation). This would allow directed loading of the complex on DNA. Previous studies established that *Drosophila* CMG (DmCMG) can bind to forked DNA substrates that contain overhangs of 40 thymidines ('polyT') on either both strands or on one strand only ('one-armed'), and to ssDNA with thymidine stretches. Yet, no binding could be observed to dsDNA (Ilves et al., 2010). Equivalent DNA substrates were designed with conjugated fluorescein groups, confirming that CMG can bind to all three listed substrate geometries as long as they offer ssDNA thymidine nucleotides. In this case, CMG does not distinguish between ssDNA on the 3' or 5' strand (Pesavento et al., 2013, manuscript in preparation). However, Moyer et al., (2006) have shown that DmCMG translocates only in the 3' to 5' direction. The forked substrate with thymidine stretches on both strands would allow CMG to bind on both. One of these species would be oriented in the correct way to initiate the unwinding of the substrate (3' binding) while with the other species (5'binding), CMG would fall off the substrate as soon as translocation initiated. The necessity evolved for substrates that allow a unique orientation of the CMG on DNA to guarantee that all CMG molecules bind the same way and separate the substrate. The same substrates with 50bp of dsDNA region and two 40nt ssDNA tails (our regular 'fork' substrate) were prepared while substituting one of the thymidine stretches with a G/C-rich nucleotide stretch at a time (Pesavento et al., 2013, manuscript in preparation). The binding efficiency of the CMG to the different substrates was compared and it was further tested whether CMG could unwind these G/C-rich substrates.

The affinity of CMG was not altered when the 3' or 5' polyT tail was substituted with a G/C-rich tail. However, only the substrate with 3'polyT and 5'G/C-rich region could be unwound by CMG, whereas the substrate with 3'G/C-rich region and 5' polyT did not get

displaced (Pesavento et al., 2013, manuscript in preparation). This suggests that the CMG is bound in the 'wrong' direction on the 5' polyT tail and falls off the substrate after translocation is initiated. No CMG appears to be bound on the G/C-rich ssDNA (Figure 4.2B) (Pesavento et al., 2013, manuscript in preparation*)*.

I tested these observations by designing different substrates that contain either one thymidine and one G/C-rich ssDNA tail or only one ssDNA GC-rich overhang (Figure 4.2A). CMG does not bind these substrates when only G/C nucleotides are available (Figure 4.2B) (Pesavento et al., 2013, manuscript in preparation). The data show that CMG preferably binds to thymidines and has no (or extremely weak) affinity for G/C rich strands for its loading. A preference for T-rich DNA substrates has been previously shown for other helicases (You et al., 2003).

These findings allowed the design of substrates that can distinguish between contacts made by the CMG with the leading strand versus those established with the lagging strand. This new substrate has a 50bp duplex, 40nt of polyT on 3', and 40nt G/C rich sequence on the 5' strand (Figure 4.2A). Only one CMG molecule can bind to the 3' strand and unwind the two strands.

To validate the hypothesis that exclusively one CMG is bound on this substrate, I tested this substrate side-by-side with the 'fork' substrate (3' and 5' polyT) in an electrophoretic-mobility-shift-assay (EMSA) while titrating the amounts of protein (Figure 4.2C). The replacement of the second thymidine-rich stretch with a G/C-rich stretch in this substrate changes the behavior of the fork at higher protein concentrations with this new substrate. A second super-shifted species is apparent, suggesting that two CMG molecules bound to the same fork (Figure 4.2C). This super-shifted band, indicative for a second CMG molecule bound to the same substrate, is only apparent at high protein/DNA ratios. The 'leading' strand substrate does not show this second shifted species even at higher protein/DNA ratios, confirming that CMG only binds to thymidines and that, on this substrate, only one CMG molecule is bound on the 3' strand.

In order to analyze the contacts of leading versus lagging strands with the CMG, fluorescein (FITC)-conjugated nucleotides were included at certain positions to allow for UV-induced protein:DNA crosslinks (Pesavento et al., 2013, manuscript in preparation). For testing contacts with the 'leading' strand, nucleotides at positions 3,6,9,15 were substituted with the modified FITC-labeled nucleotides on the 3' strand. For the 'lagging' strand contact points, these modified nucleotides were introduced at positions 5,10,15 on the 5' strand (Figure 4.2A).

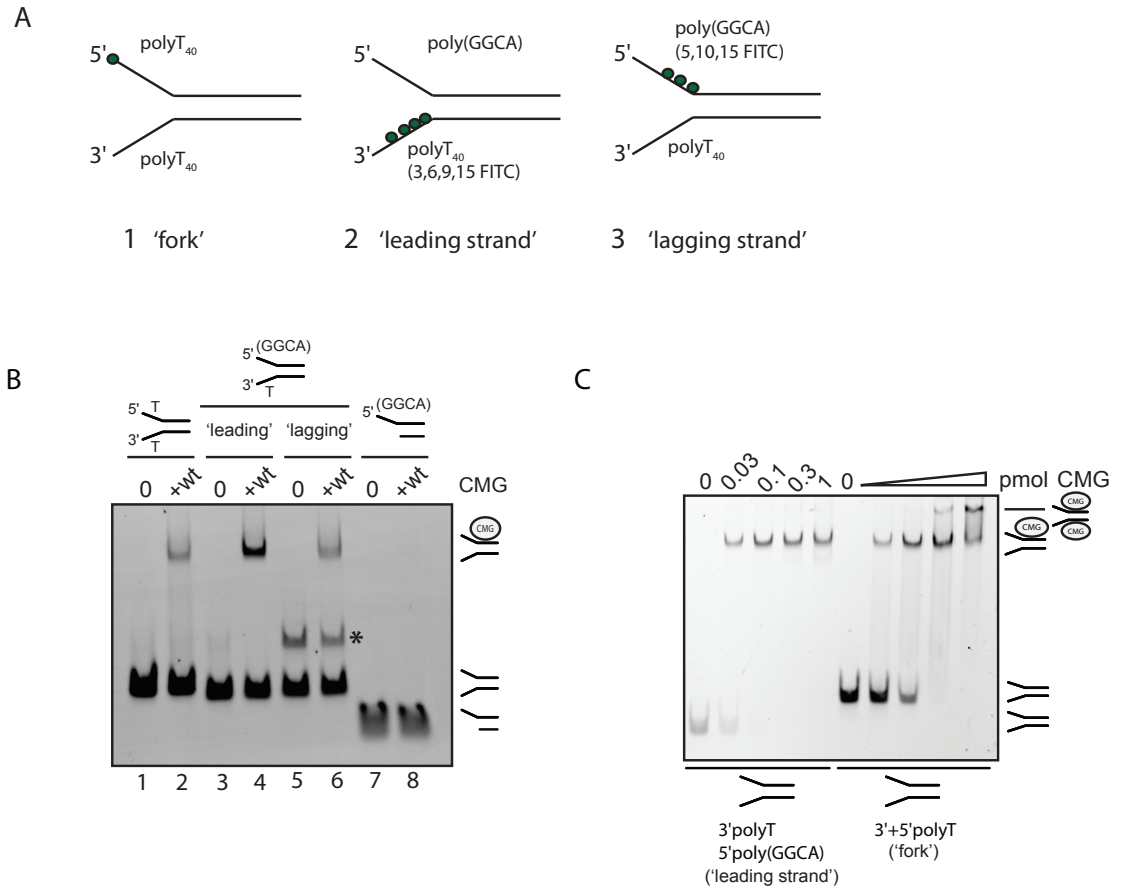

Figure 4.2. CMG loads onto thymidines.
(A) Schematic of different DNA substrates used in the following assays. The green dot in each substrate represents the position of the FITC-group on the nucleotide and induces cross-linking of DNA to proteins under UV; (1) 'fork' substrate contains 40nt of single-stranded poly thymidine extensions on 3' and 5' strands in addition to the double-stranded (50bp) region and one FITC group at the end of the 5' strand; 'leading strand' and 'lagging strand' contain the same double stranded region as (1) and both contain a polyT overhang on 3' strand, but a GC-rich sequence on the 5' strand. The difference between these two substrates are the amounts and positions of the FITC-modified nucleotides: (2) 'leading strand' has 4 FITC groups on the 3' strand at positions 3,6,9,15 away from the ss-ds intersection, (3) 'lagging strand' has 3 FITC groups on the 5' strand at positions 5,10,15 away from the ss-ds.
(B) Different forked substrates were tested for CMG binding in an EMSA using 100fmol of DNA and 100fmol of protein. '0' indicates free substrates without protein. CMG binds to forked substrates with 3'+5'polyT single-stranded extensions of the 50bp duplex DNA ('fork') (lane 2). Substituting the 5' polyT tail with a (GGCA)-rich extension still allows the CMG to bind to the fork (lanes 4 and 6). A one-armed substrate offering only a 5'(GGCA) extension does not allow CMG to bind (lane 8). Since CMG translocates in 3' to 5' direction the CMG can be directed to the 'correct' 3' to 5' binding on a forked substrate with a 3' polyT and 5' (GGCA) overhang that leads to subsequent unwinding upon nucleotide hydrolysis. The asterisk indicates an unspecific band. 'Leading' and 'lagging' strand substrates differ in the positions of the fluorescent groups. The varying intensities can be attributed to the differing number of fluorescent groups on the different substrates used in this assay.
(C) The indicated amounts of wtCMG protein were incubated with 50fmol of 'leading' strand substrate or 100fmol of '3'+5' polyT' ('fork') substrate. The 'fork' substrate displays a second super-shifted species in the gel-shift assay, indicating a second CMG complex bound to the same substrate. The super-shift is not present in the 'leading' strand substrate that presents only one polyT tail for the CMG to bind. Signals were detected by FITC fluorescence signals, and positions of free forked DNA substrate and CMG bound substrate are indicated on the side.

Cdc45 within the context of the CMG contacts the leading strand

A central question of interest in the field of DNA replication is to understand the path of DNA in the CMG complex. A study from the Walter lab demonstrated that, for the *Xenopus laevis* (Xl) helicase, the leading strand enters the central channel, whereas the lagging strand is excluded (Fu et al., 2011). The structural analysis of the *Drosophila* (Dm) CMG suggested that the central channel is too narrow to encompass double-stranded DNA and revealed a second smaller channel formed by Cdc45/GINS on the side of the Mcm2-7 ring (Costa et al., 2011). As Dm CMG shows no binding to duplex DNA, I hypothesized that, like in *Xenopus*, the *Drosophila* Mcm2-7 motor encircles one strand while the other is excluded from the central channel. The lagging strand could be guided through the side channel by the auxiliary replication factors as was previously hypothesized (Costa et al., 2011).

To this end, it was necessary to develop an assay that captures the individual DNA contacts made by CMG. This assay would further help us investigate the ability of the auxiliary factors Cdc45 and GINS to bind DNA in the context of the active CMG helicase and to understand, which, if any, strand they might contact. I used the designed 'leading' and 'lagging' strand substrates with the modified fluorescein isothiocyanate (FITC) groups at specific residues along the 3' or the 5' strand to analyze the binding to either strand separately (Figure 4.2A and Figure 4.3). Inducing protein:DNA cross-linking of these groups by UV allowed me to monitor which proteins in the CMG establish contact points with the DNA substrates. If our previous model were correct, one would have anticipated Cdc45 and GINS to bind to the lagging strand.

Therefore, first the binding of the leading strand was looked at, where the majority of the CMG binding is anticipated to be contributed by the Mcm2-7 ring. Previous reports from the Botchan lab established that the CMG shows ATP-dependent DNA binding and that upon nucleotide addition the Mcm2-7 ring closes and tightens the central channel (Ilves, et al., 2010, Costa et al., 2011). To test the two structural modes the cross-linking was performed in the absence and presence of nucleotide side-by-side. Even loading of proteins was confirmed by staining the gel after SDS-PAGE with SYPRO® red (Figure 4.3B). Without nucleotide, only a few Mcm2-7 subunits cross-linked to DNA, while in the presence of DNA more subunits established contacts with the DNA (Figure 4.3A). This observation is anticipated due to the tightening of the central channel, which would lead to stronger engagement of the leading strand by the Mcm2-7 ring.

The same experimental setup with the lagging strand revealed a similar switch upon nucleotide addition, but the pattern of the cross-linked bands was different, suggesting that different subunits contact the two strands (Figure 4.3A). Furthermore, with both substrates only the Mcm2-7 subunits were detected upon short exposure, suggesting that

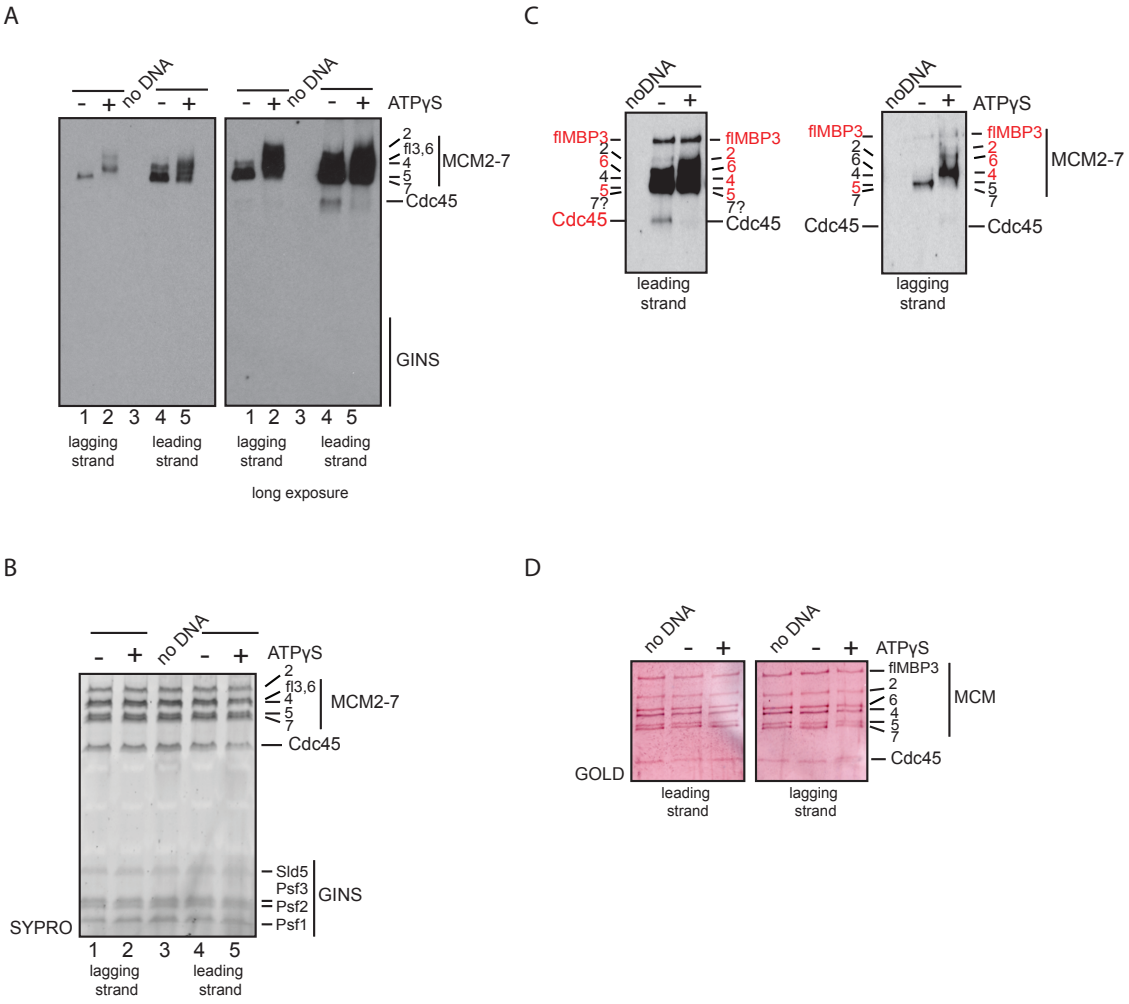

Figure 4.3. Cdc45 within the CMG complex contacts the leading strand.
(A) Western blot analysis of forked FITC-labeled 'leading' and 'lagging' DNA substrates cross-linked to wt CMG by UV in the absence and presence of ATPγS. Reactions were separated by SDS-PAGE (10%) after digestion of cross-linked DNA by micrococcal nuclease. The immunoblot was probed with a monoclonal α-FITC antibody to detect cross-linked DNA-protein bands. Specific MCM proteins cross-linked to either the 'leading' or 'lagging' strand substrates. The long exposure (*right panel*) shows that Cdc45 exclusively cross-links to the 'leading' strand substrate. GINS are not detected to cross-link to either substrate.
(B) SDS polyacrylamide gel (10%) stained with SYPRO® red to visualize proteins after the western transfer in Figure 4.3A.
(C) Western blot analysis of FITC-labeled 'leading' and 'lagging' DNA substrates cross-linked to CMG with MBP-FLAG-Mcm3. Reactions were separated on a SDS polyacrylamide gel (10%), and bands were detected with an α-FITC antibody. Proteins cross-linked to DNA are highlighted in red on either side of the immunoblot (*left panel* '-' ATPγS, *right panel* '+' ATPγS).
(D) Immunoblots of Figure 4.3C were stained with AuroDye™ (GE Healthcare) to visualize protein bands.

the majority of the DNA binding and strongest affinities of the CMG complex are to the Mcm2-7 subunits. To check if Cdc45 and GINS contribute with lower affinities to the DNA binding on either strand, longer exposures of the Western blot were looked at. Neither Cdc45 nor GINS cross-linked to the lagging strand in the absence or presence of the non-hydrolyzable nucleotide analogue ATPγS (Figure 4.3A, lanes 1, 2, long exposure). However, surprisingly, it was found that Cdc45 contacts the leading strand in the absence of nucleotide (Figure 4.3A, lane 4, long exposure). Upon addition of ATPγS in the DNA binding assays, a dramatic decrease of cross-linked Cdc45 was observed (Figure 4.3A, lane 5, long exposure). This observation fits with our model that there is a closure of a gap present between Mcm2 and 5 in the Mcm2-7 ring in the larger CMG complex (Costa et al., 2011). These data suggest that Cdc45 binds the leading strand, which is engaged by the central channel of the Mcm2-7 motor, but that this contribution only occurs when the ring is not entirely closed in the absence of nucleotide and allows a direct contact to the leading strand via the ring opening. Upon '2/5 gate' closure, the leading strand is fully encircled by the Mcm2-7 ring, which now tightly locks down on the DNA and excludes Cdc45 from making contacts.

As previously mentioned, only a subset of Mcm2-7 subunits was observed to contact leading and lagging strands without nucleotide, and several more establishing contact when nucleotide was added to the reaction. Our wildtype (wt) CMG complex contains a FLAG-tagged version of Mcm3 and co-migrates with Mcm6 on an SDS-PAGE gel (Figure 4.3A). I wished to separate the two bands in order to distinguish which subunit of the two was binding DNA. For this purpose, I purified a CMG with an MBP group upstream of FLAG-Mcm3. The cross-linking with leading and lagging strand substrates was repeated and confirmed that the Mcm2-7 proteins bind to both strands, whereas Cdc45 only binds to the leading strand without nucleotide (Figure 4.3C). Mcm5 is the predominant subunit that cross-links to both strands without ATPγS. The leading strand is also contacted by Mcm3 and Mcm6 proteins. Upon addition of nucleotide, a binding of all other Mcm2-7 subunits to DNA is observed, as well as an exclusion of Cdc45 analogous to the previous observations with wtCMG. In the case of the lagging strand, there is a switch from Mcm5 (and to a lesser extent Mcm3) to mainly Mcm4 (and to a lesser extent Mcm2, Mcm3, and Mcm6) upon ATPγS binding. Even loading of protein was confirmed by AuroDye™ (GE Healthcare) staining after immunoblotting (Figure 4.3D). Because the band detected for the Mcm5- leading strand DNA cross-link on the Western blot is very strong and close to the Mcm7 protein band, I could not rule out the possibility that Mcm7 also makes contacts with DNA. In order to distinguish between these two proteins, a CMG complex containing an mCherry at the C-terminus of Mcm5 was purified by James Pesavento

(Pesavento et al., 2013 manuscript in preparation). The addition of this tag allows for separation of Mcm5 from Mcm7. Upon cross-linking of this CMG complex to the leading strand, he could confirm that Mcm7 also binds to DNA in the absence and presence of ATPγS, but to a lesser extent than Mcm5 (Pesavento et al., 2013 manuscript in preparation). Moreover, the experiments thus far indisputably confirmed that Mcm5 is the strongest DNA binding Mcm2-7 subunit in absence of nucleotide.

Mutations in the Mcm2-7 subunits affect DNA binding of Cdc45 in the context of the CMG

An unanticipated finding was that none of the auxiliary factors Cdc45 or the GINS proteins contact the lagging strand. This discovery led me to rule out our previous hypothesis that the lagging strand might be guided through the second channel. Also, I further ruled out that the GINS proteins contribute to any direct DNA binding as part of the CMG complex.
But, then, what role does the contact between Cdc45 and the leading strand in the absence of nucleotide play? Does Cdc45 help guide the leading strand into the central channel? Our hypothesis was revised for the role of Cdc45 as a 'guardian' of the Mcm2/5 gate. In this capacity, Cdc45 would be required when the lagging strand is excluded from the central channel after the duplex is melted or whenever the gate opens during translocation, if this is the case. Cdc45 might help 'catch' the leading strand to ensure it remains in the central channel and preventing it from slipping out.
To test this hypothesis, I made CMG mutants that have impaired the DNA binding affinity by introducing mutations in DNA binding motifs. If these mutant CMG complexes shift the binding equilibrium away from the central cavity, I anticipate that the leading strand should be more likely moved towards the outside when the gate is open. The first mutant tested was a CMG complex that contains alanine substitutions in a conserved lysine in the Pre-Sensor1 β-hairpin (PS1-hp, Figure 4.4A). This hairpin was previously reported to participate in the helicase activity of an archael MCM helicase (McGeoch et al., 2005). Mutation of this residue in archaea decreased DNA binding and abolished helicase activity, suggesting a role for this hairpin in the transduction of ATP hydrolysis to translocation of MCM on DNA. This suggests that, if this residue is important for DNA binding to the leading strand by the DmMcm2-7 ring within the CMG, assays that test the binding to DNA should reflect a diminishment. Upon substitution of these conserved residues in all six *Drosophila* Mcm2-7 subunits with alanines ('6xPS1' CMG) and testing the mutant CMG in a gel-shift assay, I found that the DNA affinity of this '6xPS1' CMG complex was significantly reduced (Figure 4.4C, about 7-8% of wildtype). Next, I analyzed the helicase activity of this complex. As anticipated from the abolished DNA binding affinity of this complex, I found that its helicase activity was also abrogated (Figure 4.4D, about 5-6% of wildtype).

Compared with the mild defects in DNA binding caused by the PS1-hp mutations in in archaea, the degree of defectiveness in DmCMG suggests that the targeted positive residues in the PS1 hairpins strongly contribute to the DNA binding affinity of DmCMG, and also to the translocation of the Mcm2-7 motor along DNA. I performed a cross-linking assay with the FITC-labeled leading strand substrate using the '6xPS1' mutant CMG. Equal loading was checked by staining of the Western blot AuroDye™ (GE Healthcare) (Figure 4.4F).

Compared to the wildtype CMG, I observed a significant decrease in cross-linking of the Mcm2-7of ring in the '6xPS1' CMG complex to about 20-30% of wildtype with or without ('-/+') ATPγS. Surprisingly, not only was the cross-linking of Cdc45 to the leading strand enhanced within the 6xPS1 CMG, but most strikingly, this mutant CMG complex maintained the contact to DNA even after addition of the nucleotide analogue, unlike wildtype CMG (Figure 4.4E, lanes 4 and 5).

I wished to understand if this changed cross-linking behavior of Cdc45 to the leading strand was due only to the decreased DNA binding affinity of the 6xPS1 mutant CMG. My assumption is that we might observe a shift in the binding equilibrium for the leading strand from inside the Mcm2-7 central channel to outside of it and that the leading strand might be 'pushed' through the Mcm2/5 gate when open. If, however, the reduced DNA affinity of the mutant CMG was not solely responsible for the increased Cdc45-DNA interaction, the altered pattern of Cdc45 could indicate an additional structural defect that could influence the closure of the Mcm2/5 gate. The state of this gate opening might become prolonged when these mutations are introduced, allowing enhanced interactions of Cdc45 with the leading strand. This shift and gate defect then is reflected in the binding patterns of the leading strand.

To test this idea, I analyzed other mutant complexes that were entirely defective for DNA binding. The introduction of alanine substitutions in the conserved lysine residues in the ATP binding Walker A motifs (previously reported in Ilves et al, 2010, Figure 4.4B) of all six Mcm2-7 subunits simultaneously ('CMG 6xKA') resulted in abolished DNA binding (Figure 4.4C) and, hence, abrogated helicase activity (Figure 4.4D). In the cross-linking assay, the six-fold KA mutant CMG displayed a decrease in cross-linking of the Mcm2-7 subunits with the leading strand to similar levels as the six-fold PS1 CMG (Figure 4.4E, lanes 6 and 7). This observation confirmed the abrogated DNA binding affinity of this '6xKA' CMG mutant as observed by gel-shift assays (Figure 4.4C). However, the cross-linking of Cdc45 to the leading strand within the '6xKA' CMG complex showed no increase in signal in the absence or presence of ATPγS as compared with wt CMG (Figure 4.4E, lanes 6 and 7, long exposure).

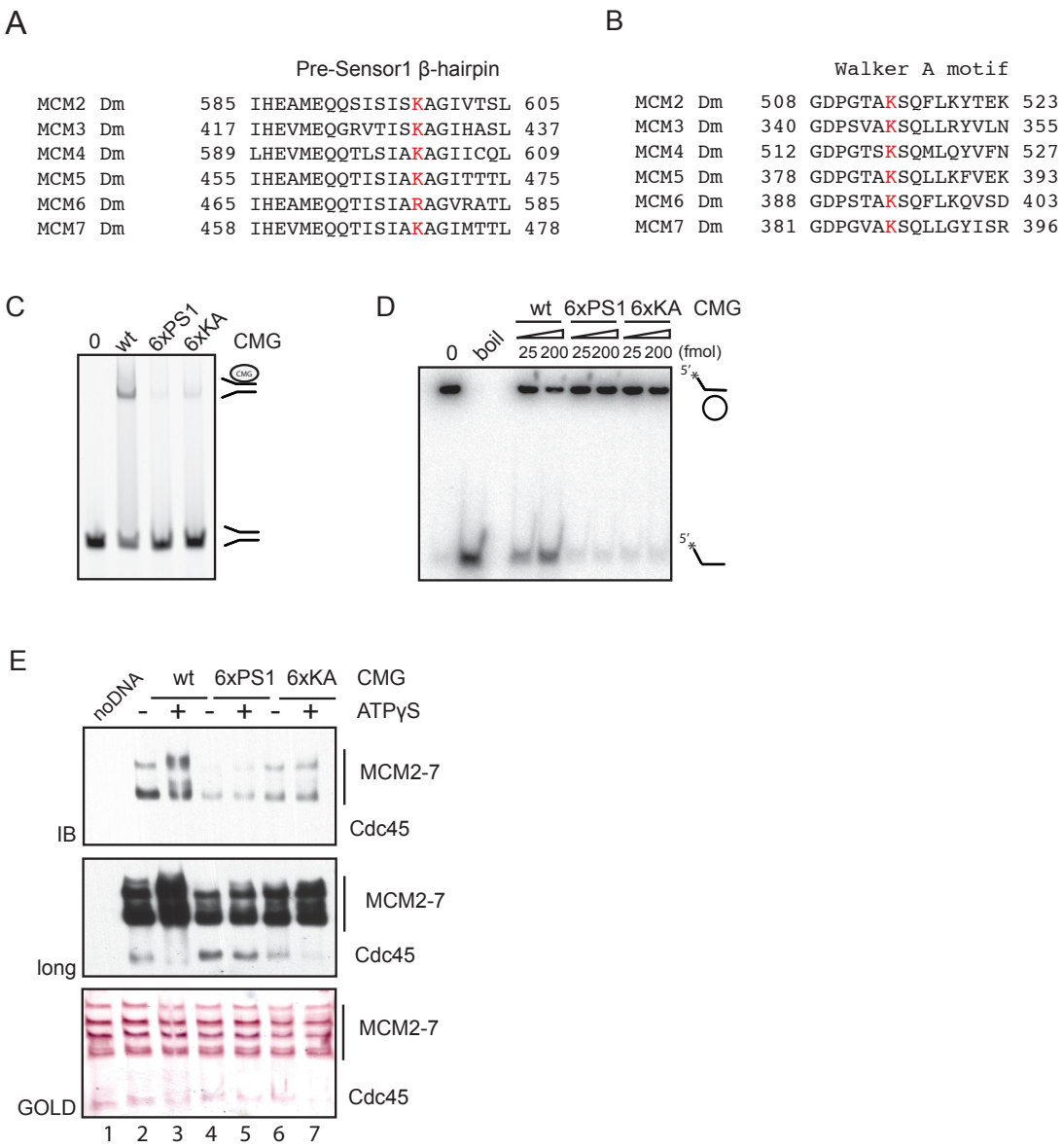

Figure 4.4. Mutations in Mcm2-7 affect the DNA interaction of Cdc45 in the CMG.
(A) Sequence alignment of the Pre-Sensor1 β-hairpin regions of all six *Drosophila* (Dm) Mcm2-7 proteins. A critical basic residue of the hairpin (K or R in *red*) has been substituted with an alanine.
(B) Sequence alignment of the Walker A box of all six *Drosophila* Mcm2-7 proteins. The conserved lysine of the ATP binding motif (K in *red*) has been mutated to an alanine.
(C) EMSA of wtCMG and CMG complexes with mutations in the Pre-Sensor1 β-hairpin of all six Mcm2-7 proteins ('6xPS1') or mutations in the Walker A motifs of all six Mcm2-7 subunits ('6xKA'). '0' indicates a control lane without protein. All lanes contain 100fmol of FITC-labeled forked DNA substrate (3'+5'polyT, 'fork') and, where indicated, 100fmol of purified CMG protein in the presence of 10μM ATPγS. Reactions were separated on a native TBE polyacrylamide gel (4%), and signals were detected by FITC fluorescence on Typhoon™9400. Positions of free 'fork' substrate and CMG-bound substrate are indicated on the side of the gel.
(D) Autoradiograph of a helicase activity assay performed on a circular DNA substrate based on the M13 ssDNA. The positions of free circular substrate and displaced oligonucleotide in the gel are indicated on the side. Two amounts of purified CMG proteins (25, 200fmol) were added to 1fmol of DNA substrate and 300μM ATP. Reaction products were separated by TBE-PAGE (8%). '0' and 'boil' contain controls of substrate without protein or boiled substrate, respectively.
(E) Western blot analysis of FITC-labeled 'leading' strand substrate cross-linked to the indicated purified CMG proteins by UV. Reactions were separated by SDS-PAGE (8%). The immunoblot was probed with monoclonal α- FITC antibody to detect cross-linked DNA:protein bands. The Western blot was stained with AuroDye™ (GE Healthcare) to visualize the proteins and is shown below.

Taken together, the results of the two six-fold CMG mutants in the Walker A and Pre-Sensor1 β-hairpin motifs suggest that the decrease in the DNA affinity alone is not responsible for the enhanced cross-linking of Cdc45 to the leading strand in the context of the '6xPS1' CMG. Rather, the striking effect of the Pre-Sensor1 β-hairpin mutations on Cdc45's contacts to the leading strand most likely results from a combination of two factors. In addition to shifting the binding equilibrium of the leading strand from within the Mcm2-7 inner channel to outside the channel, the mutant CMG might be defective in closing the gate created by Mcm2/5. Such a defective gate could account for the inability of the Mcm2-7 ring to efficiently close down on the DNA substrate and, therefore, account for a stronger participation of Cdc45 in contacting the strand by catching the exiting leading strand. Thus, I hypothesize that one role of Cdc45 is as a 'guardian of the gate' that holds the leading strand in place *in vivo* either while the lagging strand is excluded after the origin melting or when the '2/5 gate' is actively opened, which could occur during translocation or when the helicase encounters stalling.

The lagging strand wraps around the exterior of the Mcm2-7 ring

Since the new findings from the cross-linking excluded that Cdc45 and GINS help in guiding the lagging strand through the side channel, I sought to understand the path of this segregated strand outside the internal Mcm2-7 channel.

Several studies have shown that the lagging strand is excluded from the central channel while the MCM motor tracks on the leading strand and have culminated in the attractive 'steric-exclusion' unwinding model (Brewster et al., 2008; Graham et al., 2011). Data from archaeal organisms have suggested that this excluded lagging strand wraps around the outer surface of the MCM ring. I wished to test the validity of such a model for the DmCMG complex to understand if this was the path of the lagging strand captured in the cross-linking assays. The overall affinity of the lagging strand for the CMG members appeared weaker compared to that of the leading strand, suggesting that the greatest contributor to the overall DNA binding affinity is the binding of MCM subunits to the leading strand (Figure 4.3A, compare lanes 1,2 to 4, 5). In keeping with this model, the CMG does not unwind the one-armed substrate, which offers the necessary 3' thymidine overhang but lacks a 5' ('lagging strand') overhang (Pesavento et al., 2013, manuscript in preparation). Given the fact that CMG binds equally efficient to substrates with 3' and 5' overhangs or one overhang alone as long as thymidine stretches are present, one possibility is that the lagging strand might play a crucial role in stabilizing the complex during translocation. Additionally, the MCMs could help in guiding the lagging strand after separation of the parental strands along a discrete path that is required for establishing the correct physical

environment for other proteins at the progressing replication forks. Unfortunately, we have no assay to directly test the stability of the complex on DNA during translocation, but I wished to test the latter possibility. If the excluded strand wraps around the outside of the Mcm2-7 ring and is, thus, guided during fork movement, I anticipate a certain path for the DNA. Previous studies in archaea have suggested residues that contribute to the binding of the lagging strand to the outside of the MCM ring (Graham et al., 2011). All six *Drosophila* Mcm2-7 subunits possess an external β-hairpin on their outer surface. Based on the data from archaea, I targeted a conserved basic residue in the external β-hairpin of all six Mcm2-7 subunits simultaneously and purified this CMG mutant complex '6xEXT' (Figure 4.5A and B). First, DNA binding of this mutant complex was tested in a gel-shift assay. Given the observation that the majority of the affinity for DNA binding is established through the internal channel of the Mcm2-7 proteins, I reasoned that the DNA affinity should remain unaffected with these mutations that alter residues on the outside of the Mcm2-7, unless the mutations induce a structural defect on the Mcm2-7 interior. As anticipated, this '6xEXT' mutant CMG showed wildtype DNA binding (Figure 4.5C). However, when I tested the DNA helicase activity, I found the activity reduced to about 40-50% of the wildtype activity on a circular M13-based DNA substrate (Figure 4.5D). This observation, together with the fact that CMG does not unwind substrates lacking the 'lagging' strand, corroborate a hypothesis that the lagging strand is indeed required to stabilize the CMG during translocation. To test this hypothesis further, I assayed the '6xEXT' CMG for cross-linking to leading and lagging strands. Interaction with the leading strand was not affected; however, the cross-linking to the lagging strand showed a significant decrease in the Mcm2-7-DNA contacts (Figure 4.5E and F). Taken the results together, the data suggest that mutations in the conserved lysines in all of the Mcm2-7 external β-hairpins drastically weaken the interactions of this excluded strand with the outer surface of the Mcm2-7 ring. Given that cross-linking is not entirely abolished, I, however, hypothesize that there might be alternate residues in close proximity that account for additional binding and that one would need to mutate all of these residues to entirely abrogate contacts between the lagging strand and the Mcm2-7. My data strongly suggest that not only does the lagging strand wrap around the outside of the Mcm2-7 ring, but specifically that it wraps from the '2/5' gate toward Mcm6 and Mcm4. Without nucleotide, Mcm5 establishes the main contact with the lagging strand, suggesting that the strand is kept in close proximity to the 'gate' which could be the scenario after the initial melting step when the lagging strand needs to be excluded from the central channel. After the addition of ATPγS and gate closure, I observe a switch in the main interacting partner to the Mcm4 subunit (Figure 4.3A). While this still does not resolve the question of the

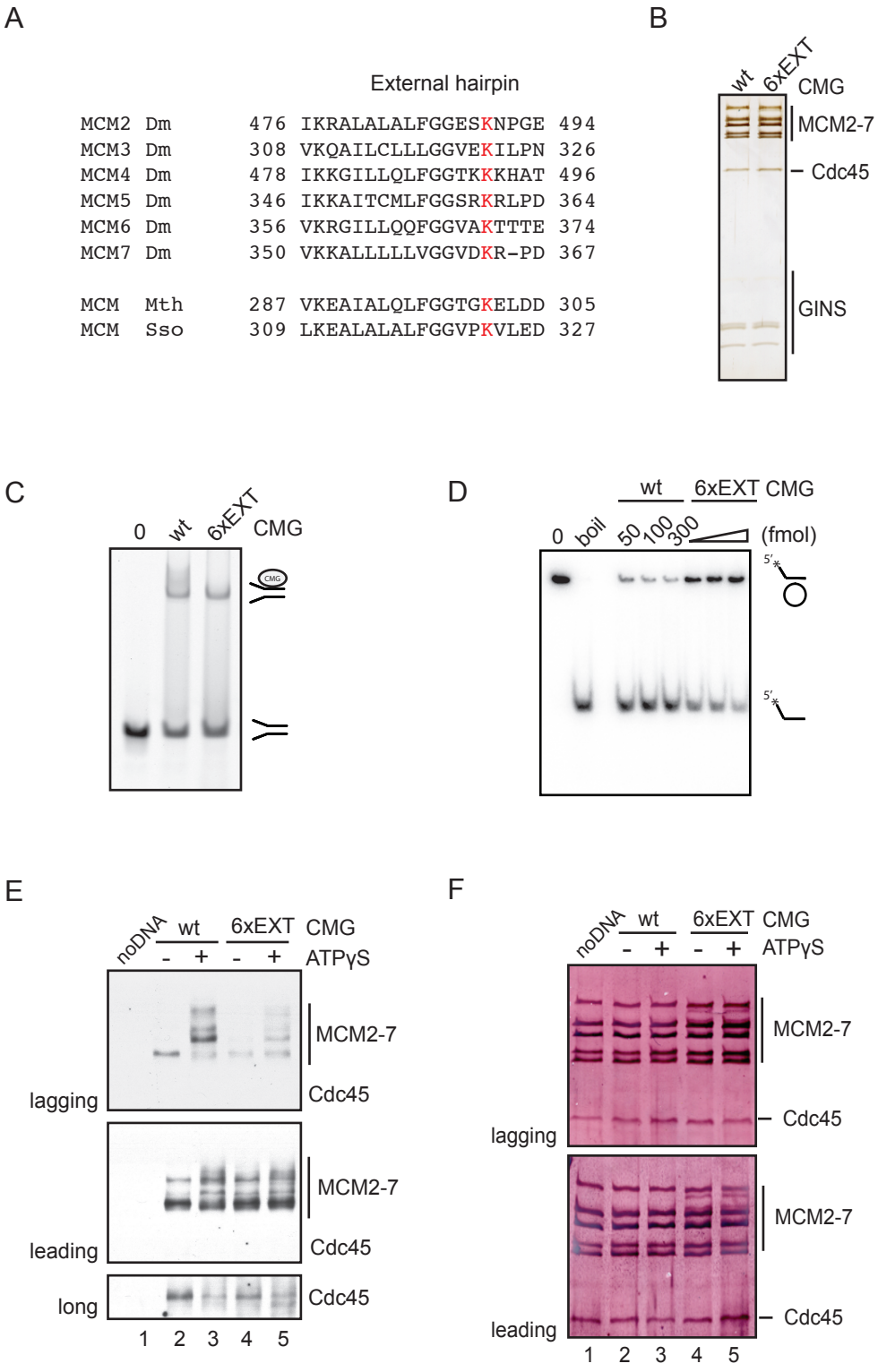

Figure 4.5. The lagging strand wraps around the external surface of Mcm2-7.
(A) Sequence alignment of the external ß-hairpin ('EXT-hp') regions of all six *Drosophila* (Dm) Mcm2-7 proteins and archaeal MCMs of *Methanobacter thermoautotrophicus* (Mth) and *Sulfolobus solfataricus* (Sso). A conserved lysine in EXT-hp was substituted with an alanine.
(B) Purified wt and '6xEXT' (alanine substitutions in all six Mcm2-7 subunits) rCMG were separated on a SDS polyacrylamide gel (10%) and silver-stained. Sld5 protein stains weakly, but its presence was previously confirmed through SYPRO® red staining (Invitrogen) (data not shown).

