Cryo-EM Studies of Licensing and Activation of a Eukaryotic Origin of DNA Replication

Ferdos Abid Ali

Imperial College London
and
The Francis Crick Institute
PhD Supervisor: Dr. Alessandro Costa

A thesis submitted for the degree of
Doctor of Philosophy
Imperial College London
September 2018

Declaration

I Ferdos Abid Ali confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researches must make clear to others the licence terms of this work.

Abstract

Eukaryotic DNA replication is initiated during the G1 phase of the cell cycle, where the core of the replicative helicase, Minichromosome Maintenance complex (MCM), is loaded onto origins as an inactive double hexamer (origin licensing). During S-phase, the two MCMs are converted into two active Cdc45-MCM-GINS (CMG) holo-helicases that unwind DNA bi-directionally. This step, termed 'origin firing' is governed by two cell cycle regulated kinases, Dbf4-dependent kinase (DDK) and cyclin dependent kinase (CDK), which in turn mediate the recruitment of 'firing factors' (Cdc45, Sld3/7, Sld2, Dpb11, GINS, Polε and Mcm10). The active CMG helicase acts as the leading edge of the advancing replisome, a multi-subunit macromolecular machine that includes a primosome (Polymerase α-primase) as well as leading strand polymerase ε and lagging strand polymerase δ. CMG and Polε interact directly and GINS-Polε are simultaneously recruited during helicase assembly. Overall, several reconfigurations of the MCM ring have to occur to enable replication fork establishment, including stable recruitment of the helicase co-activators Cdc45 and GINS, double hexamer separation, DNA melting and activation of ATP hydrolysis. However, the molecular mechanism of CMG activation and replisome progression is poorly understood.

In this study, I used electron microscopy and biochemical techniques to visualise several MCM intermediates and inform the molecular basis for helicase activation. I showed that the inactive double hexamer forms a stable complex of two tilted rings and encircles bent, double-stranded DNA. Phosphorylation of N-terminal MCM tails by DDK creates a symmetrical docking platform, which may be used for downstream recruitment Cdc45 via the Sld3/7 chaperone. Double hexamer separation occurs during the GINS recruitment stage of CMG assembly prior to helicase activation, and this step is concomitant with melting of the DNA, through an MCM-mediated duplex stretching mechanism that untwists the double helix. Cdc45 and GINS bind to the side of the MCM ring at the N-terminal tier, stabilising the key Mcm2-5-3 sites that are essential for helicase activity. The active CMG helicase motor translocates on single-stranded DNA, excluding the complementary strand from the main channel. Moreover, the CMG unwinds DNA with its N-terminal MCM tier ahead of the motor

domain, meaning that the two helicases cross paths upon activation to ensure that all origin DNA is unwound for genome stability maintenance. A cryo-EM structure of the CMG-Polε complex reveals that the essential, non-catalytic Polε subunits bind to the Mcm2-5-3 ATPase domains, stabilizing this region for productive fork engagement by the CMG. Polε is predominantly bound to the translocating CMG and bridges the GINS and MCM components, providing a rationale for the direct role of the leading strand polymerase in CMG assembly. Thus, regulated post-translational modifications and replisome partner protein interactions promote distinct helicase structural transitions. This flexibility is integral for MCM to fulfil its varied roles during origin licensing, helicase assembly, DNA melting and replication fork progression.

Acknowledgement

Firstly, I would like to thank my supervisor Dr. Alessandro Costa for the exciting scientific projects, guidance and constant support throughout these four years. I am very grateful for the considerable time Alessandro spent patiently teaching me electron microscopy skills and how to navigate challenging experiments. The lessons learnt are invaluable and I hope to carry them with me throughout my career.

I would also like to thank my thesis committee members Dr. John Diffley, Prof. Xiaodong Zhang and Dr. Hasan Yardimci for their helpful advice and for always making sure the projects are on the right track. Thank you to Boehringer Ingelheim Fonds for funding my PhD studentship, conferences and providing extensive training.

A big thank you to all Costa lab members for your experimental advice and friendship. In particular, thanks to the awe-inspiring Dr. Julian Gannon, Dr. Adelina Davies, Dr. Ludovic Renault for patiently teaching me in my early days. Thanks to Dr. Panchali Goswami and Dr. Jin Chuan Zhou for all your help and providing strength throughout my PhD. I wish to thank Dr. Julia Locke, not only for her extensive contributions to my cryo-EM projects and ensuring the lab runs smoothly, but also for instilling confidence in me and being highly supportive. Thanks to Drs. Marcus Wilson and Thomas Miller for always being willing to help and suggesting useful ideas for experiments. Thanks to Patrik Eickhoff for being a great lab neighbour and bright scientist; tack for allowing me to reminisce about life in Sweden and always making sure I am caffeinated enough. Last but definitely not least, thanks to Dr. Paolo Swuec for being my EM/life coach, cooking me the best food ever, laughing with me and not at me but most of all thanks for being hugely influential.

Massive thank you to Dr. Max Douglas for a highly enjoyable collaboration and the creative solutions provided to scientific problems. Thank you to Dr. John Diffley for initiating a long-standing collaboration and for offering a unique perspective, which has always helped drive the projects forward. Thanks to Drs. Hailey Gahlon and David Rueda at Imperial College London for performing the single molecule experiments and Dr. Abhay Kotecha (Oxford) for collecting cryo-EM data. I wish to

thank Drs. Andrea Nans and Peter Rosenthal for setting up cryo-EM data collection at the Crick. Thanks to Dr. Aga Janksa and Anne Early for creating the Polε strains used in this study. Thank you to Drs. Valerie Pye and Andrew Purkiss for their modelling expertise.

A special thanks to all the STP scientists who have been instrumental in creating a productive work environment, including Dr. Raffa Carzaniga and Dr. Lucy Collinson (EM STP), Dr. Andrea Nans, Dr. Philip Walker and Dr. Andrew Purkiss (structural biology STP), Namita Patel, Alireza Alidoust and Damini Patel (Fermentation STP) and Paul Willis (Cell production STP).

398 thanks to my friends Aga, Corella, Dan, Diana, Jake, Julienne, Khalid, Paolo and Sam (and partners!). We truly shared something special at Clare Hall and you made me feel like I had a second family away from home. Thanks to Oliver for all the laughs and for coping with my endless troubleshooting questions. To Andrea, thank you for always cheering me up and for being an attentive friend – I will miss you.

To my four brothers, thank you for your unwavering support and faith in me. Thanks also to my extended family for always giving me something to smile about and being so understanding in my absence. Mama and Baba, none of this would have been possible without your belief, sacrifice and generosity; I am eternally appreciative of the work you put in to make this happen.

Finally, I thank my fiancé Chris for his selfless, patient and active encouragement of all my endeavours. Thanks for bearing with my particle picking and structure spinning. I am very thankful to have shared my PhD experience with you.

Table of Contents

Abstract **4**
Acknowledgement **6**
Table of Contents **8**
Table of figures **11**
List of tables **15**
Abbreviations **16**
Chapter 1.Introduction **18**
1.1 Overview **18**
1.1.1 Activation of helicase requires extensive MCM remodelling 20
1.2 Structure and function of replicative helicases **24**
1.2.1 Structural and functional conservation of replicative helicases 24
1.2.2 Functional and structural characterisation of Mcm2-7 26
1.2.3 The AAA+ tier harnesses energy from ATP hydrolysis into unwinding 27
1.2.4 ATPase insertions couple ATP hydrolysis with translocation 29
1.2.5 The N-terminal non-catalytic tier of the MCM ring 32
1.2.6 Structure of the Cdc45-Mcm2-7-GINS helicase 34
1.2.7 The structural basis for Mcm2-7 activation by Cdc45 and GINS 34
1.3 Origin licensing **37**
1.3.1 Origins of replication and ORC 38
1.3.2 Cdc6 42
1.3.3 Mcm2-7/Cdt1 43
1.3.4 Structures of MCM loading intermediates and mechanism of pre-RC formation 45
1.4 Origin firing **51**
1.4.1 DDK 51
1.4.2 Sld3/7 53
1.4.3 Cdc45 54
1.4.4 CDK and the pre-LC complex 56
1.4.5 Mcm10 63
1.4.6 Mechanism of origin firing 65
1.5 The CMG helicase at the replication fork **68**
1.5.1 Mechanistic models of helicase unwinding 69
1.5.2 The mechanism of helicase strand engagement 72
1.5.3 Polarity of the helicase at the replication fork 74
1.5.4 The CMG helicase forms the core of the replisome 75
1.6 Aims **79**
Chapter 2.Materials & Methods **80**
2.1 Reagents **80**
2.1.1 *S. cerevisiae* strains 80
2.1.2 Oligonucleotides 82
2.1.3 Purification and reconstitution buffers 83
2.1.4 Media for budding yeast cells 86
2.1.5 Media for insect cells 86
2.2 Protein expression **86**
2.2.1 Expression using the baculovirus-insect cell system 86

2.2.2 Expression of yeast strains 87
2.3 Protein purification 87
2.3.1 *D. melanogaster* CMG purification 87
2.3.2 *S. cerevisiae* CMG purification 88
2.3.3 Polymerase epsilon (wild-type, ∆cat and exo-) purification 89
2.3.4 Purification of replication factors for *in vitro* reconstitution 90
2.4 Reconstitution reactions 96
2.4.1 DmCMG-DNA-ATPγS reconstitution 96
2.4.2 ScCMG-Polε-DNA-ATPγS reconstitution 96
2.4.3 Helicase assays 97
2.4.4 Single molecule DmCMG-DNA experiments 97
2.4.5 *In vitro* reconstitution of MCM double hexamers 98
2.4.6 *In vitro* reconstitution of CMG assembly 99
2.4.7 *In vitro* replication assays 99
2.5 Electron microscopy sample preparation 100
2.5.1 Negative stain grid preparation 100
2.5.2 Positive stain grid preparation 101
2.5.3 Cryo-EM grid preparation 101
2.6 Electron microscopy image acquisition 102
2.6.1 Negative-stain data collection and cryo-grid screening 102
2.6.2 Cryo-EM data collection 102
2.7 Electron microscopy image processing 103
2.7.1 A general pipeline for negative-stain image processing 103
2.7.2 Cryo-EM image processing and analysis 104
Chapter 3.The Structure of the MCM double hexamer and insights into CMG Activation 111
3.1 Introduction 111
3.2 Structure of the MCM Double Hexamer 112
3.2.1 Image processing of the MCM double hexamer 113
3.2.2 Overall architecture of the MCM double hexamer 113
3.2.3 ATPase state of the MCM double hexamer 115
3.3 Structure of MCM double hexamer bound to DNA 123
3.3.1 Using signal subtraction to retrieve DNA bound particles 123
3.3.2 The double hexamer binds duplex DNA 128
3.3.3 Mapping the DNA-MCM contacts in the double hexamer 128
3.4 Effect of DDK phosphorylation on the double hexamer 130
3.5 The transition from double hexamer to active CMG 132
3.5.1 Testing the suitability of the origin firing assay for structural studies 133
3.5.2 Establishing the temporal order of helicase activation 134
3.6 Conclusions 136
Chapter 4.Architecture of the active CMG helicase 138
4.1 Introduction 138
4.1.1 Background 138
4.1.2 Aims 140
4.2 Results 141
4.2.1 Purification and negative stain EM analysis of CMG 141
4.2.2 Sub-nanometre structure of the CMG helicase 145

4.2.3 The active MCM complex in the CMG holo-helicase 151
4.2.4 The role of Cdc45 and GINS in the CMG complex 155
4.2.5 The AAA+ domain adopts relaxed and compact ATPase conformations 159
4.2.6 The CMG helicase translocates on single-stranded DNA 163
4.2.7 Locating the lagging strand template 166
4.2.8 Single-stranded DNA in the CMG is stabilised in an extended conformation 167
4.2.9 Comparison of MCM to CMG suggests a model origin DNA melting 171
4.2.10 The orientation of the CMG helicase at the replication fork 173
4.3 Conclusions 180
Chapter 5.Cryo-EM structure of the CMG helicase bound to Polε 182
5.1 Introduction 182
5.2 Results 184
5.2.1 Purification and EM analysis of recombinant yeast CMG 184
5.2.2 Purification of yeast polymerase epsilon variants 188
5.2.3 Optimisation and reconstitution of a CMGε complex 191
5.2.4 Sub-nanometre cryo-EM structure of CMGε on fork DNA 193
5.2.5 Structure of the essential, non-catalytic domain of Polε 198
5.2.6 DNA binding and nucleotide occupancy of CMGε 200
5.2.7 Analysis of the CMGε interaction network 203
5.2.8 A model for the role of C-Pol2 and Dpb2 in CMG formation 205
5.3 Conclusions 208
Chapter 6.Discussion 210
6.1 Introduction 210
6.2 Key conclusions 210
6.3 Origin licensing 212
6.3.1 Structure of the licensed origin 212
6.3.2 The eukaryotic MCM double hexamer binds duplex DNA 215
6.4 CMG assembly 217
6.4.1 DDK phosphorylation stabilises the N-terminal tails of Mcm4,6 to create a recruitment platform for origin firing factors 217
6.4.2 GINS recruitment finalises holo-helicase assembly and causes double hexamer separation 219
6.4.3 GINS and Polε are recruited together to the helicase 220
6.5 CMG activation and replication fork establishment 220
6.5.1 Mcm10 converts assembled CMG complexes to active holo-helicases 220
6.5.2 The role of Cdc45 and GINS in the active CMG helicase 221
6.5.3 The active MCM ring binds single-stranded DNA 223
6.5.4 MCM reconfigurations from origin licensing to replication fork establishment inform origin melting and DNA translocation 224
6.5.5 ATPase remodeling in the CMG 226
6.5.6 Polε stimulates productive fork engagement by CMG 227
6.5.7 Flexibility of N-Pol2 is important for polymerase function 227
6.5.8 Polarity of the CMG helicase at the replication fork 228
6.6 A provisional model for helicase assembly and activation 231