(C) Gel-shift of wt and '6xEXT' CMG complexes. '0' indicates the control lane without protein. All lanes contain 100fmol of 'leading' strand substrate, and, where indicated, 100fmol of purified CMG protein in the presence of 10μM ATPγS.
(D) Autoradiograph of a helicase activity assay performed on the circular DNA substrate. Increasing amounts of purified CMG proteins (50, 100, 300fmol) were added to 1fmol of DNA. Unwinding was initiated by the addition of 300μM ATP and reaction products were separated by TBE-PAGE (8%). '0' and 'boil' are controls without protein or boiled substrate, respectively. Positions of substrate and displaced oligonucleotide are indicated on the side.
(E) Western blot analysis of FITC-labeled 'leading' and 'lagging' strand substrates cross-linked by UV to '6xEXT' CMG complex. Reactions were separated by SDS-PAGE (8%). DNA-protein crosslinks were detected with a monoclonal α- FITC antibody.
(F) The Western blot shown in Figure 4.5E was stained with AuroDye™ (GE Healthcare) to visualize the CMG proteins.

precise way the strand wraps around the Mcm2-7, I can postulate the directionality of wrapping based on the fact that I never observe any Cdc45 or GINS cross-linking to the lagging strand. Hence, I hypothesize that if the strand were to wrap in the Mcm5-3-7-4 direction one would detect contacts with the auxiliary factors as well as with Mcm3 and Mcm7. Instead, mainly interactions with Mcm5-2-6-4 are captured, suggesting that in this 'snapshot' with the non-hydrolyzable ATP analogue, the lagging strand wraps away from Mcm5 and around Mcm2, 6 and 4 (Figure 4.5E). It should be emphasized that one cannot rule out the possibility that the CMG establishes contacts between Cdc45/GINS and the lagging strand once translocation is initiated and the CMG begins to move along DNA.

Specific residues in Cdc45 within the CMG complex directly contact the leading strand

As discussed, we found that Cdc45 directly contacts the leading strand; I further wished to identify the residues responsible for this DNA binding.

Homology modeling of Cdc45 N- and C-terminal lobes onto the *Thermus thermophiles* (Tth) RecJ crystal structure shown in Figure 4.1 helped identify two putative DNA binding residues in Cdc45, K363 and R419. These two residues superimpose onto two ssDNA binding residues in TthRecJ, R350 and H394 (Figure 6A, highlighted in red) (Yamagata et al., 2002). Primary sequence alignment of Cdc45 between different species (*Homo sapiens*, *Xenopus laevis*, *Mus musculus* and *Drosophila melanogaster*) showed that these two putative DNA binding residues, K363 and R419, are absolutely conserved, and I wished to test if they directly bind to the leading strand (Figure 4.6B, highlighted in red). To this end, I substituted the two residues in Cdc45 with alanines and purified CMG complexes containing either single ('K363A' or 'R419A') or double ('45-2x', K363A+R419A) mutations (Figure 4.6C).

As I believe that the majority of DNA binding is established by the leading strand interacting with the Mcm2-7 ring, and that the lagging strand contributes little, if anything, to DNA affinity, I anticipated that the Cdc45 mutants should have little or no effect on overall DNA binding affinity, unless these alanine substitutions induce a conformational change in the Mcm2-7 or alter the gate.

As shown in Figure 4.6D, the CMG complexes with the Cdc45 mutations displayed wildtype DNA binding in the gel-shift assays. I further measured the on-rate of the different complexes to test if they follow the same kinetics as wtCMG and if the Cdc45 K363A and R419A introduce any defects in loading the CMG onto the leading strand. These on-rate experiments demonstrated that the Cdc45 mutations did not affect the loading kinetics of the CMG (Figure 4.6E). Based on these results, I argue that the Mcm2-7 ring is unaffected within these CMG complexes, and I wanted to test the Cdc45 mutant

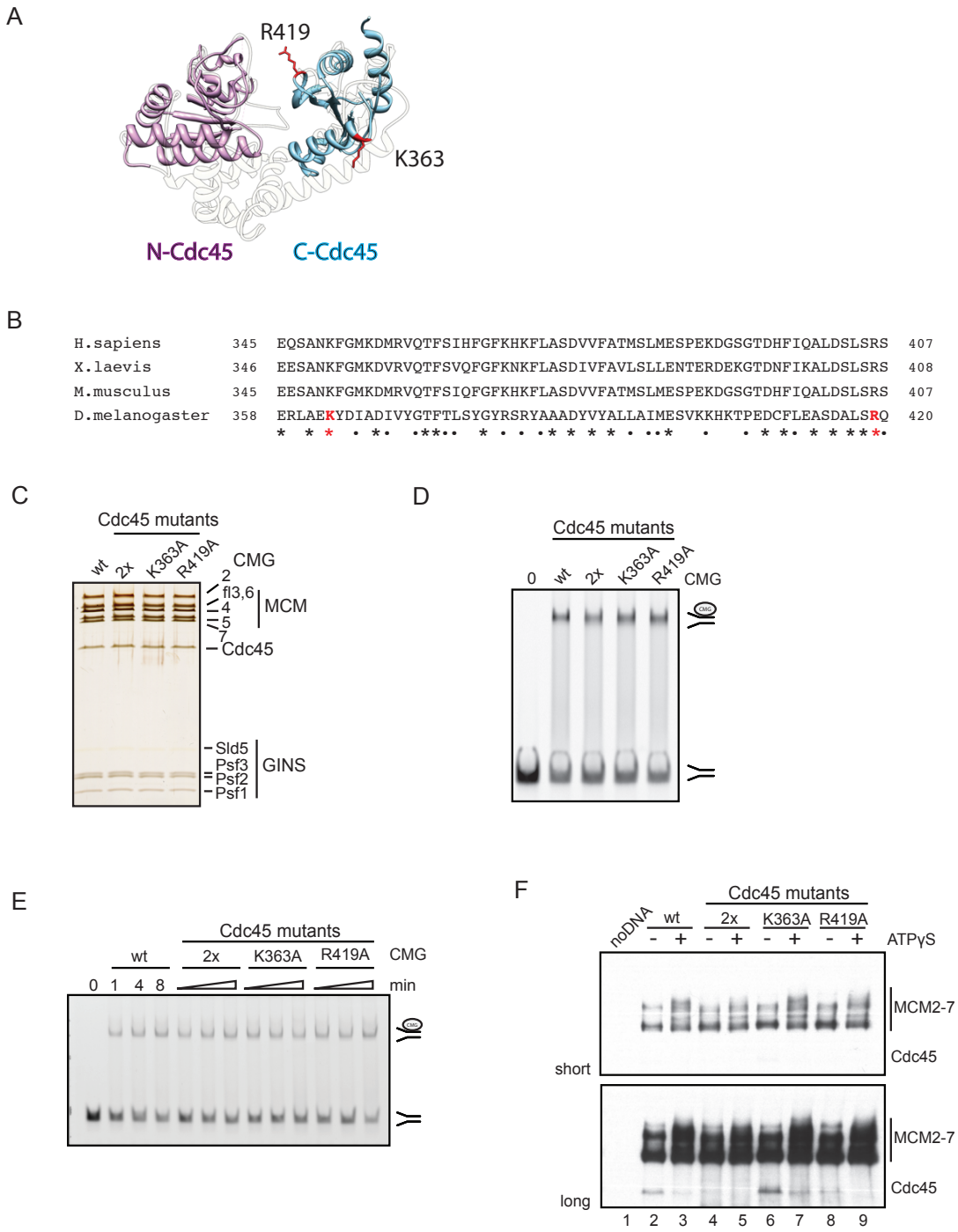

Figure 4.6. Specific Cdc45 contact the leading strand in the CMG.
(A) N- (*magenta*) and C-terminal (*cyan*) lobes of *Drosophila* Cdc45 homology models are superimposed onto the crystal structure of *Thermus thermophilus* RecJ (Tth) (*translucent*). Two putative DNA binding residues K363 and R419 are depicted as sticks (*red*) and superimpose with the two TthRecJ DNA binding residues R350 and H394, respectively.
(B) Sequence alignment of Cdc45 proteins from *Homo sapiens* (Hs), *Xenopus laevis* (Xl), *Mus musculus* (Mm) and *Drosophila melanogaster* (Dm) shows regions where point mutants in Cdc45 were introduced. The red asterisks indicate the two basic residues that have been mutated to alanines (K363 and R419) in DmCdc45 (in *red*). Both residues are absolutely conserved (asterisks indicate sequence conservation, dots indicate similar residue).
(C) Silver-stained SDS polyacrylamide gel (10%) of purified wtCMG and CMG complexes containing mutations in Cdc45. '2x' indicates the mutant CMG complex with alanine substitutions in both Cdc45 residues K363 and R419.

(D) EMSA of wildtype and CMG complexes containing Cdc45 mutants. '0' indicates a control lane with substrate alone, and all other lanes contain 100fmol of FITC-labeled 'fork' with 100fmol of CMG protein in the presence of 10μM ATPγS. Reaction products were separated on a native TBE gel (4%). The positions of free fork substrate and substrate bound by CMG are indicated on the side of the gel.
(E) Loading of various CMG complexes (wt and Cdc45 point mutants) onto a forked DNA substrate was tested in a gel-shift assay. Binding reactions were incubated for 1, 4, or 8 minutes, and reactions products were separated on a native TBE gel (4%).
(F) Western blot of cross-linked 'leading' strand DNA substrate to CMG containing single point mutations in the putative DNA binding sites of Cdc45 (K363A or R419A), or in the two conserved residues in Cdc45 ('2x') simultaneously. The Western blot was probed with α-FITC antibody to detect the DNA-protein interactions. The long exposure (*lower panel*) visualizes the Cdc45 bands.

CMG complexes for DNA-protein cross-linking. This assay allows us to take a look at specific subunits within the CMG and test their DNA binding individually. Using this assay, I sought to understand if these Cdc45 mutants abolish direct interactions of Cdc45 with DNA. First, I confirmed that the binding of the Mcm2-7 subunits to the leading strand substrate was not altered compared to wildtype CMG (Figure 4.6F, short exposure). Strikingly, however, the '45-2x' mutant CMG entirely abolished the binding of Cdc45 to the leading strand in the absence of nucleotide (Figure 4.6F, lane 4, long exposure). To test if both residues contribute to this abolished DNA binding equally, I checked CMG complexes, in which only one of the residues was mutated at a time. I found that K363A displays wildtype binding of Cdc45 to the leading strand substrate in the cross-linking assay, whereas R419A abrogates DNA binding of Cdc45 within the CMG complex (Figure 4.6F, compare lanes 6,7 to 8,9). These data suggest that R419 in Cdc45 establishes contact with DNA in the context of CMG, as its mutation accounts for the entire abolishment of the contact to the leading strand.

Thus, I have identified at least one residue of Cdc45, R419, that directly contacts the leading strand of DNA within the context of the CMG complex. If the CMG actively uses the '2/5 gate' to bring in the leading strand into the center of the Mcm2-7 ring during the loading of CMG onto the substrate in our assays, or possibly exclude the lagging strand after initial melting, the R419 residue of Cdc45 may be critical for keeping the leading strand engaged in the Mcm2-7 central channel during helicase translocation.

The identification of Cdc45 as a RecJ orthologue raised the question of whether its exonuclease activity is conserved during the evolution. If so, this would imply that Cdc45 has analogous roles in replication as have been reported for RecJ. The initial finding of sequence conservation between RecJ and the N-terminus of Cdc45 led to extensive analysis of the typical catalytic motifs known from the DHH family as well as motifs implicated in metal binding. I aligned the N-terminal primary sequences of Cdc45 from various organisms and found changes in some of the motifs similar to those that have been reported by Silvia Onesti's lab in the human Cdc45 (Krastanova et al., 2012). Figure 4.7 shows four conserved motifs from the DHH family in the N-terminal region of Cdc45. Only a subset of the motifs that coordinate the Mn^{2+} ion is retained, such as the aspartate residue in motif 1, which is important for metal ion coordination, and some residues in motifs 2 and 4. On the other hand, key residues for binding of the metal ion are not conserved (the aspartate in motif 3 to asparagine, and the aspartate in motif 4 to glutamine), as well as the 'DHH' motif 3 is changed to 'DSH'. These data suggest that DmCdc45 has lost a RecJ-like exonuclease activity. Pesavento et al., (2013) indeed report the absence of such activity for the purified DmCdc45 in isolation.

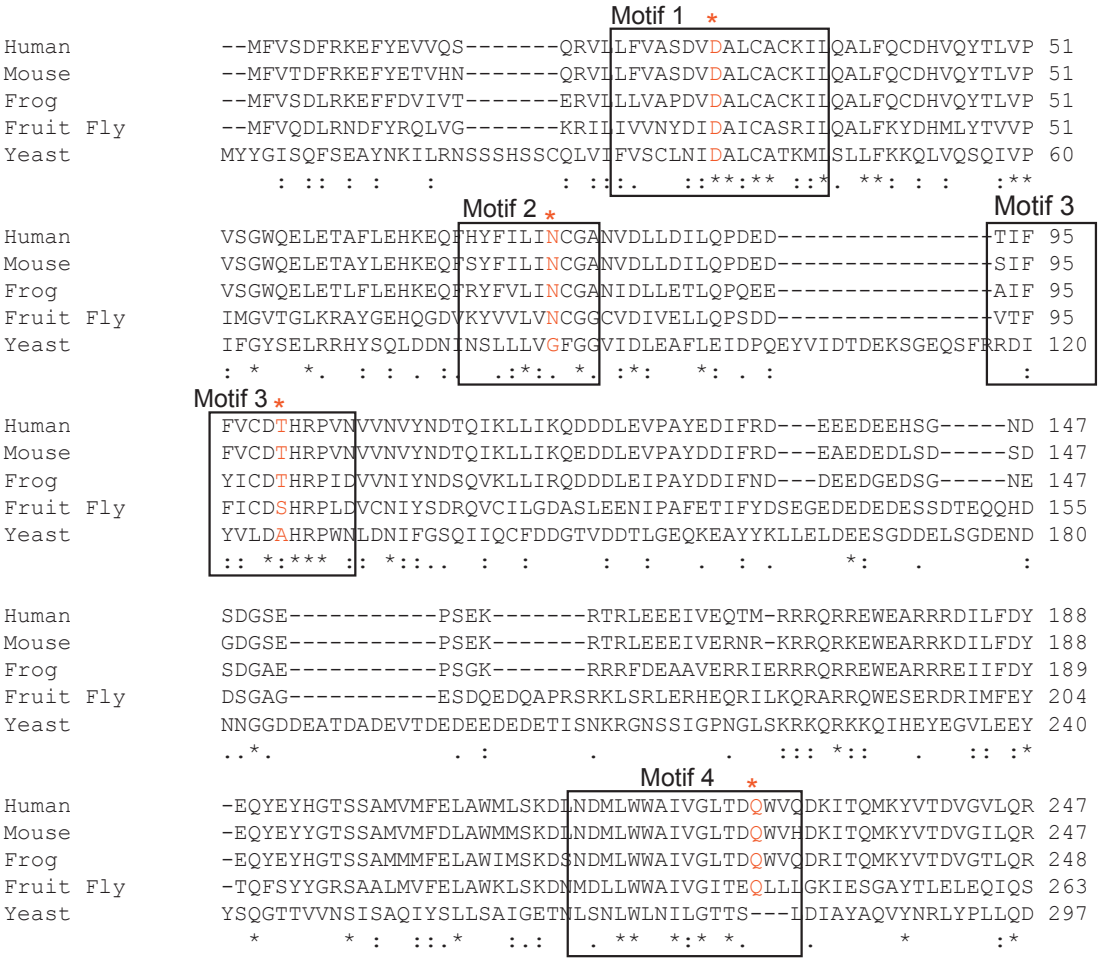

Figure 4.7. Alignment of Cdc45 N-terminus.
Multiple sequence alignment between human (*Homo sapiens*), mouse (*Mus musculus*), frog (*Xenopus laevis)*, fruit fly (*Drosophila melanogaster)*, and yeast (*Saccharomyces cerevisiae*) Cdc45 N-termini using ClustalW2. Characteristic catalytic RecJ motifs, required for the hydrolysis of phosphodiester bonds via the two metal ion mechanism, are highlighted with black boxes as in Krastanova et al., (2012). Highlighted in red are the residues in these motifs of *Drosophila* Cdc45 that are not conserved with the residues in the corresponding motifs in RecJ.

The study of isolated human Cdc45 also showed that it is a ssDNA binding protein (Krastanova et al., 2012). It was of interest to test if *Drosophila* Cdc45 also binds to ssDNA in isolation, and further if this R419 residue accounts for the identified contact to DNA. For this purpose recombinant wtCdc45 and Cdc45 containing the R419A substitution were purified. Cdc45's binding was tested in EMSAs revealing a very weak DNA affinity of Dm wtCdc45 in isolation (Pesavento et al, 2013, manuscript in preparation). Given that the CMG shows a preference for thymidines, James Pesavento next tested if Cdc45 in isolation also displays a sequence and geometry preference for substrates. Single-stranded DNA, one-armed or forked DNA substrates with either polyT or G/C-rich overhangs were tested for binding to Cdc45 alone in gel-shift assays. James Pesavento found that that Cdc45 also prefers polyT stretches for binding. It bound equally efficiently to the different tested substrate geometries as long as thymidine extensions were available for binding.

Comparison of the DNA binding affinities of wt and 'R419A' Cdc45 surprisingly showed that they were similar. These data suggest that Cdc45 must bind differently in the context of CMG than it does in isolation. When the protein is incorporated into the active CMG helicase, the positioning of residue R419 must be such as to account for its putative role as a 'guardian of the gate'.

But why then does the mutation of this residue not affect the on-rate of the CMG complex onto DNA? It is possible that additional residues contribute to this role.

The finding that the R419A mutation abolishes the binding of Cdc45 to DNA within the CMG raises the immediate question, does Cdc45 have an active role during translocation, and thus contribute directly to the helicase activity? To this end, I tested the double Cdc45 mutant ('45-2x') side-by-side with the Cdc45 single mutant CMGs (K363A, R419A) for helicase activity on a circular DNA substrate (Figure 4.8A, B). All three mutants show helicase activities similar to wt CMG, suggesting that neither residue contributes to the translocation mechanism. We cannot rule out the possibility that additional neighboring residues might contribute to Cdc45's role in translocation during cycles of ATP hydrolysis. By adding the non-hydrolyzable ATPγS nucleotide substitute, we might be capturing the CMG complex in a 'locked' conformation, in which only residue R419 establishes contact with the leading strand. There might be movements and contributions of proximal residues around R419 that come into play during translocation.

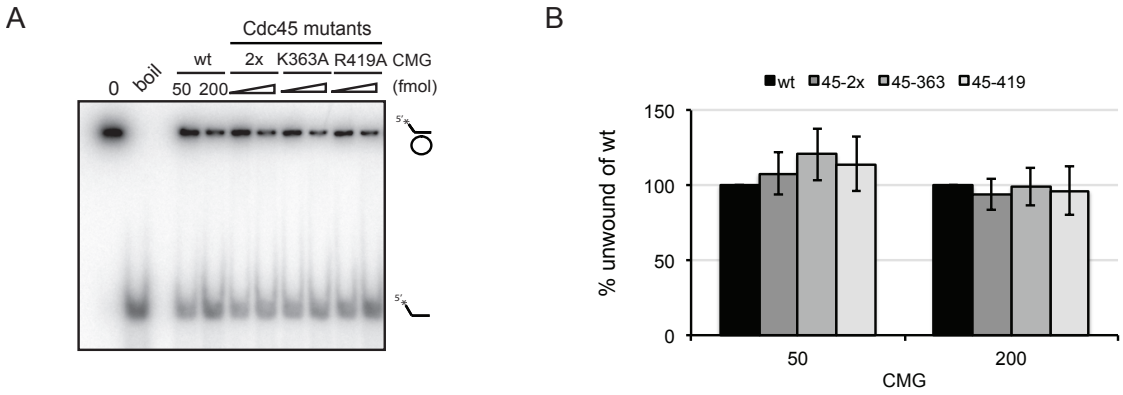

Figure 4.8. Mutation of R419 in Cdc45 does not alter the helicase activity of CMG.
(A) Autoradiograph of a helicase assay of wt and CMG complexes with the '2x', 'K363A', and 'R419A' Cdc45 mutants. Lanes '0' and 'boil' correspond to controls either without protein or with heat-denatured substrate. Two different amounts of proteins (100, 200fmol) were added to 1fmol of circular DNA substrate in presence of 300μM ATP. Reaction products were separated by TBE-PAGE (8%) and visualized on a PhosphorImager Screen. The positions of free circular DNA substrate and displaced oligonucleotide are indicated on the side of the gel.
(B) Quantification of the helicase assay shown in (A). Average values of unwinding (expressed as percentage of wt) were plotted and standard deviations of six independent series were included.

The Cdc45 loop affects processivity of the CMG

One way to test if the R419 residue of Cdc45 has an effect on CMG activity during translocation would be to capture the DNA contacts during movement of the helicase along DNA. In this case, I anticipate wtCdc45 to contact the DNA regularly if it has a direct role during translocation. In accordance with our previous findings, then Cdc45 containing the R419A mutation, should not make these DNA contacts as the CMG translocates. Unfortunately, we do not have an assay to test this directly. However, I wished to explore the idea that residues proximal to R419 that could touch DNA during translocation and contribute to the DNA binding role of R419. First, I analyzed the secondary structure predicted for this region of Cdc45 and found that residue R419 resides within a loop between two predicted helices (Figure 4.9A, loop is highlighted in red including K422, R419 highlighted with asterisk). Reasoning that the role of the R419 residue during translocation might be supported by other residues in its close proximity I decided to mutate all amino acids within the predicted loop and close by, including K422, to alanines. Primary sequence analysis shows that there are two additional basic residues proximal to R419 within the predicted loop (H421 and K422) that might substitute for R419. I purified the recombinant CMG containing the five amino acid residues, 'SRQHK', mutated to alanines (Figure 4.9B) and named this mutant '45-5A'. First, I wished to confirm that this '45-5A' CMG also had abolished DNA binding in the DNA-protein cross-linking experiment as seen with the CMG with the Cdc45 'R419A' mutant (Figure 4.6F). Equal loading of wt and '45-5A' CMG was confirmed by staining the Western blot with AuroDye™ (Figure 4.9C). The DNA-protein cross-linking confirmed that this mutant Cdc45 has almost entirely lost its interaction with the leading strand (Figure 4.9C, compare lane 2 to lane 4). Next, the DNA binding affinity of the '45-5A' mutant CMG was tested as a function of protein concentration and I confirmed that the mutant CMG has a binding affinity on par with wildtype CMG (Figure 4.9D). Given that the majority of the DNA binding is established by the Mcm2-7 ring the data suggest that the Cdc45 mutant does not induce any conformational changes in the Mcm2-7 ring that would then affect its affinity for DNA. If the hypothesis is correct that the loop surrounding R419 contributes to DNA contacts during translocation of the helicase, a defect in the helicase activity of '45-5A' CMG would be anticipated. For this purpose, I tested the '45-5A' CMG for helicase activity on our M13-derived circular DNA substrate. The unwinding of this 50bp duplex DNA substrate will likely require several cycles of ATP binding and hydrolysis. In agreement with the proposed model, I found, the helicase activity was reduced by 40-50% (Figure 4.9E). This mutant should display an even more drastic impairment of helicase activity on longer DNA substrates. To test this view, I prepared longer substrates by Klenow extension of the

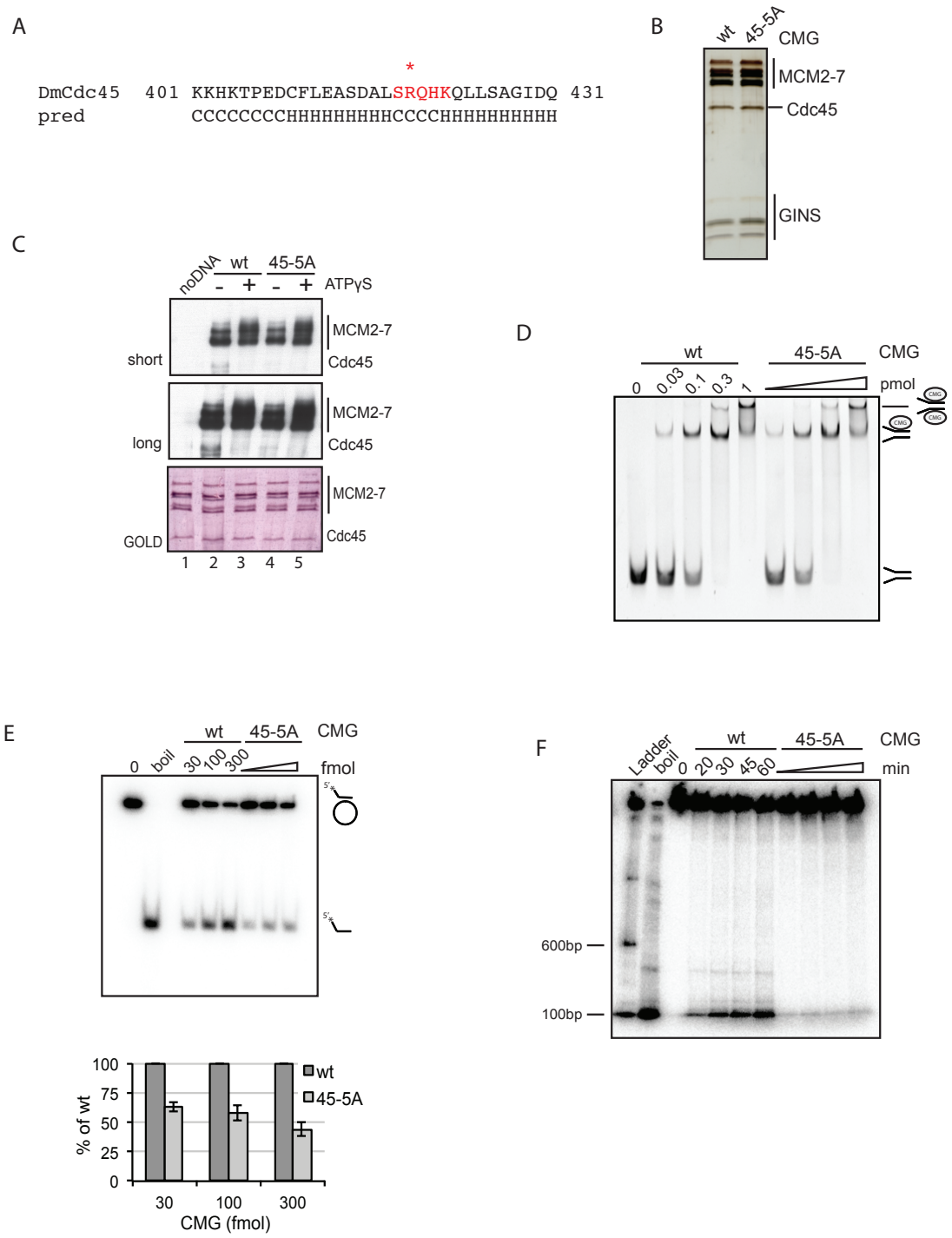

Figure 4.9. A loop in Cdc45 affects the processivity of the CMG.
(A) Secondary structure prediction of the region around R419A (highlighted with an asterisk) in Dm Cdc45. 'C' indicates random coiled and 'H' helical regions based on PsiPred. Residue R419 is predicted to reside within a loop between two helices. Introduced alanine substitutions in all five residues of this Cdc45 loop were named '45-5A' ('SRQHK' to 'AAAAA').
(B) Purified wtCMG and '45-5A' protein complexes were separated by SDS-PAGE (10%) and silver-stained.
(C) Western blot analysis of cross-linked wt and '45-5A' CMG complexes to the 'leading' strand substrate. The Western blot was probed with an α-FITC antibody. The long exposure (*lower panel*) is shown to visualize the weaker Cdc45 bands. The Western blot was stained with AuroDye™ (GE Healthcare) to visualize the protein bands.
(D) EMSA of protein titrations of wildtype and 'Cdc45-5A' loop mutant CMGs.

Control lane ‘0’ indicates free substrate alone, while all other lanes contain binding reactions of 100fmol of ‘fork' with 100fmol of CMG protein in the presence of 10μM ATPγS. Bound and free substrates were separated through a native TBE gel (5%) and their positions are indicated on the side of the gel.
(E) Autoradiograph of helicase activity assay with increasing amounts of purified CMG proteins (30, 100, 300fmol) on a circular DNA substrate. Translocation was initiated by addition of 300μM ATP and unwound reaction products were separated from substrate by TBE-PAGE (8%). ‘0’ and ‘boil’ are substrate controls without protein or boiled substrate, respectively. Quantification of the helicase activity is shown below with average values and standard deviations of two independent series.
(F) Processivity assay with wt and ’45-5A’ loop mutant CMGs. Unwinding reactions were incubated for increasing time points (20, 30, 45 and 60min) and stopped by addition of EDTA and SDS. The products were separated by SDS-PAGE (6%). A labeled DNA ladder is used to identify product sizes. ‘0’ and ‘boil’ are controls of prepared processivity substrate without protein or boiled, respectively.

M13-based circular substrate and tested the processivity of '45-5A' CMG. Consistent with the hypothesis, the '45-5A' CMG helicase activity is heavily compromised and unable to unwind the longer substrates compared to wt CMG, which could unwind substrates of up to ~500bp (Figure 4.9F).

Previously, I discussed the possibility that Cdc45 might play a direct role during translocation. These observations raise the possibility that the Cdc45 loop plays a role in the processivity of the CMG. Could '45-5A' CMG have an increased off-rate compared to wt Cdc45 and thus fall off of the substrate more easily?

Our model here is that Cdc45 acts as a 'guardian of the gate', where it catches the leading strand and keeps it inside the central channel to prevent its exit when necessary. This role might become important when CMG encounters various challenges during unwinding that impose a threat to DNA engagement. One example of such obstacle occurs right after melting of the duplex DNA when the lagging strand is excluded from the central channel of the Mcm2-7 ring while the leading strand remains within the central channel. Another important step when this role as a 'guardian of the gate' needs to be executed properly, is when the CMG stalls during translocation and needs to avoid falling off the DNA. To test this hypothesis, I wished to measure the off-rate on a DNA substrate that does not allow unwinding. I reasoned that, on such a substrate, the off-rate could be measured in the presence of hydrolyzable ATP, and, thus, while the Mcm2-7 motor is working and undergoing cycles of ATP binding and hydrolysis. Since CMG has a very low off-rate, I included competitor DNA for the measurement of the off-rate kinetics.

As a substrate, I chose a 50bp duplex region with a 3' 40nt polyT extension that would allow for the CMG to bind. Pesavento et al., (2013) report that the CMG efficiently binds to this substrate, but that the missing 5' tail does not allow the helicase to unwind. Because the unwound strand cannot be well separated from the one-armed substrate with this substrate on a native TBE gel used in this assay, an excess of polyT ssDNA was included as competitor that contains the complementary sequence to the 3'polyT strand of the one-armed substrate. Thus, if any one-armed substrate gets unwound, it subsequently anneals to the polyT ssDNA and forms the 'fork' substrate that separates well from the one-armed substrate on this native gel. Hence, unwinding can be assessed as a function of the formation of 'fork' DNA (Figure 4.10A). I had previously tested this in a proof of principle experimental using the 'leading' strand substrate and including as competitor the complementary sequence, allowing for the formation of a 'fork' substrate. The 'fork' and 'leading' strand substrates separate distinctly on the native gel, and confirm that the unwound strand can anneal to the complementary competitor DNA during these reactions (Figure 4.10B).

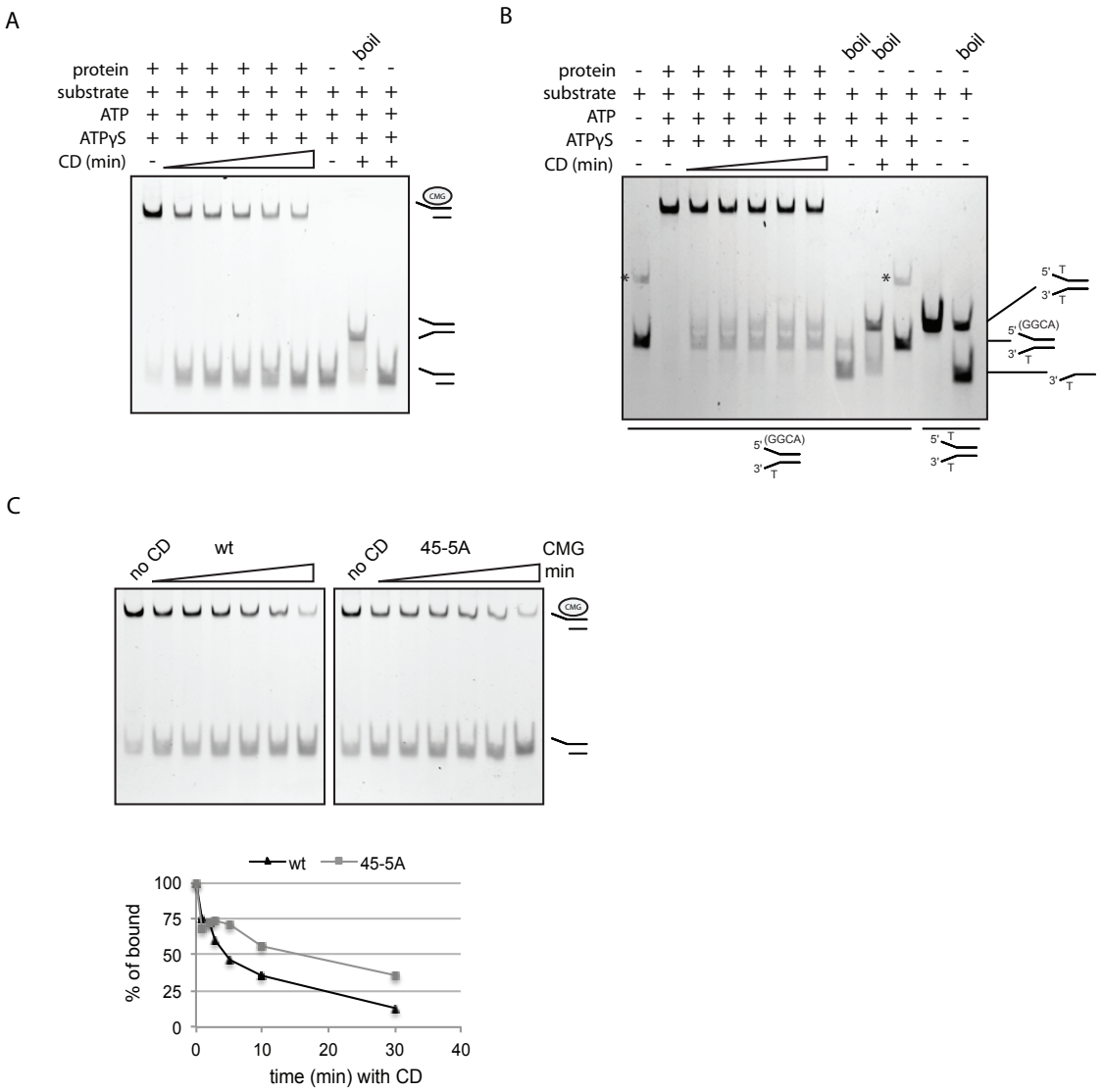

Figure 4.10. Off-rate kinetics remain unaltered with the '45-5A' CMG.
(A) and (B) EMSAs of competition assays of wt CMG using either a one-armed 3'polyT substrate in the presence of ATP (A) or 'leading strand' substrate in the presence of non-hydrolyzable ATPγS (B). Competitor DNA (complementary sequence to dsDNA region of substrate with poly T tail) was added and reactions were incubated for various time points (15, 30, 45, 60, 90, 120min) to measure the off-rate kinetics of CMG.
(C) EMSA of competition assay of CMG wt and '45-5A' mutant using a one-armed 3'polyT substrate that does not allow unwinding by CMG. Reactions were carried out in presence of ATP and incubated for various amounts of time (1, 2, 3, 5, 10, 30min) with competitor DNA (90nt polyT ssDNA). Products were separated on native TBE gel (5%). Reactions were quantified by plotting the various incubation periods in presence of competitor DNA (CD) against the initial amount bound by CMG before CD was added (set as 100%).

Using this assay and this 3' polyT one-armed substrate that does not support unwinding by the CMG, I could test the off-rate of the '45-5A' in the presence of ATP (Figure 4.10C). However, no increase in the off-rate with the '45-5A' CMG was found. One concern in this experimental setup is that, while the motor undergoes cycles of ATP binding, hydrolysis and release, actual translocation is never initiated. It will be important to design experiments that allow us to test whether the off-rate might increase for this mutant when the CMG moves regularly through the duplex region. In future experiments this will require the design of substrates that initiate CMG translocation but cause it to stall due to roadblocks, or, similar obstacles.

4.4 Discussion

Two previous studies showed that two copies of the Mcm2-7 complex are concertedly loaded onto double-stranded DNA in a head-to-head fashion (Evrin et al., 2009; Remus et al., 2009). This assembly at the replication origins occurs right before the cells enter the S phase of the cell cycle. Intriguingly, a biochemical study of the recombinant *Drosophila* CMG helicase showed that it cannot bind onto duplex DNA (Ilves et al., 2010). Together, these data implicate the necessity of a melting event in the center of the Mcm2-7 motor with major remodeling of the DNA structure. The association of the auxiliary factors GINS and Cdc45 with the pre-assembled Mcm2-7 hetero-hexamer, together with the action of two critical cell cycle-dependent kinases cyclin dependent kinase (CDK) and Dbf4-dependent kinase (DDK), mark the transition from G_1 to S phase of the cell cycle (Heller et al., 2011; Sclafani and Holzen, 2007). This activation process of the helicase involves extensive structural re-arrangement of the Mcm2-7 motor, leading to the closure of the discontinuity between Mcm2/5. Additionally, upon nucleotide binding, the Mcm2-7 central pore constricts, leading to the tighter association of DNA through this channel and simultaneously to the formation of another lateral channel by Cdc45/GINS (Costa et al., 2011). The appearance of this alternate channel led us to the hypothesis that the allosteric factors help guide the lagging strand while the leading strand threads through the central channel by the Mcm2-7 motor. Once the right architecture has been established for an active helicase and the leading strand is fully engaged through the Mcm2-7 ring, it is absolutely pivotal that the strand remains inside the Mcm2-7 pore; sliding off of the DNA would have detrimental consequences. Moreover, the presence of this discontinuity between Mcm2 and Mcm5 raises the possibility that it could transiently open during replication, allowing the leading strand to exit. How can the helicase prevent such an occurrence and ensure that the leading strand is properly engaged at all time?

Recent findings reported the discovery of an evolutionary relationship between Cdc45 and phosphoesterases belonging to the DHH family, such as the bacterial RecJ protein (Krastanova et al., 2012; Sanchez-Pulido and Ponting, 2011). In this present study, the focus lay on understanding what role, beyond initiation, Cdc45 might play as a stable part of the CMG helicase and whether this protein possesses any RecJ–like 5' to 3' exonuclease activity.

Similar to the reported findings for human Cdc45, *Drosophila* Cdc45 also represents a RecJ orthologue. We found that the N-terminus shows high primary sequence homology with RecJ, whereas the C-terminus was found to be homologous to a RecJ-related exopolyphosphatase through structural homology predictions.

A biochemical assay was developed that allowed us to induce protein-DNA crosslinks and thus analyze the contributions of individual CMG subunits to DNA binding. The finding that CMG only displays affinity for thymidines, without binding G/C-rich nucleotide sequences, allowed us to develop various substrates that orient the CMG in a specific way onto the DNA (Pesavento et al., 2013, manuscript in preparation). We designed forked substrates that cause the CMG complex to only bind in its 3' to 5' tracking direction and subsequently unwind the duplex DNA upon the initiation of translocation. By including fluorescein isothiocyanate (FITC) groups conjugated to thymidines at specific positions on the 3' extension of the substrate, we were able to capture and visualize CMG contacts with the leading strand using UV-induced cross-links. Alternatively, the DNA contacts to the lagging strand could be investigated by providing such cross-linking groups on the 5' extension.

Using this experimental approach, we uncovered DNA interactions of the individual Mcm2-7 subunits with either leading or lagging strands, and assessed the differences in the contacts that the two separated strands made in the absence and presence of nucleotide. Structural studies have suggested that the '2/5 gate' remains unlocked without nucleotide, and closes upon ATP binding (Costa et al., 2011). The main DNA interactions with both strands are established by the Mcm2-7 subunits within the CMG. In the absence of nucleotide, the leading strand is mainly bound by the Mcm5 subunit (and to a lesser extent Mcm3). In contrast to the Mcm2-7 ring alone, the CMG requires nucleotide for its DNA binding. Consistent with the observed tightening of the Mcm2-7 ring structure upon ATP addition, ATP binding induced all of the remaining Mcm2-7 subunits to grasp firmly the DNA (Costa et al., 2011).

Interestingly, Mcm5 also bound to the lagging strand in the absence of nucleotide. When ATP was added to the binding reactions, we observed a switch between the subunits that interact with the excluded strand. Unlike the leading strand, however, only a subset of Mcm2-7 proteins cross-linked to the lagging strand in the presence of nucleotide. Now, the main contact was formed with Mcm4 (and to lesser extent with Mcm2 and Mcm6), which suggests that only one half of the Mcm2-7 ring interacts with the lagging strand under these conditions. While the findings with the leading strand are consistent with the models for archaeal and eukaryotic organisms, which posit that it passes through the Mcm2-7 central channel, the observations with the lagging strand raise further questions regarding its precise path (Brewster and Chen, 2010; Fu et al., 2011).

A conserved residue in the external β-hairpin of the archaeal MCM was shown to be important for its DNA binding (Graham et al., 2011). In accordance with this report, I had previously observed that the simultaneous mutation of a conserved lysine in the β-external hairpins of all six Mcm2-7 subunits of the CMG significantly diminished the helicase activity of the complex. However, DNA binding by this same mutant CMG to a forked substrate in the gel-shift assays was not altered, suggesting that the strong Mcm2-7 binding to the leading strand was not affected by these mutations. In some unknown way, thus, these mutations had abolished the helicase activity of the CMG. As expected from the EMSA result with the fork substrate, I confirmed that the '6xEXT' CMG displayed wildtype binding to the leading strand. Using the novel cross-linking assay, however, the DNA contacts to the lagging strand were strongly reduced in this mutant, suggesting that these hairpins might guide the excluded DNA strand around the exterior surface of the Mcm2-7 ring.

Together with the finding that the CMG requires a 5' overhang for unwinding, reported in Pesavento et al., (2013), these data lead us to hypothesize a mechanistic model for the eukaryotic CMG analogous to the archaeal 'steric exclusion and wrapping' model (Graham et al., 2011). Mutations introduced in the external β-hairpin of individual Mcm2-7 proteins, or in only a subset of MCM subunits simultaneously, resulted in minor defects. It is likely that there is a certain path of the DNA around the external Mcm2-7 surface that needs to be entirely weakened to observe drastic defects, or that additional residues contribute to the binding, or a combination of both. Previous studies from *S. solfataricus* MCM have reported that this complex moves with the C-terminus towards the duplex DNA, engaging the 3'-tail through the MCM center while binding the 5'-tail around the exterior (McGeoch et al., 2005; Rothenberg et al., 2007).

To validate the hypothesis that the lagging strand might wrap around the Mcm2-7 ring and does not pass through the alternate side channel formed by Cdc45 and GINS, as was previously suggested, it was absolutely crucial to clarify whether any of the DNA binding affinity of the CMG to the lagging strand was contributed by Cdc45/GINS.

In contradiction to our previous hypothesis, a direct interaction between Cdc45 and DNA, was only found for the leading strand. Cdc45 only cross-linked to DNA in absence of ATP, when we anticipate there to be an open '2/5 gate', implying that the leading strand is in close proximity to this auxiliary factor. The DNA binding affinity of Cdc45 vanishes upon the addition of nucleotide, suggesting that the closure of the gap prevents the leading strand from contacting Cdc45.

Taking these results together, I suggest a 'steric exclusion and wrapping' model for *Drosophila* CMG. Specifically, I propose that the lagging strand leaves the Mcm2-7 central channel though the gap between Mcm2 and Mcm5 after duplex melting and is subsequently guided around subunits Mcm2, 6 and 4. No interaction was observed with the auxiliary factors, but I cannot rule out the possibility that binding of lagging strand to these proteins is established during subsequent cycles of ATP hydrolysis and translocation.

How can one understand Cdc45's participation in touching the leading strand? This surprising finding supports a model for Cdc45 as a 'guardian of the gate' in which Cdc45 needs to catch the leading strand and prevent its exit from the central Mcm2-7 pore whenever the gate is open. Such an opening of the gate would be necessary after the initial melting of the duplex DNA to allow the lagging strand to exit the central channel. An alternative scenario would be that the gate opens from time to time during helicase translocation or upon encountering an obstacle consistent with my mutations in the Cdc45 loop that diminish the helicase activity.

I tested the first model by analyzing CMG complexes that abolish DNA binding through the central channel of the Mcm2-7 ring and, thus, shift the binding equilibrium of the leading strand from the internal Mcm2-7 channel to the exterior. Alanine substitution of a conserved lysine in the Pre-Sensor1 β-hairpins of all six Mcm2-7 subunits ('6xPS1') strongly reduced the DNA affinity and abolished the helicase activity of the CMG complex. As anticipated, cross-linking reflected a reduced affinity of the Mcm2-7 subunits for the leading strand; in contrast, the affinity of Cdc45 for the leading strand increased for these mutants, a scenario one would anticipate if the leading strand now was no longer tightly bound through the central pore. This strand might be pushed out through the gate and captured more often in close proximity to Cdc45. Surprisingly, even after the addition of nucleotide, the strong interaction between the leading strand and Cdc45 was retained with this '6xPS1' CMG. One possibility that would account for the sustained Cdc45-DNA interaction is that the introduced point mutations in the PS1-hp affect the closure of the '2/5 gate', inducing a permanently open state.

Analysis of a second CMG mutant that showed a similarly strong reduction in DNA affinity (reflected in the DNA cross-linking of the Mcm2-7 subunits) and abolishment of helicase activity did not display an increase in Cdc45 affinity for the leading strand. This mutant CMG contained alanine substitutions of a conserved lysine in the Walker A boxes of all six Mcm2-7 subunits ('6xKA'). One interpretation is that the major difference between these two mutants lies in the nature of the mutations. While introduction of mutations in the

ATP binding motifs most likely prevents any DNA binding at all, mutation of the Pre-Sensor1 β-hairpins might still allow for some binding to DNA to occur within the Mcm2-7 motor. In addition to this, the latter mutation might impair the '2/5 gate' closure. Combined, the decreased affinity for the leading strand in the central channel and the open state of the gate shift the binding equilibrium of this DNA strand.

If Cdc45's role is to guard the '2/5 gate' and prevent the exit of the leading strand, I wished to find the residues responsible for this function.

Independent homology modeling of the N- and C-termini of Cdc45 allowed the superimposition of the models on the TthRecJ crystal structure and, thus, the identification of specific residues in Cdc45 that correspond to known ssDNA binding residues in RecJ (Yamagata et al., 2002). In this way, two basic residues putatively important for DNA binding by Cdc45 within the CMG complex were identified and mutated. These mutations allowed me to uncover at least one residue that is critically involved in directly contacting the leading strand. Alanine substitution of the basic residue R419 abrogates DNA cross-linking of Cdc45 to the leading strand in the absence of ATP. Interestingly, loading and translocation of the mutant CMG complex remain unaffected by this mutation. My model predicts that the helicase activity remains unaffected because there are additional neighboring basic residues that might contribute to the protein's tasks in loading and translocation. Given the fact that the CMG complex is locked onto DNA without being able to start the MCM motor under these conditions, I cannot rule out the possibility that Cdc45 might be positioned in a particular geometry that only allows the R419 residue to contact the leading strand. During cycles of ATP binding, hydrolysis and release, I propose that the region proximal to R419 becomes important due to perhaps small movements. This fact could account for the reduced effect on the helicase motor when only the R419 residue is targeted. In the future, I would like to develop an assay that would allow for the cross-linking of Cdc45 to the leading strand in the presence of hydrolyzable ATP. In agreement with my hypothesis, one would anticipate an unaffected DNA-Cdc45 interaction with the R419 mutation in the presence of ATP. Unfortunately, our experimental setup does not support cross-linking to DNA in presence of ATP. We observed extensive protein-protein cross-linking, induced by the presence of hydrolyzable nucleotide, which no longer allow for the unique identification of the DNA binding residues of individual proteins within the CMG.

Further primary sequence analysis of the R419 region in Cdc45 revealed that this critical residue resides within a predicted loop. Alanine substitutions of the entire loop ('45-5A') dramatically decreased the helicase activity and abrogated the processivity of the enzyme.

These data are in agreement with the model that the loop participates in the interaction of Cdc45 with DNA as part of the CMG complex. The loading and DNA binding of the '45-5A' mutant CMG were not affected, suggesting that the Mcm2-7 ring remained unaltered by introducing the mutant. So, what defects does the alanine substitution of this loop cause then? I propose two hypotheses for the role of this region. First, the mutations might increase the off-rate of the CMG complex, thus, explaining the decrease in helicase activity. However, off-rate measurements on forked substrates in the presence of ATPγS and competitor DNA showed no alteration in its kinetics. This raised the question of whether this loop in Cdc45 might play a critical role in CMG processivity. In this case, the Mcm2-7 subunits would need to undergo several cycles of ATP binding and hydrolysis to reveal such a defect in off-rate kinetics. For this purpose, the off-rate was measured on a substrate with 50bp of duplex DNA and a 3' thymidine extension. This substrate allowed CMG to bind, but could not be displaced. In this case, the off-rate was measured in presence of ATP, but again no difference was observed between mutant and wtCMG. The possibility still remains that the provision of ATP alone does not fully allow the proper movements of the motor during translocation in this *in vitro* set up. The CMG complex does not initiate tracking on the leading strand with this one-armed substrate, but rather remains locked on the ssDNA arm. The challenge for future experiments will be to develop DNA substrates that allow the CMG to initiate unwinding but introduce roadblocks that cause stalling of the complex during its translocation mode. The development of such a substrate will allow the testing of the model that the introduction of the five alanines in the loop of Cdc45 results in an increased off-rate. Since the role of Cdc45 as the 'guardian of the gate' is thought to be abolished in this mutant CMG, the leading strand could slip out of the Mcm2-7 central channel when the '2/5 gate' opens from time to time. Based on the findings, an alternate hypothesis remains plausible, in which Cdc45 directly contributes to the translocation by interacting with the leading strand regularly. In this scenario, the role of this allosteric factor would not only be structural for the initial gate closure, but would account for a regular gate opening during translocation on the DNA, possibly caused by a certain movement within the Mcm2-7 ring. This model would suggest a direct role for Cdc45 in unwinding. If the off-rate remains unaltered on the 'actively' unwound substrates, the need will arise to develop assays that allow the capture of DNA-Cdc45 interactions during translocation and test an active participation of the auxiliary factor.

Based on the findings, I hypothesize that Cdc45 plays a role in guarding the '2/5 gate', ensuring that the leading strand remains in the central pore of the Mcm2-7 ring at any given time. These DNA-Cdc45 interactions might become crucial during progression

through or at stalled replication forks, in which case Cdc45 might have a putative role during replication stress reminiscent of its archael homolog, RecJ. However, due to the mutations of the distinct exonuclease motifs found in RecJ, it is reasonable to argue that Cdc45 has lost its exonuclease activity and likely evolved to a new role in the eukaryotic DNA replication (Lovett and Kolodner, 1989; Sutera et al., 1999; Wakamatsu et al., 2010; Yamagata et al., 2002). While Cdc45 has retained the ssDNA binding affinity of RecJ, I believe that it has acquired the diverged function of ensuring the path of the leading strand. Other RecJ orthologues might also be found in future studies of evolutionary relationships. It is very probable that such orthologous proteins play roles in the eukaryotic DNA repair pathways, required for RecJ-like exonuclease acitivities in homologous recombination, base excision repair or rescue of stalled DNA forks. If they are not repaired, DNA damage can lead to mutations, cell death or loss of genomic stability. Thus, the proteins involved in such DNA repair pathways are absolutely pivotal for any growing and dividing organisms (Courcelle et al., 2006; Courcelle et al., 2003; Courcelle and Hanawalt, 1999).

Based on previous studies in other organisms and our findings in *Drosophila*, a model is presented here for the role of Cdc45 within its biological context that depicts the paths of the two melted and subsequently segregated parental DNA strands (Figure 4.11). Two Mcm2-7 rings are loaded as double hexamers in a head-to-head manner onto double stranded DNA. Recruitment of Cdc45 and GINS causes melting of the duplex DNA by an unknown mechanism. The lagging strand then exits the central channel through possibly a gap between subunits Mcm2 and Mcm5 while Cdc45 could serve as a 'cow-catcher' to prevent the leading strand from slipping out. Upon full association of these auxiliary factors and nucleotide binding, the '2/5 gate' is closed. The activated helicase engages the leading strand in the Mcm2-7 core while the excluded lagging strand wraps around the Mcm2-7 on the outside.

The hypothesis is that during cycles of ATP hydrolysis and translocation, there might be an equilibrium between 'open' and 'closed' states of the '2/5 gate' in the Mcm2-7 ring. The from time to time, transient opening of the gap would require Cdc45 to ensure that the leading strand remains tightly engaged through the Mcm2-7 motor. In the presence of ATPγS this equilibrium might be strongly shifted towards longer periods in the locked state of the Mcm2-7 ring and would explain why the DNA-Cdc45 contacts are not captured under these reaction conditions. The presence of the non-hydrolyzable ATP analog would artificially cause a prolonged closure of the gate and would mask the effects of the Cdc45 point mutations in the DNA binding residues.

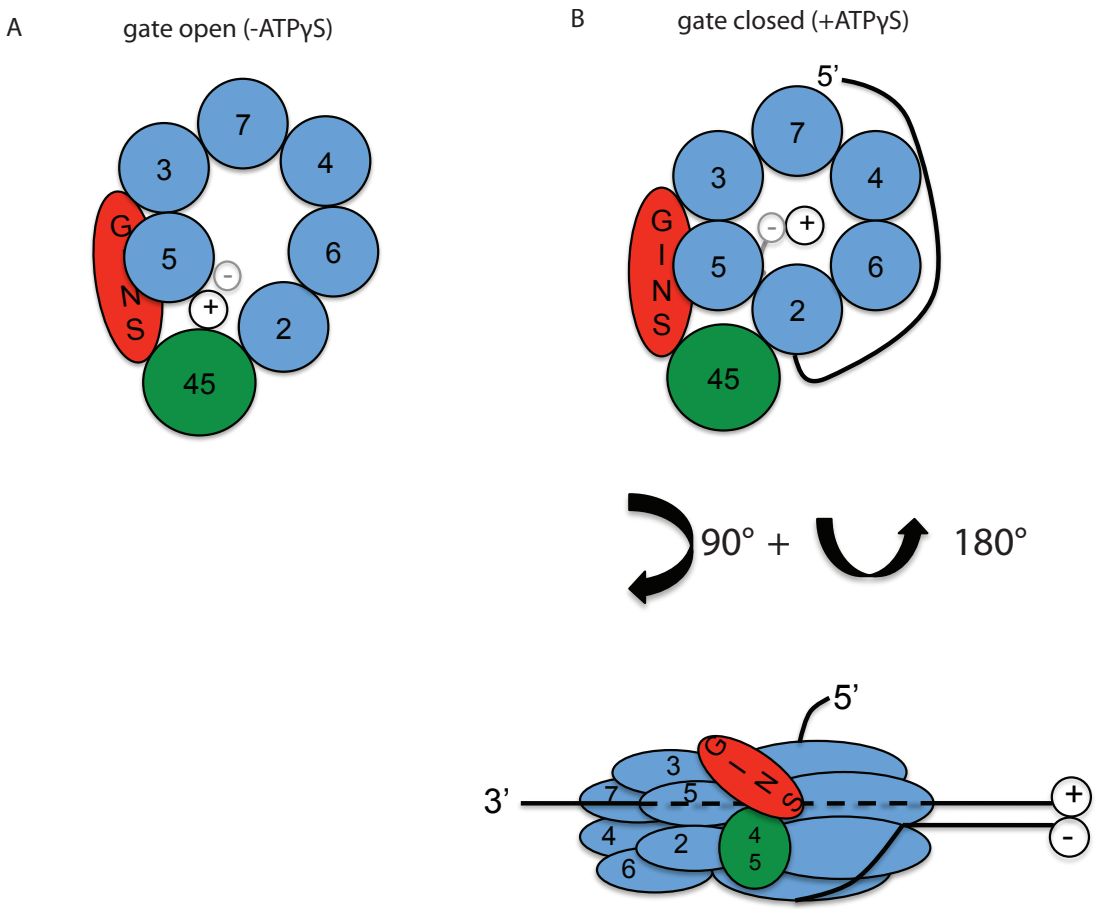

Figure 4.11. Model
After the initial melting of the duplex DNA, the separated strands are spatially segregated and lead to a final arrangement, in which the 3'-leading strand ('+') passes through the Mcm2-7 central channel while the 5'-lagging strand ('-') does not enter the central cavity. The model here shows the arrangement according to the binding of leading and lagging strand of a 'fork' substrate that contains 40nt extensions on the 3' and 5' strand. (A) In absence of nucleotide, the discontinuity between Mcm2 and Mcm5 within the Mcm2-7 ring is not properly tightened, and Cdc45 can establish contact to the leading strand, whereas the lagging strand is positioned on the exterior and remains outside of the central cavity (indicated by the light grey '-'). In such scenario, Cdc45 acts as the 'guardian' of the opened gate, preventing an accidental DNA exit if the helicase encounters breaks or short strands, and ensures that the leading strand remains encircled within the Mcm2-7 center at any times. (B) Upon nucleotide binding the discontinuity between Mcm2/5 is closed, and the central channel constricts leading to a better engagement of the leading strand. Unlike the leading strand, the lagging strand does not enter the central pore during translocation and wraps around the exterior surface of Mcm2-7 as shown in the top view from N-terminal side in (B, *top panel*) and in the side view in (B, *lower panel*). The CMG translocates along DNA with a 'steric exclusion and wrapping' model.

4.5 Materials and Methods

Cloning, protein expression and purification

Proteins were prepared according to the protocol provided in Ilves et al., (2010) and were described in detail in Chapter 3.

Mutations in the Walker A boxes, arginine finger motifs, Pre-Sensor1 β-hairpins, external β-hairpins of all six MCM subunits, as well as the point mutants in Cdc45 were introduced by PCR.

Alanine substitutions of the conserved lysine residues in all six Walker A motifs were previously described in Chapter 3 and in (Ilves et al., 2010).

The alanine substitutions of the conserved basic residue in the Pre-Sensor1 β-hairpins targeted: in MCM2 K598, in MCM3 K430, in MCM4 K602, in MCM5 K468, in MCM6 R478 and in MCM7 K471. Alanine substitutions were introduced in the two basic residues of Cdc45: K363 and R419.

The R to A mutations in the arginine finger of the six MCM subunits were introduced in Chapter 3.

The alanine substitutions of the conserved lysine in the external β-hairpins targeted following residues: K490 in MCM2, K322 in MCM3, K492 in MCM4, K360 in MCM5, K370 in MCM6, and K364 in MCM7.

Additional amino acids in Cdc45 R419A were substituted with alanine at residues S418, Q420, H421, and K422 to generate the '45-5A' loop mutant.

DNA helicase and EMSA assays

The detailed description of helicase and EMSA assays has been provided in the 'Material and Methods' section of Chapter 3.

UV-induced cross-linking assay

Cross-linking was enabled by introduction of modified FITC-labeled (fluorescein isothiocyanate) nucleotides at specific positions in the fork DNA substrate.

To cross-link CMG proteins specifically to the leading or lagging strand of a forked DNA substrate, the protein was bound to the substrate under the same reaction conditions as used in the electrophoretic mobility shift assays (EMSA) described by Ilves et al., (2010) and Chapter 3. Binding reactions were mixed and incubated for 30 min at 30°C in the presence or absence of 10μM ATPγS. Cross-linking was induced by UV irradiation for 5min, and the DNA was digested with 180 gel units of Micrococcal nuclease (NEB) for 1

hour at 37°C. Proteins were separated by SDS-PAGE (8%) and analyzed by Western blotting with a monoclonal α-FITC antibody (Abcam).

The forked 'leading strand' DNA substrate was prepared by annealing
Oligo1 of the following sequence:
5' GGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCA
GGTTGGCCGATCAAGTGCCCAGTCACGACGTTGTAAAACGAGCCCGAGTG 3'
to the fluorescently labeled Oligo2 of the following sequence:
5' CACTCGGGCTCGTTTTACAACGTCGTGACTGGGCACTTGATCGGCCAACC
TT/iFluorT/TT/iFluorT/TT/iFluorT/TTTTT/iFluorT/$(T)_{25}$ 3'

For preparation of the forked 'lagging strand' DNA substrate, I annealed
Fluorescently labeled Oligo3 of the follwoing sequence:
5' GGCAGGCAGGCAGGCAGGCAGGCAC/*iFluor*T/GGCG/*iFluor*T/CGGG/*iFluor*T/CGGC
GGTTGGCCGATCAAGTGCCCAGTCACGACGTTGTAAAACGAGCCCGAGTG 3'
To Oligo 4 of the following sequence:
5' CACTCGGGCTCGTTTTACAACGTCGTGACTGGGCACTTGATCGGCCAACC$(T)_{40}$ 3'

Both substrates were PAGE purified (8% TBE), ethanol precipitated and dissolved in TE buffer (pH 8.0). DNA concentrations were measured by the 260nm absorbance on a Nanodrop (Thermo Scientific). Single-stranded DNA contamination was estimated to be less than 5% of the forked DNA substrate.

Processivity assay

For preparation of DNA substrates used to analyze the processivity of CMG complexes, the 3' of the M13-based circular substrate described in Moyer et al., (2006) and Chapter 3 was extended with the Klenow fragment under limiting amounts of [α^{32}P]dATP (10μCi, Perkin Elmer) and 8μM of cold dGTP, dTTP and dCTP for 2 min at room temperature (RT). For the preparation of longer circular substrates, an excess of cold dNTPs was added at a concentration of 80μM and supplemented with 20μM of ddGTP and the elongation reactions were incubated for an additional 15min. For shorter DNA substrates, the M13-based circular substrate was extended with Klenow in the presence of [α^{32}P]dATP (10μCi) and 7.5μM of cold dGTP, dTTP, dCTP, supplemented with 2.5μM ddGTP for 15 min at RT. Both substrates were purified over three consecutive MicroSpin S-400 HR columns (GE Healthcare) and mixed in a 1:1 ratio.

For the processivity assay, the indicated amounts of protein were mixed with the processivity substrate and the unwinding reactions were performed in buffer A (25mM

Hepes pH 7.6, 10% glycerol, 50mM sodium acetate, 10mM magnesium acetate, 1mM DTT, 250 µg/ml insulin) in a total volume of 10 µl. Initial binding of CMG to the substrate was allowed in the presence of 10µM ATPγS for 15 min at 30°C. The unwinding reaction was then started with the addition of 300µM ATP. The reactions were stopped at different time points by addition of 0.2% SDS and 100mM EDTA after incubation at 30°C. The reaction products were separated for 1 hour at 120V on 6% SDS-PAGE. The gels were dried, exposed to a PhosphorImager Screen (FujiFilm), and signals were detected with either the Typhoon™9400 (GE Healthcare) or Pharos FX (BioRad). Results were analyzed using the Image Quant or ImageJ software.

Competition assay/Off-rate kinetics

To measure off-rate kinetics of the CMG from DNA on our forked substrates, I performed EMSA assays under the regular binding conditions as described in Chapter 3 were performed. After the initial binding of the protein complex to DNA for 30 minutes at 30°C in the presence of ATPγS, competitor DNA was added and incubated for various time points. Poly-thymidine single-stranded DNA was used as the competitor in the presence of ATPγS. To measure the off-rate in the presence of hydrolyzable ATP, a one-armed forked substrate containing a 3' polyT tail but lacking the 5' overhang, was used by annealing Oligonucleotide 1 of the following sequence:
5' GGTTGGCCGATCAAGTGCCCAGTCACGACGTTGTAAAACGAGCCCGAGTG 3'
to Oligonucleotide 2 of the follwoing sequence:
5' CACTCGGGCTCGTTTTACAACGTCGTGACTGGGCACTTGATCGGCCAACC
TT/*iFluorT*/TT/*iFluorT*/TT/*iFluorT*/TTTTT/*iFluorT*/$(T)_{25}$ 3'

First, the CMG's ability to unwind any of the substrates used for the competition assay were tested. If the native electrophoresis did not separate the displaced oligonucleotide from the substrate, an excess of an alternative competitor ssDNA was included. This oligo contains the 40nt thymidine stretch as well as a 50nt region that is complementary to the double stranded region of the used substrate. In case of unwinding, the displaced species anneals to the competitor oligo and forms a regular forked substrate that causes a distinct shift in the gel and allows the clear identification of the substrate species formed during the reaction.

Acknowledgements

I want to thank Alessandro Costa and James J. Pesavento for the collaboration that gave rise to detailed analysis of Cdc45 and established the cross-linking assay. Further, I would like to thank Jingdan Liang for help with the mutagenesis in the Cdc45 protein.

Chapter 5

Mutations of the CMG helicase internal channel β-hairpins reveal an ordered path for translocation on the leading strand

5.1 Abstract

The CMG complex, consisting of Cdc45/GINS/Mcm2-7 proteins, represents the activated form of the eukaryotic replicative helicase. In contrast to archaeal helicases, which contain six copies of one MCM monomer that assemble into a toroid, the eukaryotic MCM motor possesses six different MCM polypeptides, numbered 2-7, that oligomerize into a hexameric ring. Various models for DNA unwinding have been proposed, in which either dsDNA enters into the central channel with the lagging strand leaving through a side channel after separation, or in which the leading strand is engaged through the interior of the Mcm2-7 ring while the lagging strand is excluded to the outside. Since *Drosophila* CMG exclusively binds to ssDNA, the latter 'steric exclusion model' is favored for its unwinding mechanism. At the same time, the question of the direct path of both DNA strands arises. The crystal structure of the near-full-length *Sulfolobus solfataricus* MCM hexamer revealed a total of four candidate β-hairpin motifs in each Mcm2-7 subunit that are implicated in DNA binding. The N-terminal β-hairpin (NT-hp) forms the narrowest point in the central channel of the Mcm2-7 ring and is located in the N-terminal domain of the protein. Two of the β-hairpins, the Pre-Sensor 1 (PS1-hp) and Helix-2-Insert (H2I-hp) β-hairpins, are positioned in the interior of the AAA+ domain, whereas the fourth External β-hairpin (EXT-hp) resides on the exterior surface of the MCM hexamer. Archaeal helicases represent a simplified version of the MCM helicase enzyme and have been previously used as model to analyze and identify critical residues within the hairpins. To better understand and dissect each hairpin's role in DNA binding and subsequent unwinding by the *Drosophila* CMG complex, systematic mutational analysis of different critical residues within the four hairpins was performed. The biochemical characterization of all mutant-hairpin CMG complexes demonstrated the different effects of specific basic residues in the four hairpins on DNA binding, helicase activity, and direct interactions to leading or lagging strands. Several of these positively charged residues were tested for the DNA binding by CMG in comparison to that of the isolated Mcm2-7 complex to understand differences and compliances between the inactive and active form of the helicase. Further, the results identified critical residues required for the enzyme's activities and implicate an analogy to the 'steric exclusion and wrapping' model suggested for SsoMCM where the leading strand enters the central channel while the lagging strand is excluded and 'wrapped' around the outside of the Mcm2-7 ring. This mutant analysis confirmed the functional asymmetry of the Mcm2-7 ring within CMG with regard to the DNA binding motifs and suggests a discrete start site for the 'power stroke'.

5.2 Introduction

A genetic screen in budding yeast gave in the 1980's rise to the name of a new family of proteins, several of which were paralogs with implications in DNA replication (Forsburg, 2004). This MCM protein family revealed defects in minichromosome maintenance and was named thereafter (Maine et al., 1984). A central role is played by a subgroup of these polypeptides, of which Mcm2-7 form together a hetero-hexameric ring complex that has been shown to constitute the helicase motor in eukaryotic cells . Its enzymatic malfunction was correlated to loss of genomic integrity as well as to rise of various carcinomas (Bochman and Schwacha, 2008; reviewed in Forsburg, 2004; Lau et al., 2007). For many years, however, the isolated Mcm2-7 complexes failed to demonstrate helicase activity *in vitro*, unlike the archaeal counterpart consisting of six identical copies of one MCM protein working together as a homo-hexamer to robustly unwind duplex DNA (Chong et al., 2000; Moreau et al., 2007; Sakakibara et al., 2008; Schwacha and Bell, 2001). Weak helicase activity *in vitro* has long been only demonstrated for a dimer of the Mcm4/6/7 trimer, which led to the proposal of these subunits as the true 'catalytic' core while Mcm2, 3, 5 were attributed negative regulatory functions during translocation (Ishimi, 1997; Lee and Hurwitz, 2001; Schwacha and Bell, 2001; You and Masai, 2005). These observations contradicted the earlier findings showing that, in cells, the entire Mcm2-7 hetero-hexamer is most commonly present, whereas the Mcm4, 6, 7 sub-complex was only found during biochemical purifications of the complex. It is only recent that this caveat was resolved as it was finally demonstrated that the helicase activity of the budding yeast Mcm2-7 complex existed under distinct reaction conditions (Bochman and Schwacha, 2008). In case of *Drosophila* or human helicases, a robust and vastly enhanced unwinding could be detected only after association of auxiliary factors Cdc45/GINS into the higher-molecular weight CMG complex (Cdc45/Mcm2-7/GINS) that is believed to represent the activated from of the eukaryotic helicase (Aparicio et al., 2006; Ilves et al., 2010; Kang et al., 2012; Moyer et al., 2006).

Various aspects of the active CMG complexes have been investigated *in vivo* or by biochemical assaying of recombinant proteins *in vitro*. However, insights into the molecular mechanism have remained elusive heretofore due to the absence of high-resolution structural studies. Because archaeal organisms possess a simpler version of this enzyme, in which six identical polypeptides display robust helicase activity *in vitro*, it has been used as a model system to gain clues about the structure and function. Archaea possess only one MCM gene, and best insights have been obtained from *Methanothermobacter thermoautotrophicus* MCM (MthMCM) and *Sulfolobus solfataricus*

MCM (SsoMCM). Both helicases are formed by assembly of six copies of the MCM polypeptide into a toroid, analogous to the suggested Mcm2-7 ring structure and contain DNA binding affinities to both single- (ss) and double-stranded (ds) DNA, intrinsic ATPase activities, and translocate on DNA with a specific directionality (3' to 5') *in vitro* (Carpentieri et al., 2002; Chong et al., 2000; Davey et al., 2003; Kelman et al., 1999; Moreau et al., 2007; Pape et al., 2003; Pucci et al., 2007; Shechter et al., 2000).

In contrast to the eukaryotic DNA replication, where structural knowledge remains in stagnancy, studies from archaeal helicases have prominently progressed and provided helpful information to understand the basis of its unwinding mechanism. Because of many similarities in common, drawn analogies allow biochemical characterizations for the more complicated eukaryotic enzymes. Thus, models and implications can be derived from archaeal findings, to reveal possible mechanisms of nucleotide-hydrolysis driven conformational changes and consecutive strand separation.

To date, crystallographic structures of the N-termini of MthMCM and SsoMCM, as well as the near full-length structures of SScoMCM and *Methanopyrus kandleri* (MkaMCM2) have been determined (Bae et al., 2009; Brewster et al., 2008; Fletcher et al., 2003; Liu et al., 2008). Further, structural insights have been obtained from several low-resolution electron microscopic studies of MthMCM (Chen et al., 2005; Costa et al., 2006; Gomez-Llorente et al., 2005; Pape et al., 2003).

Taking the structural data of the archaeal MCM together, several common features of the overall MCM arrangement become evident. The six MCM monomers oligomerize to form a two-tiered ring in which the amino-terminal domains form a ring that is stacked on top of the ring formed by the carboxy-terminal AAA+ domains below. This topology leaves an internal channel that provides one possible path for the DNA. Besides its structural organization and implication in the hexamerization of the complex, the N-terminal domain has mainly been considered necessary for *in vitro* processivity and 3' to 5' directionality of the full complex (Barry et al., 2007; Fletcher et al., 2003; Kasiviswanathan et al., 2004; Liu et al., 2008). Further, a conserved loop named the allosteric control loop (ACL), is located in this domain and establishes communication between the domain and the C-terminal domain of the adjacent MCM protein upon nucleotide hydrolysis (Barry et al., 2009; Brewster et al., 2008; Sakakibara et al., 2008). Different from the N-terminal domain, most of the sequence conservation between MCM orthologs and paralogs has been shown for its AAA+ domain contained between different species and harbors all required ATPase active site motifs (discussed in Chapter 3) (reviewed in Kearsey and Labib, 1998; Koonin, 1993; Neuwald et al., 1999). Extensive biochemical work has been carried out on MthMCM and

SScoMCM that helped shed light onto the tasks of the different domains and functional motifs, as well as on the roles of various specific residues mainly with regard to DNA binding and the enzyme's intrinsic ATPase and helicase activities, but also with regard to inter- and intra-subunit crosstalk (Barry et al., 2009; Barry et al., 2007; Jenkinson and Chong, 2006; Jenkinson et al., 2009; Kasiviswanathan et al., 2004; McGeoch et al., 2005; Moreau et al., 2007; Pucci et al., 2007; Pucci et al., 2004; Sakakibara et al., 2008).

Molecular information from X-ray studies of archaeal MCM have not only provided detailed knowledge about discrete sites, but also heavily shaped our understanding about the three-dimensional structure of the MCM hexamer. Even though the full-length (FL) SsoMCM as well as MkaMCM2 polypeptides were only monomeric in the crystals, the N-terminal domains of MthMCM and SsoMCM crystallized as double and single hexamers, respectively, thus allowing modeling for the hexameric FL SsoMCM (Brewster and Chen, 2010; Brewster et al., 2008). The general shape of the hexamer, in agreement with electron microscopic models, shows a characteristic 'dumbbell' shape with a constriction between the N- and C-terminal tiers (reviewed in Brewster and Chen, 2010; Costa et al., 2006; Gomez-Llorente et al., 2005; Pape et al., 2003).
Structures from the N-terminal MCMs showed that the central channel is constricted by six β-hairpins each protruding from one MCM monomer, and additionally similar constrictions through positively charged β-hairpins in the AAA+ domain became obvious from the full-length crystal structures (Brewster et al., 2008; Fletcher et al., 2003). The dimensions of the central cavity in archaeal MCM are large enough to accommodate ssDNA or dsDNA, and six 'side' channels between two adjacent MCM subunits leave this center proximal to the ATPase active sites, leading to the exterior and establishing a direct connection between internal channel and the exterior (Brewster et al., 2008; Fletcher et al., 2003).

A closer look at each MCM monomer shows a total of four characteristic β-hairpins (hp): one is located around the central channel in the N-terminal domain named N-terminal β-hairpin (NT-hp), while the remaining three are positioned in the AAA+ domain; two around the central pore named Pre-Sensor1 β-hairpin (PS1-hp) and Helix-2-Insert β-hairpin (H2I-hp), and one proximal to the side channel called Exterior β-hairpin (EXT-hp) (Figure 5.1) (Brewster et al., 2008).

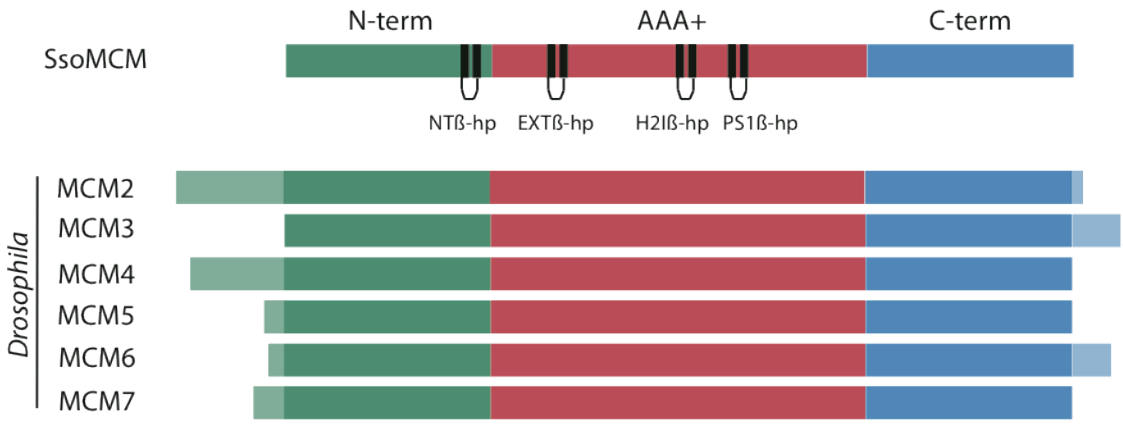

Figure 5.1. Schematic of sequence alignment of *Drosophila* Mcm2-7.
A multiple sequence alignment of all six *Drosophila* Mcm2-7 was created by ClustalW2 and is schematically shown (introduced in Chapter 3). The relative position of each of the four β–hairpins is indicated. One of these is located in the N-terminal domain and therefore named N-terminal β-hairpin (NT-hp). Three β–hairpins reside within the AAA+ domain named Pre-Sensor1 β–hairpin (PS1-hp), Helix-2-Insert β–hairpins (H2I-hp), and Exterior β–hairpin (EXT-hp).

The six NT-hp constrict the central cavity and have been demonstrated to bind DNA, where mutations of specific basic residues strongly impaired the DNA binding affinity and subsequently the helicase activity (Fletcher et al., 2003; McGeoch et al., 2005). Genetic approaches in yeast showed that alanine substitutions of the basic residues in this precise hairpin of Mcm5 displayed defects in the transition from G1 to S phase of the cell cycle *in vivo*, in initiation of DNA replication as well as in the recruitment of the Mcm2-7 hetero-hexamers to replication origins, suggesting pivotal roles in the initial stages of Mcm2-7 loading (Leon et al., 2008).