6.7 Future directions 233
Reference List 235

Table of figures

Figure 1.1: The Mcm2-7 helicase undergoes significant conformational changes during activation 22
Figure 1.2: Hexameric helicases of the SF3-6 superfamilies adopt a similar architecture 24
Figure 1.3: MCM domain architecture, monomer structure and active site formation 28
Figure 1.4: DNA binding elements in MCM 30
Figure 1.5: Mcm2-7 binds Cdc45 and GINS to shut the Mcm2-5 gate and enhance ATPase activity 35
Figure 1.6: Schematic overview of *S. cerevisiae* ARS1 origins 39
Figure 1.7: Cryo-EM structure of *S. cerevisiae* ORC bound to ACS and B1 element 41
Figure 1.8: Negative stain EM analysis of isolated Mcm2-7 and Mcm2-7/Cdt1 44
Figure 1.9: Cryo-EM structure of *S. cerevisiae* ORC-Cdc6-Cdt1-Mcm2-7 45
Figure 1.10: Origin licensing models 47
Figure 1.11: Cryo-EM structure of the DNA-free S. cerevisiae double hexamer at 3.8 Å resolution. 49
Figure 1.12: Domain organisation of human Cdc45 and structural comparison to RecJ and GAN 55
Figure 1.13: Crystal structure of human GINS 59
Figure 1.14: Domain architecture and structural characterisation of individual Polε domains 61
Figure 1.15: A model for step-wise assembly and activation of the eukaryotic replicative helicase 67
Figure 1.16: Models for ATP hydrolysis by hexameric helicases. 70
Figure 1.17: Co-crystal structures of hexameric helicases bound to single-stranded DNA or RNA. 71
Figure 1.18: Models for CMG DNA fork engagement 72
Figure 1.19: DNA binding polarity of hexameric helicases 75

Figure 1.20: Structural characterisation of a minimal replisome 76
Figure 3.1: Sample preparation and cryo-EM structure of the MCM double hexamer 114
Figure 3.2: Image processing workflow of the unmodified MCM double hexamer 118
Figure 3.3: Image processing workflow for the DDK phosphorylated MCM double hexamer 120
Figure 3.4: Cryo-EM of the unmodified MCM double hexamer 121
Figure 3.5: Cryo-EM of the DDK phosphorylated MCM double hexamer 122
Figure 3.6: Subunit comparison of unmodified and DDK phosphorylated MCM double hexamer 122
Figure 3.7: Signal subtraction of AAA+ motor allows visualisation of DNA bound particles 124
Figure 3.8: Cryo-EM of the unmodified and DDK phosphorylated DNA bound double hexamers 126
Figure 3.9: The MCM double hexamer bound to double stranded DNA. 127
Figure 3.10: Effect of DDK phosphorylation on the double hexamer 131
Figure 3.11: EM analysis of reconstituted origin firing and 3D reconstruction of active CMG 134
Figure 3.12: Structural characterisation of salt stable replicative helicase activation intermediates 135
Figure 4.1: Overview of DmCMG expression and purification strategies 143
Figure 4.2: Purification of DmCMG 144
Figure 4.3: Negative stain EM analysis of DmCMG 145
Figure 4.4: Sample preparation, data collection and initial image processing of CMG-DNA 146
Figure 4.5: Single-particle CMG-DNA-ATPγS 3D reconstruction and classification overview. 149
Figure 4.6: CMG-DNA-ATPγS Compact ATPase angular distribution and resolution estimation 150
Figure 4.7: CMG-DNA-ATPγS Relaxed ATPase angular distribution and resolution estimation. 151
Figure 4.8: Cryo-EM structure of DmCMG at 7.4 Å resolution 152
Figure 4.9: Transitions in the DNA interacting collar from MCM to CMG. 153

Figure 4.10: Domain architecture of Cdc45 and structural role in guarding the Mcm2-5 gate 156
Figure 4.11: Structure and interactions of GINS with the MCM and Cdc45 158
Figure 4.12: GINS binding may cause destabilisation of the MCM double hexamer 159
Figure 4.13: CMG bound to DNA and ATPγS exists in two main configurations.. 160
Figure 4.14: Remodelling of the CMG ATPase motor. 162
Figure 4.15: The CMG helicase binds to single-stranded DNA 165
Figure 4.16: Locating the lagging strand template in a CMG-DNA complex 167
Figure 4.17: Comparison of CMG-DNA-ATPγS with helicases co-crystallised with nucleic acid 169
Figure 4.18: DNA deformation by the CMG helicase..... 170
Figure 4.19: A model for DNA melting 172
Figure 4.20: The CMG translocates with its N-terminal tier ahead of the AAA+ engine 175
Figure 4.21: Trains only form in the presence of roadblocks and Mcm10 177
Figure 4.22: Cryo-ET of MCM trains confirms that the active CMG helicase translocates N-tier first..... 179
Figure 4.23: A model for CMG helicase activation and translocation. 180
Figure 5.1: Purification of recombinant *S. cerevisiae* CMG 185
Figure 5.2: EM analysis of yeast CMG and co-purification of endogenous Ctf4 .. 186
Figure 5.3: Purification of *S. cerevisiae* Polε-exo- 189
Figure 5.4: Negative-stain EM analysis of Polε exo- shows that the leading strand polymerase can switch between compact and extended conformations..... 190
Figure 5.5: Reconstitution and EM analysis of CMG-Polε variant complexes on fork DNA 191
Figure 5.6: Cryo-EM structure of CMGε at 5.0 Å resolution 194
Figure 5.7: Single-particle CMG-DNA-Polε-ATPγS 3D reconstruction and classification overview 196
Figure 5.8: CMG-DNA-Polε-ATPγS cryo-EM, angular distribution and resolution estimation 197
Figure 5.9: Structure of the non-catalytic lobe of Polε 199
Figure 5.10: DNA binding in CMGε..... 201

Figure 5.11: Polε mediated compaction of MCM ring and nucleotide occupancy in CMGε 202
Figure 5.12: Analysis of the CMGε interaction network 204
Figure 5.13: C-Pol2/Dpb2 are the minimal Polε subunits required for CMG assembly and replication 206
Figure 5.14: Proposed model for the concerted GINS and Polε incorporation onto MCM-Cdc45 207
Figure 6.1: Comparison of ATPase state and DNA binding modes of different MCM intermediates 215
Figure 6.2: DDK phosphorylation of Mcm4,6 creates a docking platform for Sld3/7-Cdc45 218
Figure 6.3: Polarity of CMG helicase suggests a model for origin activation 229
Figure 6.4: Revised polarity of the CMG suggests an efficient replisome architecture 230
Figure 6.5: Provisional model for helicase activation 232

List of tables

Table 2.1: Strains used in this study. 80
Table 2.2: Oligonucleotides used in this study 82
Table 2.3: Buffers used in this study 83
Table 2.4: Cryo-grid preparation parameters 101
Table 2.5: Movie acquisition parameters 102
Table 2.6: Cryo-EM image processing parameters 104
Table 2.7: Model refinement parameters 105

Abbreviations

3HM	3 helix motif
AAA+	ATPases associated with diverse cellular activities
Abf	ARS binding factor
ACL	allosteric communication loop
ACS	ARS consensus sequence
APC/C	anaphase promoting complex / cyclosome
ARS	autonomously replicating sequence
ATP	adenosine 5'-triphosphate
ATPγS	adenosine 5'-O-(3-thio)triphosphate
BAH	bromo-adjacent homology
Bp	base pairs
BRCT	BRCA1 C-terminus
CBP	calmodulin binding peptide
Cdc	cell division cycle
CDK	cyclin-dependent kinase
Cdt	Cdc10 dependent transcript
Clb	cyclin B
CMG	Cdc45-Mcm2-7-GINS
CMGε	Cdc45-Mcm2-7-GINS-polymerase ε
CTD	C-terminal domain
Ctf	chromosome transmission fidelity
Dbf	dumbbell former
DDK	Dbf4-dependent kinase
DNA	deoxyribonucleic acid
Dpb	DNA polymerase B possible subunit
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol tetraacetic acid
EM	electron microscopy
GAN	Gins associated nuclease
GINS	Go-Ichi-Ni-San (Sld5-Psf1-3)

H2i	Helix-2 insert
Kb	kilobase
L-tag	large T-antigen helicase
MCM	minichromosome maintenance
Mrc	mediator of the replication checkpoint
MSSB	MCM single-stranded DNA binding
Nt	nucleotide
NTD	N-terminal domain
NTP	nucleoside triphosphate
OB	oligosaccharide/oligonucleotide binding
OCCM	ORC-Cdc6-Cdt1-MCM
ORC	origin recognition complex
PBS	phosphate buffer saline
PCNA	proliferating cell nuclear antigen
Pol	polymerase
pre-IC	pre-initiation complex
pre-LC	pre-loading complex
pre-RC	pre-replicative complex
PS1	pre-sensor 1
Psf	partner of Sld5
RecQL	RecQ protein-like
RPA	replication protein A
RPC	replisome progression complex
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SF	superfamily
Sld	synthetically lethal with dpb11-1
SV40	Simian virus 40
TICRR	topBP1 interacting checkpoint and replication regulator
TopBP	topoisomerase II binding protein
Zf	Zinc-finger

Chapter 1. Introduction

1.1 Overview

Precise replication and transmission of genetic information is essential for propagation of life. A pre-requisite for genome duplication is the unwinding of the DNA double helix to provide the parental templates that replicative polymerases use for daughter strand synthesis. In all organisms, DNA strand separation is catalysed by a replicative helicase, a molecular machine that coordinates ATP binding and hydrolysis with translocation on nucleic acids. As well as unwinding the DNA, the replicative helicase forms the organising centre of the advancing replisome, a dynamic multi-protein complex that comprises primases, polymerases, processivity and scaffolding factors (Burgers and Kunkel, 2017). Despite conservation of the basic framework of genome duplication, replisome components have evolved twice independently, with bacteria using distinct enzymes compared to archaea and eukaryotes (Leipe et al., 1999, Yao and O'Donnell, 2016).

In eukaryotes, the assembly of the replisome and completion of DNA replication is tightly coordinated with progression through the cell cycle (Siddiqui et al., 2013). Genome duplication mainly occurs in synthetic (S)-phase ahead of chromosome segregation and cell division in mitosis (M)-phase. DNA replication begins at specific genomic loci called origins (Robinson and Bell, 2005). Whereas bacteria mostly initiate genome duplication from a single start site, eukaryotes have multiple origins for their larger genomes to ensure that DNA replication is completed within physiological time constraints (Parker et al., 2017). For example, *S. cerevisiae* comprise approximately 400 origins for their 12 Mbp genome and human genomes contain 30,000-50,000 origins within 3 Gbp of total DNA (Mechali et al., 2013).

A strict requirement in eukaryotes is that genome duplication only occurs once within the S-phase of the cell cycle as re-replication can lead to gene amplification, genomic instability and tumorigenesis (Arias and Walter, 2007). Thus, DNA replication has to occur on an extraordinary scale and rate while maintaining chromosome integrity. This dictates a requirement for high levels of regulation, which is facilitated principally

through a conserved two-step DNA replication initiation mechanism in eukaryotes (Blow and Laskey, 1988, Diffley et al., 1994, Remus and Diffley, 2009).

In the first step, an initiator recruits the replicative helicase to origins during late mitosis and early G1 phase of the cell cycle (Bleichert et al., 2017, Costa et al., 2013). In eukaryotes, the core of the replicative helicase is the ring-shaped Minichromosome maintenance (MCM) complex, formed of six related subunits (Mcm2-7) that belong to the AAA+ superfamily (ATPases associated with diverse cellular activities) (Bochman and Schwacha, 2009, Forsburg, 2004, Dutta and Bell, 1997). Initially, MCM is loaded onto origins as an inactive double hexamer through the coordinated action of the initiator, termed the origin recognition complex (ORC), and two other proteins, Cdc6 and Cdt1 (Remus et al., 2009, Evrin et al., 2009). This step is called 'origin licensing' or pre-replicative complex (pre-RC) formation (Diffley et al., 1994, Blow and Laskey, 1988, Bell and Stillman, 1992).

Whereas assembly of pre-RC during G1 is governed by low cyclin-dependent kinase (CDK) activity and elevated levels of the E3 ubiquitin ligase anaphase promoting complex / cyclosome (APC/C), Mcm2-7 motor activation upon the switch into S-phase depends on high CDK levels and low APC/C activity (Siddiqui et al., 2013). A second kinase, Dbf4-dependent kinase (DDK) works in concert with CDK to recruit numerous 'firing factors' to licensed origins (Labib, 2010). This results in the assembly of two single, active 11-subunit Cdc45-Mcm2-7-GINS (CMG) holo-helicase complexes that unwind DNA (Moyer et al., 2006, Ilves et al., 2010) and this step is called 'origin firing' (Yeeles et al., 2015). The high activity of CDK during S-phase is not only required for replicative helicase assembly but also safeguards against new origin licensing events during this cell cycle phase to prevent re-replication (Siddiqui et al., 2013).

Unwinding by the CMG provides the template strands used by polymerase α (Polα) -primase to prime the DNA templates prior to extension by DNA polymerase ε (Polε) and polymerase δ (Polδ) on the leading and lagging strand, respectively (Frick and Richardson, 2001, Burgers and Kunkel, 2017). During fork progression, CMG organises the replisome to maintain coupling of unwinding and synthesis functions (Sun et al., 2015, Gambus et al., 2006, Gambus et al., 2009). In this context, CMG

directly binds Polε and GINS-Polε are recruited together during helicase assembly (Muramatsu et al., 2010, Sengupta et al., 2013). Additionally, CMG recruits scaffolding factors including the homo-trimeric molecular bridge Ctf4, which enables helicase binding to Polα-primase (Gambus et al., 2009, Simon et al., 2014). Within the replisome, a large number of proteins also ensure efficient replication, including the fork stabilization factors Mrc1 and Csm3/Tof1, the single-stranded DNA binding protein RPA, processivity factors such as PCNA and the histone chaperone FACT (Gambus et al., 2006, Yeeles et al., 2015, Kurat et al., 2017). Coordinated association of the replisome in this way is key to both enhancing CMG activity and preventing the accumulation of fragile single-stranded DNA segments to maintain genome stability (Gambus et al., 2009, Yeeles et al., 2017).

1.1.1 Activation of helicase requires extensive MCM remodelling

While the identity of factors required for helicase assembly is established, how Mcm2-7 gets activated at the G1/S transition remains mechanistically poorly understood (Deegan and Diffley, 2016, Yeeles et al., 2015). How the active eukaryotic replicative helicase establishes replication fork progression is similarly unclear. Multiple studies of the double hexamer and the CMG indicate that the inactive and active helicases are distinct in composition, conformation and DNA binding mode (Remus et al., 2009, Evrin et al., 2009, Costa et al., 2011, Fu et al., 2011, Yardimci et al., 2010).

The biochemical reconstitution of the pre-RC and electron microscopy (EM) imaging of this complex has informed the architecture of the inactive helicase (Evrin et al., 2009, Remus et al., 2009). The double hexamer is formed of two Mcm2-7 rings connected through their N-termini, while the AAA+ domains face outwards (Remus et al., 2009, Li et al., 2015). Rotary shadowing EM experiments indicate that inactive Mcm2-7 complexes assembled onto circular templates are seemingly bound to relaxed duplex DNA, as opposed to the supercoiled DNA that would be expected upon single-stranded DNA formation (Remus et al., 2009).

In contrast, the active CMG is believed to function as a single holo-helicase that binds single-stranded DNA at the replication fork. This is supported by two main

biochemical studies in *Xenopus* egg extracts (Yardimci et al., 2010, Fu et al., 2011). In the first study, Yardimci *et al.* used single-molecule techniques to discern whether the two sister replisomes remain together or uncouple for fork progression (Yardimci et al., 2010). A coupled model predicts DNA pumping towards the joined replication machineries and that newly replicated DNA is displaced laterally in between the interface of the two replisomes. Therefore, in this model, if two DNA ends are fixed in a glass slide and genome duplication is initiated, then tension would accumulate in the unreplicated regions of the nucleic acid, rendering DNA replication inefficient. However, upon addition of *Xenopus* egg extracts to the double-tethered DNA, extensive DNA replication was observed by one pair of oppositely translocating replisomes, indicating that the double hexamer separates into two single bi-directionally unwinding CMG helicases upon activation. In support of this notion, active recombinant *Drosophila* and *S. cerevisiae* CMG form stoichiometric 1:1:1 Cdc45-Mcm2-7-GINS structures (Costa et al., 2011, Sun et al., 2015, Zhou et al., 2017, Costa et al., 2014).

In a separate study, Fu *et al.* examined the mechanism of replisome translocation on plasmid templates containing site-specific roadblocks to differentiate between a single or double-stranded DNA helicase binding mode (Fu et al., 2011). Roadblocks were either placed on the leading strand template (3' end) or the lagging strand template (5' end). If the active helicase is a double-stranded DNA translocase, then replisome progression should be impeded by blocks on either strand. However, the lagging strand template blocks were bypassed and only leading strand template blocks completely prevented fork progression, providing evidence that CMG tracks on single-stranded DNA.

Thus, if these models are correct, the eukaryotic replicative helicase has to undergo a number of structural and topological reconfigurations to remodel the inactive Mcm2-7 double hexamers from encircling duplex DNA into two active CMG helicases at the replication fork (**Figure 1.1**).

First, origin melting must take place to provide the early DNA strand separation that eventually matures into a replication fork (Gai et al., 2010, Bleichert et al., 2017). In bacteria, this event is well established; the bacterial DnaA initiator binds, stretches

and melts DNA in an ATP dependent manner prior to loading of the DnaB helicase onto single-stranded DNA (Duderstadt et al., 2011, Bleichert et al., 2017). In eukaryotes, the mechanism of origin melting is unknown but is thought to be mediated by the MCM complex itself rather than ORC (Li et al., 2015, Heller et al., 2011, Yeeles et al., 2015). This notion is evidenced by a recent cryo-EM structure of ORC bound to origin DNA showing no duplex melting by the initiator (Li et al., 2018), and two biochemical studies of DNA replication *in vitro*, which show that assembly and activation of the replicative helicase is necessary for initial unwinding (Yeeles et al., 2015, Heller et al., 2011). Notably, a cryo-EM structure of the inactive DNA-free Mcm2-7 double hexamer suggests that melting could occur upon origin licensing. The double hexamer shows a constricted central channel due to an offset stacking of the two rings that is predicted to deform the DNA (discussed in *section 1.3.2.3*) (Li et al., 2015). Whether origin melting occurs at the pre-RC or later on during helicase activation remains to be demonstrated.

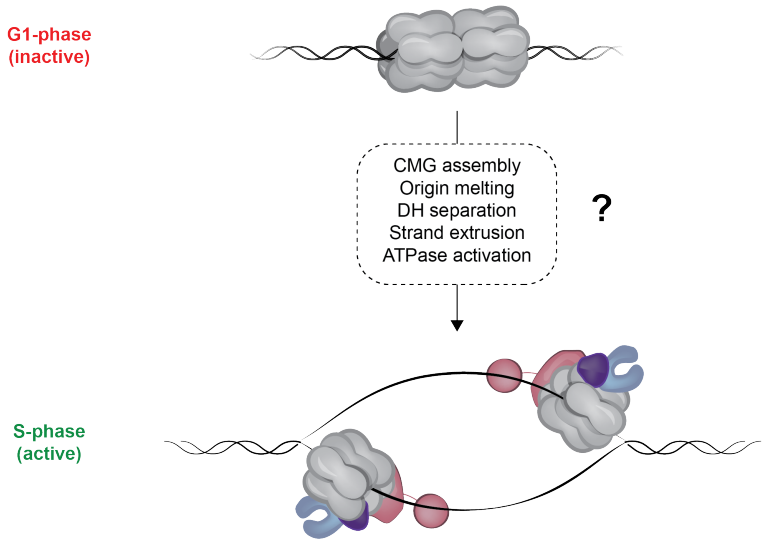


Figure 1.1: The Mcm2-7 helicase undergoes significant conformational changes during activation
The molecular basis for the predicted helicase activating steps (highlighted in the box) is largely unknown. Blue cartoon is Cdc45, pink is Polε and purple is GINS. Figure inspired by (Deegan and Diffley, 2016)

Second, Cdc45 and GINS have to associate with the Mcm2-7 to form a fully assembled holo-helicase (Yeeles et al., 2015, Costa et al., 2011, Ilves et al., 2010, Moyer et al., 2006). As will be shown in *section 1.2.6* we now have an understanding of the overall architecture of the CMG but it is not entirely clear how the Cdc45 and

GINS cofactors get recruited to the helicase or how they remodel and stabilise Mcm2-7 for activation.

Third, during or after CMG assembly, the helicases have to separate from their double hexamer state, poised for bi-directional fork establishment. A loosely tethered CMG double helicase intermediate has been previously observed by EM studies (Costa et al., 2014). Although the physiological relevance of this structure is unclear, it suggests that CMG assembly can occur prior to helicase dissociation. A key candidate for splitting the two rings is the origin firing factor Mcm10, which has recently been shown to bind the double hexamer and is strictly required for helicase activation (Yeeles et al., 2015, Douglas and Diffley, 2016, Quan et al., 2015).

Fourth, each active Mcm2-7 ring is presumed to open and close around the leading strand template while partitioning the lagging strand template from the main channel (Fu et al., 2011). A gate at the Mcm2-5 site has been identified that is used for initial loading of the Mcm2-7 helicase during pre-RC assembly and may also constitute the escape route for the excluded strand (Bochman and Schwacha, 2007, Samel et al., 2014, Costa et al., 2011).

Finally, active CMG must couple strand separation activity with replication functions within the replisome. In the case of the latter, Polε has been shown to directly bind and greatly stimulate CMG *in vitro* (Kang et al., 2012, Langston et al., 2014) and is also essential for GINS incorporation into the helicase in the first place (Muramatsu et al., 2010, Sengupta et al., 2013). Therefore, in addition to strand synthesis, the leading strand polymerase plays integral roles in helicase activation although the mechanistic basis for Polε mediated holo-helicase stabilisation is unclear.

The recent reconstitution of DNA replication *in vitro* has revealed the identity of the origin firing factors that are essential and sufficient to activate an origin (Yeeles et al., 2015). However, very little is known about the structural changes that accompany helicase activation and the spatiotemporal order of the predicted MCM conformational changes (**Figure 1.1**). With my thesis work, I aim to address key mechanistic questions regarding eukaryotic helicase activation using a primarily structural approach (aims outlined *in section 1.6*).

In this introduction, I will review what is presently known about Mcm2-7 loading, holo-helicase assembly and unwinding. Initially, I will give a functional and structural overview of helicases, with particular emphasis on the archaeal and eukaryotic MCM complex. Thereafter, I will introduce current paradigms of origin licensing and origin firing in the *S. cerevisiae* model system, where these events are best characterised. Finally, I will provide a synopsis of helicase mechanisms of ATP binding and hydrolysis, DNA translocation and replisome organisation.

1.2 Structure and function of replicative helicases

1.2.1 Structural and functional conservation of replicative helicases

Helicases are molecular motors that coordinate the protein reconfigurations caused by cycles of nucleotide binding and hydrolysis with translocation along nucleic acids (Singleton et al., 2007). Based on sequence homology and mechanistic differences, helicases are divided into six superfamilies (SF1-6). Whereas SF1 and SF2 helicases are mostly monomers involved in recombination, repair and translation (Fairman-Williams et al., 2010), the SF3-6 groups include hexameric replicative helicases. In this context, bacterial helicases belong to SF4 and archaeal/eukaryotic-based helicases belong to SF3,6 (**Figure 1.2**) (Singleton et al., 2007).

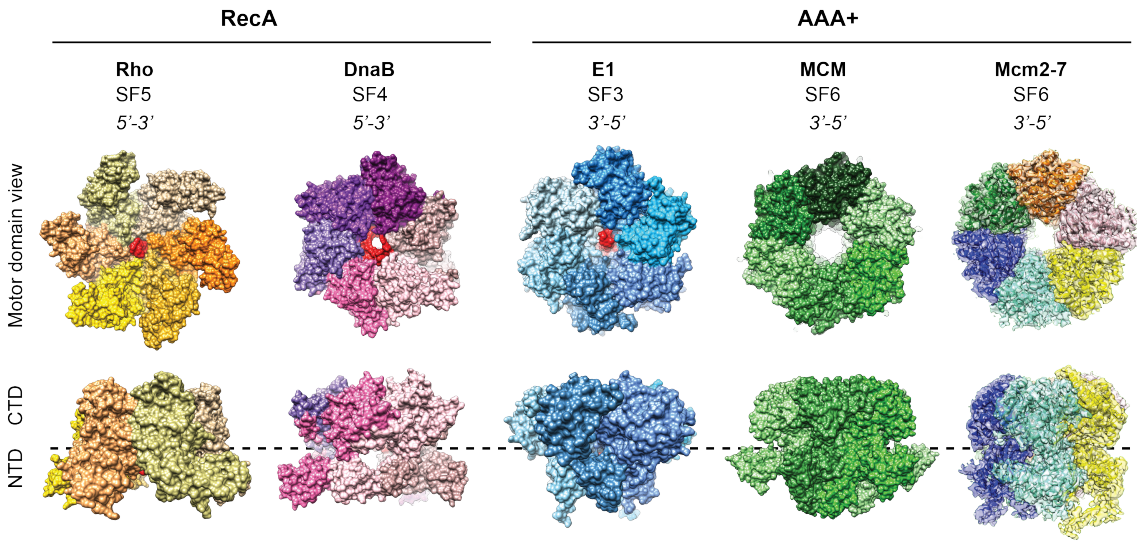


Figure 1.2: Hexameric helicases of the SF3-6 superfamilies adopt a similar architecture
Surface depiction of motor domains and side views of co-crystal structures of Rho (PDB: 3ICE), DnaB (PDB: 4ESV), E1 (PDB: 2GXA). The archaeal MCM crystal structure in green is from a hybrid *Sulfolobus solfataricus* NTD and *Pyrococcus furiosus* AAA+ hybrid construct (PDB: 4R7Y). Mcm2-7 ring is extracted from the double hexamer and shown with docked coordinates (EMD: 6338, PDB: 3JA8). Figure design adapted from (Trakselis, 2016).

The SF3-6 superfamilies also comprise helicases that are not involved in DNA replication but have provided extensive mechanistic insight into how replicative helicases function through shared architectures. This is largely exemplified by the Rho transcription terminator (SF5) which translocates on RNA (Thomsen and Berger, 2009) and will be referred to for comparison in this thesis.

In most cases, helicases form homo-hexameric assemblies, including the papilloma virus E1 helicase (SF3) (Enemark and Joshua-Tor, 2006), SV40 Large T-antigen (L-tag) (SF3) (Gai et al., 2004), the bacteriophage T7 gp4 (SF4) (Singleton et al., 2000), bacterial DnaB (SF4) (Itsathitphaisarn et al., 2012), Rho (SF5) (Thomsen and Berger, 2009) and archaeal MCM replicative helicase (SF6) (Miller et al., 2014). In contrast, the eukaryotic Mcm2-7 helicase (SF6) is a hetero-hexamer (**Figure 1.2**) (discussed in more detail in *section 1.2.2*).

Helicases either contain ATP sites based on the RecA-type motif or a AAA+ ATPase core, which together belong to a broad additional strand catalytic glutamate (ASCE) P-loop family of ATPases (Lyubimov et al., 2011). In both RecA-type and AAA+ helicases, the active site is situated at the interface of two adjacent subunits with ATPase hydrolysis motifs contributed by both protomers. Both classes of enzymes contain the *cis* Walker A and B sequence motifs, which coordinate binding to the nucleotide phosphate and an essential magnesium ion co-factor; and a *trans* acting arginine finger, which binds and hydrolyses ATP (Iyer et al., 2004).

Despite these similarities, RecA-type and AAA+ classes of helicases have evolved independently and exhibit significant divergence (Leipe et al., 1999). This is reflected in (*i*) differences in orientation of the ASCE folds within the respective RecA-type and AAA+ domain (Lyubimov et al., 2011) (*ii*) RecA-type helicase translocation with a 5'-3' polarity (bacterial helicases SF4,5) compared to the opposite 3'-5' directionality of AAA+ helicases (archaeal, eukaryotic-based helicases SF3,6) (Enemark and Joshua-Tor, 2008) and (*iii*) distinct locations of DNA binding elements within the ATP folds (Lyubimov et al., 2011).

Nevertheless, bacterial and eukaryotic helicases have converged on a similar overall architecture and function. First, each helicase subunit approximates a dumbbell

shape with an N-terminal domain (NTD) and C-terminal domain (CTD). Oligomerisation of each subunit gives rise to a two-tiered ring with a central channel, which topologically encircles single-stranded nucleic acids (first three structures in **Figure 1.2**) (Patel and Picha, 2000, Trakselis, 2016). This toroidal architecture is thought to increase the processivity of the enzyme but means that the helicase has to rely on DNA-loading co-factors to initially assemble onto origins (Bleichert et al., 2017, Wigley, 2009, Trakselis, 2016) (discussed further in *section 1.3*). Second, the ATPase activity resides at the interfaces of CTD domains, whereas the NTD domains are mostly non-catalytic. Third, pore loops project towards the central channel for interaction with nucleic acids during unwinding (Enemark and Joshua-Tor, 2008). These pore loops coordinate DNA binding with ATP hydrolysis as evidenced by co-crystal structures of DnaB, E1, and Rho bound to single-stranded nucleic acids (Enemark and Joshua-Tor, 2006, Itsathitphaisarn et al., 2012, Thomsen and Berger, 2009). However, the molecular basis by which the eukaryotic replicative helicase unwinds DNA is unknown. Therefore, analysis and comparison of the DnaB, E1 and Rho structures may provide important insights into how Mcm2-7 functions at the replication fork. The models for Mcm2-7 coordination of ATP hydrolysis and DNA unwinding will be discussed in *section 1.5* after providing an overview of what is currently known about Mcm2-7 structure and function.