While the mutation of a critical lysine in the PS1-hp only mildly weakened DNA binding, it was demonstrated to be essential for the helicase activity (McGeoch et al., 2005). Interestingly, in the hairpins of each of the six subunits of the papillomavirus E1 helicase, equivalent to PS1-hp, it was demonstrated that the adjacent combination of an acidic and basic residue form a 'staircase' that sequentially pulls the oligonucleotide through the central cavity (Enemark and Joshua-Tor, 2006; reviewed in Enemark and Joshua-Tor, 2008). This tracking mechanism is enabled by positional changes of the hairpins induced by cycles of ATP-binding, -hydrolysis and –release. Further, equivalent hairpins in simian virus 40 large T antigen not only also protrude into the central channel, but also show striking movements in response to nucleotide binding and hydrolysis (Brewster and Chen, 2010; Gai et al., 2004; Li et al., 2003). Also, in archaeal MCM the PS1-hp hairpin from one

monomer was shown to interact with the ACL from an adjacent subunit. Thus, a communication not only between N-terminal and C-terminal domains of the same subunit was established, but also an additional inter-subunit crosstalk, giving an idea about the complexity to be expected due to inter- and intra-subunit communications (Bae et al., 2009; Barry et al., 2009; Brewster et al., 2008; Fletcher et al., 2003; Sakakibara et al., 2008).

Unlike the PS1-hp that are slightly recessed from the center and located marginally towards the side channel, the H2I-hp completely extend into the central channel of the AAA+ domain. This hairpin is unique to MCM and other H2I-containing subfamily of AAA+ proteins (Brewster et al., 2008). In contrast to PS1-hp, deletion of residues in this hairpin vastly stimulated the DNA binding affinity, but, analogous to PS1-hp, abrogated unwinding activity (Jenkinson and Chong, 2006). This motif was attributed a potential role as 'plowshare' in separation of duplex DNA during translocation as its intrinsic ATPase activity was only enhanced by addition of dsDNA, while remaining unaffected by addition of ssDNA (Jenkinson and Chong, 2006).

Differently, the EXT-hp is located at the exterior surface of each MCM monomer adjacent to the side channel. Introducing mutations in specific residues within this hairpin resulted in striking impairment of helicase activity, suggesting a crucial but unknown role during the unwinding mechanism (Brewster et al., 2010; Brewster et al., 2008). Additionally, recent studies of this hairpin from SsoMCM have not only confirmed the latter findings, but further proposed that this hairpin with its location on the outside of the MCM hexamer serves to grasp the separated and excluded strand that wraps around the MCM ring surface (Graham et al., 2011; Rothenberg et al., 2007).

Two central questions of utmost importance for the eukaryotic DNA replication field remain unanswered: How does the helicase achieve initial melting of the duplex and what is the precise molecular mechanism during unwinding?
Studies in yeast demonstrated that in the recruitment step two Mcm2-7 are loaded onto duplex DNA in a head-to-head manner in the G1 phase (Evrin et al., 2009; Gambus et al., 2011; Remus et al., 2009). These double hexamers can slide along the duplex, but remain inactive. In contrast to this, biochemical analysis of the recombinant *Drosophila* and human CMG complexes, shown to represent the activated form of the eukaryotic helicase, revealed that this form of the helicase only binds to ssDNA in the S phase of the cell cycle (Ilves et al., 2010; Kang et al., 2012). Taken these data together, a key event must take

place to comply with the final observed DNA organization at the progressing replication fork, leading to strand separation and major structural remodeling. A single particle electron microscopic study of the DmCMG complex further provided insights from the structural perspective by showing that the central channel within the Mcm2-7 ring of the entire complex is strikingly tightened upon nucleotide binding and its dimension unlikely large enough to support a dsDNA passage through it (Costa et al., 2011). In the past years various models have been proposed for the path of the DNA through the central pore of Mcm2-7 during the unwinding action. For the DmCMG helicase, however, models in agreement with the observed final DNA architecture are strongly favored, ruling out those that suggest dsDNA entry into the central channel. Most attractive is the 'steric exclusion' model; one unwinding mechanism proposed for single hexameric helicases, in which the enzyme engages and translocates on one strand while the separated strand is excluded from the interior (reviewed in Brewster and Chen, 2010; Brewster et al., 2008; Kaplan et al., 2003). Such a model is in agreement with the unwinding model derived from the crystal structure of papillomavirus E1 bound to ssDNA in which one strand is "escorted" through the central pore with a possible steric exclusion of the separated one (Enemark and Joshua-Tor, 2006). Further, work from the Walter lab introduced roadblocks on either leading or lagging strand or on both strands simultaneously and succeeded elegantly to demonstrate that the CMG indeed represents a ssDNA translocase encircling the leading strand while the lagging strand remains excluded (Fu et al., 2011). In support of this unwinding mechanism, recent findings from SsoMCM reported dynamic binding and release events of the 5' excluded lagging strand with the exterior surface of the MCM polypeptides, proposing a wrapping around the outer surface (Graham et al., 2011; Rothenberg et al., 2007).

5.3 Results

Drosophila Mcm2-7 possess four β-hairpins analogous to archaeal MCM proteins

On the basis of the near-full-length SsoMCM crystal structure, homology models for each of the six *Drosophila* Mcm2-7 subunits can be derived using SWISS-MODEL, and it was found that they can be nicely threaded onto the existent crystal structure (Figure 5.2A) (Brewster et al., 2008). Electron microscopic and crystallographic studies from several archaeal organisms have provided detailed structural information about their MCM polypeptides. A schematic of the assembly of the six MCM monomers into a toroid shows several characteristic features expected for all MCM and Mcm2-7 ring helicases. The MCM proteins form a two-tiered ring assembly with the N-termini of all six subunits building the top tier stacked on the bottom tier provided by the circle of their AAA+ domains (Figure 5.2A and B). The hexamerization of the MCM subunits leaves a channel in the center running from N- to C-terminal domains of the longitudinal axis of the ring, and further six side channels each leaving the central pore radially towards the exterior of the MCM at the constriction between N- and C-terminal domains at the interface of two subunits (Figure 5.2B). Along the two channels constrictions of the pores become evident, arising from protruding β-hairpins of each MCM polypeptide. In the N-terminal domain the hairpin motif points straight into the central channel and is thereafter names N-terminal β-hairpin (NT-hp) (Figures 5.2B and C). The center of the AAA+ domain is constricted by the Helix-2-insert β-hairpin (H2I-hp), and, in the same domain resides the Pre-Sensor1- β-hairpin (PS1-hp), whose position is mildly shifted from the central pore towards the side channel (Figures 5.2B and D). Unlike these three β-hairpins, the External β-hairpin (Ext-hp) is located at the external surface of each subunit near the side channel between two adjacent MCM subunits (Figures 5.2B and D).

Biochemical and electron microscopic analysis of free Mcm2-7 complexes as well as of those within the higher-molecular weight CMG complex both confirmed a stable hetero-hexameric assembly (Costa et al., 2011; Ilves et al., 2010; Moyer et al., 2006). Thus, it is anticipated that the six different Mcm2-7 subunits in eukaryotic helicases adopt similar structures as the ones solved for the archaeal MCM homo-hexamers. *Drosophila* Mcm2-7 proteins can be modeled based on their homology to archaeal MCMs to identify the positions of the four introduced conserved hairpin motifs, allowing analysis of specific residues of interest in DmMcm2-7 models threaded on the MthMCM structure (Figures 5.2E and F). The only published mutational analyses of these DNA binding hairpin motifs were performed with archaeal MCMs that contain six identical copies of only one MCM

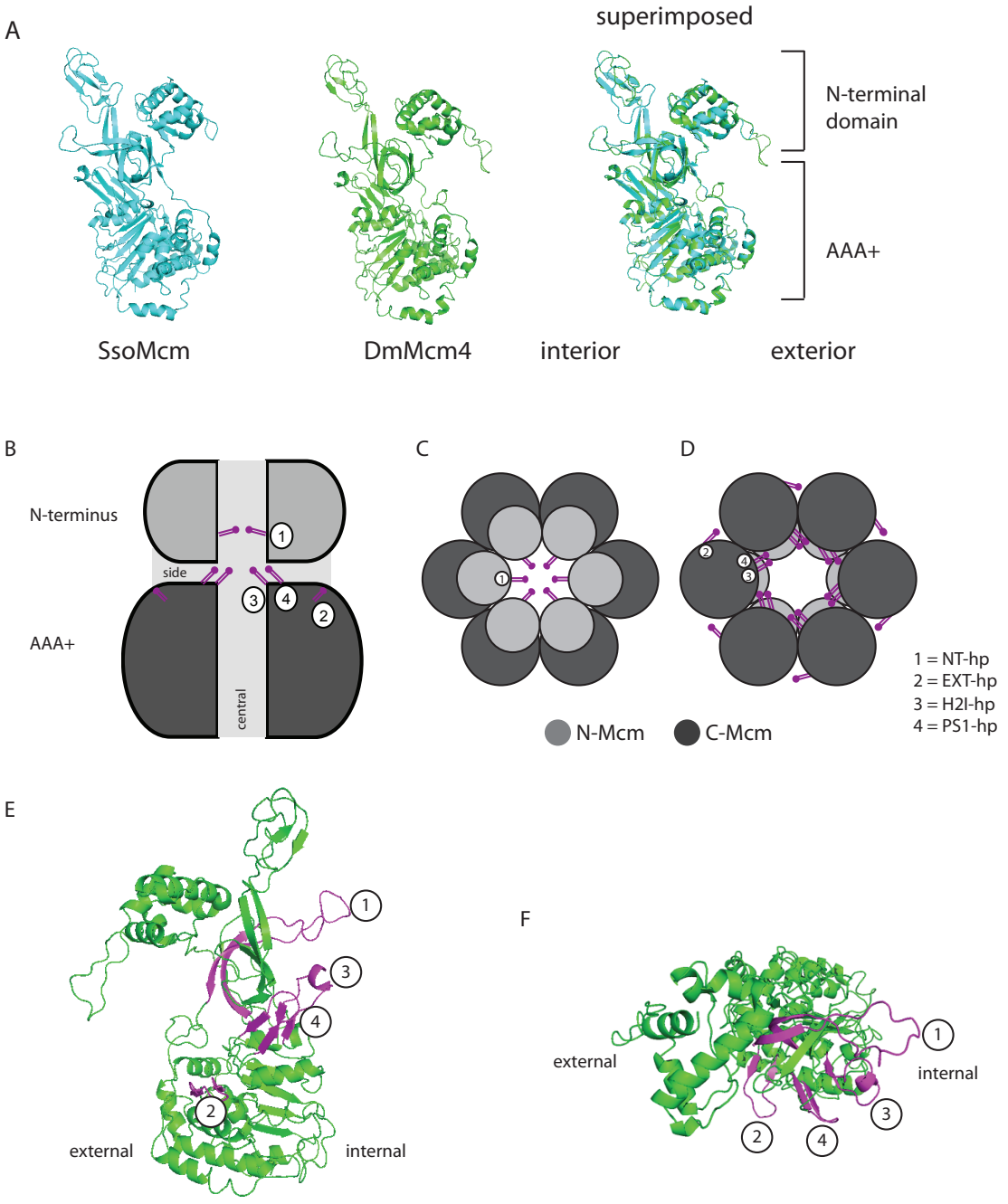

Figure 5.2. Schematic of the β-hairpins within the Mcm2-7 hexameric ring.
(A) Crystal structure of *Sulfolobus solfataricus* (Sso) MCM (PDB entry: 3F9V) in cyan. Homology model of *Drosophila* (Dm) Mcm4 based on SsoMCM structure in green; superimposition of DmMcm4 model onto SsoMCM crystal structure reveals high analogy.
(B) Profile of a schematic Mcm2-7 ring displaying two monomers facing each other. The oligomerization of the six subunits results in formation of a central channel in the middle of the ring, and six side channels each between two adjacent polypeptides proximal to the isthmus between N- and C-terminal domains (*light and dark grey*, respectively) of each subunit. The four hairpins are indicated through numbers from amino to carboxy termini (NT-hp as 1, EXT-hp as 2, PS1-hp as 3, and H2I-hp as 4). Profile figure (*left panel*) was modified after Brewster et al, (2008).
(C) *Top view (N-terminus)* of the schematic Mcm2-7 ring depicting the constriction of the central channel within the N-terminal domain by the NT-hp (numbered with 1) from each of the six Mcm2-7 subunits that point to the center. The β–hairpins are highlighted in *magenta*. In the entire MCM hexamer the N-terminal domains (*light grey*) form a ring that 'stacks' upon the ring formed by the C-terminal domains (*dark* grey) of all six MCM polypeptides.

(D) *Bottom view (AAA+ domain)* of the schematic Mcm2-7 ring showing the three hairpins located in the AAA+ domain. The H2I-hp (3) protrude directly into the central channel, while the PS1-hp (4) are slightly recessed from the center to the side channel. Contrary to these, the EXT-hp (2) are positioned on the exterior of the MCM proteins.
(E) *Side view* of a homology model of DmMcm4 (in *green*) based on the crystallographic structure of *Sulfolobus solfataricus* MCM (PDB entry: 3F9V) as example for Mcm2-7 polypeptides and its four hairpins (in *magenta*). The numbering of the hairpins is explained in (B).
(F) *Top view* of the DmMcm4 model (in *green*) highlighting the hairpins (in *magenta*).

polypeptide that assemble into a hexameric ring structure, competent for robust unwinding. However, this also entails that introduction of a mutation in the MCM is simultaneously present in all six copies of the protein, and, thus, enables only the analysis of sixfold mutant MCMs. Various residues reported to be required for either DNA binding or unwinding activities in archaeal MCM helicases were targeted in this study in individual Mcm2-7 subunits within the *Drosophila* CMG complex to analyze their effects and understand the contribution of single MCM subunits.

Drosophila Mcm6 subunit posseses a long NT-hp

Earlier reports from MthMCM and SsoMCM demonstrated the importance of the NT-hairpin motif positioned in the amino-terminal domain of the protein. The crystal structure of the dodecamer of N-terminal fragments of MthMCM proteins emphasized the positive charges in this hairpin motif suggesting a role in DNA binding (Fletcher et al., 2003). These motifs from each of the six monomers point into the central channel, causing a drastic constriction, and modeling of dsDNA showed that even a duplex could be accommodated at this narrowest point of the cavity (Fletcher et al., 2003). The hypothesis that MthMCM binds dsDNA in the central channel was confirmed by alanine substitutions of two basic residues at the tip of the hairpin (Arg227 and Lys229), which entirely abrogated the binding to dsDNA, and also to ssDNA. However, the effect of such mutations in the context of the full-length MCM helicase remained elusive. Subsequent mutagenesis in the corresponding NT-hp of full-length SsoMCM introduced alanine substitutions of Lys246 and Arg247 at the tip of the motif, showing that the DNA binding was strikingly reduced and, thus, also decreased subsequently the unwinding activity (McGeoch et al., 2005).

On the other hand, the eukaryotic helicase not only consists of six homologous, but different MCM proteins numbered 2 through 7 (Mcm2-7), but also requires additional factors (Cdc45 and GINS) for its activation to a robust helicase activity (Aparicio et al., 2006; Ilves et al., 2010; Im et al., 2009; Moyer et al., 2006). Sequence alignment of all DmMcm2-7, MthMCM, and SsoMCM by ClustalW2 helped identify the location of their NT-hp motifs (Figure 5.3A). Almost all Mcm2-7 subunits (besides Mcm2) contain several positively charged residues throughout the hairpin motif. With the help of the Mth N-MCM as well as the near-full-length SsoMCM crystal structures, the two critical basic residues from the corresponding archaeal MCM could be identified at the tip of the hairpin in several Mcm2-7 subunit (Figures 5.2B and C)(Brewster et al., 2008; Fletcher et al., 2003). Both residues project straight into the putative central channel, but strikingly only three of the six Mcm2-7 subunits contain the combination of an adjacent arginine and lysine

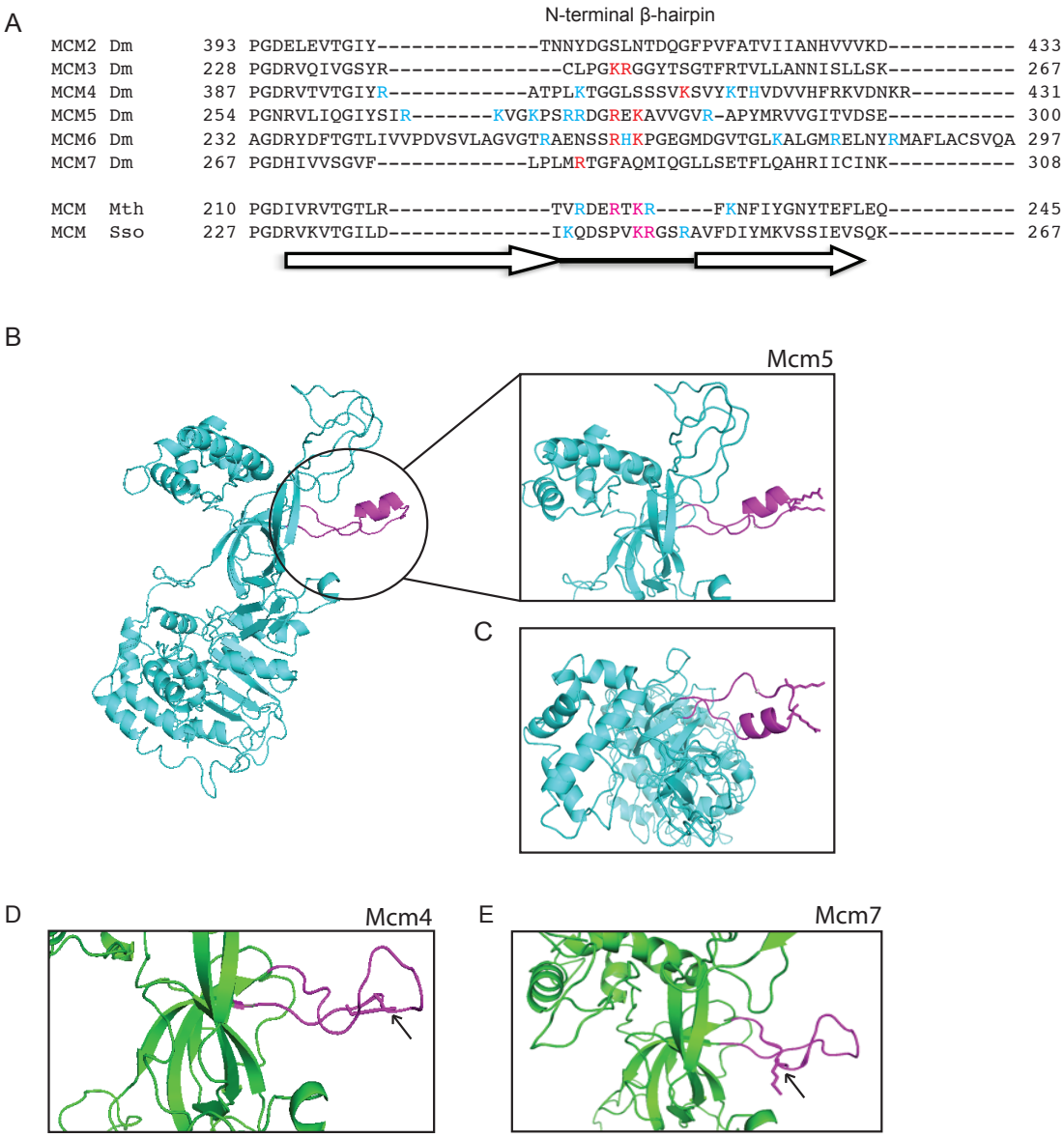

Figure 5.3. Sequence alignment of N-terminal β-hairpin region.
(A) Multiple sequence alignment of *Drosophila* Mcm2-7, *M. thermoautotrophicus* (Mth) and *S. solfataricus* (Sso) MCM using ClustalW2. Shown is the region of the NT-hp, and all basic residues within each subunit's NT-hairpin loop are highlighted in color (in *cyan, red* and *magenta*). The two basic residues (lysine K and arginine R) have been previously mutated in the archaeal Sso MCM (in *magenta*), and residues at corresponding positions in the Mcm2-7 NT-hp have been substituted with an alanine (in *red*). All remaining basic residues within the loop of the hairpin (in *cyan*) have been left unmodified. Arrows below indicates the β-strands and loop regions for Sso MCM.
(B) *Side view* of a homology model of DmMcm5 (in *cyan*) based on the crystal structure of Sso (PDB entry: 3F9V). The location of the NT-hp is highlighted (in *magenta*) and the two mutated basic residues (Arg278 and Lys280) within the hairpin are represented as sticks in the close-up view of this hairpin.
(C) *Top view* of the homology DmMcm5 model depicting the NT-hp with the two mutated residues (K and R as sticks).
(D) *Side view* of NT-hp loop of *Drosophila* Mcm4 highlighting basic residue Lys412 (*stick*).
(E) *Side view* of NT-hp loop of DmMcm7 highlighting the positively charged residue Arg282 (*stick*).

residue (Figures 5.3A and C). For the *Drosophila* helicase motor, such arginine/lysine pairs could only be identified in Mcm3, Mcm5 and Mcm6. Interestingly, Mcm3 possesses Lys244 and Arg245 directly next to each other, reminiscent of the analogous adjacent residues in the SsoMCM NT-hp, while the two basic residues at the tip of NT-hp in Mcm5 (Arg278 and Lys280) and Mcm6 (Arg263 and Lys265) are separated by a residue, as can be observed in a similar way for the MthMCM NT-hp (Arg227 and Lys229) (Figures 5.3A, B, C). DmMcm4 and DmMcm7 do not contain such Lys/Arg pairs, and possess only Lys or Arg residues, which are slightly recessed to the base of the hairpin motif (Figures 5.3A, D, and E). Targeted residue in Mcm4 (Lys412) was based on the closest location to the analogous basic residues in Mcm 3, 5, or 6, and in Mcm7 the only basic residue was mutated. In contrast to this, no positively charged residues are present in the NT-hp of DmMcm2.

Also, a closer look at the multiple sequence alignment and the loops of NT-hp makes another striking difference apparent. The length of the N-terminal β-hairpin loops varies significantly between the six DmMcm2-7 subunits as well as the amount of positively charged residues contained throughout this motif. While Mcm6 possesses, compared to SsoMCM and the other DmMcm2-7, the longest NT-hp loop, it also contains most of the positively charged residues at the tip of the hairpin (Figures 5.3A and 5.4A, B, C). Following Mcm5 with its eight positively charged amino acids within the loop, seven such residues can be found throughout the loop of Mcm6 NT-hp (Figures 5.3A).

NT-hp is required for DNA binding by CMG

While corresponding residues have been shown to be necessary in the isolated N-terminus of MthMCM and in the full-length SsoMCM, so far the effect of analogous mutations in the NT-hp of eukaryotic MCM helicases remained unanswered (Fletcher et al., 2003; McGeoch et al., 2005). Therefore, *Drosophila* CMG complexes with alanine substitutions of the lysine and arginine residues within the NT-hp of either single, a subset or all of the Mcm2-7 proteins have been purified and subsequently tested for DNA binding (Figure 5.5A). The ability of these β-hairpin mutant CMG proteins to bind to DNA was analyzed in electrophoretic-mobility-shift assays (EMSA) using a forked substrate with a 50bp duplex and 40nt unpaired poly-thymidine extension on 3' and 5' strands. The CMG complexes with NT-hp mutations in Mcm3, 5 and 6 remained of particular interest, because they possess adjacent lysine and arginine amino acids as those seen in archaeal MCM helicases. Besides CMG with mutations in the NT-hp of these individual subunits ('3Nthp', '5Nthp', and '6Nthp), simultaneous mutations in the NT-hp of all three proteins and, thus, one-half of the Mcm2-7 ring ('356Nthp') were introduced (Figure 5.5A). To test if the NT-hp in Mcm2-7 within the CMG complex were similarly defective in DNA binding

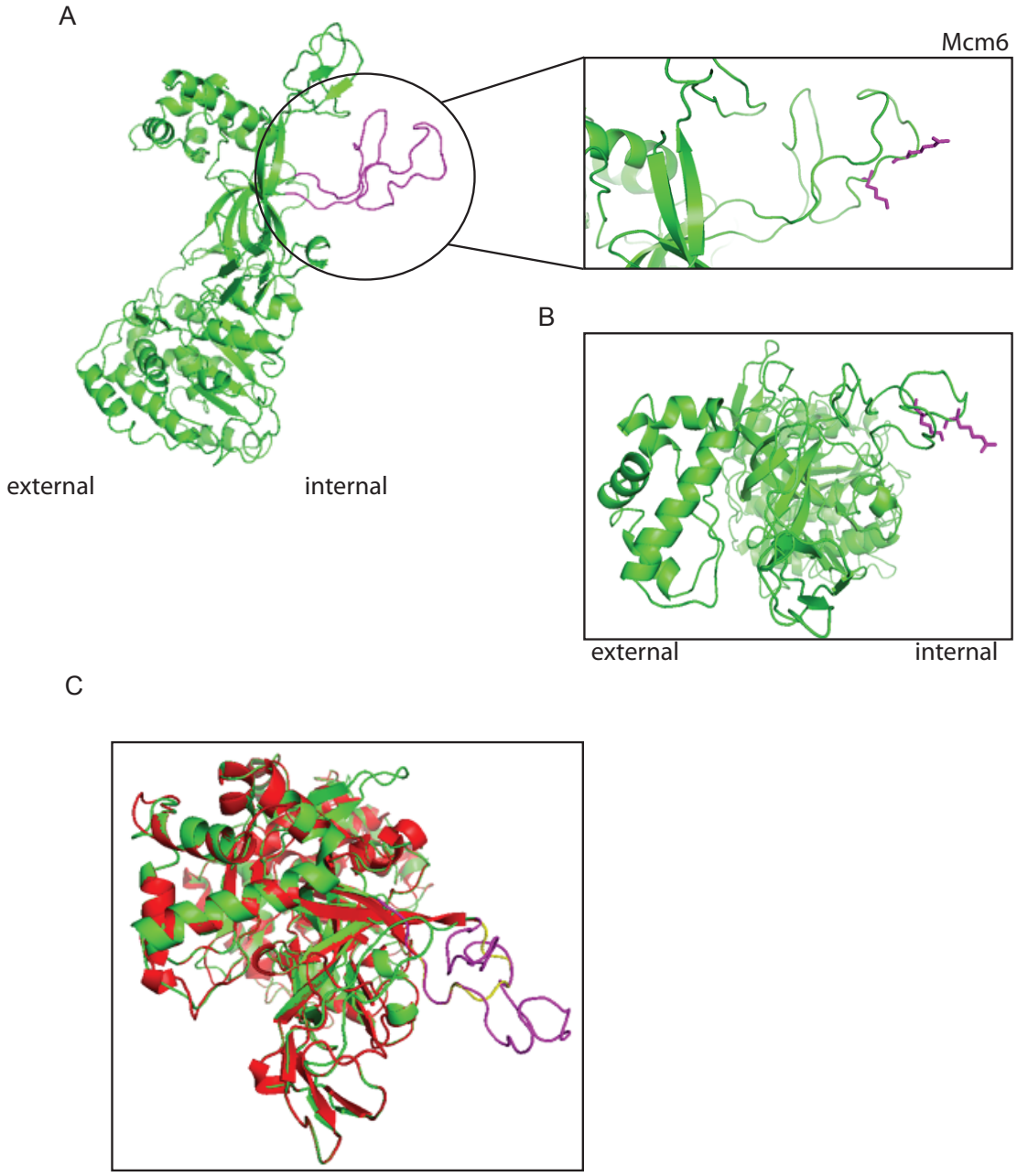

Figure 5.4. Homology modeling of DmMcm6 reveals a long NT-hp.
(A) *Side view* of a homology model of DmMcm6 (in *green*) based on the crystal structure of Sso MCM (PDB entry: 3F9V). The NT-hp is highlighted (in *magenta*) and the two mutated basic residues within the hairpin are shown as sticks in the close-up view of this hairpin (Arg263 and Lys265).
(B) *Top view* of the homology DmMcm6 model depicting the NT-hp with the two mutated residues (K and R as sticks).
(C) Homology model of DmMcm6 (in *green*) threaded onto the crystal structure of SsoMCM (in *red*). Both NT-hp are highlighted (Dm in *magenta*, Sso in *yellow*) showing the differences in length.

when eliminated, as was earlier reported for archaeal MCM NT-hp, a complex containing mutations of the basic residues in NT-hp of Mcm3, 4, 5, 6, and 7 ('5xNthp') was purified and tested side-by-side. Since Mcm2 does not contain any positively charged residues in its NT-hp, the '5xNthp' CMG was prepared in presence of wildtype Mcm2 protein. While the mutations of NT-hp in Mcm3 (86% of wt), Mcm5 (90% of wt) or Mcm6 (98% of wt) alone retained mostly wildtype DNA binding, the '5xNthp' CMG complex resulted in a striking reduction of its binding to about only 32% of wt DNA binding (Figures 5.5 C).

Given the fact that, compared to archaeal MCMs, only a subset of DmMcm2-7 proteins contain similar motifs at the tip of the hairpin, I hypothesized that this observed defect in DNA binding predominantly arises from the mutant NT-hp motifs of subunits Mcm3, Mcm5, and Mcm6, while assuming that the NT-hp of Mcm4 and Mcm7 might rather play a subsidiary role. In support of this view, the DNA binding of CMG with NT-hp mutations in '356Nthp' was significantly diminished to about 58% of the wildtype binding, whereas mutations in '47Nthp' displayed an unaltered wildtype DNA binding (Figures 5.5C). In archaeal organisms defects of the NT-hp mutants were only tested for the MCM helicase activity, as no archaeal counterpart of the CMG complex has thus far been reported. On the other hand, different from most eukaryotic helicases, the isolated archaeal MCM homo-hexamer already displays robust helicase activity. Together, the observations from archaeal and *Drosophila* helicases suggest that the NT-hp within the amino-terminal domain of MCM in both cases play a comparable role, making these hairpin motifs a requirement for DNA binding, either in the isolated MCMs or the MCMs within the CMG complex. However, to validate this point, *Drosophila* Mcm2-7 complexes were purified containing '5xNthp', '356Nthp', or '47Nthp' mutations and tested for DNA binding (Figure 5.5B and D). As reported for these mutations within the CMG complex, the same pattern of defects in the DNA binding was observed for the isolated Mcm2-7 complexes. Different from the CMG helicase, all three mutant Mcm2-7 complexes displayed significantly reduced DNA binding affinities (Figure 5.5D). Consistent with the data observed in CMG, the mutations in NT-hp of Mcm4 and Mcm7 showed the weakest defects. This data suggest that, while the same NT-hp seem indispensable for the complexes' binding to DNA, the CMG complex has likely undergone strong structural re-arrangements that possibly led to a proper alignment of other redundant DNA binding motifs, making the overall defects of the NT-hp mutations less prominent.

However, testing of NT-hp in *Drosophila* Mcm2-7 and CMG complexes confirmed an unequal contribution of these hairpin motifs in the six Mcm2-7 subunits to the DNA binding. Compared to the remaining monomers, three of the subunits of the Mcm2-7 ring seem to contribute more importantly to the DNA binding within the central channel.

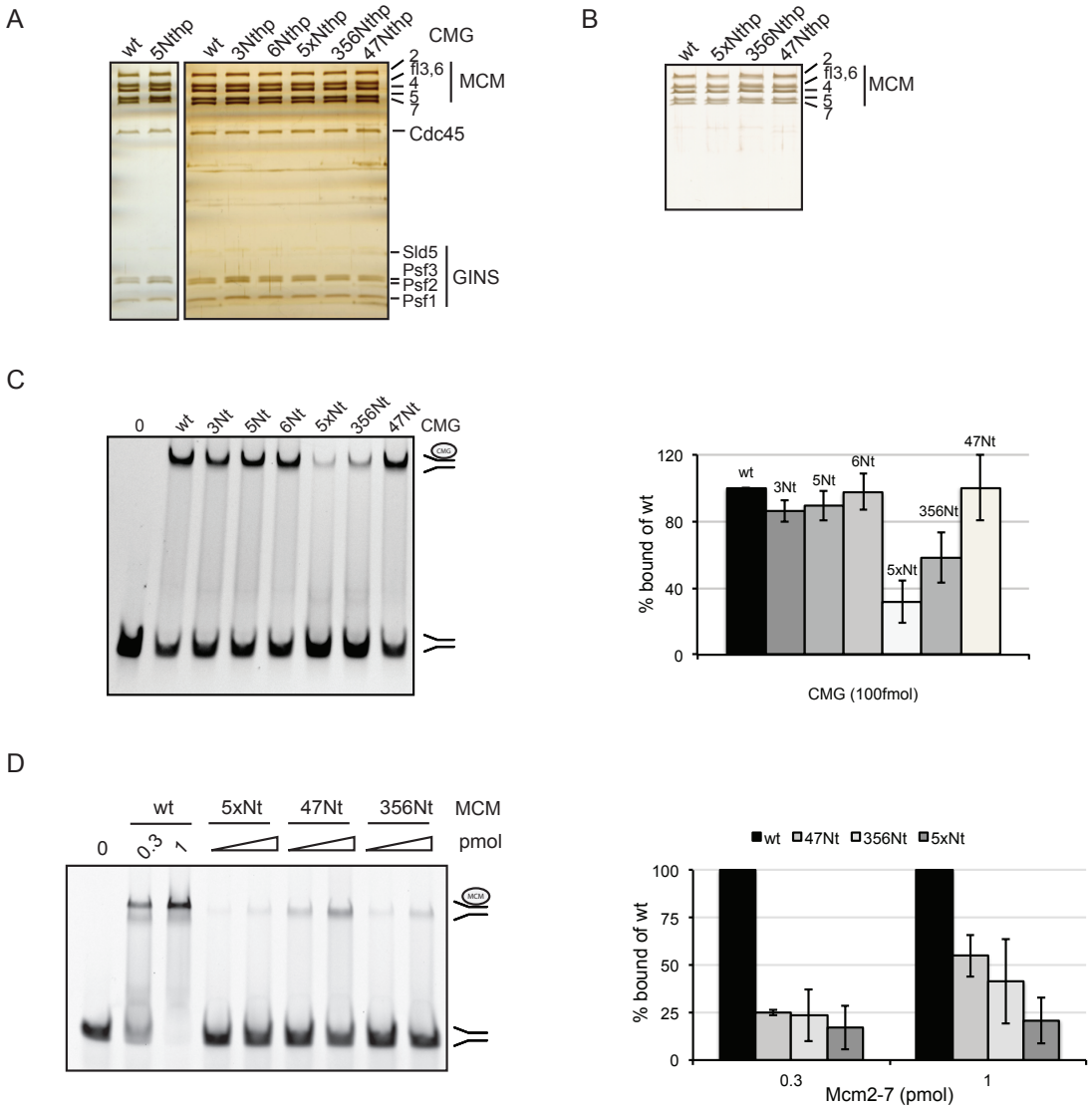

Figure 5.5. NT-hp in Mcm2-7 are required for DNA binding of the CMG.
(A) Purified rCMG complexes containing alanine substitutions in the basic residues of their NT-hp were separated by SDS-PAGE (10%) and proteins were silver-stained for visualization. The numbers indicate which of the Mcm2-7 subunits contains the NT-hp mutation, e.g. Mcm3Nt stands for the NT-hp mutation in the Mcm3 subunit within the CMG complex.
(B) Mcm2-7 side-fractions from CMG purification have been subsequently further purified and separated by SDS-PAGE (10%). Proteins were silver-stained for visualization.
(C) EMSA of wt and NT-hp mutant CMG complexes. Per reaction 100fmol of purified protein were added to 100fmol of FITC-labeled 'fork' substrate (3'+5' polyT tails) in presence of 10μM ATPγS. Positions of free and CMG-bound substrate are indicated on the side. Quantification of gel-shifts with standard deviations from two independent series are depicted (*right panel*).
(D) EMSA of wt and NT-hp mutant Mcm2-7 complexes. Two different protein concentrations (0.3pmol, 1pmol) were added to 100fmol of a fork substrate with polyT extensions of 3' and 5' strands. Binding reactions were performed in absence of nucleotide as Mcm2-7 binding to DNA is ATP-independent. Positions of free and MCM-bound substrate are shown on the side of the gel. Quantification of EMSA with standard deviation from two independent series were included (*right panel*).

As this motif in the N-terminal domain of Mcm2-7 was absolutely essential for DNA binding, it was crucial to understand in a next step the effects that these mutations might have on the unwinding activity of the CMG complex. All previously described NT-hp mutations in single and multiple Mcm2-7 subunits within the CMG have been tested on a circular substrate with a 40bp duplex region and a 30nt 5' poly-thymidine extension. In agreement with their defects on DNA binding, all three tested NT-hp mutations in individual Mcm2-7 subunits (Mcm3, Mcm5, or Mcm6) as well as the combined mutations in Mcm4 and Mcm7 together ('4/7Nthp') displayed wildtype helicase activity (Figures 5.6A and B). In contrast to this, mutations in the NT-hp of either all Mcm2-7 subunits ('5xNthp') or in Mcm3, Mcm5, and Mcm6 simultaneously ('356Nthp') showed a strongly reduced, but detectable helicase activity to levels consistent with the diminished DNA-binding activity (Figures 5.6A and B).

Alternatively, to test if the NT-hp in Mcm2-7 established direct contact to the leading strand in the central pore of the CMG complex, I used the 'leading' strand forked substrate with modified fluorescein isothiocyanate (FITC) groups at specific residues along the 3' strand to analyze the binding. This substrate was introduced in Chapter 4 and contains FITC-modified nucleotides on the 3' thymidine strand, while the 5' strand consists of a GC-rich sequence that prevents CMG to bind on that strand. UV-induced protein:DNA cross-linking of these groups allowed me to monitor the proteins in the CMG that establish contact points to the DNA substrates (see for details Chapter 4). While in this assay mainly Mcm5 as well as Mcm3/Mcm6 – that co-migrate on SDS-PAGE of the wtCMG - cross-link to the DNA in absence of nucleotide, all remaining Mcm2-7 subunits participate in DNA-protein contacts upon ATP binding (see Chapter 4, Figures 5.3A and B).

In support of the observed reduction of DNA binding to a forked DNA substrate in gel-shift assays and the proposal that particularly NT-hp of Mcm3, Mcm5 and Mcm6 seem to account for this, the cross-linking assay further revealed that CMG '5xNthp' and '356Nthp' mutant CMG complexes also altered the direct protein-DNA interactions. In both mutant complexes, Mcm3/Mcm6 failed to establish any direct interaction to the leading strand in absence of ATPγS when the NT-hp were mutated (Figure 5.6C). However, both CMG complexes show the same cross-linking pattern of all Mcm2-7 subunits upon nucleotide binding, suggesting that binding of ATPγS induces re-arrangements within the central channel through which additional motifs can engage the DNA. Taken these observations together, it leads me to propose that among all six Mcm2-7 proteins, predominantly Mcm3, Mcm5, and Mcm6 bind the leading strand in the initial recruitment; a hypothesis corroborated by the additional finding, that CMG '47Nthp' does not show any altered DNA-protein contacts after induced cross-linking.