1.2.2 Functional and structural characterisation of Mcm2-7

In eukaryotes, the core of the replicative helicase is the Mcm2-7 toroidal hexamer comprising six distinct but homologous polypeptides (Bochman and Schwacha, 2009, Li et al., 2015). First identified in two *S. cerevisiae* genetic screens for mutations with impaired minichromosome maintenance or cell cycle progression (Moir et al., 1982, Maine et al., 1984) all Mcm2-7 subunits were subsequently shown to have orthologues in every eukaryotic genome annotated (Liu et al., 2009, Forsburg, 2004), underscoring the importance of the hetero-hexameric organisation in this domain of life. The eukaryotic MCM genes are thought to have been diversified into six orthologues from one single ancestral archaeal MCM gene (Kelman and Hurwitz, 2003, Kelman et al., 1999). Indeed, all archaea studied to date have at least one MCM member but unlike the eukaryotic counterpart form homo-hexameric helicase assemblies (Miller and Enemark, 2015, Barry and Bell, 2006).

After initial discovery, multiple lines of evidence suggested that Mcm2-7 constitutes the core of the eukaryotic replicative helicase (Bell and Labib, 2016). The first was *in vivo* chemical crosslinking experiments which led to the identification of MCM proteins as part of both pre-RCs and the replication fork (Aparicio et al., 1997). Second, tracking of Mcm4 in live yeast cells revealed the replication-specific nuclear localisation of this protein during G1 and S-phase but its nuclear exclusion from the end of S-phase until the end of M phase (Labib et al., 1999). Third, experiments *in vivo* using heat inducible degrons revealed that depletion of any of the MCM subunits after origin activation led to the irreversible block of replication fork progression (Labib et al., 2000). Finally, Mcm2-7 has *in vitro* helicase activity (Bochman and Schwacha, 2008, Ilves et al., 2010) and is greatly stimulated by two key activators, Cdc45 and GINS (Ilves et al., 2010), which co-purify as a the full CMG replicative helicase complex as evidenced by three seminal papers in 2006 (discussed further in *section 1.2.6*) (Moyer et al., 2006, Gambus et al., 2006, Pacek et al., 2006).

Mcm2-7 subunits associate in a defined order: Mcm5, Mcm3, Mcm7, Mcm4, Mcm6 and Mcm2 (Davey et al., 2003, Costa et al., 2011). Much of what we know about the function and domain architecture of the Mcm2-7 complex stems from early biochemical and structural studies of the archaeal homo-hexameric MCM ring (Fletcher et al., 2003, Froelich et al., 2014, Brewster et al., 2008, Miller et al., 2014) and the recent cryo-EM structure of the yeast double hexamer (Li et al., 2015). Each MCM subunit shares a similar domain organisation that is highly conserved across archaeal and eukaryotic species (**Figure 1.3a**): a AAA+ domain, which belongs to the sensor 2 insert clade (clade VII) of the superfamily (Erzberger and Berger, 2006), sandwiched between an NTD and winged-helix domain (WHD) (**Figure 1.3b**).

1.2.3 The AAA+ tier harnesses energy from ATP hydrolysis into unwinding

Previous studies in archaea have shown that the AAA+ tier of the MCM ring can unwind DNA without the N-terminal tier, although the latter is required for improved processivity and substrate selection (Barry et al., 2007). As described in *section 1.2.1*, the ATPase sites in MCM proteins are formed at the interface between two adjacent subunits. In this arrangement, a *cis* domain, comprising a sequence of loops

connecting β strands (called the P loop) contributes motifs that bind ATP (Walker A box) and position a water molecule (Walker B and sensor 1). A second *trans* domain, called the 'lid' contains an arginine finger and sensor 2 element that bind the ATP γ phosphate (**Figure 1.3c**) (Moreau et al., 2007, Erzberger and Berger, 2006, Iyer et al., 2004).

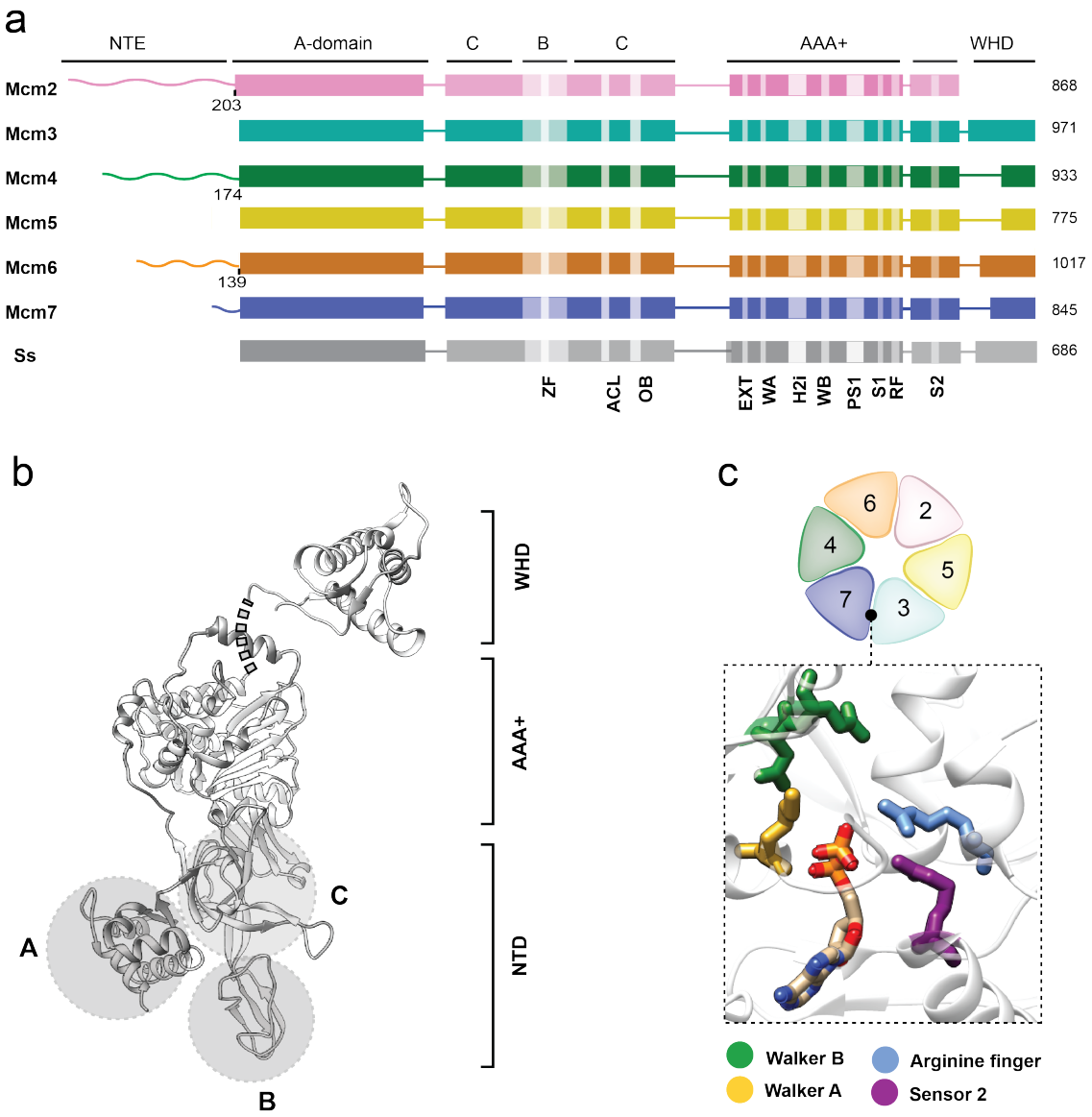


Figure 1.3: MCM domain architecture, monomer structure and active site formation
(a) Domain architecture of the *S. cerevisiae* Mcm2-7 subunits based on sequence alignment and structural characterisation of the double hexamer (Li et al., 2015) and comparison to the *Sulfolobus solfataricus* (SS) MCM domain organisation. **(b)** Crystal structure of near-full length *SS* MCM (PDB: 3F9V), highlighting the NTD subdomains with grey circles. Mcm6 winged helix domain structure (PDB: 2KLQ) tentatively modelled on top of AAA+. **(c)** The Mcm3-7 active site with modelled ADP from *S. cerevisiae* double hexamer structure (PDB: 3JA8).

Presence of bipartite active sites generalizes to other AAA+ machines and is deemed important for sensing and transmitting conformational changes around the ring to perform mechanical work (Erzberger and Berger, 2006). However, a unique pre-sensor 2 element exits in MCM that is absent in prototypical AAA+ ATPases and causes domain swapping of helices α5 and α6 onto the neighbouring subunit (Costa and Onesti, 2009). This enables the positioning of the sensor 2 arginine at the active site as a *trans* acting motif and may in turn act to stabilize inter AAA+ contacts (Erzberger and Berger, 2006, Moreau et al., 2007).

1.2.4 ATPase insertions couple ATP hydrolysis with translocation

As well as the active site elements, MCM AAA+ domains also comprise several β-hairpin insertions, including the helix-2 insert (H2i) and pre-sensor 1 (PS1), which protrude towards the channel (Li et al., 2015, Miller et al., 2014, McGeoch et al., 2005) (**Figure 1.4**). It is predicted that both PS1 and H2i transduce the energy harnessed from ATP hydrolysis to unwinding (Jenkinson and Chong, 2006). Consistent with this notion, PS1 and H2i are closely positioned (**Figure 1.4**) (Li et al., 2015, Brewster et al., 2008), which has led to the suggestion that they function as a single ensemble (Beattie and Bell, 2011).

1.2.4.1 Pre-sensor 1

The PS1 hairpin is a highly conserved motif involved in DNA binding and sensing nucleotide state as exemplified by crystal structures of the AAA+ L-Tag and E1 homo-hexamers (Gai et al., 2004, Enemark and Joshua-Tor, 2006). In MCM, PS1 inserts between helix 3 and strand 4 of the AAA+ domain and contains a lysine that is universally conserved from archaea to humans (Lys246 in archaea) (McGeoch et al., 2005). Substitution of this lysine with alanine abrogates helicase activity and causes an approximately 3-fold reduction of DNA binding in archaea (McGeoch et al., 2005). A similar detrimental effect was observed for equivalent mutations in *Drosophila* MCM of the CMG (Petojevic et al., 2015) while a yeast mutant of PS1 Mcm3 leads to a loss of viability (Lam et al., 2013). Collectively, these data underscore the importance of the PS1 element for helicase activity.

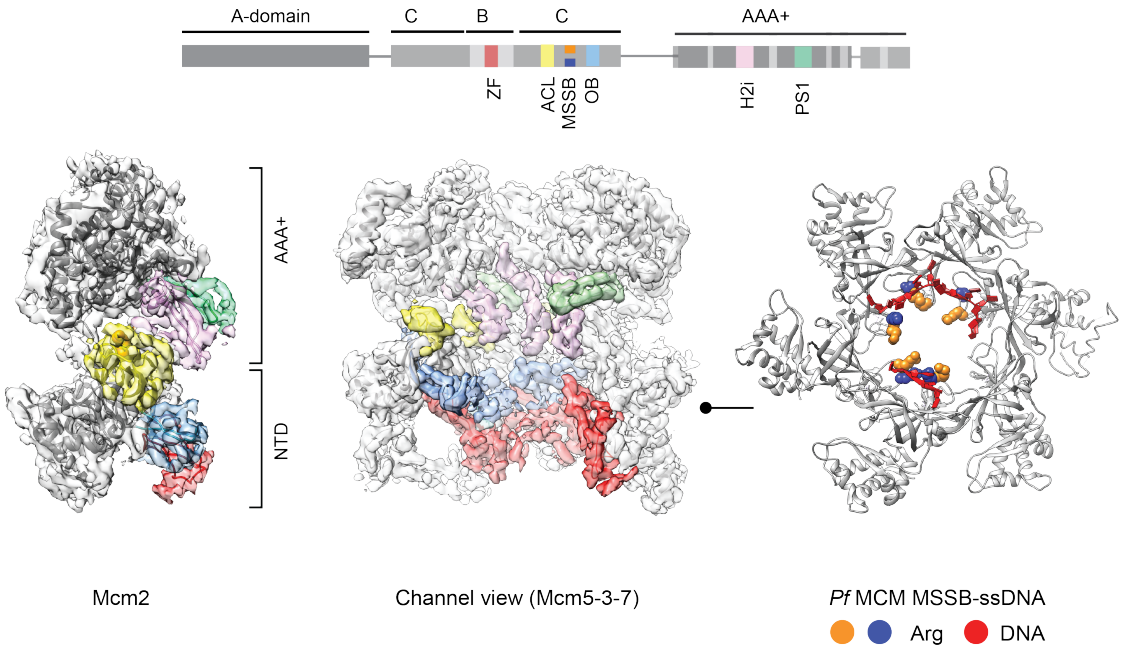


Figure 1.4: DNA binding elements in MCM
Each MCM subunit contains a set of conserved loops that project towards the channel. First two panels show Mcm2-7 subunits from *S.cerevisiae* double hexamer (PDB: 3ja8). Right panel shows *Pyrococcus furiosus* MCM NTD tier with single-stranded DNA captured by the conserved Arg124 and 186 residues of the MSSB (PDB:4pog).

In line with a predicted nucleotide sensing function by PS1, ATP binding by L-Tag results in a long range (17 Å) movement of the PS1 loop (Gai et al., 2004). As discussed in *section 1.5.1*, PS1 loops are also thought to be the main drivers of the E1 helicase 'escort' translocation mechanism (Enemark and Joshua-Tor, 2006). Whether MCM uses these PS1 loops for DNA binding in a similar fashion to E1 remains to be investigated.

Independent of DNA binding, PS1 has also been identified to bind the NTD *in trans* via an allosteric communication loop (ACL), which is thought to reconfigure the two MCM tiers upon ATP binding to promote DNA melting and unwinding (Barry et al., 2009). In this role, ACL modulates ATPase activity rather than directly affecting DNA binding (Sakakibara et al., 2008).

1.2.4.2 Helix-2 insert

The H2i is an α-β-α insertion that projects towards the channel further than PS1 and acts as the division point between the NTD and AAA+ tier (Miller et al., 2014, Li et al., 2015, Brewster et al., 2008)(**Figure 1.4**). In the presence of nucleotide and DNA,

a key tryptophan residue, which is conserved in archaea, repositions within the channel as evidenced by steady state fluorescence measurements (Jenkinson and Chong, 2006) akin to what has been observed for PS1 (Barry et al., 2009). Although overall sequence conservation is low, the H2i comprises a charged amino acid patch that is always 15 residues long across archaeal and eukaryotic MCMs (Parker et al., 2017). Deletion of all H2i residues in archaeal MCM enhanced ATP hydrolysis and DNA binding but completely abrogated DNA unwinding (Jenkinson and Chong, 2006). The enhanced DNA binding observed in the absence of H2i may indicate that this loop normally functions to destabilise protein-DNA contacts to enable DNA movement (Enemark and Joshua-Tor, 2008). In summary, abolished helicase activity upon both H2i and PS1 mutation indicates that both these loops play essential roles in unwinding.