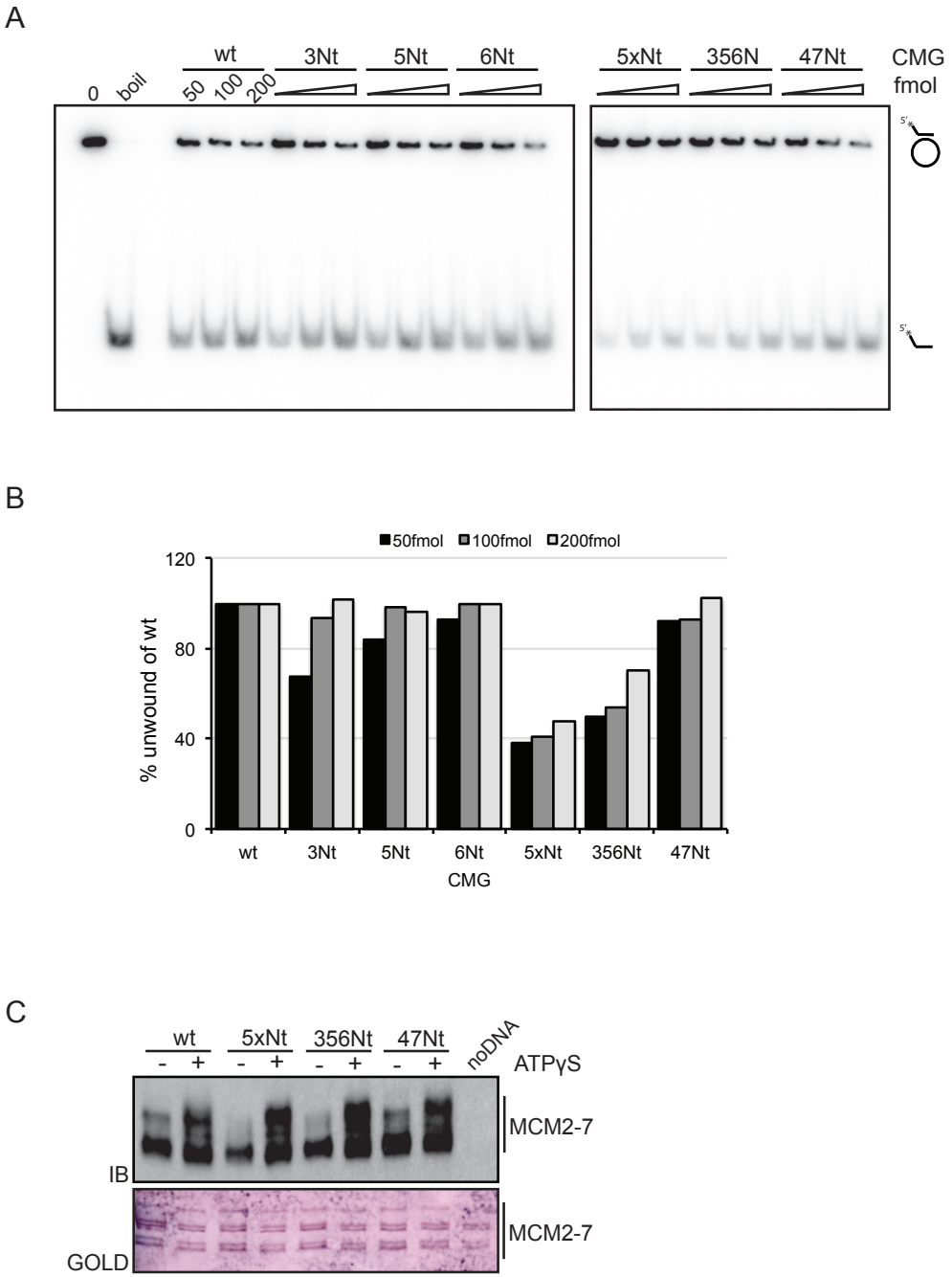

Figure 5.6. Helicase activity of 'NT-hp' CMG complexes is diminished.
(A) Autoradiograph of helicase activity assay on a circular substrate with wt CMG or CMG with mutations in the NT-hp residue in either single Mcm2-7 subunits or several Mcm2-7 proteins simultaneously. Position of circular substrate and displaced oligonucleotide alone are indicated on side. Increasing amounts of purified CMG proteins (50, 100, 200fmol) were added to 1fmol of DNA substrate in presence of 300μM ATP. Unwound products were separated by TBE-PAGE (8%). '0' and 'boil' indicate controls with substrate alone or heat-denatured substrate, respectively.
(B) Quantification of helicase activity assays from (A).
(C) Western blot analysis of forked FITC-labeled 'leading' strand DNA substrate cross-linked to wt and NT-hp mutant CMG by UV in absence and in presence of ATPγS. Reactions were separated by SDS-PAGE (10%). Immunoblot was probed with monoclonal α- FITC antibody to detect cross-linked DNA-protein bands. An even loading of various CMG proteins was monitored by staining the western blot with AuroDye™ (GE Healthcare) to visualize the protein bands.

PS1-hp of Drosophila Mcm2-7 posses an invariant lysine in its loop

The hypothesis that Mcm2-7 proteins must engage the leading strand within the central cavity tighter upon nucleotide binding was first postulated for the *Drosophila* CMG through the direct observation of such constriction in a structural study (Costa et al., 2011). A second line of evidence derives from biochemical findings on how the Mcm2-7 subunits establish direct contact to the leading strand. In absence of nucleotide, only a subset of Mcm2-7 proteins within the CMG interact with the leading strand that passes through the central pore, whereas all six subunits individually contact the DNA when ATP is bound suggesting a tighter encircling (see Chapter 4) (Pesavento et al., 2013, manuscript in preparation). While the NT-hp constitutes one of the crucial DNA-binding motifs within the N-terminal domain of MCM proteins, it is obvious that further motifs for DNA-binding and subsequent –unwinding exist.

A second candidate β-hairpin involved in DNA binding and helicase activity was found in the AAA+ domain, slightly recessed from the central channel. Structural studies demonstrated in the SsoMCM crystals of the near-full-length protein the location of this hairpin motif (Brewster et al., 2008). The characteristic AAA+ domain of MCM proteins that encompasses the ATPase active site motifs contains an insertion upstream of the sensor 1 motif, which itself is implicated in sensing the status of bound nucleotide and interacting with the γ–phosphate of ATP (reviewed in Hanson and Whiteheart, 2005; Hattendorf and Lindquist, 2002). This insert has high sequence conservation between archaeal and eukaryotic MCM polypeptides, and crystal structures of MCM proteins show the formation of a β-strand at the corresponding position, thereafter named Pre-Sensor 1 β-hairpin (PS1-hp) (Brewster and Chen, 2010; Brewster et al., 2008; Hanson and Whiteheart, 2005). Further, structural studies of SV40 Tag or papillomavirus E1 that belong to the Superfamily-3 (SF3) helicases, demonstrated that also these proteins contain a similar insertion at this location that forms a β-hairpin, which extends into the central channel (Abbate et al., 2004; Enemark and Joshua-Tor, 2006; Gai et al., 2004; Li et al., 2003).

Multiple sequence alignments of all *Drosophila* Mcm2-7 using ClustalW2 disclosed the high sequence conservation of the PS1-hp region (Figure 5.7A). Homology modeling of the DmMcm2-7 based on the crystal structure of SsoMCM (PDB entry: 3F9V) enabled a threading of these polypeptides onto the SsoMCM structure, thus allowing the identification of the discrete location of a conserved basic residue (Figure 5.7B). An invariant lysine at the tip of the loop was previously demonstrated to mildly decrease DNA binding but entirely disrupt helicase activity in archaeal MCM (McGeoch et al., 2005).

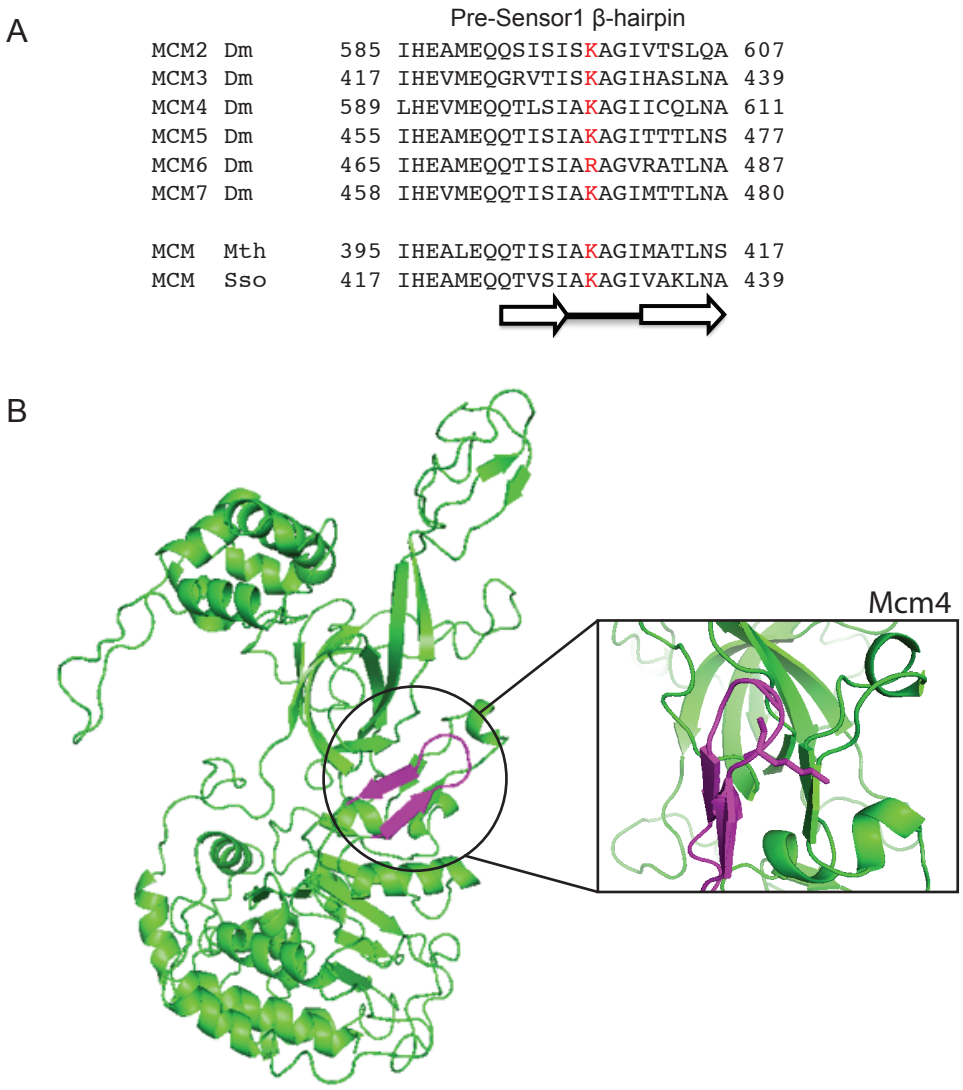

Figure 5.7. Sequence alignment of Pre-Sensor1 β-hp of *Drosophila* Mcm2-7.
(A) Multiple sequence alignment of *Drosophila melanogaster* (Dm) Mcm2-7, *Sulfolobus solfataricus* (Sso) MCM, and *Methanobacter thermoautotrophicus* (Mth) MCM. Shown is the region around the Pre-Sensor1 β-hp (PS1-hp) highlighting the critical invariant lysine (K in *red*). Arrows and line below indicate the β-strands and loop regions for Sso MCM.
(B) *Side view* of a homology model of DmMcm4 (in *green*) based on the crystal structure of SsoMCM (PDB entry: 3F9V) is shown. The position of the PS1-hp is highlighted (in *magenta*), and the conserved lysine residue at the tip of the hairpin is represented as a stick in the close-up view.

The invariant basic residue in PS1-hp is critical for DNA binding and helicase activity

To test whether this β-hairpin motif might participate in the unwinding mechanism, an alanine substitution of this positively charged residue was introduced in the PS1-hp of all Mcm2-7 subunits (Figures 5.7A). Similar to the mutational analysis of this NT-hp, introducing the mutation in the PS1-hp of Mcm2-7 did not affect the formation of rCMG complexes, yielding complexes of equal stoichiometry and wildtype purification feature (Figure 5.8A). Complexes with alanine substitutions in single Mcm2-7 subunits within the CMG were purified (named 'PS1'), and their ability to bind DNA was subsequently tested. The same 3'+5' poly-T forked DNA substrate, as described for the NT-hp studies, was used, and the gel-shift assays instantly revealed that almost all PS1-hp mutant CMG complexes displayed a reduced DNA binding (Figure 5.8B). Only the single mutation in the PS1-hp of Mcm6 retained about 100% of the wtCMG activity. Even though all other mutant PS1 CMG complexes showed defects in DNA binding, the extent of reduction strongly varied, with mutations of Mcm4PS1 and Mcm3PS1 being the most affected and leading to 55% and 48%, respectively, reduced DNA binding when compared to wtCMG. The remaining CMG complexes retained ~73-76% of wtCMG-DNA binding (Figure 5.8B).

Neither the NT-hp nor PS1-hp within the Mcm2-7 seem absolutely essential for DNA binding of the CMG complex when either of the hairpin motifs was mutated in individual Mcm2-7 subunits. To test if the findings from SsoMCM were also true for the *Drosophila* helicase and if these PS1-hp mutations would result in a complete abrogation of unwinding activity, the helicase activity was measured on a circular DNA substrate (McGeoch et al., 2005). Besides mutations of Mcm5PS1, all remaining CMG complexes with corresponding mutations in the other five Mcm2-7 subunits experience diminished unwinding activity (Figure 5.8C). While most diminished unwinding resulted in a slightly increased reduction compared to the levels as were observed for the DNA binding activity, the mutation in PS1-hp of Mcm7 most strikingly decreased the helicase activity (16-36% of wt unwinding compared to 76% of wt DNA binding) (Figure 5.8C).

Taken the results together and assigning the defects to the unique Mcm2-7 order around the ring, an asymmetric contribution from the Mcm2-7 subunits both in DNA binding as well as helicase activity becomes evident (Figure 5.9). CMG with mutation in Mcm4 PS1-hp displays the weakest DNA binding, and PS1-hp mutations in Mcm3, in Mcm4, and in Mcm7 reduce the helicase activity the most (Figure 5.9 A-D). These findings emphasize a strong asymmetry of the Mcm2-7 with one-half of the ring (Mcm3, 4, 7) being more important for DNA binding and the unwinding mechanism through the PS1-hp (Figure 5.9 B and D).

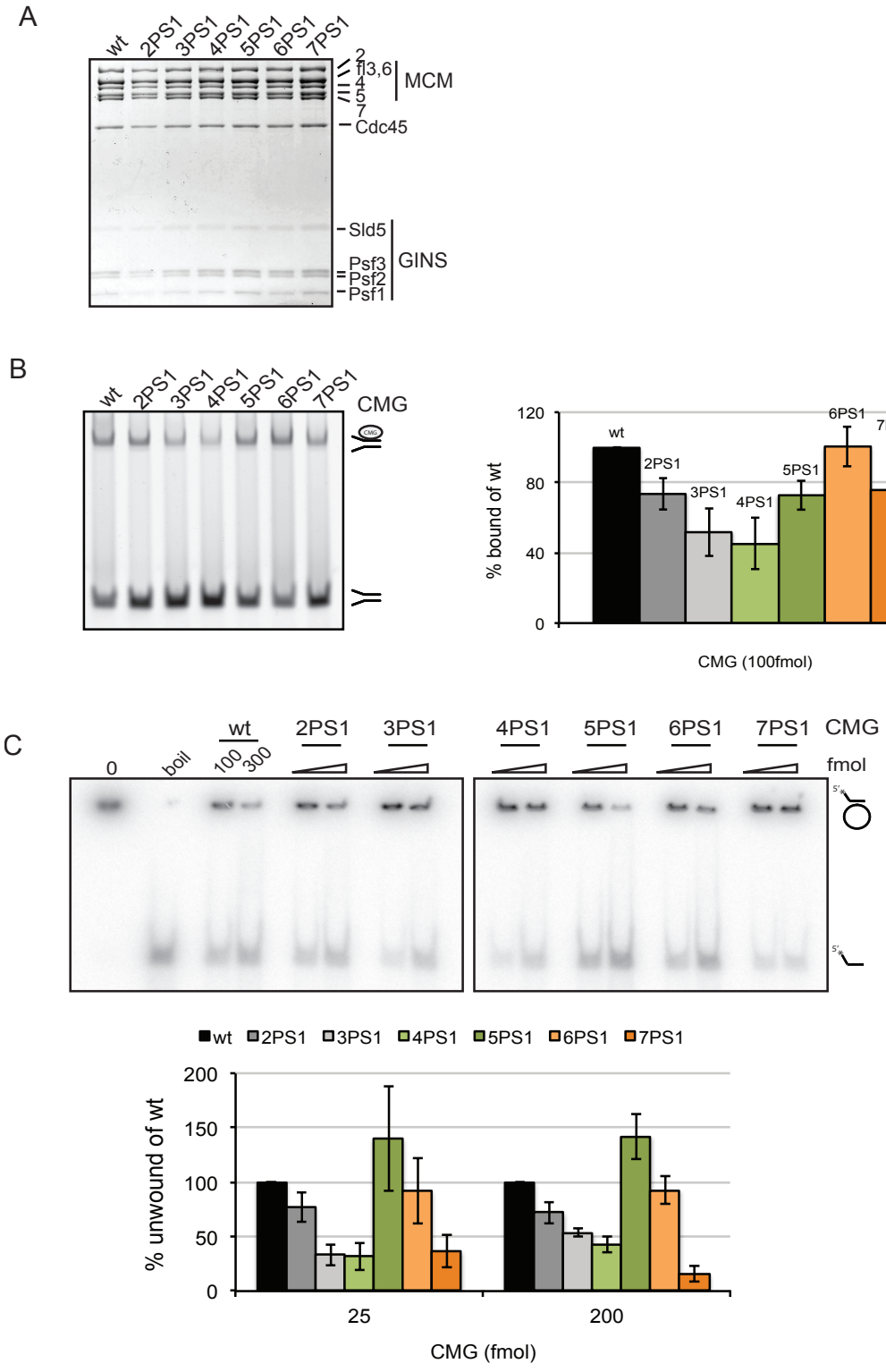

Figure 5.8. Mutations of PS1-hp in Mcm3/4/7 reduce helicase activity.
(A) rCMG complexes containing an alanine substitution of the conserved lysine in the Pre-Sensor1 β-hairpin of one of the Mcm2-7 subunits accompanied by five wt Mcm proteins were purified and separated by SDS-PAGE (10%). Which subunit is mutant is indicated by the number, e.g. '2PS1' introduces the mutation in the PS1-hp of Mcm2 within the CMG complex. The proteins were visualized by subsequent silver-staining.
(B) EMSA of wtCMG and 'PS1-hp' CMG complexes. All lanes contain 100fmol of FITC- labeled forked DNA substrate (3'+5'polyT) and where indicated 100fmol of purified CMG protein in presence of

10μM ATPγS. Binding reactions were incubated for 30min and then separated on a native TBE polyacrylamide gel (4%). Signals were detected by FITC fluorescence of the labeled DNA. Positions of free 'fork' substrate and CMG bound substrate are indicated on the side of the gel. Quantification of the gel-shift assays is shown on the right. Average values, expressed as percentage of DNA bound by wtCMG (set as 100%), are plotted with standard deviations of five independent series.
(C) Autoradiograph of helicase activity assay on a circular DNA substrate. Two different amounts of purified CMG proteins (100, 300fmol) were added to 1fmol of DNA substrate in presence of 300μM ATP. Unwinding was allowed for 30 min and products were separated by TBE-PAGE (8%). '0' and 'boil' contain substrate controls without protein or boiled substrate, respectively. Positions of free circular substrate and displaced oligonucleotide are indicated on the side. Four independent series of helicase assays were quantified, and average values of unwinding were expressed as percentage of wtCMG including the standard deviations of four independent series.

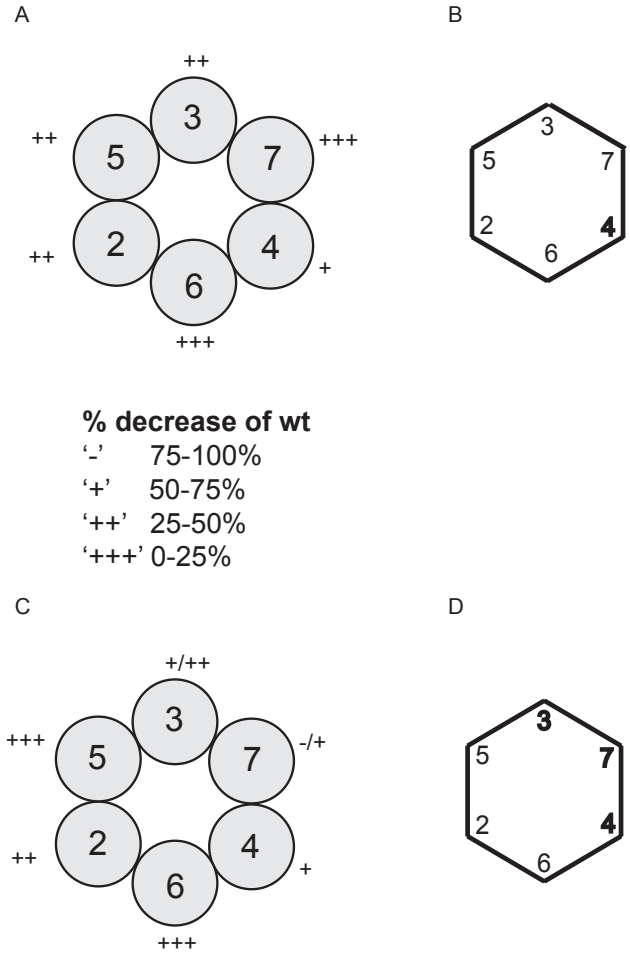

Figure 5.9. Schematic summary of PS1-hp effects on DNA binding and helicase activity.
(A) Schematic of Mcm2-7 ring in CMG showing the unique order of all six Mcm2-7 subunits. The schematic summarizes the relative effects of point mutations in the PS1-hp on the DNA binding of the CMG complexes. The percentages of decreased binding after introducing the alanine substitutions of the conserved lysine are shown for all MCM subunits around the ring. Hereby '-' indicates 75-100%, '+' 50-75% , '++' 25-50%, and '+++' 0-25% diminishment of wtCMG DNA binding.
(B) Summarizing schematic of the effects listed in (A). The strongest affected MCM subunits with ≥ 50% diminished binding are highlighted with thicker lines.
(C) Schematic as in (A), but this schematic illustrates the relative effects of the PS1-hp point mutations on the helicase activity. The decreased unwinding activities are shown as percentages of the wtCMG activity for all individual MCM subunits around the ring.
(D) Summarizing schematic of the effects listed in (C). The strongest affected MCM subunits are highlighted with thicker lines.

Asymmetric contributions of the PS1-hp in Mcm2-7 within the CMG to the DNA binding

Since none of the point mutations in PS1-hp of single Mcm2-7 subunits within the CMG entirely abrogated the DNA binding or helicase activity of the full complex, a CMG complex with mutations in the PS1-hp of all six Mcm2-7 polypeptides ('6xPS1') was tested (see also Chapter 4, Figures 5.4C and D). Similar to the observations from the archaeal SsoMCM helicase that contain the corresponding PS1-hp mutation in all subunits, no unwinding activity was present (Figure 5.10D). To understand if this observed abrogation of helicase activity absolutely required the simultaneous mutation of PS1-hp in all six Mcm2-7 subunits, or if it resulted from mutation in only a subset of Mcm2-7 proteins, CMG complexes were purified that possess different combinations of this hairpin mutation (Figure 5.10A). These mutations pointed to the strongest defective helicase activity in CMG when PS1-hp in Mcm3, Mcm4 or Mcm7 were mutated, while mutations in the other subunits either displayed none or very weak effects (Figures 5.8C and 5.9C).

To test this apparent importance of the PS1-hp motif in this half of the Mcm2-7 ring for DNA binding and unwinding, CMG was purified with PS1-hp-mutant Mcm3, Mcm4, and Mcm7 ('347PS1'), the two most affected subunits Mcm4 and Mcm7 alone ('47PS1'), or Mcm2, Mcm5, and Mcm6 ('256PS1') (Figure 5.10A). Analysis of DNA binding by gel-shift assays showed that all three mutant CMG complexes lead to a decreased DNA binding, in which the reduction by the CMG '347PS1' and '47PS1' are more prominent and diminish the wildtype binding activity by 79% and 75%, respectively (Figure 5.10B). Simultaneously introduced alanine substitution in the PS1-hp of Mcm2, 5, and 6 weakened the DNA binding activity by about 50% (Figure 5.10B). Next, the unwinding activity of these complexes was tested on a circular DNA substrate. At the same time, several additional CMG complexes were purified that introduce mutant PS1-hp in different MCM pairs around the ring (Figure 5.10A). The pairs were chosen such to be adjacent to the Mcm4 PS1-hp mutation as this CMG complex displayed the weakest enzymatic activity (Mcm4 and Mcm6 as '46PS1', and Mcm3 and Mcm7 as '37PS1') (Figure 5.10A).

The majority of the CMG complexes hosting two or three mutant PS1-hp in the Mcm2-7 subunits completely inactivated the helicase activity (Figure 5.10D). Unlike the '37PS1' CMG complex, the helicase activity '47PS1' CMG showed less reduced unwinding to about 23% of the wildtype activity. Even though DNA binding was comparably abolished between all these PS1-hp CMG complexes, this data suggest, that for the entire ablation of the CMG activity, the minimum of two mutant PS1-hp in Mcm3 and Mcm7 are required, and attribute a special role for translocation to these two subunits (Figure 5.10C).

When alanine substitutions were introduced into the N-terminal hairpin motifs of DmMcm2-7 subunits, a comparison of their effects on DNA binding of isolated Mcm2-7

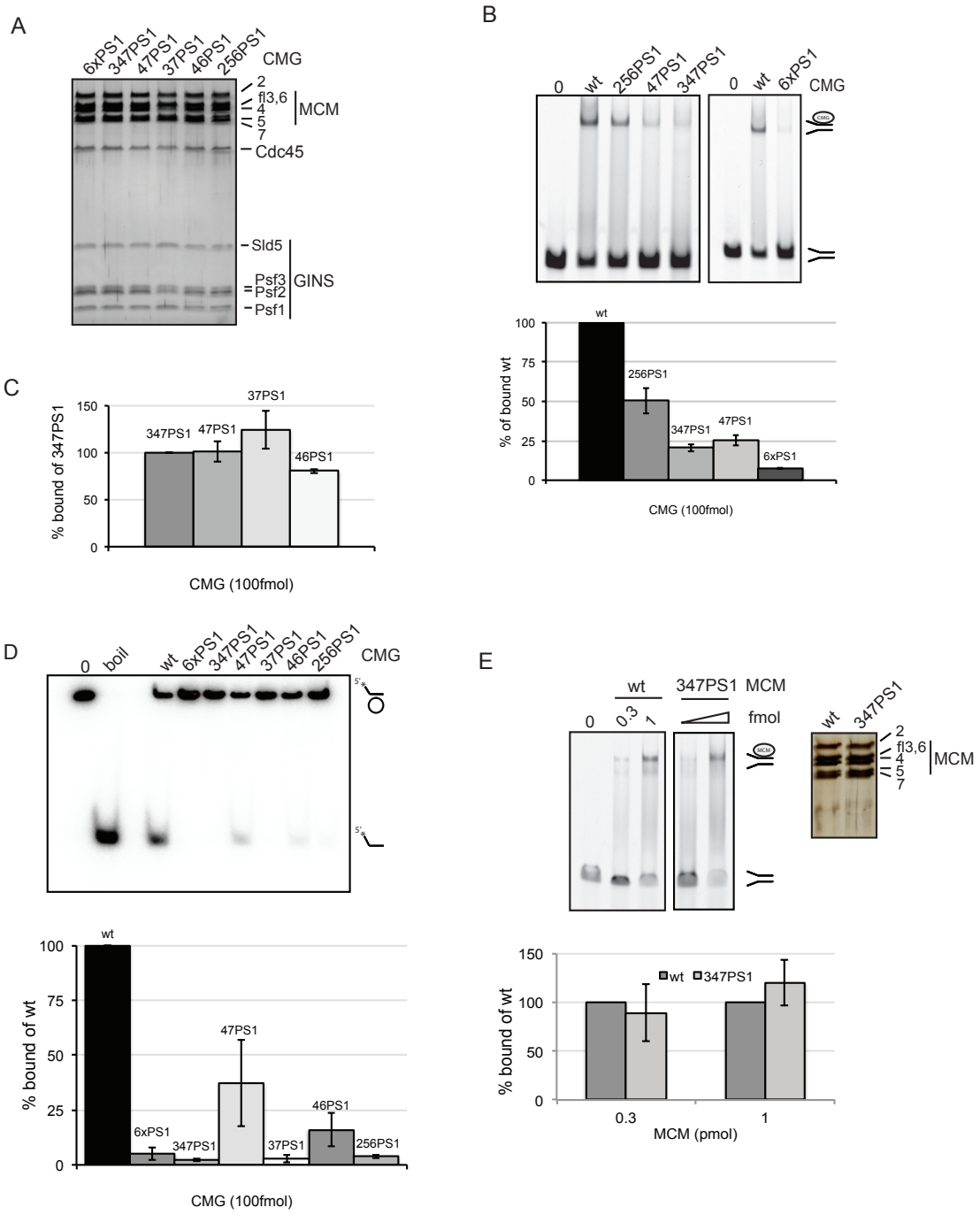

Figure 5.10 PS1-hp in Mcm3/4/7 are required for DNA binding of CMG.
(A) rCMG complexes containing an alanine substitution of the conserved lysine in the PS1-hp of several Mcm2-7 subunits simultaneously were purified and separated by SDS-PAGE (10%). The proteins were visualized by subsequent silver-staining. The number indicates the MCM subunit in which the PS1-hp was mutated, e.g. '347PS1' substituted the conserved lysine in the PS1-hp with alanines in Mcm3, Mcm4, and Mcm7 simultaneously.
(B) EMSA (*top panel*) and quantification (*lower panel*) of two independent series of gel-shifts of wt and mutant CMG complexes with mutations in the PS1-hp of several Mcm2-7 proteins ('256PS1', '47PS1', or '347PS1'). Binding reactions were performed with 100fmol of FITC- labeled '3'+5'polyT' forked DNA substrate and 100fmol of purified CMG protein in presence of 10μM ATPγS. Average values expressed as percentage of wildtype binding (set as 100%) were plotted with standard deviations of two independent series.
(C) Quantification of EMSA of mutant CMG complexes that contain simultaneous mutations in the PS1-hp of Mcm3, Mcm4 and Mcm7 ('347PS1'), Mcm4 and Mcm7 ('47PS1'), Mcm3 and Mcm7

('37PS1'), or Mcm4 and Mcm6 ('46PS1'). Results are compared to the DNA binding of '346PS1'CMG, and standard deviations from two independent series are shown.
(D) Autoradiograph of the helicase activity assay testing various combinations of CMG 'PS1-hp' mutations on a circular DNA substrate (*top panel*). The results were quantified and expressed as percentages of the wtCMG activity with standard deviations of two independent series (*lower panel*).
(E) DNA binding of wt and '347PS1' mutant Mcm2-7 complexes was tested in a gel-shift assay with the Y-shaped 'fork' DNA substrate (3'+5'polyT) at two different protein concentrations (0.3 and 1pmol) (*top, left panel*). The two Mcm2-7 complexes were separated by SDS-PAGE (10%) and silver-stained for visualization of protein bands (*top, right panel*). Quantification of the EMSA is shown below with averages and standard deviations of two independent series (*lower panel*).

and CMG complexes led to the observation, that the overall trend of defects was similar. This data raised the question of a possible role of the NT-hp motifs for the initial recruitment of the MCMs to the replication origins. To test this possibility for the PS1-hp motifs, Mcm2-7 containing these hairpin mutations in Mcm3, Mcm4, and Mcm7 ('347PS1') were tested for DNA binding in the gel-shift assay (Figure 5.10E). However, different from the findings with NT-hp mutations, this mutant Mcm2-7 complex did not affect the DNA binding even though the same combination of Mcm2-7 PS1 mutants entirely abrogated the DNA binding of the CMG (Figures 5.10E and B). Together, these data suggest that, opposite to NT-hp, the PS1-hp might not be used for DNA binding of isolated Mcm2-7 in the same way as they are used within the CMG complex. Additionally, impressive re-arrangements of the Mcm2-7 ring have been structurally observed when the auxiliary factors Cdc45 and GINS associate to activate the MCM motor (Costa et al., 2011). It is likely that several hairpin motifs within the central channel might undergo movements, also upon nucleotide binding that is required for CMG's binding to DNA. As a result, these hairpins might become properly aligned as required for subsequent binding after the duplex, to which Mcm2-7 are initially recruited, is melted and the Mcm2-7 helicase activated in form of the CMG complex (Evrin et al., 2009; Remus et al., 2009).

Lastly, I wished to analyze with the PS1-hp set of mutants if each of the six Mcm2-7 subunits majorly contributed to some of the direct DNA-protein contacts necessary during the translocation on DNA. Testing of the DNA binding activity in EMSAs suggested that mutations of PS1-hp motifs in Mcm3, Mcm4 and Mcm7 particularly resulted in an overall decrease of DNA binding by the CMG complex. To specifically test these residues and if mutations in the PS1-hp of some Mcm2-7 subunits abolished the direct interaction to DNA, cross-linking of CMG to the leading strand substrate was performed (detailed described in Chapter 4). Each of the PS1-hp mutations in single Mcm2-7 proteins within the CMG complex were tested for cross-linking to the leading strand in presence and in absence of ATPγS (Figures 5.11A and B). Besides a slight decrease in the most defective '4PS1' CMG, none of the individual Mcm2-7 PS1-hp mutations caused a diminished DNA-protein cross-linking (Figure 5.11B, lanes 4 and5). Also the alanine substitution in the hairpin of Mcm4 did not directly abolish the interaction between DNA and the Mcm4 subunit, but rather caused an overall mild decrease in the cross-linking efficiencies.

Two CMG complexes with PS1-hp mutations in several subunits were included in the cross-linking assay to test their effects and understand if the combination of defective hairpins in several Mcm2-7 subunits might lead to an entire abrogation of the direct DNA-protein interactions. The mutation in Mcm4 and Mcm7 PS1-hp ('47PS1') caused a substantial reduction in DNA binding, while Mcm2, Mcm5, and Mcm6 ('256PS1') together

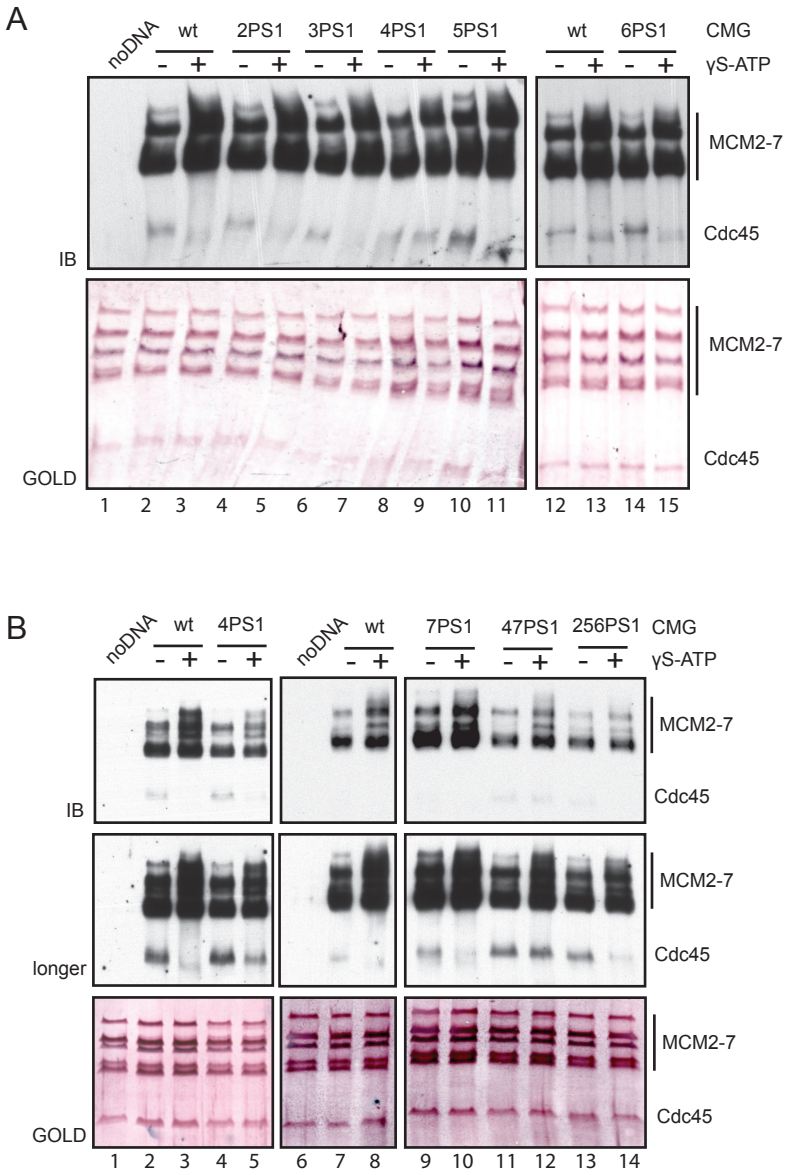

Figure 5.11. Mutations of PS1-hp in Mcm4 and Mcm7 affect the gate of the ring.
(A) Western blot analysis of forked FITC-labeled 'leading' DNA substrate cross-linked by UV to either wtCMG or CMG with point mutations in PS1-hp of individual Mcm2-7 subunits in absence and in presence of ATPγS. Reactions were separated by SDS-PAGE (10%) after digestion of DNA by micrococcal nuclease. Immunoblot was probed with monoclonal α- FITC antibody to detect cross-linked DNA-protein bands and subsequently stained with Aurodye™ (GE Healthcare) to visualize all proteins.
(B) Same experimental set-up as in (A), repeat of cross-linking of leading strand to CMG containing the PS1-hp in Mcm4, and of CMG complexes that contain point mutations in PS1-hp of either Mcm7 alone, in Mcm4 and Mcm7 simultaneously, or in Mcm2, Mcm5 and Mcm6 at the same time.

led to a medium defect (Figure 5.10B). Both observed reductions in DNA binding from gel-shift assays were confirmed in the cross-linking assay here. However, in both mutant CMG complexes none of the interactions between Mcm2-7 members and DNA was altered or abolished, leading only to an overall reduction in the cross-linking intensity (Figure 5.11B, lanes 11-14).

Strikingly, another significant defect of PS1-hp in certain Mcm2-7 subunits became obvious. Chapter 4 extensively demonstrated and discussed the potential role of Cdc45 as a 'guardian' of the gate located between Mcm2 and Mcm5 proteins within the ring. When mutations in PS1-hp of all six Mcm2-7 subunits were introduced at the same time, the CMG complex resulted in not only an abrogation of the DNA binding, but also in what is believed to indicate a defective 'gate' closure (see Chapter 4 for details, and Figure 5.4E). Generally, in absence of nucleotide, a direct interaction of Cdc45 to the leading strand is observed, which is prevented when ATP is bound and, thus, the 'gate' closed. However, mutations that cause a dysfunction of such active gate closure will likely induce a permanent opening. In this case the direct interaction of Cdc45 to the leading strand persists even upon nucleotide addition since the leading strand can directly interact with Cdc45 that 'guards' the gate. An analogous disruption of the gate function became evident when the PS1-hp of Mcm4 and Mcm7 were simultaneously mutated (Figure 5.11B, lanes 11-12).

To further dissect if both subunits independently, or in combination, contributed to this defect, the single point mutations in the hairpins of either Mcm4 or Mcm7 were analyzed for the Cdc45:DNA contacts. While the '7PS1' CMG complex displayed a wildtype behavior with regard to the cross-linking of Cdc45 to DNA, and suggested a proper function of the '2/5gate', mutations in Mcm4 PS1-hp led to a similar, but slightly weaker Cdc45-cross-linking abnormality as described for the '47PS1' CMG (Figure 5.11B, lanes 4-5, and 9-10). That this behavior is specific to the PS1-hp mutations introduced in this MCM pair was further confirmed by testing the triple PS1-hp mutation in Mcm2, Mcm5, and Mcm6 ('256PS1') for cross-linking. Mutations of PS1-hp of three consecutive MCM subunits within the Mcm2-7 ring did not affect the Cdc45:DNA interactions, suggesting a normal 'gate' closure and emphasizing a role of the Mcm4/7 PS1-hp in inter-subunit cross-talk that ensures that the '2/5 gate' properly closes (Figure 5.11B, lanes 13-14).

The second narrowest point in the Mcm2-7 central channel is formed by the Helix-2-insert β-hairpin

In addition to the narrowest point in the N-terminal domain, a second direct constriction of the central cavity in the Mcm2-7 is formed by the Helix-2-insert β-hairpin located in the

AAA+ domain. Different from the slightly recessed PS1-hp, this hairpin motif protrudes directly into the central channel (Brewster et al., 2008). This hairpin consists of an inserted sequence motif named 'Helix-2-insert' (H2I) within the AAA+ domain of MCM proteins and its deletion in MthMCM resulted in complete abolished helicase activity (Jenkinson and Chong, 2006).

The corresponding PS1-hairpin motif of the six monomers in papillomavirus E1 was demonstrated to account for the 'staircasing' of the ssDNA and sequentially 'pull' nucleotides through the central channel by a combination of adjacent basic/acidic residues; a feature that is absent from the PS1-hp in Mcm2-7 proteins (Enemark and Joshua-Tor, 2006). The only basic/acidic residues in MCM proteins was found in the H2I-hp which itself is absent in E1 and other SF-3 helicases leading to the suggestions of its possible role during translocation (Enemark and Joshua-Tor, 2008).

The previously described homology models for DmMcm2-7 proteins confirmed a predicted hairpin motif at the corresponding position based on the SsoMCM crystal structure and the location of H2I-hp in the central channel (Figure 5.12A and B, shown for DmMcm4) (Brewster et al., 2008).

Sequences of the H2I-hp motifs in *Drosophila* Mcm2-7, SsoMCM, and MthMCM were aligned for detailed analysis (Figure 5.12A). At the tip of the H2I-hairpin of several Mcm2-7 subunits, a basic and acidic residue was found reminiscent of such similar motif described for E1 PS1-hp (Enemark and Joshua-Tor, 2006). Mutational analysis of the H2I-hp in MthMCM reported that the deletion of this hairpin motif showed a 12-fold increase in ATPase activity upon dsDNA addition. This data suggested a key role of this MCM motif in coupling the energy derived from ATP-hydrolysis to separation of duplex DNA by MCM proteins (Jenkinson and Chong, 2006).

Mutation of the basic/acidic residue in Mcm4 reduces the unwinding activity of CMG

Due to the suggested link of ATP hydrolysis and the H2I-hp, I wished to target H2I-hp in Mcm2-7 subunits that belong to the two different groups of ATPase defects (described in Chapter 3). Mutations in the active site elements of Mcm3 and Mcm5 were shown to lead to abolished helicase activity of the CMG complex, while such corresponding mutations in MCM subunits on the other side of the Mcm2-7 ring (e.g. Mcm4, Mcm7) retained mostly wildtype activity (Chapter 3, Figure 3.7A and B). To test whether H2I-hp motifs in the *Drosophila* CMG were essential for unwinding, alanine substitutions of the basic/acidic residues in the H2I-hp Mcm4 (K551 and D552) and Mcm5 (K417 and D418) were introduced. Since this combination of positive and negative charges next to each other was

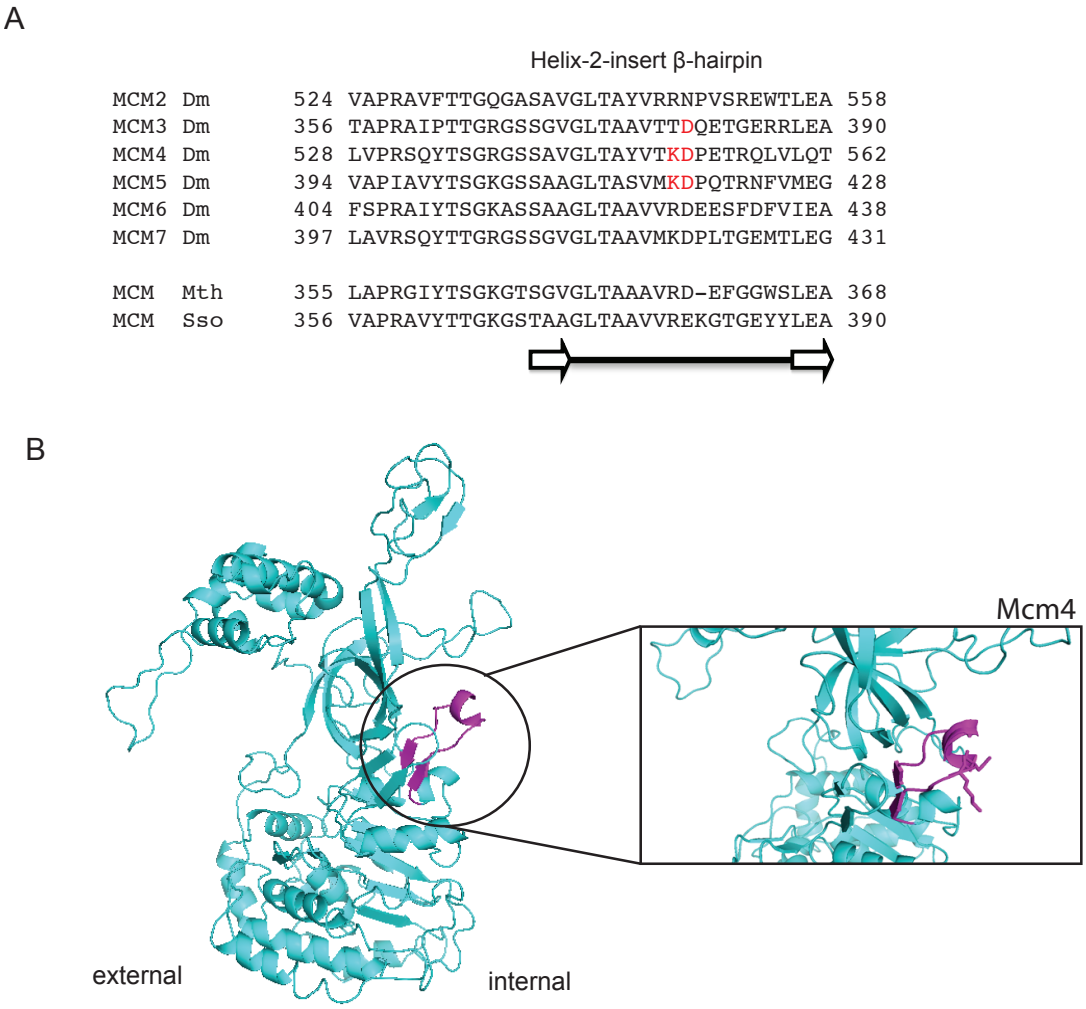

Figure 5.12. Alignment of Helix-2-insert β-hairpin of *Drosophila* Mcm2-7.
(A) Multiple sequence alignment of DmMcm2-7, MthMCM and SsoMCM using ClustalW2. Shown is the region surrounding the Helix-2-insert β-hairpin (H2I-hp). The basic and acidic residues mutated within this hairpin of Mcm3, Mcm4, and Mcm5 are highlighted (in *red*). Arrows and line indicate β-strands and loop of SsoMCM H2I-hp.
(B) *Side view* of a homology model of DmMcm4 (in *cyan*) based on the crystal structure of SsoMCM (PDB entry: 3F9V) is shown. The H2I-hp is highlighted (in *magenta*), and the basic and acidic residues (Lys551 and Asp552) are represented as a stick in the close-up view.

absent in Mcm3, a mutation was introduced in the present acidic residue of Mcm3 alone (D380) (Figure 5.12A).

I purified CMG complexes that contain the alanine substitutions in H2I-hp (named 'H2I') in one of the three Mcm2-7 subunits (Mcm3, Mcm4 or Mcm5) individually (Figure 5.13A). The DNA binding was tested on a forked substrate that contains a duplex with unpaired 3'- and 5'-extensions. Mutations in the H2I-hp of Mcm3, Mcm4, and Mcm5 resulted in reduced average DNA binding to about ~70% of wildtype activity (Figure 5.13B and C). The strand displacement was subsequently measured on the circular substrate with a 5'-extended tail. While CMG complexes with mutant H2I-hp in Mcm3 and Mcm5 retained mostly wildtype unwinding activity, with an average displacement of 96% of wt for the CMG 5H2I-hp and 83% of wt for the CMG 3H2I-hp complexes (Figure 5.13B and D). Differently, the CMG complex containing the alanine substitution in the H2I-hp of Mcm4 led to a more significant reduction with 42% of wt helicase activity (Figure 5.13D). Interestingly, the Mcm4 subunit of the DmCMG helicase, which showed the weakest defects in ATPase and helicase activity when the ATPase sites were mutated, now displayed the strongest defective helicase upon mutation of the basic/acidic residues in H2I-hp.

Since the DNA binding was not significantly reduced in any of these three H2I-hp mutant CMG complexes, it seemed reasonable to expect wildtype behavior in the cross-linking to the leading strand, unless the H2i-hp unexpectedly establishes the only direct contact between protein and DNA in any of the six Mcm2-7 subunits. To test this, '3H2I', '4H2I', or '5H2I' CMG proteins were cross-linked to the leading strand (Figure 5.13E). The DNA binding was captured in absence and in presence of nucleotide, and showed wildtype cross-linking confirming that none of the mutant CMG complexes led to abrogation of the Mcm2-7-DNA interaction (Figure 5.13E). This data, together with the cross-linking results from the PS1-hp mutant CMG complexes suggest that neither of the altered residues in these two hairpin motifs is solely responsible for the interaction of the individual Mcm2-7 subunits to the leading strand, and that likely multiple contacts are established.

The Drosophila Mcm2-7 ring possesses a conserved hairpin motif on its external surface

Besides the highly positively charged β–hairpin in the N-terminal domain, and the two discussed hairpin motifs within the central channel of the AAA+ domain, a fourth conserved loop, forming a β–hairpin structure, becomes evident when the homology models of *Drosophila* Mcm2-7 are examined (Figures 5.14A-D). Unlike the other internally positioned loops, this motif is located on the external surface of the hetero-hexamer, close to each subunit's side channel (Figure 5.14B).

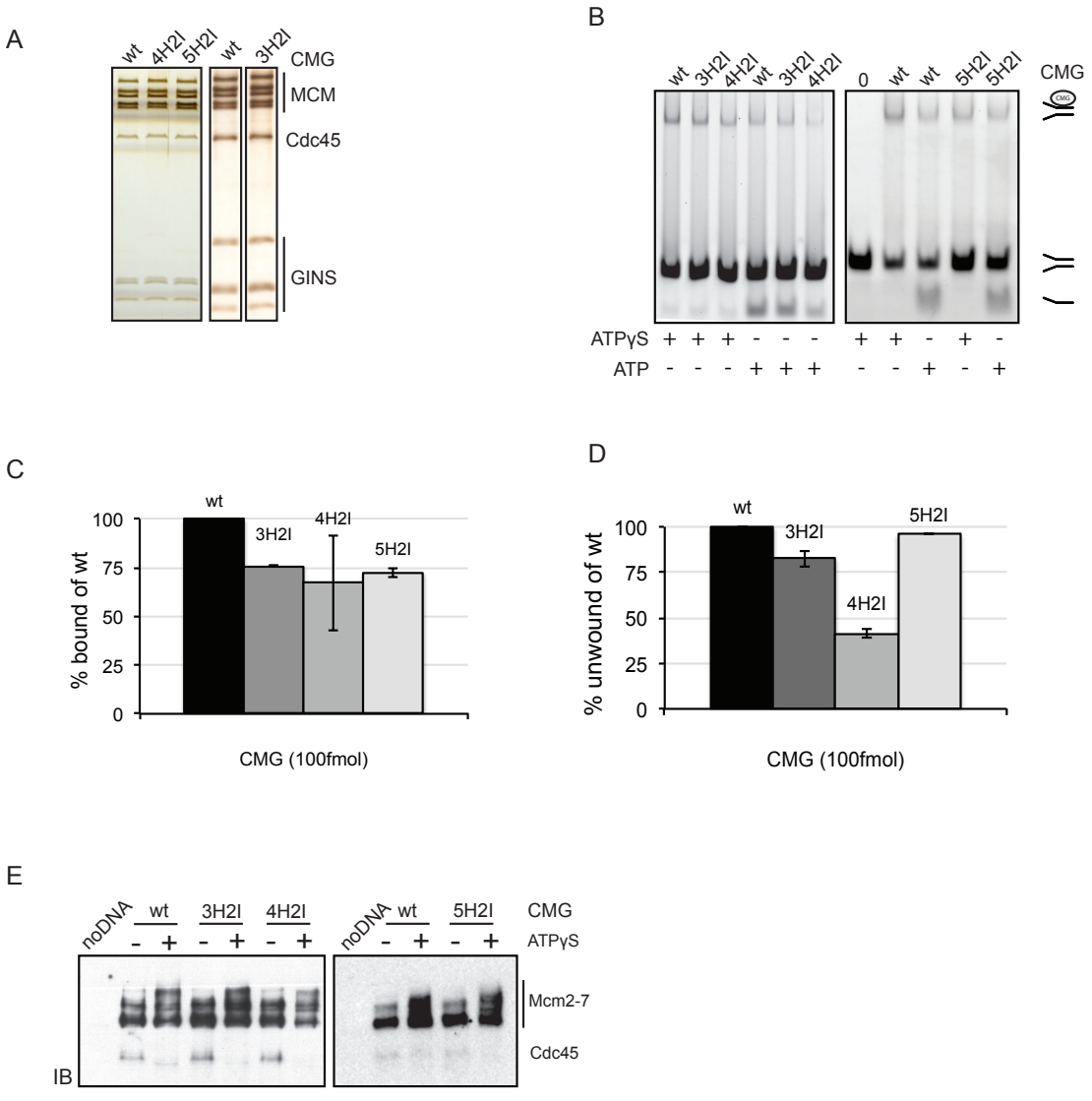

Figure 5.13. Mutation of H2I-hp in Mcm4 diminishes the helicase activity.
(A) Purified rCMG complexes with mutations in H2I-hp of DmMcm3, Mcm4 or Mcm5 individually were separated by SDS-PAGE (10%) and silver-stained for visualization. The Sld5 protein band stains weakly by silver. The '3H2I' gel was previously SYPRO-stained which enhances the subsequent Sld5 protein signal on the silver-staining.
(B) Gel-shift and unwinding were measured in parallel by incubating various CMG complexes with 'fork' DNA substrate (3'+5'polyT) for 30 minutes in presence of ATPγS, or in case of the helicase activity test, for 30 minutes in presence of ATP. Reaction products were separated on a native TBE-gel (4%), and signals were detected by the fluorescence of the FITC-modified nucleotides in the DNA substrate.
(C) Quantification of the DNA binding of CMG complexes with average values expressed as percentage of wt DNA binding shown in (B). Standard deviations from two independent series were plotted.
(D) Quantification of the measured helicase activity in (B). Average values and standard deviations from two independent series are plotted as percentages of the wtCMG activity.
(E) Western blots of CMG complexes with mutations in the H2I-hp of specific Mcm2-7 subunits were cross-linked to the 'leading' strand substrate by UV. Reactions were separated on a polyacrylamide gel (10%) and subsequently transferred onto a membrane to detect DNA-protein bands by a monoclonal α- FITC antibody.

Systematic mutational analysis of the Chen group has dissected the various basic, hydrophobic, and acidic residues throughout the β–hairpin and reported that some of the altered amino acids led to a reduced helicase affinity of the SsoMCM (Brewster et al., 2010). Especially the alanine substitution of a lysine residue at the base of the hairpin on the β–strand (K323), located on the outside, caused a ~60% reduction in unwinding activity on a Y-shaped forked DNA substrate (Brewster et al., 2010). Following studies of SsoMCM have suggested a dynamic interaction between the separated excluded 5'-strand and the MCMs, and shown that this interaction confers a higher stability of the MCM complex on DNA (Rothenberg et al., 2007). This latter group also confirmed that this K323 residue is essential for the unwinding activity of the SsoMCM, showing that no helicase activity is detectable when this residue is mutated (Graham et al., 2011).

I have already discussed in Chapter 4 that simultaneous alanine substitution of a conserved lysine (K) in the EXT-hp of all six Mcm2-7 subunits of *Drosophila* CMG (named 'EXT') resulted in highly reduced helicase activity while the DNA binding remained mostly unaffected (Figure 5.14A and Chapter 4 with Figures 5C and D). This residue corresponds to the K323 in SsoMCM and is similarly located at the base of the EXT β–hairpin as can be visualized by threading the homology models onto the crystal structure of SsoMCM (Figures 5.14B and D). Further, cross-linking of CMG proteins to the lagging strand of a forked DNA substrate identified a subset of Mcm2-7 subunits that specifically interact with the 5'-tail upon binding of ATPγS (Chapter 4 Figure 3A and B). Such interactions and the necessity for the presence of a 5'-tail have earlier been reported by single-molecule fluorescence resonance energy transfer (spFRET) studies that hypothesized a binding of the lagging strand via the MCM external side (Rothenberg et al., 2007).
Since CMG does not bind to dsDNA, the classic 'steric exclusion' model for unwinding is favored for the *Drosophila* helicase complex, in which the duplex is separated prior to the entry into the central channel formed by the Mcm2-7 hexameric ring. While the leading strand is engaged by the internal Mcm2-7 central cavity, the lagging strand is believed to wrap around the external MCM surface (Fu et al., 2011; Graham et al., 2011). Mutations of the invariant lysine at the base of the external β–hairpin strongly reduced the contacts between the *Drosophila* Mcm2-7 hexamer and specifically lagging strand DNA, suggesting direct interactions of this excluded DNA strand with likely the external surface of the Mcm2-7, which was also discussed in detail in Chapter 4 (Figure 5.16B, lanes 4 and 5). This finding led to the suggestion of an analogy to the proposed 'steric exclusion and wrapping' model suggested for SsoMCM (Graham et al., 2011; Rothenberg et al., 2007).

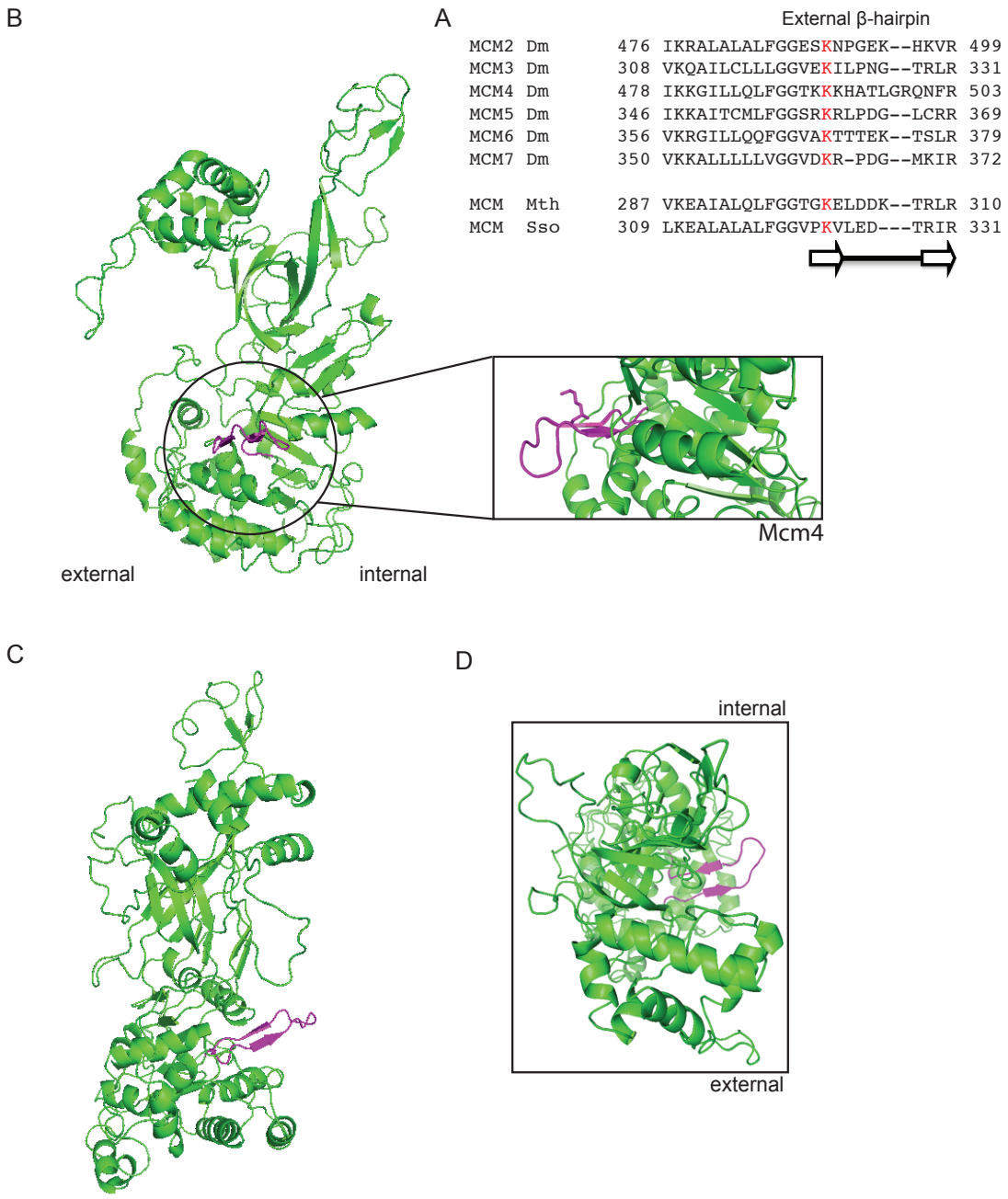

Figure 5.14. Alignment of External β-hairpin of *Drosophila* Mcm2-7.
(A) ClustalW2 was used for a multiple sequence alignment of DmMcm2-7, SsoMCM and MthMCM. Shown is the region surrounding the External β-hairpin (EXT-hp). Highlighted is the invariant lysine residue at the base of the hairpin loop (K) in each Mcm2-7 subunit that was substituted by an alanine (in *red*).
(B) *Side view* of the DmMcm4 homology model (in *green*) is shown. The modeling was based on the crystal structure of SsoMCM (PDB entry: 3F9V). The close-up view highlights the position of the EXT-hp (in magenta), and the critical lysine is represented as a stick.
(C) *Exterior view* (facing Mcm2-7 ring) of the DmMcm4 model highlighting the EXT-hp (in *magenta*).
(D) *Top view* from N-terminal domain of the DmMcm4 model displaying the EXT-hp (in *magenta*).

The cross-linking (see for details Chapter 4) of CMG to the lagging strand helped identify that upon ATP binding major interactions are established between the 5'-tail and Mcm2, Mcm4, and Mcm6, whereas Mcm5 mainly interacts with this DNA strand in absence of nucleotide (Figure 5.16A). DNA binding was measured for several mutant CMG complexes in a protein concentration dependent manner, and estimation of Kd values from the half-saturation point revealed that the '6xEXT' only very weakly reduced the DNA binding affinity by about 1.5 fold (wt Kd=~10nM, '6xEXT' Kd=~15nM) (Figure 5.15C). However, this sixfold EXT-hp mutant CMG drastically decreased the unwinding activity in the strand displacement assay by about half of the wildtype activity (Figure 5.15D, and Chapter 4, Figure 5D). Since only one-half of the Mcm2-7 establishes direct interaction through its outside in presence of the non-hydrolyzable nucleotide, I wished to test if the EXT-hp in Mcm2, Mcm4, and Mcm6 alone were responsible for these contacts. To test this hypothesis, CMG complexes with mutations in this hairpin motif in either single MCM subunits ('2EXT', '4EXT', '5EXT', and '6EXT') and in Mcm4 and Mcm6 ('46EXT'), or Mcm2, Mcm4, and Mcm6 ('246EXT') together were purified (Figure 5.15A).

DNA binding was tested in gel-shift assays with 3'-+5'-polyT-tail forked substrate. In agreement with the sixfold EXT-hp mutation (6xEXT) that caused minimal decrease of the DNA binding by CMG, none of the single or combined mutations displayed a defective binding activity (Figure 5.15B). Since the 'Mcm2/4/6' protein subset of Mcm2-7 subunits predominantly interact with the lagging strand in the cross-linking assay, the binding affinity for '6xEXT' and '246EXT' CMG complexes were estimated based on half-saturation points and yielded the wildtype range of Kd=10-15 nM (Figure 5.15C). To understand if the EXT-hp of specific Mcm2-7 contribute to the helicase activity, the effects on the unwinding were dissected by testing the single EXT-hp mutant CMG as well as the simultaneous mutations in several subunits (Figure 5.15D). The CMG helicase activity on the circular substrate remained unaffected for mutations in the EXT-hp of Mcm2 (97% of wt), or Mcm5 (94% of wt), and was only very mildly reduced for the Mcm4 EXT-hp mutation to 89% of wt CMG activity. Elimination of the conserved lysine at the base of EXT-hp in Mcm6, Mcm4/6, or Mcm2/4/6 caused moderately diminished activities by about 17-25% of wtCMG unwinding. This data suggest, that, even though only a subset of the Mcm2-7 subunits within the CMG directly contact the leading strand when nucleotide is bound, subsequent hydrolysis and link to translocation might establish further interactions with the external side of the CMG and the 5'-lagging strand.

To specifically test if this basic residue at the base of EXT-hp was alone responsible for the direct interactions to the lagging strand, several of the EXT-hp mutant CMG complexes

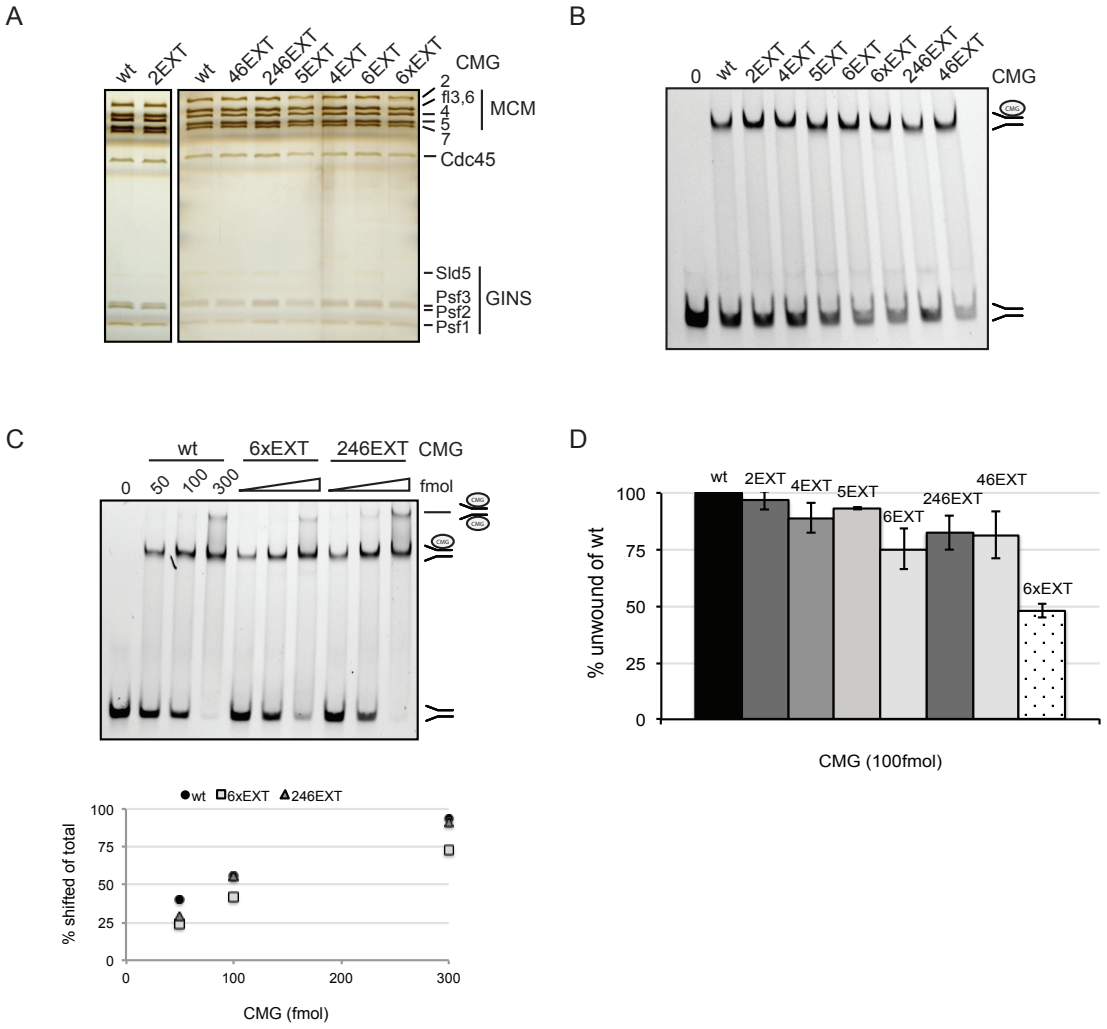

Figure 5.15. Mutations in EXT-hp of single MCM subunits display no defect on DNA binding or unwinding.

(A) rCMG complexes containing alanine substitutions of the conserved lysine in the EXT-hp of either single MCM or several Mcm2-7 subunits simultaneously have been purified and separated by SDS-PAGE (10%). Protein bands were visualized by subsequent silver-stain of the gel.

(B) EMSA of all purified CMG complexes with the mutations in the EXT-hp. Lanes contain 100fmol of FITC- labeled '3'+5'polyT' forked DNA substrate, and where indicated 100fmol of purified CMG protein in presence of 10μM ATPγS. The control lane ('0') contains only DNA. The reaction products were separated by a native TBE-PAGE (4%), and signals were detected by FITC fluorescence. Free DNA substrate and CMG-DNA positions are indicated on the side of the gel.

(C) Gel-shift assaying the DNA binding affinity of CMG in a protein concentration dependent manner (50, 100, 300fmol). These CMG complexes contain mutations in the EXT-hp of either all six Mcm2-7 subunits ('6xEXT') or in the subunits of one-half of the Mcm2-7 ring (Mcm2, Mcm4, Mcm6, '246EXT'). 'O' indicates the control lane that contains only the DNA substrate (3'+5'polyT fork). Quantification is shown in the lower panel, showing the percentage of shifted bands compared to the total DNA substrate signal.

(D) Quantification of the unwinding activity (at 100fmol) of various 'EXT-hp' CMG protein complexes. The average values were expressed as percentages of the wtCMG helicase activity and standard deviations of two independent series are plotted.

were tested in the cross-linking assay. As discussed in detail in Chapter 4, in this assay the CMG was recruited and bound to only the 3'-leading strand, while interactions were captured via FITC (fluorescein isothiocyanate) groups at specific positions on the GC-rich lagging strand (Figure 5.16A, see Chapter 4 with Figure 2).

As expected from gel-shifts that demonstrate a wildtype DNA binding of CMG complexes with mutations in the EXT-hp of individual, a subset, or all Mcm2-7 subunits, the DNA binding to the leading strand remained unaffected for the '6xEXT' CMG (Figure 5.16B, *top*). Since the Mcm4 protein represents the major subunit that binds to the lagging strand upon nucleotide addition, '4EXT' CMG was tested for its direct interaction with the 5'-strand (Figures 5.16B). The Mcm2-7 protein band pattern obtained after cross-linking of this CMG complex to the leading strand remained unaltered compared to wtCMG, and further it also did not abrogate the interaction of Mcm4 with the excluded lagging strand (Figure 5.16B, *lower panel*, lanes 6 and 7, and C). This suggests that either Mcm4 possesses additional residues on its outside that enable binding to the lagging strand, or that elimination of the EXT-hp residue in one Mcm2-7 subunit alone is to weak to fully ablate the path of this strand around the external CMG surface. The 5'-tail, thus, might still remain mostly within its regular 'wrapping' path and in close proximity to the proteins and allow UV-induced DNA-protein cross-links. In agreement with this is the observation of a highly dynamic nature of the interactions between the Mcm2-7 outside and the 5'-tail, in which weakening of one single Mcm2-7-lagging strand contact should not cause the lagging strand to fall off, but the DNA might instead grasp other proximal binding residues (Rothenberg et al., 2007).

Since Mcm4 represents the strongest binding subunit for the lagging strand, this observation might additionally suggest more than one residue that contacts DNA. And indeed, the sequence of the EXT-hp in Mcm4 displays a stretch of four consecutive positive amino acids, flanking the invariant lysine that was targeted in this study (Figure 5.14A).

Therefore, I wished to further test if the mutated basic residue in EXT-hp might fully abrogate another Mcm2-7 subunit's weaker interaction to the lagging strand. For this purpose, '6EXT' CMG was cross-linked to the 5'-strand and analyzed. Conversely, even this weaker interaction was sustained and none of the Mcm2-7-DNA interactions altered (Figure 5.16D, lanes 4 and 5). This data is in agreement with the hypotheses that the lagging strand either might interact with additional positively charged residues around the Mcm2-7 ring within the CMG or that its path around the Mcm2-7 external surface is established through collaborative contacts to all Mcm2-7 proteins, so that elimination of one residue could be absorbed by the residual intact ones. However, given that mutations in EXT-hp of three consecutive Mcm2-7 subunits around the ring (Mcm2, Mcm4, and

Mcm6) did not greatly reduce the unwinding activity of the CMG, and that this activity was even with the '6xEXT' CMG only reduced to about 50% of wtCMG, the former suggestion is favored. Future mutagenesis of other candidate DNA-binding residues might clarify the lagging strand interactions of *Drosophila* CMG in more detail.

And indeed, mutations of a second positive residue (R440) in SsoMCM inactivated the complex's helicase activity, as well as the combined mutations of the acidic tip (ED326/327) and a basic residue (R329) that has been shown to drastically decrease strand separation of this MCM helicase suggest further DNA binding motifs on the exterior surface of the MCM proteins (Brewster et al., 2008; Graham et al., 2011).

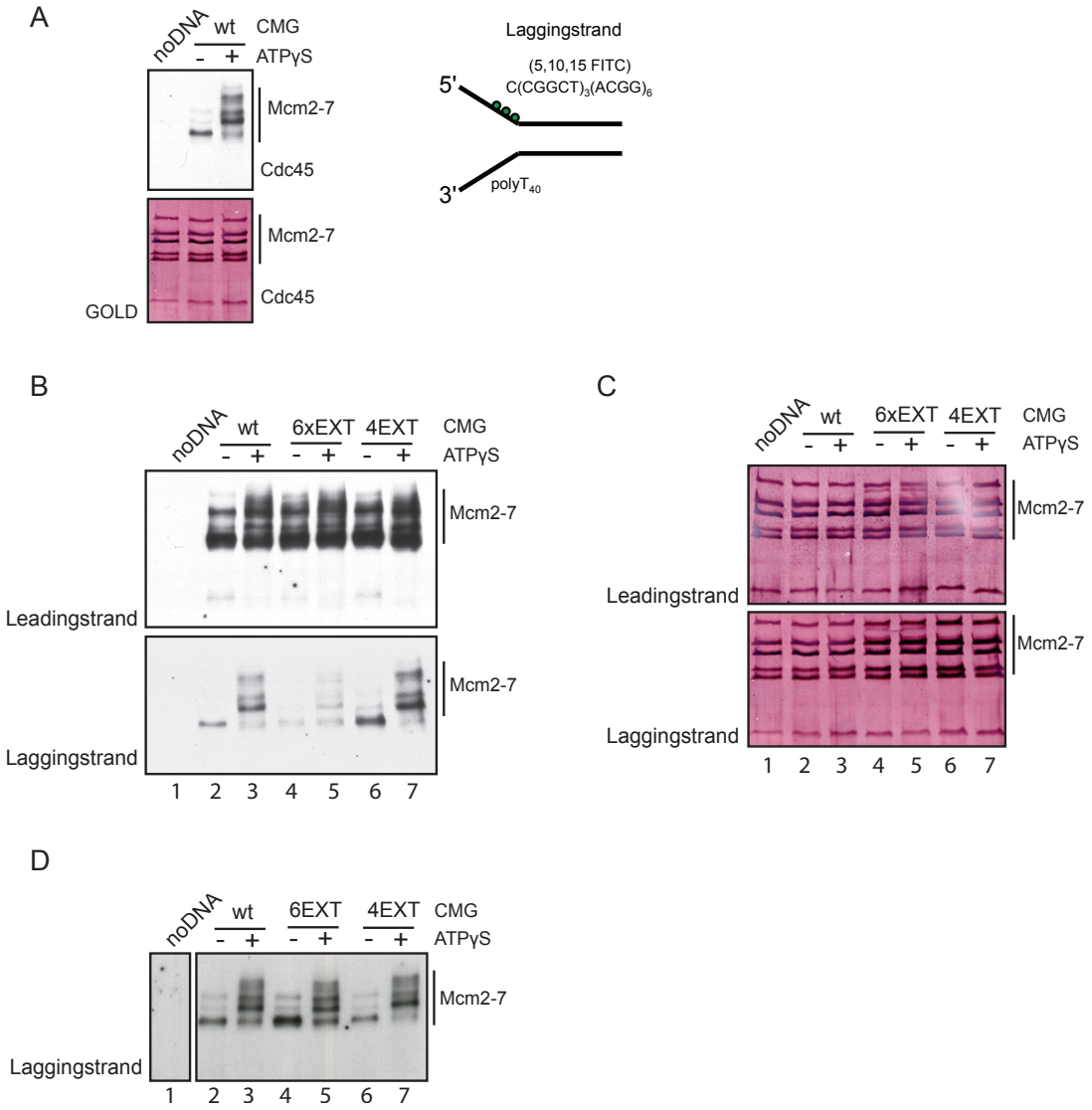

Figure 5.16. Mcm4 EXT-hp is not the only residue responsible for Mcm4-DNA contact to the lagging strand.

(A) Western blot analysis of forked FITC-labeled 'lagging' DNA substrate cross-linked to wtCMG by UV in absence and in presence of ATPγS. Reactions were separated by SDS-PAGE (10%) after digestion of DNA by micrococcal nuclease. Immunoblot was probed with monoclonal α- FITC antibody to detect cross-linked DNA-protein bands. Western blot was subsequently stained with AuroDye™ (GE Healthcare) to control an even protein loading (*lower panel*). A schematic of the 'lagging' strand substrate displays the positions of the cross-linking FITC groups on the 5' strand (lagging strand) and the sequence composition of the 3' and 5' extensions (*right panel*).

(B) Cross-linking of wtCMG and CMG complexes with point mutations in the EXT-hp of either all six Mcm2-7 proteins ('6xEXT'), or in Mcm4 alone ('4EXT'). CMG was cross-linked to either the leading strand (*top panel*) or the lagging strand (*lower panel*) substrate.

(C) Staining of the western blots from (B) with AuroDye™ (GE Healthcare).

(D) Cross-linking of CMG complexes with single alanine substitution of the conserved lysine in the EXT-hp in Mcm4 or Mcm6 to the lagging strand.