1.2.4.3 The winged helix domain

A stretch of amino acids extending from the AAA+ domain is predicted to fold into a WHD, which consists of two structural elements: a helix-turn-helix motif and a wing made up of a β-loop-β fold (Harami et al., 2013). WHDs can serve as double-stranded DNA binding modules and are found in approximately a third of nucleic acid interacting proteins (Aravind et al., 2005). In eukaryotes, all MCM subunits apart from Mcm2 contain this appendix (Li et al., 2015). A solution structure of Mcm6 WHD revealed that the domain itself is well ordered but is flexibly linked to the AAA+ (Wei et al., 2010), consistent with the Mcm2-7 double hexamer structure showing disordered WHD densities above the motor domains (Li et al., 2015) (**Figure 1.3b**). Notably, Mcm6 WHD was found not to be involved in DNA binding but rather partakes in a number of protein-protein interactions, including Cdt1 (Wei et al., 2010), Orc4 (Yuan et al., 2017), Mcm10 (Douglas and Diffley, 2016) and Mrc1 (Komata et al., 2009). The significance of the other MCM WHDs is also beginning to emerge, with Mcm3 and 4 demonstrated to bind Cdc6/ORC to stabilise a key MCM loading intermediate (Yuan et al., 2017). Therefore, Mcm2-7 WHDs mediate important structural interactions throughout helicase activation (Zhai et al., 2017b).

1.2.5 The N-terminal non-catalytic tier of the MCM ring

The MCM NTD can be further divided into three subdomains: A, B and C (Fletcher et al., 2003)(**Figure 1.3b**). The A-domain comprises an alpha-helical bundle which facilitates protein-protein interactions with helicase co-activators (Costa et al., 2011, Costa et al., 2014); the B-domain contains zinc-finger (ZF) motifs that form the main binding interface towards the second MCM ring in the double hexamer (Li et al., 2015) and the C-domain is an OB-fold shown to be involved in binding to single-stranded DNA in archaea (Froelich et al., 2014, McGeoch et al., 2005).

1.2.5.1 The A-domain: the peripheral helical bundle

The A-domain is a compact module made up of four short alpha helices positioned at the outer ring perimeter (Li et al., 2015, Miller et al., 2014). A short loop connects the A and C-domains and may confer flexibility of the A-domain to regulate protein-protein interactions (Miller et al., 2014). Furthermore, the A-domain has also been implicated in DNA binding as its deletion diminishes MCM interaction with both single-stranded and duplex DNA (Costa et al., 2008). This result can be partially explained by the predicted position of the lagging strand template between the AAA+ and A-domains on the MCM ring exterior (Graham et al., 2011). More recently, it has been shown that A-domains mediate inter-hexamer stabilisation in the double hexamer (Li et al., 2015). For example, a yeast specific N-terminal insertion from Mcm7 contacts the Mcm5 A-domain *in trans*, which acts as a latch to bridge the two hexamers (Li et al., 2015) (discussed further in *section 1.3.2.3*). Upon the transition to S-phase, activators Cdc45 and GINS also bind the A-domains of Mcm2-5 and Mcm5-3 of the MCM ring, respectively (Costa et al., 2014) (see *section 1.2.6*). Together, this highlights that the A-domain is a hub for both inactive and active helicase stabilisation.

1.2.5.2 The B-domain: The Zinc Finger domain

The B-domain comprises a ZF module, which binds to a zinc ion in a tetrahedral geometry via three antiparallel β strands (Fletcher et al., 2003, Brewster et al., 2008, Miller and Enemark, 2015). All MCM subunits are predicted to bind zinc apart from Mcm3 (Li et al., 2015), which lacks the four highly conserved cysteines within the $CX_2CX_nCX_2C$ sequence motif (Liu et al., 2008). Consistent with the high level of

conservation, substitution of one cysteine with a serine residue in ZF abolishes archaeal MCM DNA binding and ATPase functions (Poplawski et al., 2001). Additionally, the ZF domains have been shown to form the main double hexamer interaction interface, where six ZFs from one ring brace the six ZFs from the opposing ring in an inter-digitated manner (Li et al., 2015, Fletcher et al., 2003). This arrangement is partially responsible for the observed tilted architecture of the eukaryotic double hexamer and has been proposed to drive origin melting as discussed in *section 1.3.2.3*.

1.2.5.3 The C domain: the OB-fold

The C-domain is the most conserved out of all three NTD MCM subdomains and comprises an OB-fold, which is a common β barrel fold associated with DNA binding (Murzin, 1993). In the MCM ring, the C-domain mediates interactions between neighbouring subunits and is sufficient to oligomerise the NTD tier in isolation (Kasiviswanathan et al., 2004, Fletcher et al., 2003). Indeed, a highly conserved aromatic residue within this domain is where hydrophobic inter subunit interactions would occur in the archaeal MCM and is thought to stabilise the ring (Fletcher et al., 2003). Mutation of this residue in mouse Mcm4 (F345I, termed the 'Chaos' mutation) leads to development of adenocarcinomas and chromosomal breakage while the equivalent mutation in yeast leads to a loss of function MCM phenotype (Shima et al., 2007).

In addition to maintaining the structural integrity of the ring, the OB-fold is critical for DNA binding and helicase activity (Froelich et al., 2014, McGeoch et al., 2005). In particular, the co-crystal structure of *Pyrococcus furiosus* MCM-NTD bound to single-stranded DNA showed the nucleic acid lining the pore of the MCM ring perpendicular to the axial channel (Froelich et al., 2014) (**Figure 1.4**, right panel). The single-stranded DNA is bound to OB-fold arginines (R124 and R186), which project towards the channel to create a positively charged patch. This was called the MCM single-strand binding (MSSB) motif and is conserved in the eukaryotic Mcm2-7 hetero-hexamer (Froelich et al., 2014). In yeast, the equivalent MSSB residues are found as either an arginine or lysine in the neighbouring Mcm4, 6 and 7 subunits while Mcm2, 3 and 5 only contain one conserved positively charged residue. Whereas

yeast cells containing MSSB single subunit mutations are viable, a combination of two subunit MSSB mutations leads to a loss of cell viability, impaired helicase loading and replication (Froelich et al., 2014). Together with the observation that the DNA diameter in the co-crystal structure is much wider than canonical B-form DNA, it was postulated that this conformation of single-stranded DNA represents an origin melting intermediate that is stabilised by the MSSB. Therefore, as well as mediating helicase stabilisation, the C-domain plays a role in initial strand separation of the DNA double helix.

1.2.6 Structure of the Cdc45-Mcm2-7-GINS helicase

Unlike archaeal MCMs, which exhibit robust unwinding activity (Chong et al., 2000, Kelman et al., 1999), isolated Mcm2-7 is a very weak helicase (Schwacha and Bell, 2001, Davey et al., 2003, Bochman and Schwacha, 2008, Ilves et al., 2010, Lyubimov et al., 2012). Although a ring formed of two copies of Mcm4, 6 and 7 support helicase activity *in vitro* (Ishimi, 1997, Kaplan et al., 2003), these findings were difficult to reconcile with data that all six MCM subunits are required for origin licensing and fork progression (Prokhorova and Blow, 2000, Labib et al., 2000). As mentioned previously, parallel studies subsequently established that MCM forms a stable complex with Cdc45 and GINS tetramer to form an active 11-subunit holo-helicase (**Figure 1.5**) (Pacek et al., 2006, Moyer et al., 2006, Gambus et al., 2006).

1.2.7 The structural basis for Mcm2-7 activation by Cdc45 and GINS

Following the discovery of the eukaryotic replicative helicase, phylogenetic analysis revealed that the Mcm2-7 co-activators are conserved from archaea to eukaryotes (Makarova et al., 2012) and indeed, archaeal Cdc45 and GINS stimulate MCM both *in vivo* and *in vitro* (Xu et al., 2016). Furthermore, several studies have shown that Cdc45 and GINS are indispensable components of the replication fork (Dalton and Hopwood, 1997, Moir et al., 1982, Kanemaki et al., 2003). Cdc45, which is the catalytically inactive ancestral homologue of the RecJ exonuclease (Sanchez-Pulido and Ponting, 2011)(see *section1.4.3*), has also been shown to retain its DNA binding function and is thought to capture the leading strand template that may exit from the helicase pore during replication fork stalling (Petojevic et al., 2015)(depicted in **Figure 1.5c**).

To address the role of the helicase co-activators, isolated Mcm2-7 and the fully assembled CMG were studied by negative stain single particle EM (Costa et al., 2011, Lyubimov et al., 2012)(**Figure 1.5**). The resulting three-dimensional structures of Mcm2-7 showed that the hetero-hexamer forms a constitutively open ring, which fluctuates between a topologically open, left-handed spiral assembly and a notched, planar conformation (**Figure1.5a**). The discontinuity in the hexamer was mapped to the Mcm2-5 site, a previously identified gate in the complex (Bochman and Schwacha, 2007) that has since been shown to be an important feature for helicase loading onto DNA (Samel et al., 2014).

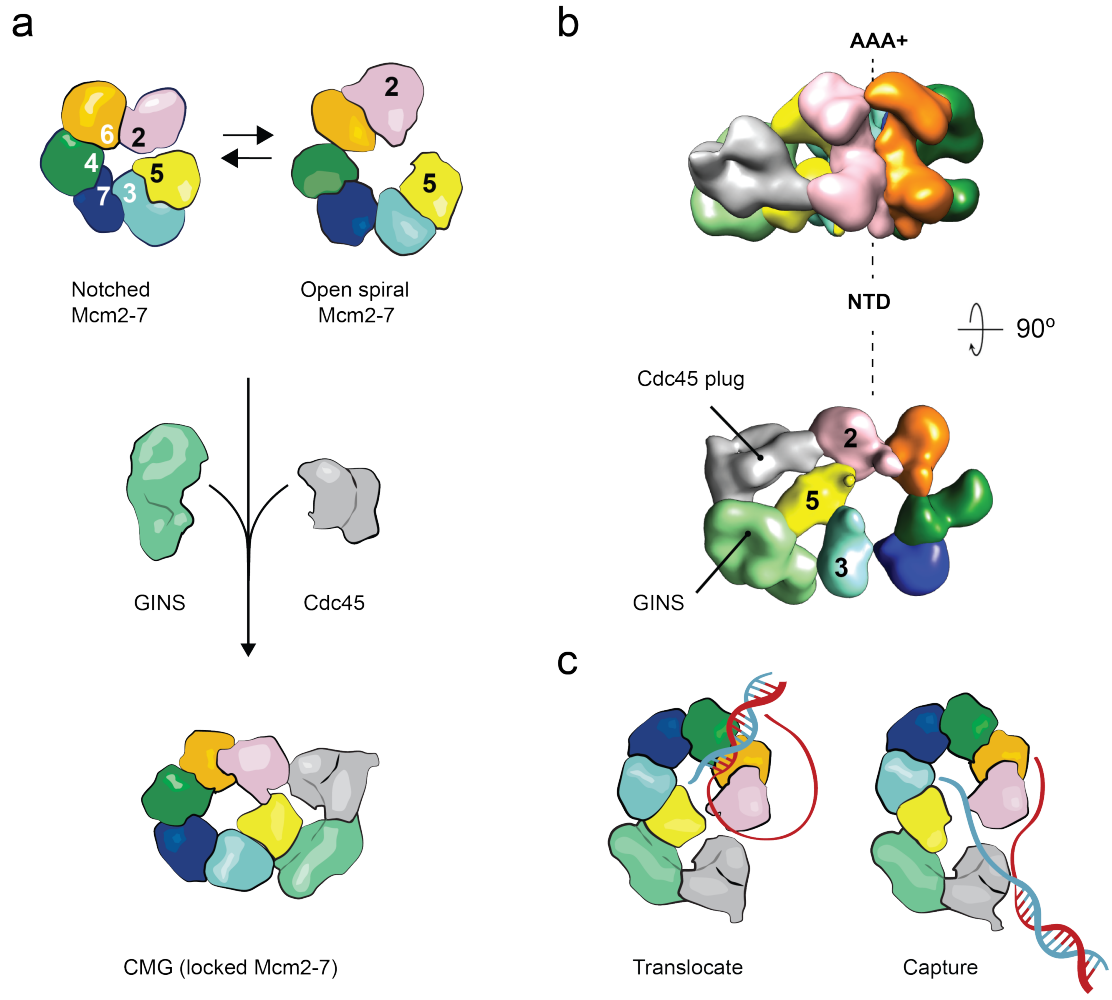


Figure 1.5: Mcm2-7 binds Cdc45 and GINS to shut the Mcm2-5 gate and enhance ATPase activity
(a) Schematic representation of results presented in structural studies of *Drosophila* CMG (Costa et al., 2011). Mcm2-7 forms a constitutively open gate at Mcm2-5, fluctuating between open spiral and planar conformations. Cdc45 binds the Mcm2-5 site whereas GINS binds to Mcm5-3, collectively promoting ATPase activity by the helicase. **(b)** Negative stain structure of the CMG showing the two-tiered architecture of the helicase and the binding sites of Cdc45 and GINS on the NTD tier of Mcm2-7 (EMD: 2772)(Costa et al., 2011). **(c)** A model for the role of Cdc45 in capturing leading strand template that may slip out of the channel upon replication fork stalling, based on crosslinking data of Cdc45 binding DNA (Petojevic et al., 2015).

Cdc45 and GINS bind at the interface of Mcm2-5-3 A-domains to shut the Mcm2-5 gate and stabilise the Mcm5-3 site (Costa et al., 2011). In this manner, Cdc45 and GINS reconstitute the two key ATPase sites that are essential for helicase activity (Ilves et al., 2010) (**Figure1.5a**). In line with such activating roles, biochemical experiments using recombinant *Drosophila* CMG revealed a helicase activity that is approximately 20-fold higher compared to the isolated Mcm2-7 ring (Ilves et al., 2010). Therefore, rather than having an enzymatic role, the Cdc45 and GINS co-activators act a stabilising 'handle' that maintain a topologically closed MCM ring to position the AAA+ domains in a productive conformation for unwinding (**Figure 1.5a**).