5.4 Discussion

The study presented here demonstrated, for the first time, the importance of the four β-hairpin motifs in the internal cavity and external surface of a eukaryotic Mcm2-7 ring within the CMG complex for DNA binding and helicase function. Several functional parallels to the homologous archaeal MCM helicase complexes could be drawn from this individual biochemical dissection of the candidate DNA-binding motifs. Structural studies of the MthN-MCM dodecamer showed that positive charges are highly present in a N-terminal fragment of the central channel (Fletcher et al., 2003). The narrowest point of the N-terminal domain in MthMCM is formed by these N-terminal β-hairpins (NT-hp) that protrude from each of the six MCM proteins directly into the central channel, large enough to accommodate ss or dsDNA (Fletcher et al., 2003; McGeoch et al., 2005). Mutations of two basic residues at the tip of this hairpin motif resulted in abrogation of DNA binding and helicase activity for N-MthMCM and full-length SsoMCM (Fletcher et al., 2003; McGeoch et al., 2005). Corresponding positively charged residues are present in five of the six *Drosophila* NT-hp of the Mcm2-7 polypeptides (besides in Mcm2) and were substituted with alanine. Since CMG does not exhibit DNA binding activity to dsDNA, all activities were subsequently tested on substrates that provide 3'-ssDNA extensions and ensure the enzyme's binding (Ilves et al., 2010). Analogous to the archaeal MCM, the DmCMG complex displayed a substantial reduction in DNA binding and helicase activity upon simultaneous mutations of all five NT-hp in Mcm2-7 ('5xNthp') in presence of wtMcm2. The presence of the motif observed in archaeal MCM with two adjacent basic residues in the NT-hp ('KR', or 'RxK') were only found in Mcm3, Mcm5, and Mcm6. Simultaneous mutations in the hairpin motifs of this subset of the Mcm2-7 ring within the CMG complex ('356Nthp') displayed a significant reduction in DNA binding and subsequent unwinding, while alteration of the basic residues in the remaining two NT-hp in Mcm4 and Mcm7 together ('47Nthp') retained wildtype activities. Reminiscent of the functional asymmetry reported with regard to the ATPase active sites of the DmCMG, mutations of the NT-hp demonstrated here that there is an unequal contribution of the DNA binding motifs positioned in the N-terminal domain of the DmMcm2-7 polypeptides. Even though Mcm3, Mcm5, and Mcm6 are not in the consecutive order within the Mcm2-7 ring, the homology modeling of DmMcm6 revealed that this subunit possesses an unusually long NT-hp and could explain how it might be able to reach the DNA if it was held proximal to the adjacent Mcm3/5 pair. Proportional to the reduced DNA binding, the helicase activities of the five-fold ('5xNthp') or triple mutant ('356Nthp') CMG complexes were decreased to a similar level and interestingly did not entirely abrogate the unwinding, suggesting a role of this hairpin motif in primarily DNA binding. This point is supported by the finding that cross-

linking of these NT-hp mutant CMG complexes to the leading strand led to a slight decrease of the Mcm5-DNA interaction, but more significantly, abolished the direct contact of Mcm3/Mcm6 to the leading strand. This data suggests, that Mcm5 might contain additional DNA binding residues that bind to the leading strand, while the NT-hp in Mcm3/Mcm6 are likely crucial for their direct interaction with the leading strand. Initially, two hexameric Mcm2-7 rings are loaded onto dsDNA in a head-to-head manner during G1 phase of the yeast cell cycle, and analysis of effects when NT-hp of several subunits in isolated *Drosophila* Mcm2-7 complexes demonstrated that the DNA binding activity was also quantitatively reduced (Evrin et al., 2009; Remus et al., 2009). The general pattern of the defects in these Mcm2-7 complexes corresponded to the one observed within the CMG; however, the '47Nthp' Mcm2-7 complex showed a more pronounced reduction in DNA binding compared to the CMG complex. At the transition of G1 to S phase of the cell cycle, the Mcm2-7 associate into the higher-molecular weight CMG complex and simultaneously experience a tremendous activation of their intrinsic activities of orders of magnitude (Ilves et al., 2010). While crystallographic data on either eukaryotic Mcm2-7 complexes or the activated CMG helicase lack behind, we hypothesized previously that these vast stimulations result from allosteric remodeling through the accessory factors GINS and Cdc45, which might cause an optimal alignment of the β-hairpin motifs within the central channel in a way essential for the subsequent productive strand displacement (Ilves et al., 2010). The data presented here show that, while the NT-hp are used in both Mcm2-7 and CMG complexes for DNA binding, a very strong asymmetry becomes evident for the binding by the CMG, suggesting that particularly the hairpins of the Mcm3-Mcm5-Mcm6 proteins of the Mcm2-7 ring are required in the DNA interactions of the CMG. Further, archaeal MCM show DNA-stimulated ATPase activities, and mutations in the NT-hp of MCM caused no increase in ATPase activity upon addition of DNA, thus, suggesting a critical role in such DNA-stimulated nucleotide hydrolysis for this hairpin motif (McGeoch et al., 2005). In contrast to this, *Drosophila* CMG displays only a very weak DNA-stimulated ATPase activity to about 1.1-1.2 fold (Aylin Goke, *unpublished data*), in par with the hypothesis of a more optimal structural arrangement of the DNA-binding motifs within the Mcm2-7 central channel in the CMG complex. However, analysis of the homology models of each of the six DmMcm2-7 subunits reveals an analogy to a second structural feature described for the N-MthMCM, in which one of the two β-strands of the NT- β-hairpin directly connects to the AAA+ domain of the protein (Fletcher et al., 2003). Fletcher et al., (2003) discuss the possibility that such link between the two domains could enable that the energy released by ATP-hydrolysis translates directly into movements of the hairpin motifs, and may, thus, provide the 'cross-talk' between subunits for a remodeling

mechanism of the initially bound duplex DNA or the translocation on ssDNA. Conversely, given that the unwinding activities in both archaeal MCM and eukaryotic CMG complexes were not inactivated upon mutations of the NT-hp in all MCM subunits, implicates that these hairpins are less likely to directly contribute to the translocation mechanism, but rather to the initial DNA binding. Aside, it is likely that other β-hairpin motifs within the AAA+ domain might contribute to the active translocation mechanism; an observation reported for archaeal MCMs.

Such a second candidate β-hairpin, the Pre-Sensor1 β-hairpin (PS1-hp), was mutated in SsoMCM by substituting a conserved basic residue at the tip of the hairpin motif. The study reported that this mutation not only reduced the DNA binding affinity by about three-fold, but it also entirely abrogated the helicase activity (McGeoch et al., 2005). Given that the unwinding was completely inactivated even though the MCMs still bound to DNA, it was reasonable to postulate a pivotal role for this β-hairpin in coupling the energy obtained from nucleotide hydrolysis with translocation of the enzyme on DNA (McGeoch et al., 2005). Mutation of the corresponding lysine (or arginine in Mcm6) at the tip of the hairpin loop in *Drosophila* Mcm2-7 within the CMG allowed not only to test if the effects of such alteration in all six Mcm2-7 subunits simultaneously ('6xPS1') caused similar defects as observed for archaeal MCM, but further enabled a detailed dissection of each subunit's contribution. Different from archaeal MCM, the substitution of the critical basic amino acid in all six Mcm2-7 subunits almost entirely abrogated the DNA binding of CMG and completely inactivated the helicase activity, suggesting that this motif is required also for DNA binding by the CMG complex. Mutations of PS1-hp, introduced in individual Mcm2-7 subunits, provided important insights into the involvement of each subunit's PS1-hp. The EMSAs showed that not all six PS1-hp contributed equally to the DNA binding of CMG; a feature already observed for the NT-hp. The highest reduction was found with PS1-hp mutations in Mcm3 and Mcm4, the same subunits that displayed the strongest decrease in DNA unwinding activity. Strikingly, the mutant PS1-hp in Mcm7 of CMG most significantly decreased the helicase activity, albeit the DNA binding was only very mildly reduced, suggesting a pivotal role of this PS1-hp in the translocation mechanism. The data together implicate that one-half of the Mcm2-7 ring with Mcm3, Mcm4, and Mcm7 is more important for the DNA binding of CMG. If, however, a sequential mode of hairpin movements upon ATP-hydrolysis, analogous to the corresponding PS1-hp in papillomavirus E1, is true for the CMG, my reasoning is that an elimination of PS1-hp in several consecutive MCMs around the ring should ablate the helicase activity of CMG. Consistent with this hypothesis, the simultaneous mutation of PS1-hp in Mcm3, Mcm4, and Mcm7 ('347PS1') diminished the DNA binding to about one quarter of the wildtype

activity, but completely inactivated the unwinding activity of the helicase complex. On the other hand, while simultaneous mutations of PS1-hp in Mcm2, Mcm5, and Mcm6 ('256PS1') also caused a complete abrogation of the helicase activity, the complex still retained about half of its DNA binding activity. Different from the mechanism proposed for E1, the positions of the six corresponding PS1-hairpins of SV40 large T antigen change at the same time based on the nucleotide state in the ATP binding sites (ATP, ADP, or free), and move up to 17Å during the cycles of ATP-binding and –hydrolysis (Gai et al., 2004). If all the six PS1-hp in DmCMG were simultaneously equally involved in nucleotide-hydrolysis driven movements and interaction with the leading strand during translocation, it would more challenging to understand why the defects of one half of the DmMcm2-7 ring would be more significant than those on the other half, thus, suggesting a sequential order of events for the DmCMG helicase.

Further mutational dissection revealed that, while the DNA binding of '47PS1' CMG was even more defective than of '37PS1' CMG, the helicase activity was completely inactivated in the latter. In contrast, '47PS1' CMG retained helicase activity of about one quarter of wt CMG, proportional to its reduced DNA binding. These findings further narrowed down the crucial role of PS1-hp in DNA binding and unwinding to the precise positions of Mcm3 and Mcm7 within the Mcm2-7 ring of the CMG complex.

Side-by-side comparison of mutations in NT-hp of critical Mcm2-7 subunits in isolated Mcm2-7 and CMG complexes demonstrated that the defects on DNA binding overall followed the same pattern in both complexes and led to the suggestion that these N-terminal hairpin motifs are required for both free Mcm2-7 and those incorporated in the CMG complex. However, the differences between the NT-hp in the various Mcm2-7 subunits were more drastic in the CMG, suggesting a stronger asymmetry in this complex for DNA binding. A comparison of PS1-hp in critical Mcm2-7 subunits of the CMG such as the '347PS1' opposed the observation for the NT-hp. Here, while the '347PS1' CMG complex drastically reduced the DNA binding to only about 25% of wildtype activity, the same combination of PS1-hp mutations in the isolated Mcm2-7 complex completely retained wt DNA binding. This suggests that the PS1-hp in free Mcm2-7 complexes are not predominantly used for DNA binding, and that other motifs, such as the NT-hp might be responsible for DNA interactions of the loaded Mcm2-7 complexes before the association of Cdc45/GINS that induces the subsequent activation within the CMG complex. It was previously discussed, that the CMG undergoes structural rearrangements and that this might lead to a more optimized alignment and correct DNA interactions of the required hairpin motifs and thus account for the vast activation. One finding from the mutational data that corroborates this hypothesis derives from the observation that PS1-hp, that are

absolutely crucial for DNA binding of the CMG complex, show no defects in the isolated Mcm2-7 complex. To understand if the basic residue in PS1-hp solely accounts for the direct contacts by the Mcm2-7 protein to the leading strand, CMG complexes that contain single mutations in the PS1-hp of Mcm2-7, were cross-linked to the leading strand. Based on the structural study of the CMG complex that reported a tightening of the central channel upon ATP addition, multiple contacts between various Mcm2-7 proteins in the CMG complex and the DNA in the central channel appear to be most likely (Costa et al., 2011). Consistent with this, the PS1-hp mutations did not abrogate any of the observed direct contacts between the Mcm2-7 subunits and the leading strand. Even simultaneous mutation of PS1-hp in several of the Mcm2-7 proteins only confirmed an overall weakened DNA binding – consistent with the gel-shifts - but did not abolish the interaction of any specific subunit to the leading strand.

Interestingly, mutations in PS1-hp of Mcm4 alone, as well as in Mcm4/7 displayed defects in the closure of the Mcm2/5 gate, a finding discussed and hypothesized in detail in Chapter 4. While the interactions between Cdc45 and the leading strand for wtCMG only exist in absence of nucleotide, and, thus, only when the discontinuity between Mcm2 and Mcm5 is not entirely locked, these captured interactions persist when the closure of the gate might be defective (see Chapter 4 for details). Cross-linking of '4PS1' and '47PS'1 CMG revealed a behavior for Cdc45 that is indicative for a putative dysfunctional '2/5gate' and demonstrated that Cdc45 retains its interaction with the leading strand even after nucleotide addition; a step that under normal circumstances entirely closes the gate (Costa et al., 2011). Mutations in other individual Mcm2-7 subunits displayed a wildtype interaction in presence and absence of nucleotide. Even the simultaneous mutation in Mcm2, 5, and 6 ('256PS1') of CMG followed the wtCMG behavior with regard to Cdc45, suggesting that this effect on the gate does not result from mutations of several PS1-hp at the same time, but specifically arises from the Mcm4 subunit, and is strengthened in combination with the adjacent Mcm7. Although the Mcm4/7 subunits are positioned opposite to the '2/5 gate', this data leads to postulate that there must be a high degree of cross-talk between the subunits regarding the nucleotide occupancy state, the location of hairpin motifs that move dependent on this, and the required closure of the discontinuity in the Mcm2-7 ring. One other discrete example of such intersubunit communication was demonstrated through the interaction between the ACL loop within the N-terminal domain and the PS1-hp of an adjacent subunit in SsoMCM and MthMCM (Barry et al., 2009; Sakakibara et al., 2008). Biochemical analysis of this loop strongly supported its crucial role in the communications between all MCM subunits crucial for cooperativity.

The importance of the PS1-hp in *Drosophila* Mcm2-7 was shown through the mutational analysis and biochemical dissection in this study. The corresponding hairpin in papillomavirus E1 was shown to be critical for the 'staircasing' of the ssDNA through the central channel. The hairpin motifs of all six subunits pull the nucleotides in a sequential manner through a combination of responsible acidic/basic residues (Enemark and Joshua-Tor, 2006). Such acidic/basic pairs are absent in DmPS1-hp, but similar motifs (basic/acidic) were only found at the tip of the Helix-2-insert β-hairpin (H2I-hp), which itself is absent in E1 and other SF-3 helicases and only found in MCM helicases (Enemark and Joshua-Tor, 2008). Mutational analysis of the H2I-hp in MthMCM described that a deletion of this hairpin motif displayed a 12-fold increase in ATPase activity upon addition of dsDNA and increased the DNA binding affinity, but inactivated the helicase activity (Jenkinson and Chong, 2006). This data suggested a key role of this motif in transducing the energy derived from the hydrolysis of nucleotide triphosphates (NTP) to the separation of DNA by MCM proteins. Therefore, alanine substitutions were introduced in the positively and negatively charged residues at the tip of H2I-hp in several DmMcm2-7 subunits. Since this hairpin motif was implicated in ATP-hydrolysis driven unwinding activity, Mcm2-7 subunits that displayed the strongest defects upon mutation of ATPase active sites were of particular interest. In *Drosophila* CMG the elimination of ATPase active site motifs (Walker A and arginine finger) at particularly the Mcm5/3 pair resulted in most defective ATPase, DNA binding and helicase activity (see also Chapter 3) (Ilves et al., 2010). With this in mind, the H2I-hp of Mcm3 and Mcm5 were mutated, and also the hairpin in Mcm4 subunit, which itself showed the weakest defects upon elimination of the ATPase active sites. All three CMG complexes containing mutated H2I-hp in one Mcm2-7 subunit, reduced the DNA binding by about one third compared to the wildtype activity. In contrast to this, a reduction of helicase activity for '3H2I' and '5H2I' was only very mild or not at all detectable, respectively. Different from this finding, an elimination of the basic/acidic residues in the H2I-hp of Mcm4 diminished the unwinding activity by about 60% of wtCMG, suggesting an important role of this subunit in the translocation mechanism. However, unlike the order of importance around the Mcm2-7 ring with regard to the ATPase activity, the mutation of this second candidate hairpin motif within the AAA+domain also pointed to key roles in unwinding for subunits located across the Mcm2/5 gate, similar to the mutational data of PS1-hp. Consistent with the DNA binding in the gel-shift assays, cross-linking of the H2I-hp mutant CMG complexes demonstrated wildtype interactions between Mcm2-7 and the leading strand. The elimination of the residues located on the tip of H2I-hp in these three tested Mcm2-7 subunits of CMG did not result in abrogation of the direct protein-DNA contacts. Together with the results from

the mutational analysis of the other two β–hairpins within the central channel, it is obvious that several interactions per MCM subunit with the encircled DNA must exist. The unequal contributions from the six different Mcm2-7 subunits with regard to DNA binding and subsequent unwinding suggest that some subunits are more important for these enzyme's activities than others, in agreement with the idea of a particular position for the 'power stroke' within the Mcm2-7 ring. Analyzing the data in a simplified straightforward way, it is tempting to speculate on such location. Given that the Mcm2-7 are loaded in a head-to-head manner onto duplex DNA, major remodeling is required to ensure that the activated helicase in form of CMG engages only the leading strand through the Mcm2-7 central cavity (Evrin et al., 2009; Fu et al., 2011; Ilves et al., 2010; Remus et al., 2009). For this transition, the exclusion of the lagging strand after initial melting is mandatory and this separated strand can be actively segregated through the '2/5 gate' of the Mcm2-7 ring. While predominantly Mcm3, Mcm5, and Mcm6 grasp the DNA strand through their NT-hp, in case of a sequential mechanism mode similar to E1, the nucleotides of the leading strand should then be consecutively pulled from subunit to subunit upon nucleotide-binding, –hydrolysis, and -release, requiring the additional grasping through hairpin motifs in the AAA+ domain. While the lagging strand is guided to the outside, Cdc45 might display the role of a 'guardian of the gate' and ensure that the leading strand remains properly engaged and escorted to one side of the gate to initiate the translocation by the AAA+ domain after the two strands have been separated. Theoretically such guidance to one starting point could be on either side of the gate, but in agreement with the mutational dissection and the understanding of the different contributions of the six Mcm2-7 proteins, I postulate that the translocation initiates close to the Mcm3/7 pair. The orientation of archaeal MCMs has been demonstrated on forked DNA substrates to be such, that the AAA+ tier of the MCM faces the dsDNA during translocation (McGeoch et al., 2005). Recent findings for the *Drosophila* CMG confirmed the same orientation (Alessandro Costa, *unpublished data*). With CMG moving with its C-terminus towards the duplex during unwinding, one could envision a scenario, in which the hairpin motifs in the AAA+ domain consecutively 'pull' the DNA through the center, while the NT-hp hold the leading strand in place as it exits the central channel at the N-terminus of the Mcm2-7 ring (Figure 5.18). If the principle unwinding mechanism is in accordance to E1, each Mcm2-7 subunit should pull one nucleotide at a time during cycles of ATP-binding, -hydrolysis, and –release (Enemark and Joshua-Tor, 2006). This would require several such cycles for unwinding of the 40bp duplex DNA used in our helicase activity assays. Mutations in the initial start site might, thus, display a more detrimental defect on the enzyme's activity if the start of the motor is impaired and hung up. On the other hand, mutations at a later

point in the sequence of events might become more readily overcome and therefore display less profound defects on the overall activity. Since the values are averaged from many protein complexes and measured after multiple cycles of nucleotide-protein interactions, one cannot assess the direct effects, which the described mutations would have, in one immediate cycle of translocation on DNA. Further, a high degree of inter- and intra-subunit cross-talk, impedes such simplified interpretation of the results. Crystallographic or single molecule studies capturing one CMG complex on DNA and measuring the precise effects of each of these hairpin motif mutations around the ring will be required to gain resolution and to further test my hypothesis about the discrete start site for the translocation initiation. Recently, Alessandro Costa has obtained structural insights into DmCMG complexes bound to DNA by electron microscopy (Figure 5.17) (Costa, 2013, manuscript in preparation). Surprisingly, the rather planar structure of CMG, previously reported in absence of nucleotide, must undergo vast structural re-arrangements leading to the final adoption of a spiral structure when DNA is bound through the central cavity of Mcm2-7 (Figure 5.17A) (Costa et al., 2011). Unlike the AAA+ domain of CMG, the N-terminal collar remains planar (Figure 5.17B). The representation of the central PS1-hp loops alone shows that the staircase is formed by Mcm5-Mcm3-Mcm7-Mcm4-Mcm6 from top to bottom, but further shows that Mcm2 is distal to an interaction with the leading strand and slightly recessed from the center (Figure 5.17B). While the 18Å resolution in this study cannot provide information about the discrete DNA path, it visualizes the dramatic distortion of the Mcm2-7 subunits from a planar to a spiral arrangement reminiscent of such spiral structures observed for other AAA+ and RecA families of helicases as well as for the base of the 19S regulatory particle of the proteasome (Enemark and Joshua-Tor, 2006, Arias-Palomo et al., 2013, Lander et al., 2012). One could speculate that the leading strand is sequentially either guided in such staircase arrangement itself through the central cavity with a static spiral arrangement of the Mcm2-7 or pulled straight through a dynamic spiral alignment of the pore loops as in E1 (Enemark and Joshua-Tor, 2006). Since this structural information represents a 'snapshot' in an ATPγS bound state, it cannot be ruled out that CMG undergoes switches between the planar and spiral states in a regular manner during translocation, or otherwise uses this switch mechanistically for the initial melting step, and retains a planar conformation during unwinding. However, together with my mutational data, it is tempting to speculate on a scenario in which the Nt-hp guide the leading strand through the exit at the N-terminal end of the complex through mainly an interaction of Mcm3, Mcm5, and Mcm6, while the hairpin motifs in the AAA+ domain sequentially pull the nucleotides (Figure 5.18). I have discussed in Chapter 4 that in absence of nucleotide the

leading strand is mainly held in place by Mcm5 protein in absence of ATP, but that upon ATP addition a major shift becomes apparent, through which the remaining Mcm2-7 subunits establish tight contacts with the DNA with Mcm2 displaying the weakest direct interaction to the leading strand. Thus, one can hypothesize that the switch occurs in such spiral manner as visualized by EM, and this discrete arrangement leads to a site-specific start for the unwinding in which the leading strand is initially held in place by Mcm5 and then subsequently 'pulled' by the adjacent subunit when the motor is started. Further high-resolution experiments (e.g. on the single-molecule level) will help to shed light onto the question of interest whether this conformation remains static or switches between different adaptations, and if a specific start site is utilized for the 'power stroke' with a sequential participation of the six Mcm2-7 subunits.

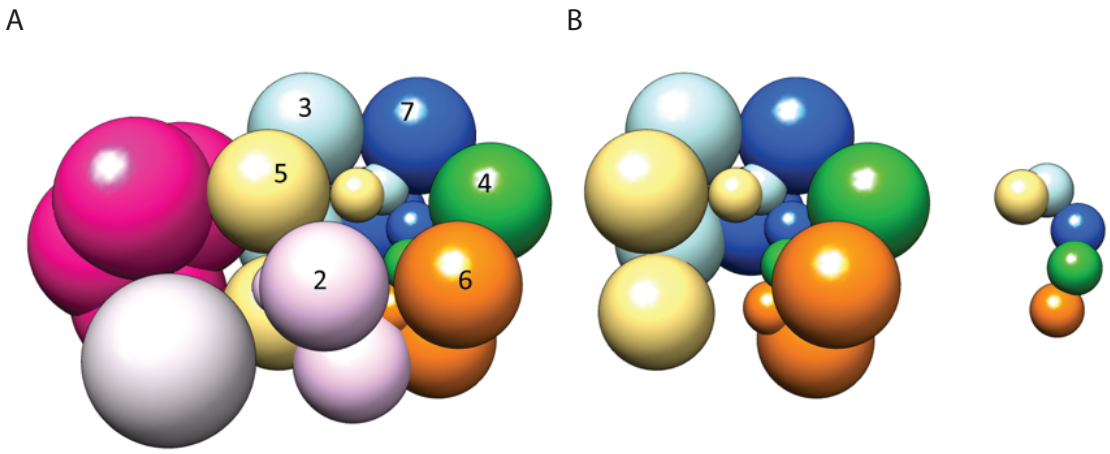

Figure 5.17. Schematic of EM of CMG complexes bound to DNA.
(A) Shown is a schematic diagram of the structure of CMG complexes bound to DNA obtained from electron-microscopic studies at 18Å resolution. The Mcm2-7 subunits are numbered and color-coded. The N- and C-terminal domains of each MCM subunit are depicted as spheres (top view on AAA+ domain, slightly tilted). The side-handle formed by Cdc45 and GINS is shown in gray and magenta, respectively (Figure was generously provided by Alessandro Costa) (Costa, 2013, manuscript in preparation).
(B) Cdc45, GINS and Mcm2 subunit are removed from the schematic. The view shows that the AAA+ domain of CMG adopts a right-handed spiral structure upon DNA binding while the N-terminal collar remains planar (*left panel*). The central Pre-Sensor1 β-hairpin motifs are shown separately and emphasize the spiral structure (*right panel*).

The proposed spatial segregation of the lagging strand to the outside raises the immediate question of its path when excluded from the central channel. Work presented from SsoMCM showed a dynamic interaction between the 5'-lagging strand and the MCM (Rothenberg et al., 2007). Additionally, a meticulous dissection of each of the amino acids

within the fourth hairpin, named External β–hairpin (EXT-hp) due to its location on the exterior of MCM, identified residues of interest with regard to DNA binding. Particularly the mutation of a conserved lysine (K323) at the base of the external β–strand was demonstrated to significantly impair the helicase activity of SsoMCM (Brewster et al., 2010). Successive mutational analysis of positively charged amino acids on the outside of SsoMCM by another group, confirmed the importance of that invariant lysine residue within the EXT-hp, showing that mutations of these positive residues on the outside destabilize the DNA-MCM interaction and reduce the unwinding activity (Graham et al., 2011). Mutations in the corresponding lysine residue in each of the six DmMcm2-7 subunits (6xEXT') did not affect the DNA binding affinity of the *Drosophila* CMG complex, consistent with the Mcm2-7 central channel being the major DNA binder. However, the introduction of alanine substitutions in these residues dramatically reduced the helicase activity of the complex (Chapter 4, Figure 5D). These data together suggest, that the strongest binding of CMG to DNA occurs through the engagement of the leading strand by the Mcm2-7 interior, while the lagging strand is required in a supportive role during unwinding and possibly for the stability of the complex during translocation (Pesavento et al., 2013, manuscript in preparation). Specific cross-linking of DmCMG proteins to the lagging strand revealed that mutations of the conserved lysine in the EXT-hp drastically decreased the overall DNA-Mcm2-7 interactions (Chapter 4, Figure 5E). However, none of the direct interactions between the lagging strand and the individual Mcm2-7 subunits were abolished, suggesting that additional contacts likely exist. Chapter 4 discussed in detail the analogy of these observed diminished DNA interactions to the 'steric exclusion and wrapping mechanism' proposed for DNA unwinding by SsoMCM (Graham et al., 2011). In contrast to the mutational dissection of the other internal β–hairpin motifs, but in agreement with the findings for the '6xEXT'CMG complex, the introduction of EXT-hp mutations in either single or a subset of Mcm2-7 subunits, which establish the direct interaction to the lagging strand upon ATP binding (Mcm2, Mcm4, and Mcm6), did not result in reduction of DNA binding in gel-shift assays. The unwinding of these mutant 'EXT' CMG complexes was only mildly reduced to about 20-25% of wt activity in complexes with mutations in either Mcm6 alone ('6EXT'), Mcm4 and Mcm6 ('46EXT'), or Mcm2, Mcm4, and Mcm6 ('246EXT'); targeting the three Mcm2-7 subunits that establish the strongest interactions to the lagging strand when ATP is bound. Between the se three mutant CMG complexes, mutation of the EXT-hp in Mcm6 could play an important role in the guidance of the lagging strand, but it is clear from the biochemical analysis that it does not play an absolutely critical role during unwinding, and that possibly other residues might alleviate the defect. Cross-linking of CMG '6EXT' or CMG '4EXT', further did not abrogate the direct

interaction of these subunits to the DNA, suggesting that other residues on the external Mcm2-7 surface might be responsible for additional binding. The existence and importance of other candidate residues on the MCM exterior with regard to the helicase activity have been demonstrated for SsoMCM (Graham et al., 2011).

These data together lead to the hypothesis, that each Mcm2-7 subunit's interaction with the EXT-hp might rather be weak and of more dynamic nature. While the leading strand is pulled through the central channel, the lagging strand wraps around the outside of the Mcm2-7 (Figure 5.18). Such interaction should remain loose to allow the continuous unwinding process as the CMG moves along the DNA. Thus, the lagging strand might wrap around a specific path of positive residues around the Mcm2-7 surface to guide the lagging strand during translocation of CMG on DNA, and only simultaneous mutations in all these residues might weaken the DNA path to such extent that leads to an inactivated helicase activity and ablation of direct interactions.

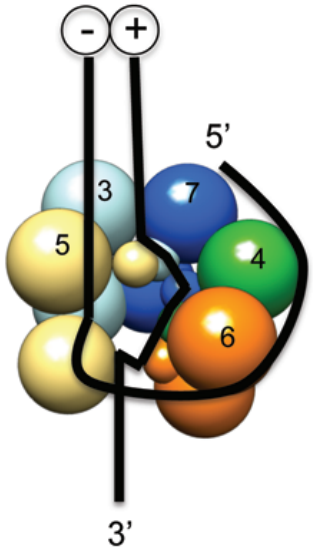

Figure 5.18. Model.
CMG moves with the AAA+ domain towards the duplex DNA and translocates in 3' to 5' direction. From the loading of the Mcm2-7 in form of a double hexamer around dsDNA, the DNA must undergo major re-modeling leading to melting of the duplex. One could hypothesize that this is mechanically achieved through adoption of the spiral arrangement of AAA+ domains within Mcm2-7 of CMG, allowing in parallel the exclusion of the lagging strand through possibly the '2/5 gate'. Subunit Mcm5 would keep both melted DNA strands in close proximity until nucleotide-binding and -hydrolysis induce re-arrangements, leading to the final binding of both strands as depicted in the model. The leading strand remains tightly engaged by the Mcm2-7 central channel, while the lagging strand is excluded to the outside and wraps around the Mcm2-7 external surface. The remaining Mcm2-7 subunits guide the leading strand in a staircasing manner through its central channel, starting the pulling of of nucleotide at the adjacent Mcm5 subunits through the PS1-hp of Mcm3 and Mcm7. During these cycles of ATP-binding, -hydrolysis, and –release, the leading strand exits the N-terminal domain of the central channel through assistance of the Nt-hp. This model depicts a sequential order of events similar as proposed for papillomavirus E1, and excludes a coordinated movement of the central channel hairpin motifs and coordinated pulling of the leading strand.

5.5 Materials and Methods

Mutagenesis of baculovirus

Viruses used in the studies were constructed following the manufacturer's manual for the Bac-to-Bac expression system from Invitrogen and were described in detail in Chapters 2 and 3.

CMG mutants were generated through PCR based site-directed mutagenesis. Sequencing was used to verify the entire protein coding regions of all generated pFastBac constructs.

In detail following specific residues were targeted and alanine substitutions were introduced:

N-terminal β-hairpin (NT-hp) mutations:

Mutations of basic residues in NT-hp were introduced in the six MCM subunits by substituting following residues with alanines: K244 and R245 in MCM3, K412 in MCM4, R278 and K280 in MCM5, R263 and K265 in MCM6, and R282 in MCM7.

Pre-Sensor1 β-hairpin (PS1-hp) mutations:

The K or R to A mutations in PS1-hp of all Mcm2-7 subunits targeted following residues: K598 in MCM2, K430 in MCM3, K602 in MCM4, K468 in MCM5, R478 in MCM6, K471 in MCM7.

Helix-2-Insert β-hairpin (H2I-hp) mutations:

Mutations of basic and acidic pairs in H2I-hp were introduced in three MCM subunits by substituting following residues with alanines: D380 in MCM3, K551 and D552 in MCM4, and K417 and D418 in MCM5.

External β-hairpin (EXT-hp) mutations:

The K to A mutations in EXT-hp of the six MCM subunits were introduced by targeting following residues: K490 in MCM2, K322 in MCM3, K492 in MCM4, K360 in MCM5, K370 in MCM6, and K364 in MCM7.

Protein expression, purification, and biochemical assays

Chapters 2-4 of this thesis contain detailed description of the methods and used materials. All biochemical assays described here have been outlined in the 'Material and Methods' of the previous chapters.

The only deviation for the described protein purification is with regard to the used Mcm2-7 complexes. The isolated Mcm2-7 complexes have been obtained from the side fractions during the CMG purifications. The side fractions containing only free Mcm2-7 complexes have been further subjected to a MonoQ column for additional clean-up and concentration, and subsequently dialyzed into buffer A (25mM Hepes pH 7.6, 15-35% glycerol, 50mM

sodium acetate, 10mM magnesium acetate, 0.2mM PMSF, 1mM DTT). The protein samples were as usual aliquoted, flash frozen in liquid nitrogen and stored at -80°C.

Homology modeling

The crystal structure of the near-full-length *Sulfolobus solfataricus* MCM (PDB entry: 3F9V) was used as template to generate homology models for *Drosophila* Mcm2-7 proteins via SWISS-MODEL. The obtained models were threaded onto the crystal structure of SsoMCM to allow identification of specific residues.

Acknowledgements

I would like to express my thank to Alessandro Costa, who generously provided some models based on his unpublished data of the CMG complex bound to DNA and, thus, allowed a comparison between the EM data and my mutational analysis. Also, I want to thank Aylin Goke, who has worked as an undergraduate student with me on some initial questions regarding this biochemical study.

Chapter 6

Conclusion

Sclafani and Holzen introduced their 2007 review on cell cycle regulation during DNA replication by quoting François Jacob from 1965:

> *"Le rêve d'une bactérie doit devenir deux bactéries* (The *dream of a bacterium is to become two bacteria)*" (Sclafani and Holzen, 2007).

With this sentence Jacob emphasized the fundamental importance of cell division to life, and, thus, the importance of DNA replication. Every proliferating cell must replicate its entire genome in a well-orchestrated and precise way. Understanding the mechanism that underlies this fundamental principle is hence of the utmost importance, and much research has been invested in deciphering the steps of this basic biological principle.
In *Drosophila melanogaster*, the Mcm2-7 helicase motor requires the incorporation of Cdc45 and GINS to activate its intrinsic activities during the switch to S phase of the cell cycle. Only then can the enzyme initiate the unwinding of double-stranded DNA, creating single-stranded templates for the polymerases to subsequently copy. The work presented in this thesis focuses on dissecting the exact participation of the auxiliary replication factors within the CMG complex, as well as that of the six different Mcm2-7 proteins to gain insights into the molecular mechanism of this activation.

The recombinant *Drosophila* CMG complex was successfully purified from insect cells using baculoviruses, each encoding one of the eleven different genes for the eukaryotic helicase complex. By introducing site-specific mutations into various CMG subunits, I dissected the ATPase, helicase and DNA binding activities of the Mcm2-7 motor, along with investigating the functional contributions made by the auxiliary factors Cdc45 and GINS.

The regulation and correct execution of each step during the cell cycle, a process proliferating cells undergo repeatedly, are mandatory in order to maintain cellular integrity. Aberration of these events can result in diseases connected to dysfunctional control and progression through the cell cycle; indeed, the set of diseases known as cancer is functionally a problem of uncontrolled cell growth. A single amino acid change in Mcm4 has even been directly linked to breast cancer in mice ($Mcm4^{Chaos3}$) and exemplifies one direct correlation between the CMG proteins and disease (Shima et al., 2007). Thus, unraveling the direct protein-protein interactions, required for CMG stability, is important to provide clues about regions within the CMG complex that might become of interest in the search for novel targets that could help control this activation step. A stable complex

formation was only found when all eleven proteins were co-expressed, suggesting that the most stable contacts are only made within the larger CMG complex.

The data presented here revealed the importance of the last 18 amino acids at the carboxy-terminus of Psf1 for the stability of the CMG. Secondary structure predictions identified an α-helical structure within the Psf1 C-terminus and alanine scanning of this helix demonstrated that the simultaneous mutation of four amino acids within this region accounted for the hindrance in CMG formation. This B-domain was shown to have a high degree of structural flexibility an its removal resulted in impairment of DNA replication (Kamada et al., 2007). Functional interactions between the GINS proteins and polymerases have been reported by multiple groups and have led to speculation whether, the accessibility of the B-domain of Psf1, with its here unraveled important role in the CMG stability, could serve as a platform to anchor replication proteins that are required beyond the initiation of unwinding (De Falco et al., 2007; Takayama et al., 2003). While specific interactions between Psf1 and other replication proteins can be investigated in future replisome assembly studies, this work has identified one essential domain in Psf1 required for CMG stability. This domain is a potential novel target for the development of cancer therapeutics, which might work to inhibit the activation of the helicase enzyme in tumor cells and, thus, prevent strand separation during one of the initial stages of DNA replication.

Aside from reports in budding yeast on ATPase and helicase activities of the isolated Mcm2-7 complexes, only *Drosophila* CMG activity has been analyzed with regard to mutations in one of the nucleotide binding sites (Walker A box) (Ilves et al., 2010). The data presented in this study demonstrated the importance of the nucleotide-bound state for the CMG helicase function. The extension of this work in studies of the nucleotide-hydrolysis 'arginine finger' motif not only confirmed the findings from the initial ATP-binding site (Walker A) studies of the CMG complex, but further substantiated the particular importance of certain active site pairs around the Mcm2-7 ring. The six ATPase active sites are formed by Walker A box motifs in *cis* and *trans*-acting arginine finger motifs, and mutations in the latter showed a reduction of ATPase rates far below the $1/6^{th}$ one would expect in the case of a stochastic model, in which all six subunits contribute equally and independently to the overall activity of the complex. A functional asymmetry of the Mcm2-7 subunits was reported for the CMG complex, in which mutations in one-half of the ring displayed a more dramatic reduction of ATPase and helicase activities while mutations in the other half had little effect. The defects of the arginine finger mutations correlated well with those of the Walker A box mutations on the helicase activity of CMG,

together pointing to Mcm2/5, Mcm5/3 and Mcm3/7 as the most critical active site pairs for the unwinding activity (Ilves et al., 2010).

While nucleotide is not required for complex assembly, ATP-binding, but not –hydrolysis, is essential for the interaction of CMG with DNA. Interestingly, both DNA binding and helicase activity display a strong nucleotide-dependence, with vast stimulation upon initial ATP addition and a decrease after the optimal nucleotide concentration is exceeded. In agreement with the ATP-dependent DNA binding of CMG, the work presented here establishes an interesting correlation between the elimination of the Walker A or arginine finger motif in the single Mcm2-7 subunits and the decrease in DNA binding. With a couple of exceptions, the ATPase active sites in the Mcm2-7 subunits that were most important for ATPase and helicase activities were also those Mcm2-7 subunits that were most important for DNA binding. Thus, this work unravels the unequal contributions of the Mcm2-7 subunits around the ring.

A major focus of the thesis lay in understanding the discrete path of the two separated strands of DNA within the activated helicase complex, as well as unraveling the contributions of the six Mcm2-7 subunits. Heretofore, only analyses of archaeal MCMs were reported, in which six copies of one MCM protein hexamerize to form the active helicase. Thus, mutations of residues in one MCM protein are present in all six MCM monomers and do not allow the distinction between individual monomers. In contrast to this, analysis of the eukaryotic CMG complex, shown here, enables a clear identification of effects that each of the six different Mcm2-7 subunits displays when specific residues are mutated. Cross-linking of CMG to labeled DNA enabled the visualization of individual protein-DNA contacts between CMG and either the leading or lagging strands separately. In the absence of ATP, both strands are in close proximity to the Mcm5 subunit. However, upon nucleotide addition major switches become apparent for both strands. Now, while all six Mcm2-7 polypeptides established contact to the leading strand, only a subset of MCM proteins (2, 4, 6) bound the lagging strand. Further, no interaction with the GINS proteins was observed, while Cdc45 bound only to the leading strand in the absence of ATP, suggesting a direct interaction between this auxiliary factor and the leading strand, which normally lies within the central channel, when the discontinuity in the Mcm2-7 ring between Mcm2 and Mcm5 is not tightly closed (named the 'Mcm2/5 gate'). Mutational analysis of Cdc45 allowed the identification of a discrete loop that is responsible for this direct contact to the leading strand. Changing the five amino acids in this loop to alanines not only abrogated the specific cross-linking of Cdc45 to the leading strand, but, moreover,

drastically decreased the helicase activity and processivity of the entire complex. For the first time, this study identified regions within Cdc45 that, even though structurally predicted to reside distal to the Mcm2-7 motor, directly induced alterations in the enzyme's activity. Mutations that hinder the proper closure of the Mcm2/5 gate retain Cdc45-DNA interactions even after ATP addition, suggesting that Cdc45 might play the role of a 'guardian' of this 2/5 gate for whenever it opens. Whether such opening occurs regularly in a dynamic fashion or only upon challenge remains to be answered. However, the abrogation of CMG's processivity upon hindering the closure of the gate is in accord with a scenario of regular gate opening during translocation. It appears that the proper gate closure, together with Cdc45's role in ensuring that the leading strand remains in the central channel, are important during movement of the CMG along DNA.