A subsequent higher resolution structure of the CMG bound to a 3' tailed duplex template and non-hydrolysable ATP revealed several new features of the holo-helicase architecture (Costa et al., 2014) (overall structure shown in **Figure 1b**). First, the co-activators make extensive contacts with each other to stabilise the CMG complex, in particular through the binding of the GINS subunit Psf1 to Cdc45. Importantly, a subdomain of Psf1 (termed the B-domain) is dispensable for GINS assembly but is essential for binding to Cdc45 and formation of the CMG complex (Costa et al., 2011, Costa et al., 2014). Second, Cdc45 and GINS wedge in between the two MCM tiers, causing a reconfiguration of the AAA+ modules into a right-handed spiral in the presence of DNA while the NTD maintains its fixed planar, topologically closed state. Therefore, rather than simply guarding the Mcm2-5 gate, Cdc45 and GINS physically re-arrange the MCM ring upon DNA binding. As the Mcm2-5 ATP hydrolysis motifs in this spiral configuration would likely be offset, it is predicted that the MCM motor fluctuates between planar and right-handed spiral configurations upon ATP turnover and translocation. Finally, the AAA+ tier of CMG was found at the single-stranded-duplex DNA junction ahead of the NTD tier, indicating that the eukaryotic replicative helicase tracks with its motor domain in front, consistent with previous FRET measurements of archaeal MCM-DNA complexes (McGeoch et al., 2005, Rothenberg et al., 2007). The polarity of the CMG helicase and how ATPase conformational changes confer movement on DNA are discussed in *section 1.5*.

1.3 Origin licensing

In all domains of life, DNA replication is initiated by a *trans*-acting initiator, which guides the loading of the helicase onto *cis*-acting origins (Robinson and Bell, 2005). In eukaryotes, the temporal separation of helicase loading and activation was first suggested by two independent seminal studies (Rao and Johnson, 1970, Blow and Laskey, 1988). In the first study, Rao and Johnson performed a series of HeLa cell hybridisation experiments in which two cells at different cell cycle stages were fused and the two resulting nuclei were analysed for DNA replication (Rao and Johnson, 1970). Upon fusion of G1 and S-phase cells, the G1 nucleus entered S-phase as assessed by bromodeoxyuridine (BrdU) pulse labelling and antibody recognition of BrdU incorporation into DNA of both nuclei. This led to the prediction that the S-phase nucleus releases an 'S-phase promoting factor' that is important for the G1 responsive nucleus to replicate. In contrast, G2 and S-phase fusions were resistant to the S-phase promoting factor and G1/G2 hybrids could not replicate unless progression through mitosis was enabled.

Building on the latter observation, Blow and Laskey used the *Xenopus* egg extract cell free system to asses which mitotic transition allowed re-replication to occur (Blow and Laskey, 1988). In this system, extracts can form intact nuclear membranes around added chromatin and replicated versus re-replicated extracts can be discerned. By disrupting the nuclear envelopes using detergents or mechanical means, G2 nuclei were able to replicate upon addition to a G1 cell extract. Therefore, permeabilisation of G2 nuclear membranes bypassed the chromosome condensation and nuclear breakdown steps of mitosis, which are normally pre-requisites for a cell to undergo replication. This led to the prediction that the nuclear membrane normally excludes an indispensable 'licensing' factor during S and G2-phases and this licensing factor only re-gains access to DNA again in the G1 phase after the nuclear envelope breakdown stage of mitosis.

The molecular basis for the two-step DNA replication initiation mechanism was subsequently established in *S. cerevisiae* genomic foot printing studies, which revealed the existence of 'pre-replicative' and 'post-replicative' origin states (Diffley et al., 1994). The former state was only observed in G1 phase blocked cells and

exhibited markedly reduced DNaseI accessibility. In contrast, asynchronous cells and cells blocked in G2/M, had smaller footprints, similar to those seen for origin bound ORC or the transcription factor Abf1. Several years later, the identity of this G1 pre-RC state was shown to be the loaded Mcm2-7 double hexamer as evidenced by depletion of MCMs in G1 irreversibly blocking replication fork progression (Labib et al., 2001). This finding is consistent with *in vivo* cross-linking ChIP experiments showing an association between Mcm4, Mcm7 and origin DNA in an ORC and Cdc6 specific manner in G1 cells (Aparicio et al., 1997). Together, these data laid the groundwork for the reconstitution of pre-RC assembly with purified budding yeast proteins, showing a strict requirement of ORC, Cdc6, Cdt1, and ATP hydrolysis for Mcm2-7 to attain its inactive origin bound double hexamer state (Remus et al., 2009, Evrin et al., 2009).

In this section, I will discuss eukaryotic replication origins and each of the factors involved in pre-RC assembly before discussing the current model for helicase loading and structure of the yeast double hexamer.

1.3.1 Origins of replication and ORC

In budding yeast, origins are mainly characterised by the presence of the autonomous replication sequences (ARS), which were first identified as a chromosomal fragment capable of initiating replication to exogenous plasmids (Stinchcomb et al., 1979). Each origin comprises modular elements, including the essential and conserved ARS consensus sequence (ACS) (Van Houten and Newlon, 1990)(**Figure 1.6**).

The ACS is an AT-rich 11 bp stretch and forms one of the composite binding sites for ORC (Diffley and Cocker, 1992, Bell and Stillman, 1992). ARS additionally contains the less well conserved B elements (B1-B3), which are only indispensable to origin function as a collective (Marahrens and Stillman, 1992). In this context, the B1 element forms the second bipartite ORC binding site (Rowley et al., 1995), the B2 element is A-rich and correlates with MCM loading to the origin (Wilmes and Bell, 2002) whilst the B3 element binds the Abf1 transcription factor to activate both replication and transcription (Diffley and Stillman, 1988).

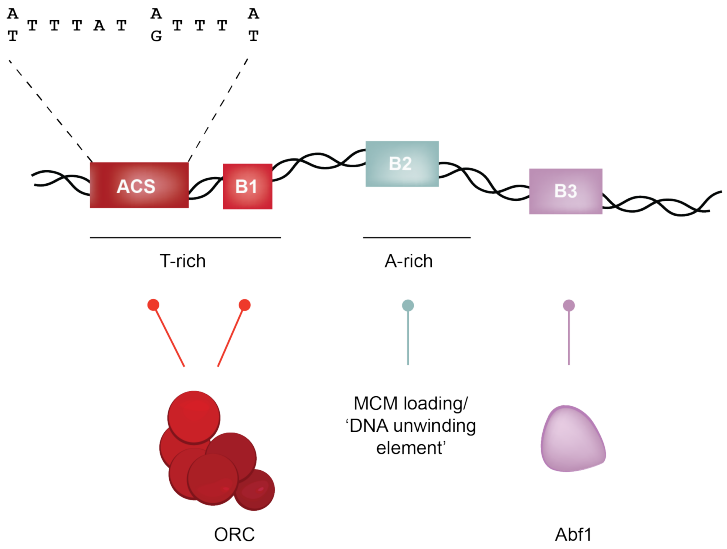


Figure 1.6: Schematic overview of *S. cerevisiae* ARS1 origins
ACS consensus sequence is from the (Eaton et al., 2010) study. ORC binds to both ACS and B1. Function of B2 is less clear but correlates with pre-RC assembly. Abf1 general regulatory factor binds to B3 to mediate transcriptional activation of DNA replication genes amongst others.

The latter binding of Abf1 to B3 is postulated to restrict multiple nucleosome positioning events on the B-domain, which in turn prevents encroachment of nucleosomes to the adjacent, indispensable ACS domain (Venditti et al., 1994).

Approximately 12,000 ACS-like sequences have been identified in the budding yeast genome but only 429 are functional (Wyrick et al., 2001). This has led to the suggestion that *S. cerevisiae* origins are further specified by the local chromatin structure. In line with this, functional origins correlate with nucleosome free regions (NFR) (Eaton et al., 2010), a feature that may be partially due to the A-T rich region of the ACS. ORC binding to the ACS further protects the NFR by physically acting as a barrier to *de novo* nucleosome assembly or sliding by existing nucleosomes (Kurat et al., 2017). However, chromatin also acts as a barrier by eliminating ORC binding to non-specific sequences (Kurat et al., 2017). Furthermore, the nucleosome-ORC relationship is integral as disruption of ORC-dependent positioning of nucleosome arrays directly impairs replication initiation (Lipford and Bell, 2001).

In contrast to budding yeast, ORC binding to origins in higher eukaryotes such as *Drosophila*, *Xenopus* and humans does not occur through a clear sequence motif (Costa et al., 2013). Although ORC often recognises AT or GC rich regions in metazoan cells (MacAlpine et al., 2010, Paixao et al., 2004), it has been

demonstrated that DNA conformation, proximity to promoters and nucleosome positioning are the primary determinants of replication start sites (MacAlpine et al., 2010, Remus et al., 2004, Eaton et al., 2010, Ghosh et al., 2004, Mechali et al., 2013). The absence of sequence preference in higher eukaryotes is exemplified by (*i*) *Drosophila* ORC, which favours actively open chromatin (MacAlpine et al., 2010) and negatively supercoiled DNA instead of relaxed DNA structures (Remus et al., 2004) (*ii*) *Xenopus* ORC which initiates DNA replication at any nucleotide sequence from a piece of microinjected DNA (Harland and Laskey, 1980) and (*iii*) humans where Orc1, via its bromo-adjecent homology (BAH) domain binds the dimethylated histone H4 tails of nucleosomes to recognise replication start sites in a sequence independent fashion (Muller et al., 2010, Kuo et al., 2012).

ORC was first discovered as an ATP dependent multiprotein assembly that is protected by DNase I digestion on *S.cerevisae* origins of replication (Bell and Stillman, 1992). Three parallel but independent studies using temperature sensitive or viable mutant strains of *orc2* showed defective entry into S-phase and an essential role for this complex in early DNA replication initiation (Bell et al., 1993, Micklem et al., 1993, Foss et al., 1993). In accordance with such an important activity, ORC subunits are highly conserved at the sequence level from archaea to human (Gavin et al., 1995, Gaudier et al., 2007) although divergences exist in the composition and stability of the complex (Li and Stillman, 2012).

Eukaryotic ORC consists of six Orc1-6 subunits, named in the order of their respective molecular weights with Orc1 being the largest subunit, partially due to its N-terminal BAH domain (Li and Stillman, 2012). Five of the six proteins (Orc1-5) belong to the same initiator clade II of AAA+ ATPases as the bacterial DnaA initiator (Erzberger and Berger, 2006). Orc1-5 subunits fold into an N-terminal AAA+ module that can bind DNA through an initiator-specific motif, and a C-terminal WHD (Li et al., 2018). Orc6 on the other hand bears little resemblance to the rest of the subunits and comprises a tandem-cyclin box domain akin to transcription factor IIB (Chesnokov et al., 2003, Costa et al., 2013, Liu et al., 2011).

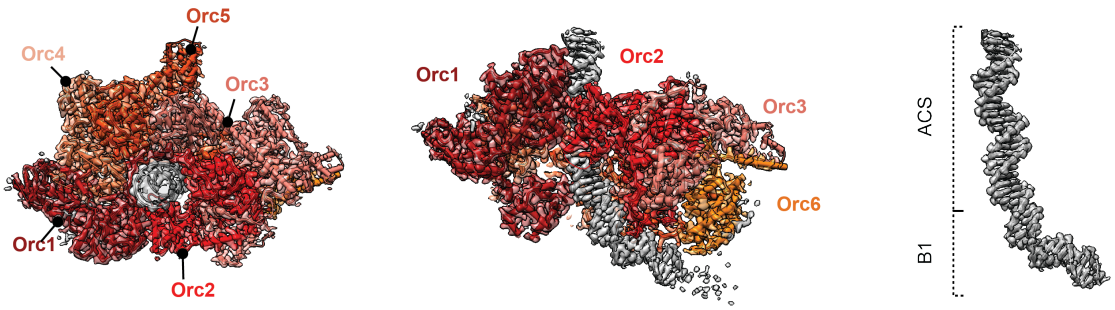


Figure 1.7: Cryo-EM structure of *S. cerevisiae* ORC bound to ACS and B1 element
Top, side and DNA-only views of ORC-DNA from (Li et al., 2018)(EMD: 6941, PDB: 5zr1). DNA bends inside the channel formed by the ORC complex. The interface between Orc1 and Orc2 is open, leaving room for DNA to enter prior to engagement of Cdc6 at this interface. See text for further observations.

A recent cryo-EM structure of ORC bound to the origin has informed the molecular basis for the bipartite ARS binding pattern and downstream helicase loading events (Li et al., 2018)(**Figure 1.7**). This work, which builds on previous structural studies of ORC in a variety of functional forms (Bleichert et al., 2015, Tocilj et al., 2017), revealed that ORC forms a notched, crescent shaped hexamer with a dynamic gap formed between Orc1 and Orc2. This gap is noteworthy as it allows initial entry of DNA and the subsequent transient interaction of Cdc6 between Orc1 and Orc2 (Yuan et al., 2017, Zhai et al., 2017b).

ORC recognises origins in an ATP-dependent manner and consistent with this, Orc1, Orc4 and Orc5 bind ATP (Orc1 can also hydrolyse ATP) (Klemm et al., 1997, Li et al., 2018). Conversely, Orc2 and Orc3 are thought to be degenerate ATPases (Bleichert et al., 2017). ORC binding to a budding yeast origin inflicts two bending points at the ACS and B1 sites, respectively, and this results in an almost 90º curvature of the double helix (Li et al., 2018)(**Figure 1.7**, right). An Orc1 basic patch and Orc4 insertion helix recognise highly conserved thymine bases of the ACS whereas the B1 element is bound by Orc2, Orc5 and Orc6. In this configuration, the Orc1 BAH domain would be free to bind nucleosomes while a region below this domain simultaneously anchors ORC to the origin. This mechanism could represent an important origin recognition mechanism in higher eukaryotes (Muller et al., 2010, Kuo et al., 2012).

By bending the DNA through these interactions, ORC is thought to deflect DNA away to create a maximal surface area for downstream Mcm2-7 docking (Li et al., 2018). Despite the significant DNA bend, and unlike the bacterial DnaA initiator (Duderstadt et al., 2011), ORC does not melt the DNA upon binding.

1.3.2 Cdc6

Cdc6 (Cell Division Cycle 6) was originally discovered in a screen of temperature sensitive *S. cerevisiae* mutants that arrested cell cycle progression (Hartwell, 1976). Subsequently Cdc6 was found to be conserved from archaea to human (Robinson et al., 2004, Williams et al., 1997). The first indication of Cdc6 involvement in DNA replication arose from experiments of temperature sensitive mutants that caused cell cycle arrest at the G1 to S-phase transition (Bueno and Russell, 1992). Thereafter, genomic foot printing strategies, PCR-based methods and chromatin fractionation experiments revealed that Cdc6 is an integral component of the pre-RC in both yeast and vertebrates (Donovan et al., 1997, Williams et al., 1997, Coleman et al., 1996, Cocker et al., 1996) and is essential for correct MCM loading (Tanaka et al., 1997, Remus et al., 2009, Evrin et al., 2009).

Like ORC, Cdc6 belongs to the AAA+ superfamily of ATPases (clade II) and modulates the activity and stability of other complexes (Herbig et al., 1999, Neuwald et al., 1999). Cdc6 physically interacts with ORC to form an ‘OC’ complex as shown from structural and biochemical studies (Speck et al., 2005, Liang et al., 1995) and indicated by earlier foot printing studies (Diffley et al., 1994). In this context, Cdc6 binding is thought to promote ATP hydrolysis by ORC and increase its specificity for origin DNA (Speck et al., 2005, Mizushima et al., 2000, Speck and Stillman, 2007) partially due to a Cdc6-mediated reorientation of the Orc1 BAH domain and the topological entrapment of DNA in the central ORC channel (Sun et al., 2012, Yuan et al., 2017, Speck and Stillman, 2007). However, the functional significance of Cdc6 ATP hydrolysis for MCM loading remains unclear. On the one hand, ATP hydrolysis has been shown to be critical for Mcm2-7 loading in genetic and *in vitro* based experiments (Randell et al., 2006, Perkins and Diffley, 1998). On the other hand, mutations in Walker B of Cdc6 can still support MCM loading both *in vivo* and *in vitro* (Coster et al., 2014, Chang et al., 2015, Kang et al., 2014). Nevertheless, the Cdc6

mutant was defective in releasing non-productively bound MCMs from the pre-RC (Coster et al., 2014) and Cdc6 ATP hydrolysis was found to be essential for yeast viability (Chang et al., 2015). Thus, a quality control model was proposed, wherein ATP turnover by Cdc6 removes MCM complexes that have not effectively engaged with the origin to ensure maintenance of genome stability (Kang et al., 2014, Coster et al., 2014).

1.3.3 Mcm2-7/Cdt1

Cdt1 (Cdc10 Dependent Transcript 1) was originally identified in *S. pombe* as a factor required for cell-cycle entry into S-phase (Hofmann and Beach, 1994) and was subsequently found to be essential for origin licensing control in the same model organism (Nishitani et al., 2000). Cdt1 is conserved across all analysed eukaryotic genomes (Maiorano et al., 2000, Whittaker et al., 2000, Wohlschlegel et al., 2000, Pozo and Cook, 2016) but the level of sequence conservation is lower than other pre-RC components. Indeed, focused sequence searches coupled with functional analysis was required to find the *S. cerevisiae* Cdt1 ortholog (Devault et al., 2002). In budding yeast, Cdt1 forms a pre-assembled heptameric complex with Mcm2-7 prior to nuclear transport at the M/G1 phase transition (Tanaka and Diffley, 2002, Frigola et al., 2017) whereas metazoan Cdt1 only forms a complex with Mcm2-7 in a chromatin dependent manner (Gillespie et al., 2001).

In contrast to ORC, Cdc6 and Mcm2-7, Cdt1 is a non-enzymatic protein (Pozo and Cook, 2016). Previous studies have identified three main functional Cdt1 domains: evolutionarily divergent N-terminal and middle domains (N and M-domains), which interact with Geminin to target the protein for proteasome mediated degradation in humans. Importantly, the M-domain also binds to the NTD of Mcm2 and Mcm6 (Frigola et al., 2017). An additional Cdt1 C-terminal WHD also interacts with Mcm6 (Pozo and Cook, 2016, Yuan et al., 2017). At the pre-RC, the role of Cdt1 was thought to be to guide MCM towards ORC-Cdc6 via a direct Cdt1-Orc6 interaction (Chen et al., 2007, Semple et al., 2006). However, since those studies a highly conserved Mcm3 C-terminus region was shown to be sufficient to direct the Mcm2-7 ring to ORC-Cdc6 in the absence of Cdt1 (Frigola et al., 2013). The current model is that Cdt1's role is to stabilise the Mcm2-7 ring in a number of ways.

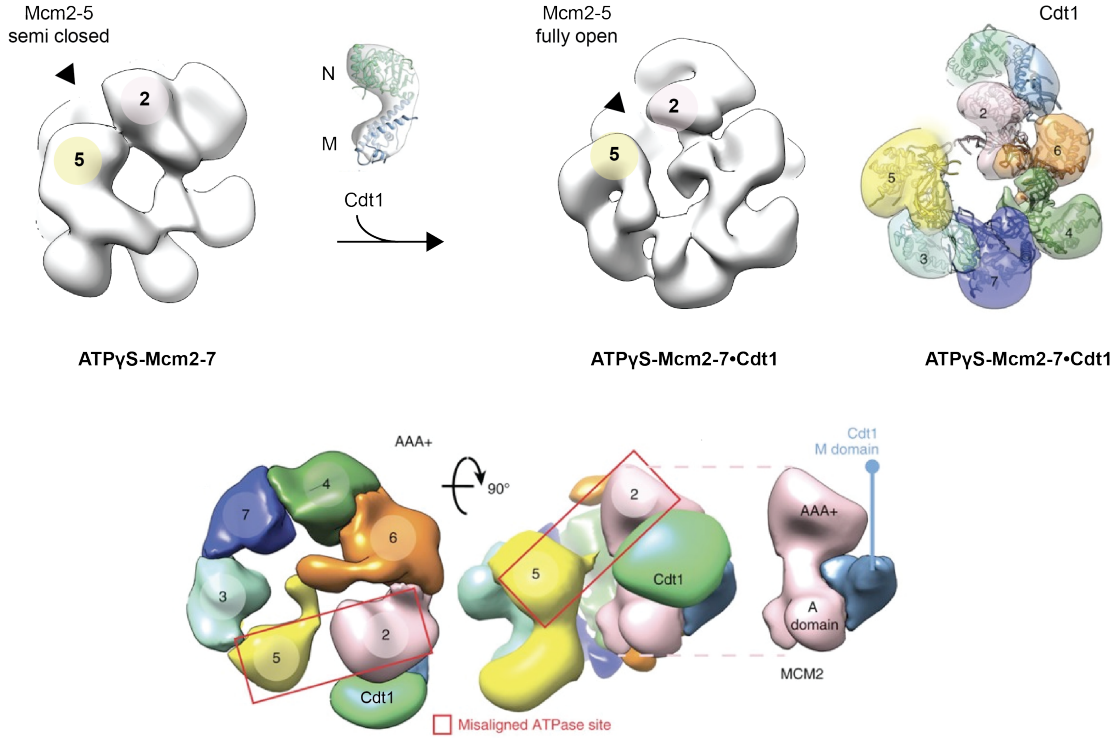


Figure 1.8: Negative stain EM analysis of isolated Mcm2-7 and Mcm2-7/Cdt1
Three-dimensional reconstructions of Mcm2-7 and Mcm2-7/Cdt1 from (Frigola et al., 2017) were all obtained in the presence of slowly hydrolysable ATP analogue. Cdt1 maintains an open Mcm2-5 gate and causes misalignment of the Mcm2-5 ATPase interface. Open gate enables loading onto DNA and prevents premature ATP hydrolysis. Figure adapted from (Frigola et al., 2017) (Mcm2-7 EMD: 3681, Mcm2-7/Cdt1 EMD: 3680 and Cdt1 PDB: 5me9 and 5mea).

First, a Cdt1 WHD interaction with the Mcm6 WHD relieves an autoinhibitory conformation of Mcm6 WHD that would otherwise cause steric clash between the MCM ring and Orc4 (Fernandez-Cid et al., 2013, Yuan et al., 2017). Second, Cdt1-Mcm2-7 forms a left-handed, notched spiral heptamer (**Figure 1.8**). In this arrangement, Cdt1 binds the A-domains of Mcm2 and Mcm6 and causes a gross misalignment of the pre-formed Mcm2-5 gate compared to the smaller gap observed in the isolated, nucleotide bound Mcm2-7 hexamer (top row, **Figure 1.8**). This serves to promote the initial loading of MCM onto DNA while keeping the helicase in an ATP bound but catalytically incompetent state (Frigola et al., 2017) (bottom row, **Figure 1.8**). Finally, on DNA, Cdt1 stabilises Mcm2,4,6 subunits to avert MCM ring disassembly upon ORC-Cdc6 ATP hydrolysis (Frigola et al., 2017, Zhai et al., 2017a, Frigola et al., 2013).

1.3.4 Structures of MCM loading intermediates and mechanism of pre-RC formation

Structural studies have provided extensive insights into the molecular mechanism of MCM loading through high resolution structures of human ORC (*section 1.3.1*), Mcm2-7 with or without Cdt1 (*section 1.3.3*), ORC-Cdc6-Cdt1-Mcm2-7 (OCCM) (Yuan et al., 2017) and the MCM double hexamer (Li et al., 2015). In particular, these studies have provided a rationale for the requirement of a series of co-factors to enable Mcm2-7 association with DNA. In the following sub-sections, I will provide a description of the OCCM and a mechanism of pre-RC formation before describing the double hexamer architecture.

1.3.4.1 The structure of OCCM

Upon Cdt1-Mcm2-7 interaction with ORC and Cdc6, a highly transient so-called OCCM structure is formed, which can only be captured and stabilised in the presence of a non-hydrolysable ATP analogue (ATPγS) (Sun et al., 2013).

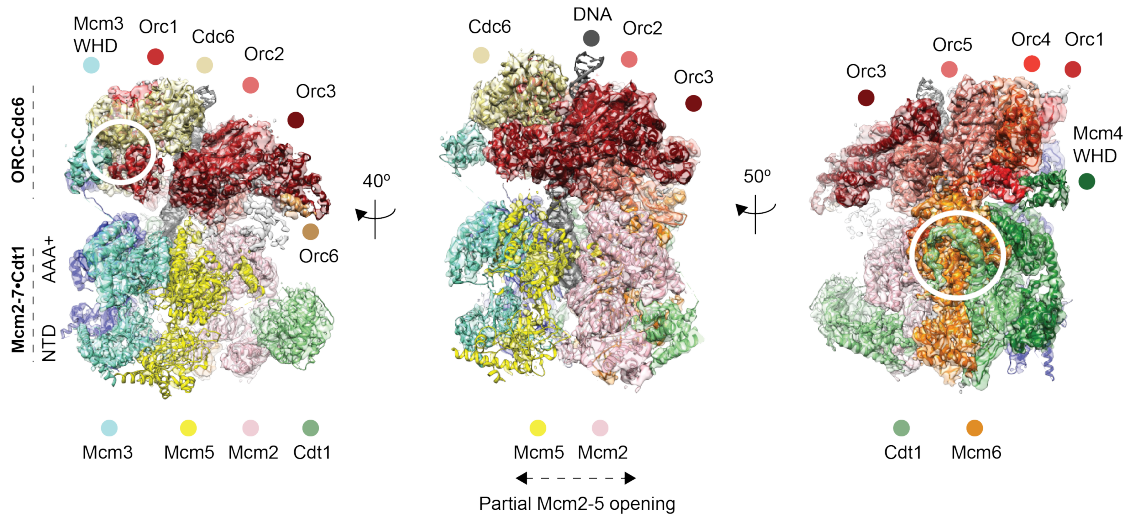


Figure 1.9: Cryo-EM structure of *S. cerevisiae* ORC-Cdc6-Cdt1-Mcm2-7
Three views of OCCM structure from (Yuan et al., 2017). White circles highlight Mcm3 WHD interaction with Cdc6 and Orc2 (left), and Cdt1 interaction with Mcm6 WHD (right).

A cryo-EM structure of this intermediate reveals a planarisation of the MCM ring upon Cdt1-Mcm2-7 interaction with ORC-Cdc6, entrapment of DNA through an Orc1-5/Cdc6 channel and the reconfiguration of Mcm2-5 towards a closed state (Yuan et al., 2017)(**Figure 1.9**). Within this structure, Cdt1 now also interacts with the Mcm6

WHD which confers stability to the MCM at this stage in the loading process. In the absence of Cdt1, the Mcm2-7 ring binds ORC-Cdc6 weakly and loses the adjacent Mcm2-6-4 subunits (Frigola et al., 2013). A rationale for this observation is provided by the OCCM structure: Cdt1 binding allows Mcm6 and 4 WHD to latch onto ORC-Cdc6 whereas the Mcm3 and Mcm7 WHDs bind the DNA loaders in a Cdt1 independent manner (right panel, **Figure 1.9**). Moreover, comparison of Cdt1-Mcm2-7 and OCCM reveals that Cdt1-Mcm6 WHD contributes to MCM ring planarisation, thereby converting the motor of the replicative helicase from an ATP hydrolysis incompetent form towards a state that can use ATP turnover to complete double hexamer formation (Zhai et al., 2017a, Zhai et al., 2017b, Frigola et al., 2017, Yuan et al., 2017). Furthermore, the Mcm3 WHD is well-resolved and binds the Orc2-Cdc6 surface (left panel, **Figure 1.9**). This strong composite binding mode explains why truncation of the Mcm3 WHD domain abolishes MCM ring docking onto ORC-Cdc6 and why a fragment of Mcm3 WHD is sufficient to bind and activate the ORC-Cdc6 ATPase activity (Sun et al., 2013, Frigola et al., 2013).

1.3.4.2 Mechanism of origin licensing

Integrating all of the above structural data with biochemical reconstitution and single molecule studies, a model for pre-RC assembly was recently proposed (**Figure 1.10**)(Zhai et al., 2017b, Ticau et al., 2015, Ticau et al., 2017, Evrin et al., 2009, Remus et al., 2009, Coster et al., 2014, Frigola et al., 2013).

First, ORC recognises origins in an ATP-dependent fashion and associates with Cdc6 to form a topologically closed complex around DNA (Speck et al., 2005, Speck and Stillman, 2007). The first Mcm2-7-Cdt1 heptamer is recruited to the ORC-Cdc6 via the Mcm3 WHD to form a transient OCCM intermediate (Sun et al., 2013, Yuan et al., 2017). At this stage, Mcm2-7 requires both ATP binding and Cdt1 for ring stability but does not yet hydrolyse nucleotide (Frigola et al., 2013), and the open Mcm2-5 gate plays a critical role in loading the hexamer onto DNA (Samel et al., 2014).