Analyses of potential DNA binding elements on the exterior surface of all six Mcm2-7 subunits revealed that the lagging strand binds to the outside of the Mcm2-7 ring within the context of the CMG. Mutations in these 'external β-hairpins' strongly reduced the cross-linking to the lagging strand and significantly reduced the helicase activity of the complex. Together with the finding that only a specific subset of Mcm2-7 subunits cross-links to the lagging strand, the data suggests a wrapping of segregated DNA strand around the external surface of the Mcm2-7.

In addition to the external β-hairpin motif, the work presented in this thesis demonstrated that three β-hairpin motifs within the central cavity of the Mcm2-7 ring are important for binding the leading strand and for the helicase function of the CMG complex. Heretofore, the only DNA binding studies reported have been from analyses of the corresponding motifs the in archaeal MCM complex, which shows many parallels to the eukaryotic replication machinery, and often serves as a simpler model. The narrowest point of the N-terminal domain in archaeal MCM is formed by the N-terminal β-hairpins (NT-hp), which protrude from each of the six Mcm2-7 proteins directly into the central channel (Fletcher et al., 2003). Reminiscent of the functional asymmetry reported for the ATPase active sites of the CMG, mutation of the NT-hp showed that there is an unequal contribution of the DNA binding motifs of the DmMcm2-7 polypeptides. Mutations in the β-hairpin motifs of Mcm3, Mcm5, and Mcm6 within the CMG significantly reduced DNA binding and subsequent unwinding by the helicase, while alteration of the positive charges in the NT-hp of the remaining two subunits, Mcm4 and Mcm7, did not affect the complexes' activities. The helicase activities of the five-fold ('5xNthp') and triple mutant ('356Nthp') CMG complexes were decreased to a level proportional to the degree of reduction in DNA

binding affinity of these mutant complexes. Interestingly, they did not abrogate the unwinding, suggesting a role of this hairpin motif in primarily DNA binding.
This point is supported by the finding that in cross-linking of these NT-hp mutant CMG complexes, the direct contact of Mcm3/Mcm6 to the leading strand is abolished. Further comparison of isolated DmMcm2-7 with CMG complexes identified that the NT-hp are required in both complexes for DNA binding, and that there is a strong asymmetry in the requirement of the NT-hp's around the complex in a similar way in both complexes.

In agreement with the suggestion from archaea that other motifs within the AAA+ domain of the MCM protein might account for the translocation mechanism, the mutation of a critical residue at the tip of the Pre-Sensor1 β-hairpin in all six Mcm2-7 proteins almost entirely abrogated DNA binding of the CMG and completely inactivated helicase activity, suggesting that this motif is required for both DNA binding and subsequent unwinding. From the biochemical dissection, it became evident that not all six PS1-hp contribute equally to the DNA binding of CMG, a feature also observed with the NT-hp. The highest reduction was found with PS1-hp mutations in Mcm3 and Mcm4, the same subunits that displayed the strongest decrease in DNA unwinding activity. Strikingly, an individual mutation in the PS1-hp of Mcm7 within CMG decreased the helicase activity most significantly, although DNA binding was only very mildly reduced, suggesting a pivotal role of this PS1-hp in translocation. Together, the data narrowed down a crucial role for the PS1-hp of Mcm3 and Mcm7 in DNA binding and unwinding. Unlike the NT-hp, side-by-side comparison of free Mcm2-7 and CMG complexes demonstrated that the PS1-hp in isolated Mcm2-7 complexes are not used for DNA binding, and that other motifs, such as the NT-hp, might be responsible for the DNA interactions of the loaded Mcm2-7 complexes before the association of Cdc45/GINS and subsequent activation within the CMG complex. These observations support a scenario, in which the Mcm2-7 ring undergoes vast structural remodeling upon binding of Cdc45 and GINS that lead to an optimal alignment of the β-hairpin motifs in the AAA+ domain and establish tighter and additional contacts with the leading strand.
However, in agreement with archaeal MCMs, the PS1-hp are crucial for translocation of the CMG along DNA. The corresponding hairpin in papillomavirus E1 was shown to be critical for the 'stair casing' of the ssDNA through the central channel. In E1, the hairpin motifs of all six subunits "escort" the nucleotides in a sequential manner (Enemark and Joshua-Tor, 2006). Whether *Drosophila* CMG follows such sequential events, or, in contrast, exhibits coordinated ATP-binding and –hydrolysis events, as shown for SV40 large T-antigen (LTag), during translocation remains to be determined (Gai et al., 2004). Given that the

motifs identified to interact with DNA do not contribute equally to the enzyme's overall activities and display strong asymmetries regarding their function, a similar sequential mechanism as described for E1 is strongly favored. The unequal contributions from the six different Mcm2-7 subunits with regard to DNA binding and subsequent unwinding suggest that some subunits are more important for the enzyme's overall activity than others, consistent with the idea of a particular position needed for the 'power stroke' within the Mcm2-7 ring.

To test this idea, future research will analyze the effects of the point mutant CMG proteins within one cycle of nucleotide binding, -hydrolysis, and –release, instead of averaging effects of molecules in bulk. Such fine-tuned molecular resolution could be achieved by the use of single-molecule techniques that visualize the CMG complex and measure the translocation steps of individual helicase complexes.

Taken together, the DNA binding studies have allowed us to understand in detail the paths of the two separated DNA strands after their initial unwinding. While the lagging strand is excluded to the outside of the motor and wraps around the exterior of the Mcm2-7 ring in a discrete path, the leading strand remains tightly bound within the central channel. Analysis of the DNA binding motifs in the center provided insights that each subunit possesses several putative interacting motifs. However, detailed analysis unraveled that the DNA binding motifs contribute unequally to binding and unwinding activities, and demonstrate that specific Mcm2-7 sites are more important than others, at least in the conditions in our assays. Thus, the detailed biochemical dissection of various activity motifs have shown *in vitro* that different subunits are required for different tasks during unwinding and provides a functional explanation why the eukaryotic replication machinery has evolved to an Mcm2-7 motor consisting of six different proteins rather than six copies of one with each of the six genes encoding for one subunit being essential.

References

Abbate, E.A., Berger, J.M., and Botchan, M.R. (2004). The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2. Genes & development *18*, 1981-1996.

Aparicio, O.M., Weinstein, D.M., and Bell, S.P. (1997). Components and dynamics of DNA replication complexes in S. cerevisiae: redistribution of MCM proteins and Cdc45p during S phase. Cell *91*, 59-69.

Aparicio, T., Ibarra, A., and Mendez, J. (2006). Cdc45-MCM-GINS, a new power player for DNA replication. Cell division *1*, 18.

Arias-Palomo, E., O'Shea, V.L., Hood, I.V., and Berger, J.M. (2013). The Bacterial DnaC Helicase Loader Is a DnaB Ring Breaker. Cell *153*, 438-448.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics (Oxford, England) *22*, 195-201.

Bae, B., Chen, Y.H., Costa, A., Onesti, S., Brunzelle, J.S., Lin, Y., Cann, I.K., and Nair, S.K. (2009). Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog. Structure *17*, 211-222.

Barry, E.R., Lovett, J.E., Costa, A., Lea, S.M., and Bell, S.D. (2009). Intersubunit allosteric communication mediated by a conserved loop in the MCM helicase. Proceedings of the National Academy of Sciences of the United States of America *106*, 1051-1056.

Barry, E.R., McGeoch, A.T., Kelman, Z., and Bell, S.D. (2007). Archaeal MCM has separable processivity, substrate choice and helicase domains. Nucleic acids research *35*, 988-998.

Bauerschmidt, C., Pollok, S., Kremmer, E., Nasheuer, H.P., and Grosse, F. (2007). Interactions of human Cdc45 with the Mcm2-7 complex, the GINS complex, and DNA polymerases delta and epsilon during S phase. Genes to cells : devoted to molecular & cellular mechanisms *12*, 745-758.

Beese, L.S., and Steitz, T.A. (1991). Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. The EMBO journal *10*, 25-33.

Bell, S.P., and Dutta, A. (2002). DNA replication in eukaryotic cells. Annual review of biochemistry *71*, 333-374.

Bell, S.P., and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128-134.

Blow, J.J., and Laskey, R.A. (1986). Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of Xenopus eggs. Cell *47*, 577-587.

Bochman, M.L., and Schwacha, A. (2007). Differences in the single-stranded DNA binding activities of MCM2-7 and MCM467: MCM2 and MCM5 define a slow ATP-dependent step. The Journal of biological chemistry *282*, 33795-33804.

Bochman, M.L., and Schwacha, A. (2008). The Mcm2-7 complex has in vitro helicase activity. Molecular cell *31*, 287-293.

Bochman, M.L., Bell, S.P., and Schwacha, A. (2008). Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability. Molecular and cellular biology *28*, 5865-5873.

Borowiec, J.A., Dean, F.B., Bullock, P.A., and Hurwitz, J. (1990). Binding and unwinding--how T antigen engages the SV40 origin of DNA replication. Cell *60*, 181-184.

Botchan, M. (2007). Cell biology: a switch for S phase. Nature *445*, 272-274.

Bowers, J.L., Randell, J.C., Chen, S., and Bell, S.P. (2004). ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. Molecular cell *16*, 967-978.

Bramhill, D., and Kornberg, A. (1988). Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the E. coli chromosome. Cell *52*, 743-755.

Brewer, B.J., and Fangman, W.L. (1987). The localization of replication origins on ARS plasmids in S. cerevisiae. Cell *51*, 463-471.

Brewster, A.S., and Chen, X.S. (2010). Insights into the MCM functional mechanism: lessons learned from the archaeal MCM complex. Critical reviews in biochemistry and molecular biology *45*, 243-256.

Brewster, A.S., Slaymaker, I.M., Afif, S.A., and Chen, X.S. (2010). Mutational analysis of an archaeal minichromosome maintenance protein exterior hairpin reveals critical residues for helicase activity and DNA binding. BMC molecular biology *11*, 62.

Brewster, A.S., Wang, G., Yu, X., Greenleaf, W.B., Carazo, J.M., Tjajadi, M., Klein, M.G., and Chen, X.S. (2008). Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase. Proceedings of the National Academy of Sciences of the United States of America *105*, 20191-20196.

Carpentieri, F., De Felice, M., De Falco, M., Rossi, M., and Pisani, F.M. (2002). Physical and functional interaction between the mini-chromosome maintenance-like DNA helicase and the single-stranded DNA binding protein from the crenarchaeon Sulfolobus solfataricus. The Journal of biological chemistry *277*, 12118-12127.

Chang, Y.P., Wang, G., Bermudez, V., Hurwitz, J., and Chen, X.S. (2007). Crystal structure of the GINS complex and functional insights into its role in DNA replication. Proceedings of the National Academy of Sciences of the United States of America *104*, 12685-12690.

Chen, Y.J., Yu, X., Kasiviswanathan, R., Shin, J.H., Kelman, Z., and Egelman, E.H. (2005). Structural polymorphism of Methanothermobacter thermautotrophicus MCM. Journal of molecular biology *346*, 389-394.

Choi, J.M., Lim, H.S., Kim, J.J., Song, O.K., and Cho, Y. (2007). Crystal structure of the human GINS complex. Genes & development *21*, 1316-1321.

Chong, J.P., Hayashi, M.K., Simon, M.N., Xu, R.M., and Stillman, B. (2000). A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. Proceedings of the National Academy of Sciences of the United States of America *97*, 1530-1535.

Chow, K.H., and Courcelle, J. (2007). RecBCD and RecJ/RecQ initiate DNA degradation on distinct substrates in UV-irradiated Escherichia coli. Radiation research *168*, 499-506.

Cocker, J.H., Piatti, S., Santocanale, C., Nasmyth, K., and Diffley, J.F. (1996). An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. Nature *379*, 180-182.

Coleman, T.R., Carpenter, P.B., and Dunphy, W.G. (1996). The Xenopus Cdc6 protein is essential for the initiation of a single round of DNA replication in cell-free extracts. Cell *87*, 53-63.

Cooper, D.L., Lahue, R.S., and Modrich, P. (1993). Methyl-directed mismatch repair is bidirectional. The Journal of biological chemistry *268*, 11823-11829.

Costa, A., et al. (2013), manuscript in preparation.

Costa, A., Ilves, I., Tamberg, N., Petojevic, T., Nogales, E., Botchan, M.R., and Berger, J.M. (2011). The structural basis for MCM2-7 helicase activation by GINS and Cdc45. Nature structural & molecular biology *18*, 471-477.

Costa, A., Pape, T., van Heel, M., Brick, P., Patwardhan, A., and Onesti, S. (2006). Structural basis of the Methanothermobacter thermautotrophicus MCM helicase activity. Nucleic acids research *34*, 5829-5838.

Courcelle, C.T., Chow, K.H., Casey, A., and Courcelle, J. (2006). Nascent DNA processing by RecJ favors lesion repair over translesion synthesis at arrested replication forks in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America *103*, 9154-9159.

Courcelle, J., and Hanawalt, P.C. (1999). RecQ and RecJ process blocked replication forks prior to the resumption of replication in UV-irradiated Escherichia coli. Molecular & general genetics : MGG *262*, 543-551.

Courcelle, J., Donaldson, J.R., Chow, K.H., and Courcelle, C.T. (2003). DNA damage-induced replication fork regression and processing in Escherichia coli. Science *299*, 1064-1067.

Crampton, D.J., Mukherjee, S., and Richardson, C.C. (2006). DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. Molecular cell *21*, 165-174.

Crevel, G., Ivetic, A., Ohno, K., Yamaguchi, M., and Cotterill, S. (2001). Nearest neighbour analysis of MCM protein complexes in Drosophila melanogaster. Nucleic acids research *29*, 4834-4842.

Crevel, G., Ivetic, A., Ohno, K., Yamaguchi, M., and Cotterill, S. (2001). Nearest neighbour analysis of MCM protein complexes in Drosophila melanogaster. Nucleic acids research *29*, 4834-4842.

Davey, M.J., Indiani, C., and O'Donnell, M. (2003). Reconstitution of the Mcm2-7p heterohexamer, subunit arrangement, and ATP site architecture. The Journal of biological chemistry *278*, 4491-4499.

De Falco, M., Ferrari, E., De Felice, M., Rossi, M., Hubscher, U., and Pisani, F.M. (2007). The human GINS complex binds to and specifically stimulates human DNA polymerase alpha-primase. EMBO reports *8*, 99-103.

Diffley, J.F., Cocker, J.H., Dowell, S.J., and Rowley, A. (1994). Two steps in the assembly of complexes at yeast replication origins in vivo. Cell *78*, 303-316.

Drury, L.S., Perkins, G., and Diffley, J.F. (2000). The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. Current biology : CB *10*, 231-240.

Elsasser, S., Chi, Y., Yang, P., and Campbell, J.L. (1999). Phosphorylation controls timing of Cdc6p destruction: A biochemical analysis. Molecular biology of the cell *10*, 3263-3277.

Enemark, E.J., and Joshua-Tor, L. (2006). Mechanism of DNA translocation in a replicative hexameric helicase. Nature *442*, 270-275.

Enemark, E.J., and Joshua-Tor, L. (2008). On helicases and other motor proteins. Current opinion in structural biology *18*, 243-257.

Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B., and Speck, C. (2009). A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proceedings of the National Academy of Sciences of the United States of America *106*, 20240-20245.

Fletcher, R.J., Bishop, B.E., Leon, R.P., Sclafani, R.A., Ogata, C.M., and Chen, X.S. (2003). The structure and function of MCM from archaeal M. Thermoautotrophicum. Nature structural biology *10*, 160-167.

Forsburg, S.L. (2004). Eukaryotic MCM proteins: beyond replication initiation. Microbiology and molecular biology reviews : MMBR *68*, 109-131.

Forsburg, S.L., Sherman, D.A., Ottilie, S., Yasuda, J.R., and Hodson, J.A. (1997). Mutational analysis of Cdc19p, a Schizosaccharomyces pombe MCM protein. Genetics *147*, 1025-1041.

Fu, Y.V., Yardimci, H., Long, D.T., Ho, T.V., Guainazzi, A., Bermudez, V.P., Hurwitz, J., van Oijen, A., Scharer, O.D., and Walter, J.C. (2011). Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. Cell *146*, 931-941.

Gai, D., Zhao, R., Li, D., Finkielstein, C.V., and Chen, X.S. (2004). Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. Cell *119*, 47-60.

Gambus, A., Jones, R.C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R.D., and Labib, K. (2006). GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nature cell biology *8*, 358-366.

Gambus, A., Khoudoli, G.A., Jones, R.C., and Blow, J.J. (2011). MCM2-7 form double hexamers at licensed origins in Xenopus egg extract. The Journal of biological chemistry *286*, 11855-11864.

Gibson, S.I., Surosky, R.T., and Tye, B.K. (1990). The phenotype of the minichromosome maintenance mutant mcm3 is characteristic of mutants defective in DNA replication. Molecular and cellular biology *10*, 5707-5720.

Gomez-Llorente, Y., Fletcher, R.J., Chen, X.S., Carazo, J.M., and San Martin, C. (2005). Polymorphism and double hexamer structure in the archaeal minichromosome maintenance (MCM) helicase from Methanobacterium thermoautotrophicum. The Journal of biological chemistry *280*, 40909-40915.

Graham, B.W., Schauer, G.D., Leuba, S.H., and Trakselis, M.A. (2011). Steric exclusion and wrapping of the excluded DNA strand occurs along discrete external binding paths during MCM helicase unwinding. Nucleic acids research *39*, 6585-6595.

Hanson, P.I., and Whiteheart, S.W. (2005). AAA+ proteins: have engine, will work. Nature reviews Molecular cell biology *6*, 519-529.

Hardy, C.F., Dryga, O., Seematter, S., Pahl, P.M., and Sclafani, R.A. (1997). mcm5/cdc46-bob1 bypasses the requirement for the S phase activator Cdc7p. Proceedings of the National Academy of Sciences of the United States of America *94*, 3151-3155.

Hattendorf, D.A., and Lindquist, S.L. (2002). Cooperative kinetics of both Hsp104 ATPase domains and interdomain communication revealed by AAA sensor-1 mutants. The EMBO journal *21*, 12-21.

Havens, C.G., and Walter, J.C. (2011). Mechanism of CRL4(Cdt2), a PCNA-dependent E3 ubiquitin ligase. Genes & development *25*, 1568-1582.

Heller, R.C., Kang, S., Lam, W.M., Chen, S., Chan, C.S., and Bell, S.P. (2011). Eukaryotic origin-dependent DNA replication in vitro reveals sequential action of DDK and S-CDK kinases. Cell *146*, 80-91.

Ilves, I., Petojevic, T., Pesavento, J.J., and Botchan, M.R. (2010). Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. Molecular cell *37*, 247-258.

Im, J.S., Ki, S.H., Farina, A., Jung, D.S., Hurwitz, J., and Lee, J.K. (2009). Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins. Proceedings of the National Academy of Sciences of the United States of America *106*, 15628-15632.

Ishimi, Y. (1997). A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex. The Journal of biological chemistry *272*, 24508-24513.

Iyer, L.M., Leipe, D.D., Koonin, E.V., and Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. Journal of structural biology *146*, 11-31.

Jackson, A.L., Pahl, P.M., Harrison, K., Rosamond, J., and Sclafani, R.A. (1993). Cell cycle regulation of the yeast Cdc7 protein kinase by association with the Dbf4 protein. Molecular and cellular biology *13*, 2899-2908.

Jacob, F., and Brenner, S. (1963). [On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon]. Comptes rendus hebdomadaires des seances de l'Academie des sciences *256*, 298-300.

Jares, P., and Blow, J.J. (2000). Xenopus cdc7 function is dependent on licensing but not on XORC, XCdc6, or CDK activity and is required for XCdc45 loading. Genes & development *14*, 1528-1540.

Jenkinson, E.R., and Chong, J.P. (2006). Minichromosome maintenance helicase activity is controlled by N- and C-terminal motifs and requires the ATPase domain helix-2 insert. Proceedings of the National Academy of Sciences of the United States of America *103*, 7613-7618.

Jenkinson, E.R., Costa, A., Leech, A.P., Patwardhan, A., Onesti, S., and Chong, J.P. (2009). Mutations in subdomain B of the minichromosome maintenance (MCM) helicase affect

DNA binding and modulate conformational transitions. The Journal of biological chemistry *284*, 5654-5661.

Kamada, K., Kubota, Y., Arata, T., Shindo, Y., and Hanaoka, F. (2007). Structure of the human GINS complex and its assembly and functional interface in replication initiation. Nature structural & molecular biology *14*, 388-396.

Kanemaki, M., Sanchez-Diaz, A., Gambus, A., and Labib, K. (2003). Functional proteomic identification of DNA replication proteins by induced proteolysis in vivo. Nature *423*, 720-724.

Kang, Y.H., Galal, W.C., Farina, A., Tappin, I., and Hurwitz, J. (2012). Properties of the human Cdc45/Mcm2-7/GINS helicase complex and its action with DNA polymerase epsilon in rolling circle DNA synthesis. Proceedings of the National Academy of Sciences of the United States of America *109*, 6042-6047.

Kaplan, D.L., Davey, M.J., and O'Donnell, M. (2003). Mcm4,6,7 uses a "pump in ring" mechanism to unwind DNA by steric exclusion and actively translocate along a duplex. The Journal of biological chemistry *278*, 49171-49182.

Kasiviswanathan, R., Shin, J.H., Melamud, E., and Kelman, Z. (2004). Biochemical characterization of the Methanothermobacter thermautotrophicus minichromosome maintenance (MCM) helicase N-terminal domains. The Journal of biological chemistry *279*, 28358-28366.

Kearsey, S.E., and Labib, K. (1998). MCM proteins: evolution, properties, and role in DNA replication. Biochimica et biophysica acta *1398*, 113-136.

Kelman, Z., Lee, J.K., and Hurwitz, J. (1999). The single minichromosome maintenance protein of Methanobacterium thermoautotrophicum DeltaH contains DNA helicase activity. Proceedings of the National Academy of Sciences of the United States of America *96*, 14783-14788.

Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L., and Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. Nucleic acids research *37*, D387-392.

Klemm, R.D., and Bell, S.P. (2001). ATP bound to the origin recognition complex is important for preRC formation. Proceedings of the National Academy of Sciences of the United States of America *98*, 8361-8367.

Klemm, R.D., Austin, R.J., and Bell, S.P. (1997). Coordinate binding of ATP and origin DNA regulates the ATPase activity of the origin recognition complex. Cell *88*, 493-502.

Koonin, E.V. (1993). A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. Nucleic acids research *21*, 2541-2547.

Kornberg, A., and Baker, T. (1992). DNA replication, 2nd edn (New York: W.H. Freeman).

Krastanova, I., Sannino, V., Amenitsch, H., Gileadi, O., Pisani, F.M., and Onesti, S. (2012). Structural and functional insights into the DNA replication factor Cdc45 reveal an evolutionary relationship to the DHH family of phosphoesterases. The Journal of biological chemistry *287*, 4121-4128.

Kubota, Y., Takase, Y., Komori, Y., Hashimoto, Y., Arata, T., Kamimura, Y., Araki, H., and Takisawa, H. (2003). A novel ring-like complex of Xenopus proteins essential for the initiation of DNA replication. Genes & development *17*, 1141-1152.

Labib, K. (2010). How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? Genes & development *24*, 1208-1219.

Lander, G.C., Estrin, E., Matyskiela, M.E., Bashore, C., Nogales, E., and Martin, A. (2012). Complete subunit architecture of the proteasome regulatory particle. Nature *482*, 186-191.

Lau, E., Tsuji, T., Guo, L., Lu, S.H., and Jiang, W. (2007). The role of pre-replicative complex (pre-RC) components in oncogenesis. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *21*, 3786-3794.

Lee, J.K., and Hurwitz, J. (2001). Processive DNA helicase activity of the minichromosome maintenance proteins 4, 6, and 7 complex requires forked DNA structures. Proceedings of the National Academy of Sciences of the United States of America *98*, 54-59.

Lei, M., Cheng, I.H., Roberts, L.A., McAlear, M.A., and Tye, B.K. (2002). Two mcm3 mutations affect different steps in the initiation of DNA replication. The Journal of biological chemistry *277*, 30824-30831.

Leon, R.P., Tecklenburg, M., and Sclafani, R.A. (2008). Functional conservation of beta-hairpin DNA binding domains in the Mcm protein of Methanobacterium thermoautotrophicum and the Mcm5 protein of Saccharomyces cerevisiae. Genetics *179*, 1757-1768.

Li, D., Zhao, R., Lilyestrom, W., Gai, D., Zhang, R., DeCaprio, J.A., Fanning, E., Jochimiak, A., Szakonyi, G., and Chen, X.S. (2003). Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. Nature *423*, 512-518.

Liu, W., Pucci, B., Rossi, M., Pisani, F.M., and Ladenstein, R. (2008). Structural analysis of the Sulfolobus solfataricus MCM protein N-terminal domain. Nucleic acids research *36*, 3235-3243.

Lovett, S.T., and Kolodner, R.D. (1989). Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the recJ gene of Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America *86*, 2627-2631.

Maine, G.T., Sinha, P., and Tye, B.K. (1984). Mutants of S. cerevisiae defective in the maintenance of minichromosomes. Genetics *106*, 365-385.

Maiorano, D., Moreau, J., and Mechali, M. (2000). XCDT1 is required for the assembly of pre-replicative complexes in Xenopus laevis. Nature *404*, 622-625.

Makarova, K.S., Koonin, E.V., and Kelman, Z. (2012). The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. Biology direct *7*, 7.

Makarova, K.S., Wolf, Y.I., Mekhedov, S.L., Mirkin, B.G., and Koonin, E.V. (2005). Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. Nucleic acids research *33*, 4626-4638.

Marinsek, N., Barry, E.R., Makarova, K.S., Dionne, I., Koonin, E.V., and Bell, S.D. (2006). GINS, a central nexus in the archaeal DNA replication fork. EMBO reports *7*, 539-545.

Masumoto, H., Muramatsu, S., Kamimura, Y., and Araki, H. (2002). S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. Nature *415*, 651-655.

Matsunaga, F., Forterre, P., Ishino, Y., and Myllykallio, H. (2001). In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. Proceedings of the National Academy of Sciences of the United States of America *98*, 11152-11157.

McGarry, T.J., and Kirschner, M.W. (1998). Geminin, an inhibitor of DNA replication, is degraded during mitosis. Cell *93*, 1043-1053.

McGeoch, A.T., Trakselis, M.A., Laskey, R.A., and Bell, S.D. (2005). Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism. Nature structural & molecular biology *12*, 756-762.

Mimura, S., and Takisawa, H. (1998). Xenopus Cdc45-dependent loading of DNA polymerase alpha onto chromatin under the control of S-phase Cdk. The EMBO journal *17*, 5699-5707.

Mimura, S., Masuda, T., Matsui, T., and Takisawa, H. (2000). Central role for cdc45 in establishing an initiation complex of DNA replication in Xenopus egg extracts. Genes to cells: devoted to molecular & cellular mechanisms *5*, 439-452.

Mizushima, T., Takahashi, N., and Stillman, B. (2000). Cdc6p modulates the structure and DNA binding activity of the origin recognition complex in vitro. Genes & development *14*, 1631-1641.

Moreau, M.J., McGeoch, A.T., Lowe, A.R., Itzhaki, L.S., and Bell, S.D. (2007). ATPase site architecture and helicase mechanism of an archaeal MCM. Molecular cell *28*, 304-314.

Moyer, S.E., Lewis, P.W., and Botchan, M.R. (2006). Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. Proceedings of the National Academy of Sciences of the United States of America *103*, 10236-10241.

Nakaya, R., Takaya, J., Onuki, T., Moritani, M., Nozaki, N., and Ishimi, Y. (2010). Identification of proteins that may directly interact with human RPA. Journal of biochemistry *148*, 539-547.

Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E.V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome research *9*, 27-43.

Nguyen, V.Q., Co, C., and Li, J.J. (2001). Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature *411*, 1068-1073.

Nishitani, H., Lygerou, Z., Nishimoto, T., and Nurse, P. (2000). The Cdt1 protein is required to license DNA for replication in fission yeast. Nature *404*, 625-628.

Nishitani, H., Sugimoto, N., Roukos, V., Nakanishi, Y., Saijo, M., Obuse, C., Tsurimoto, T., Nakayama, K.I., Nakayama, K., Fujita, M., et al. (2006). Two E3 ubiquitin ligases, SCF-Skp2 and DDB1-Cul4, target human Cdt1 for proteolysis. The EMBO journal *25*, 1126-1136.

Ogura, T., Whiteheart, S.W., and Wilkinson, A.J. (2004). Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. Journal of structural biology *146*, 106-112.

Onesti, S., and Macneill, S.A. (2013). Structure and evolutionary origins of the CMG complex. Chromosoma.

Owens, J.C., Detweiler, C.S., and Li, J.J. (1997). CDC45 is required in conjunction with CDC7/DBF4 to trigger the initiation of DNA replication. Proceedings of the National Academy of Sciences of the United States of America *94*, 12521-12526.

Oyama, T., Ishino, S., Fujino, S., Ogino, H., Shirai, T., Mayanagi, K., Saito, M., Nagasawa, N., Ishino, Y., and Morikawa, K. (2011). Architectures of archaeal GINS complexes, essential DNA replication initiation factors. BMC biology *9*, 28.

Pacek, M., Tutter, A.V., Kubota, Y., Takisawa, H., and Walter, J.C. (2006). Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. Molecular cell *21*, 581-587.

Pape, T., Meka, H., Chen, S., Vicentini, G., van Heel, M., and Onesti, S. (2003). Hexameric ring structure of the full-length archaeal MCM protein complex. EMBO reports *4*, 1079-1083.

Pesavento, J.J., Petojevic, T., et al. (2013), manuscript in preparation

Pospiech, H., Grosse, F., and Pisani, F.M. (2010). The initiation step of eukaryotic DNA replication. Sub-cellular biochemistry *50*, 79-104.

Pucci, B., De Felice, M., Rocco, M., Esposito, F., De Falco, M., Esposito, L., Rossi, M., and Pisani, F.M. (2007). Modular organization of the Sulfolobus solfataricus mini-chromosome maintenance protein. The Journal of biological chemistry *282*, 12574-12582.

Pucci, B., De Felice, M., Rossi, M., Onesti, S., and Pisani, F.M. (2004). Amino acids of the Sulfolobus solfataricus mini-chromosome maintenance-like DNA helicase involved in DNA binding/remodeling. The Journal of biological chemistry *279*, 49222-49228.

Randell, J.C., Bowers, J.L., Rodriguez, H.K., and Bell, S.P. (2006). Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. Molecular cell *21*, 29-39.

Randell, J.C., Fan, A., Chan, C., Francis, L.I., Heller, R.C., Galani, K., and Bell, S.P. (2010). Mec1 is one of multiple kinases that prime the Mcm2-7 helicase for phosphorylation by Cdc7. Molecular cell *40*, 353-363.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell *139*, 719-730.

Rothenberg, E., Trakselis, M.A., Bell, S.D., and Ha, T. (2007). MCM forked substrate specificity involves dynamic interaction with the 5'-tail. The Journal of biological chemistry *282*, 34229-34234.

Sakakibara, N., Kasiviswanathan, R., Melamud, E., Han, M., Schwarz, F.P., and Kelman, Z. (2008). Coupling of DNA binding and helicase activity is mediated by a conserved loop in the MCM protein. Nucleic acids research *36*, 1309-1320.

Sanchez-Pulido, L., and Ponting, C.P. (2011). Cdc45: the missing RecJ ortholog in eukaryotes? Bioinformatics (Oxford, England) *27*, 1885-1888.

Schwacha, A., and Bell, S.P. (2001). Interactions between two catalytically distinct MCM subgroups are essential for coordinated ATP hydrolysis and DNA replication. Molecular cell *8*, 1093-1104.

Schwob, E. (2004). Flexibility and governance in eukaryotic DNA replication. Current opinion in microbiology *7*, 680-690.

Sclafani, R.A., and Holzen, T.M. (2007). Cell cycle regulation of DNA replication. Annual review of genetics *41*, 237-280.

Shechter, D.F., Ying, C.Y., and Gautier, J. (2000). The intrinsic DNA helicase activity of Methanobacterium thermoautotrophicum delta H minichromosome maintenance protein. The Journal of biological chemistry *275*, 15049-15059.

Sherman, D.A., Pasion, S.G., and Forsburg, S.L. (1998). Multiple domains of fission yeast Cdc19p (MCM2) are required for its association with the core MCM complex. Molecular biology of the cell *9*, 1833-1845.

Sheu, Y.J., and Stillman, B. (2006). Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. Molecular cell *24*, 101-113.

Sheu, Y.J., and Stillman, B. (2010). The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. Nature *463*, 113-117.

Shima, N., Alcaraz, A., Liachko, I., Buske, T.R., Andrews, C.A., Munroe, R.J., Hartford, S.A., Tye, B.K., and Schimenti, J.C. (2007). A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. Nature genetics *39*, 93-98.

Smelkova, N.V., and Borowiec, J.A. (1997). Dimerization of simian virus 40 T-antigen hexamers activates T-antigen DNA helicase activity. Journal of virology *71*, 8766-8773.

Speck, C., and Stillman, B. (2007). Cdc6 ATPase activity regulates ORC x Cdc6 stability and the selection of specific DNA sequences as origins of DNA replication. The Journal of biological chemistry *282*, 11705-11714.

Speck, C., Chen, Z., Li, H., and Stillman, B. (2005). ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. Nature structural & molecular biology *12*, 965-971.

Stillman, B. (2005). Origin recognition and the chromosome cycle. FEBS letters *579*, 877-884.

Stinchcomb, D.T., Struhl, K., and Davis, R.W. (1979). Isolation and characterisation of a yeast chromosomal replicator. Nature *282*, 39-43.

Stoeber, K., Mills, A.D., Kubota, Y., Krude, T., Romanowski, P., Marheineke, K., Laskey, R.A., and Williams, G.H. (1998). Cdc6 protein causes premature entry into S phase in a mammalian cell-free system. The EMBO journal *17*, 7219-7229.

Su, T.T., Feger, G., and O'Farrell, P.H. (1996). Drosophila MCM protein complexes. Molecular biology of the cell *7*, 319-329.

Sutera, V.A., Jr., Han, E.S., Rajman, L.A., and Lovett, S.T. (1999). Mutational analysis of the RecJ exonuclease of Escherichia coli: identification of phosphoesterase motifs. Journal of bacteriology *181*, 6098-6102.

Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A., and Araki, H. (2003). GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. Genes & development *17*, 1153-1165.

Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y., and Araki, H. (2007). CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. Nature *445*, 328-332.

Tanaka, T., Knapp, D., and Nasmyth, K. (1997). Loading of an Mcm protein onto DNA replication origins is regulated by Cdc6p and CDKs. Cell *90*, 649-660.

Tercero, J.A., Labib, K., and Diffley, J.F. (2000). DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p. The EMBO journal *19*, 2082-2093.

Traut, T.W. (1994). Physiological concentrations of purines and pyrimidines. Molecular and cellular biochemistry *140*, 1-22.

Tye, B.K., and Sawyer, S. (2000). The hexameric eukaryotic MCM helicase: building symmetry from nonidentical parts. The Journal of biological chemistry *275*, 34833-34836.

Wakamatsu, T., Kitamura, Y., Kotera, Y., Nakagawa, N., Kuramitsu, S., and Masui, R. (2010). Structure of RecJ exonuclease defines its specificity for single-stranded DNA. The Journal of biological chemistry *285*, 9762-9769.

Wang, B., Feng, L., Hu, Y., Huang, S.H., Reynolds, C.P., Wu, L., and Jong, A.Y. (1999). The essential role of Saccharomyces cerevisiae CDC6 nucleotide-binding site in cell growth, DNA synthesis, and Orc1 association. The Journal of biological chemistry *274*, 8291-8298.

Weisshart, K., Taneja, P., Jenne, A., Herbig, U., Simmons, D.T., and Fanning, E. (1999). Two regions of simian virus 40 T antigen determine cooperativity of double-hexamer assembly on the viral origin of DNA replication and promote hexamer interactions during bidirectional origin DNA unwinding. Journal of virology *73*, 2201-2211.

Wessel, R., Schweizer, J., and Stahl, H. (1992). 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 *66*, 804-815.

Yamagata, A., Kakuta, Y., Masui, R., and Fukuyama, K. (2002). The crystal structure of exonuclease RecJ bound to Mn2+ ion suggests how its characteristic motifs are involved in exonuclease activity. Proceedings of the National Academy of Sciences of the United States of America *99*, 5908-5912.

Yardimci, H., Loveland, A.B., Habuchi, S., van Oijen, A.M., and Walter, J.C. (2010). Uncoupling of sister replisomes during eukaryotic DNA replication. Molecular cell *40*, 834-840.

Yardimci, H., Wang, X., Loveland, A.B., Tappin, I., Rudner, D.Z., Hurwitz, J., van Oijen, A.M., and Walter, J.C. (2012). Bypass of a protein barrier by a replicative DNA helicase. Nature *492*, 205-209.

You, Z., and Masai, H. (2005). DNA binding and helicase actions of mouse MCM4/6/7 helicase. Nucleic acids research *33*, 3033-3047.

You, Z., Ishimi, Y., Mizuno, T., Sugasawa, K., Hanaoka, F., and Masai, H. (2003). Thymine-rich single-stranded DNA activates Mcm4/6/7 helicase on Y-fork and bubble-like substrates. The EMBO journal *22*, 6148-6160.

You, Z., Komamura, Y., and Ishimi, Y. (1999). Biochemical analysis of the intrinsic Mcm4-Mcm6-mcm7 DNA helicase activity. Molecular and cellular biology *19*, 8003-8015.

Yu, Z., Feng, D., and Liang, C. (2004). Pairwise interactions of the six human MCM protein subunits. Journal of molecular biology *340*, 1197-1206.

Zegerman, P., and Diffley, J.F. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. Nature *445*, 281-285.

Zou, L., Mitchell, J., and Stillman, B. (1997). CDC45, a novel yeast gene that functions with the origin recognition complex and Mcm proteins in initiation of DNA replication. Molecular and cellular biology *17*, 553-563.

Publications

Ilves, I., **Petojevic, T.**, Pesavento, J.J., and Botchan, M.R. (2010). Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. Molecular cell *37*, 247-258.

Costa, A., Ilves, I., Tamberg, N., **Petojevic, T.**, Nogales, E., Botchan, M.R., and Berger, J.M. (2011). The structural basis for MCM2-7 helicase activation by GINS and Cdc45. Nature structural & molecular biology *18*, 471-477.

Manuscripts in preparation:

Pesavento, J.J.*, **Petojevic, T**.*, Costa, A., Liang, J., Wang, Z., Renault, L., Berger, J.M., and Michael R. Botchan (2013). Cdc45 guards the gate in the CMG helicase assuring leading strand engagement. *Manuscript in preparation.*

Petojevic, T., Goke, A., Tamberg N., and Michael R. Botchan (2013). Mutations of the CMG internal channel β-hairpins reveal an ordered path for translocation on the leading DNA strand. *Manuscript in preparation.*

* equal contributions