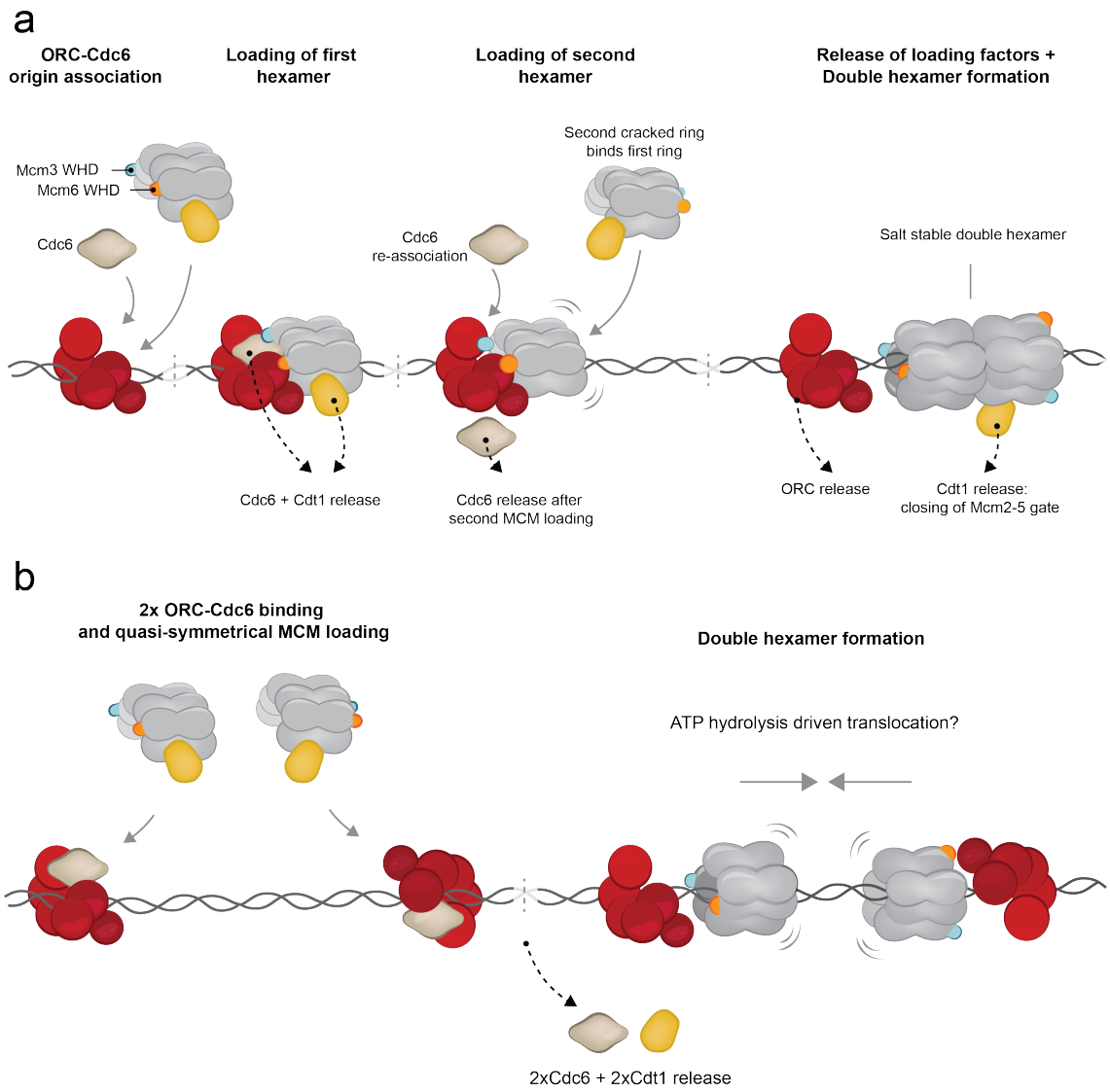


Figure 1.10: Origin licensing models
(a) One ORC model: two consecutive rounds of Mcm2-7 loading. Loading of Mcm2-7 onto DNA enabled by the open Mcm2-5 gate, a configuration maintained by Cdt1 binding to Mcm6-2. First hexamer recruited mainly through Mcm3 WHD-Cdc6 interaction to form OCCM. Subsequent Cdc6 and Cdt1 release leave behind an ORC-Mcm2-7 complex. ORC rebinds Cdc6 and the first hexamer acts as a template for second Mcm2-7 hexamer recruitment (bound to Cdt1). Ejection of Cdc6, ORC and Cdt1 leads to salt stable formation of double hexamer. Scheme primarily based on single molecule studies of MCM loading (Ticau et al., 2015, Ticau et al., 2017) and depiction of this process in (Parker et al., 2017). **(b)** Two ORC model: one round of 2xMcm2-7 loading. Two ORC-Cdc6 initiators recruit 2xMcm2-7-Ctd1 heptamers. ATP hydrolysis dependent ejection of Cdc6 and Cdt1, and ATP hydrolysis by Mcm2-7 may enable active translocation by helicases to form an NTD-NTD tethered double hexamer. This panel is based on (Coster and Diffley, 2017).

The Mcm3 WHD interaction with ORC-Cdc6 activates the ATPase activity of Cdc6, and in turn, Cdc6 releases non-productively bound hexamers (Frigola et al., 2013, Coster et al., 2014). Prior to loading of the second ring, Cdc6 is ejected alongside Cdt1 in an Mcm2-7-ATP hydrolysis dependent manner (Kang et al., 2014, Coster et al., 2014, Ticau et al., 2015). This action causes the first helicase ring to close (Ticau

et al., 2017) and is presumed to leave the Mcm2-7 bound to ORC via the WHDs of Mcm4 and Mcm6 whilst releasing the Mcm3 WHD interaction with Cdc6 (Zhai et al., 2017b, Yuan et al., 2017). The ORC-Mcm2-7 complex then re-binds to Cdc6 to recruit a second Cdt1-Mcm2-7 heptamer. The second heptamer associates with the first Mcm2-7 hexamer via an NTD-NTD interaction. Subsequent ATP hydrolysis by MCM is essential for loading (Evrin et al., 2009, Remus et al., 2009) and leads to the ordered release of Cdc6 followed by ORC and Cdt1 ejection, where release of Cdt1 is thought to drive full closure of the Mcm2-5 gates (Ticau et al., 2017) and lead to the formation of a salt stable double hexamer (Remus et al., 2009, Evrin et al., 2013).

However, how the second MCM hexamer is recruited to origins is still debated. Two single molecule studies by the Bell group indicate that loading of Mcm2-7 occurs in a step-wise manner arbitrated by the same ORC assembly and this is the model presented in **Figure 1.10a** (Ticau et al., 2015, Ticau et al., 2017). In contrast, a recent study by the Diffley laboratory indicate that two ORC molecules bind origins to recruit two hexamers by a unifying mechanism (**Figure 1.10b**) (Coster and Diffley, 2017). In this model, one ORC binds to the ACS whilst the second ORC binds to the B elements to facilitate quasi symmetrical loading of the two MCM hexamers. Whether MCM loading occurs in a strictly concerted fashion is unknown, but this mechanism posits a requirement of the two hexamers to translocate toward each other from their distally spaced bimodal binding sites, possibly in an ATP hydrolysis dependent manner (**Figure 1.10b**).

1.3.4.3 The structure of the MCM double hexamer

As revealed by a recent cryo-EM structure, the MCM double hexamer is a highly stable complex formed of two stacked, tethered rings (Li et al., 2015). The interaction between the two MCM complexes is largely mediated by B-domain (ZF) interactions that form an inter-digitated brace between the two NTD tiers (**Figure 1.11**). These inter-hexamer contacts cause a visible twist in the structure and a tilt of the rings that collectively stabilise double hexamer formation, inhibit Mcm2-5 opening and restrict the diameter of the axial channel at the interface of the helicases.

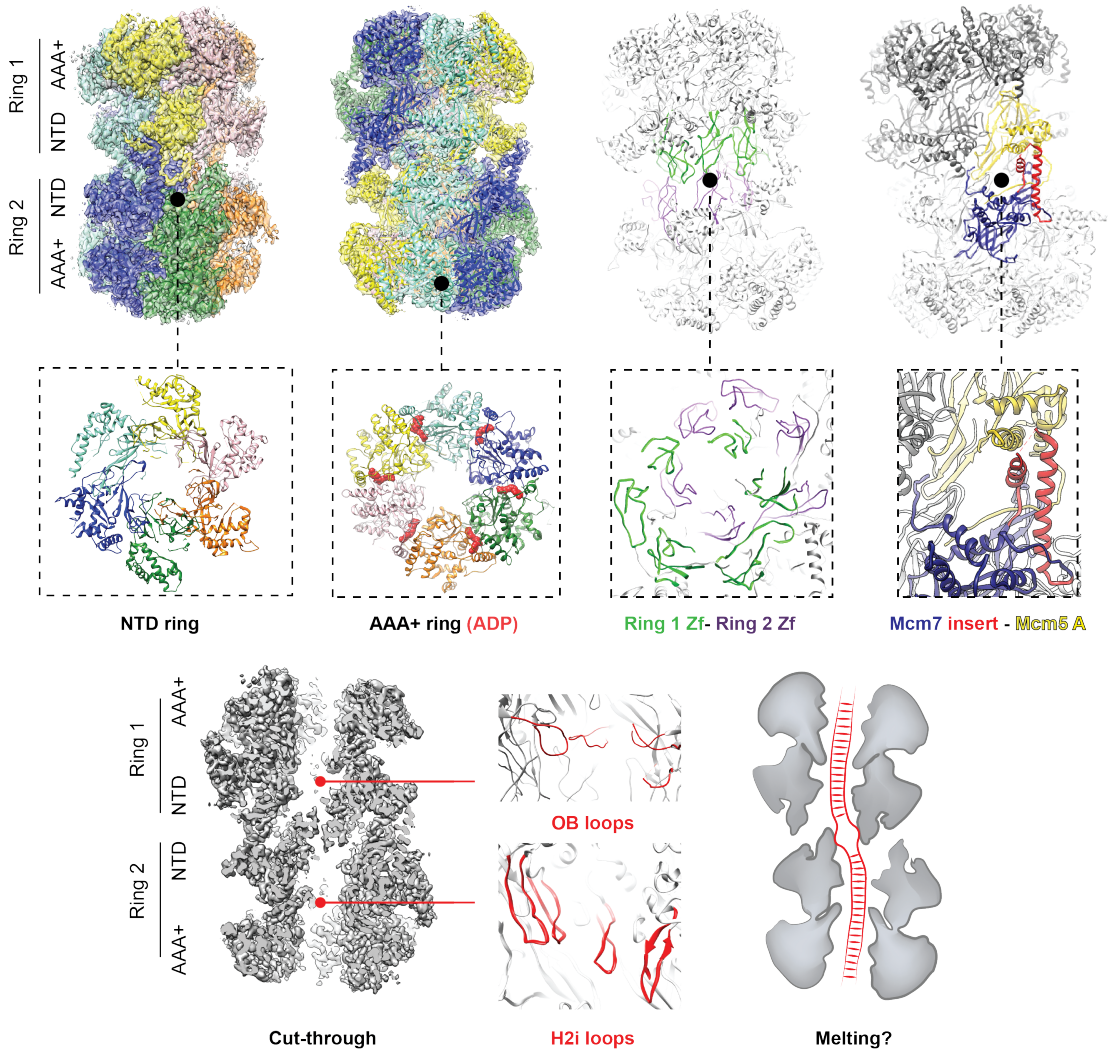


Figure 1.11: Cryo-EM structure of the DNA-free S. cerevisiae double hexamer at 3.8 Å resolution. Figure highlights (from left to right): double hexamer architecture, structure of each individual tier, the twisted NTD-NTD interface, stabilisation of double hexamer via Mcm7-5 interaction, the constriction of the DNA channel, examples of loops lining the channel and a prediction of DNA melting upon pre-RC assembly. All based on structure from EMD: 6338 and PDB: 3ja8 (Li et al., 2015).

Double ring assembly is further stabilised by eukaryotic specific insertions in the NTDs of the MCM subunits. For example, an Mcm7 alpha helix insertion interacts with the Mcm5 A-domain of the opposing ring to form a latch on the outer surface of the double hexamer (**Figure 1.11**, right panel). Within each ring, the NTD modules exhibit a left-handed twist relative to the AAA+ domains in contrast to the more co-axial architecture of the MCM hexamer within the OCCM structure (**Figure 1.9**) (Yuan et al., 2017). The functional relevance of the double hexamer twisted configuration is unclear but may serve to stabilise the ATPase motifs in an inactive conformation (Sun et al., 2014). Indeed, the AAA+ subunit interfaces bind ADP (an

example is shown in second column of **Figure 1.3c**), consistent with the double hexamer representing a post-catalytic state that is unable to unwind DNA (Remus et al., 2009, Sun et al., 2014, Kang et al., 2014, Coster et al., 2014). Moreover, distances between AAA+ modules vary, with the Mcm4-7, Mcm7-3 and Mcm2-5 sites forming a tight interaction compared to the looser Mcm6-4, Mcm5-3 and Mcm6-2 interfaces. The dynamic side of the ring is thought to enable the MCM interfaces to open up upon activation, with the Mcm6-2 site identified as a key region that could be used for strand extrusion of the lagging strand template (discussed further in *section 1.5.2.3*) (Li et al., 2015).

A cut-away view of the double hexamer reveals a series of hairpin loops that decorate the channel interior (bottom row, **Figure 1.11**). These loops have been previously characterised to bind DNA and are important for unwinding, including the H2i loops that emanate from the AAA+ modules and follow a helical path (Jenkinson and Chong, 2006), and the OB-fold loops, which emanate from the C-domain of the NTD modules (McGeoch et al., 2005, Froelich et al., 2014).

The central MCM double hexamer channel is non-uniform in diameter: 30 Å at the AAA+ and 40 Å at the NTD tier but with two constriction points at the H2i and OB-fold sites that narrow the channel to 25 Å. A further channel block is observed at the B-domain ZF double hexamer interface, causing a notable kink in the channel (bottom row, **Figure 1.11**). This led to the prediction that intervening double helix would be deformed upon pre-RC assembly and act as a nucleation site for origin melting. Computational modelling of DNA inside the channel is consistent with this notion, showing that as many as seven base pairs would be disrupted (Zhai et al., 2017b). The authors predicted that the strands separated by melting could then be extruded at the interface of the two hexamers akin to the ‘rabbit ear’ DNA pump model of L-Tag (Li et al., 2003). However, this is merely a hypothesis at this stage because the structure was determined in the absence of DNA.

An understanding of helicase activation requires a full description of the inactive licensed origin. A key focus of this study is therefore to structurally characterise the DNA bound MCM double hexamer (Chapter 3) and the conformational changes that accompany origin firing factor recruitment.

1.4 Origin firing

Activation of MCM double hexamers into CMG helicases that unwind DNA bi-directionally entails a highly regulated process that is governed by two kinases, DDK and CDK (Labib, 2010). These kinases coordinate the assembly of a transient pre-initiation complex (pre-IC) (Zou and Stillman, 1998, Tanaka and Araki, 2013) and activation of the helicase through the recruitment of firing factors to origins: Sld3/7, Cdc45, Dpb11, Sld2, GINS, Polε and Mcm10 (Yeeles et al., 2015, Heller et al., 2011).

DDK phosphorylation is required for the recruitment of the first helicase co-activator Cdc45 via the Sld3/7 chaperone (Sheu and Stillman, 2006, Francis et al., 2009, Tanaka et al., 2011a, Deegan et al., 2016). CDK phosphorylation of two essential targets, Sld2 and Sld3, enables these two proteins to interact with Dpb11 (Zegerman and Diffley, 2007, Tanaka et al., 2007) to recruit the second helicase co-activator GINS (alongside Polε) to MCM-Cdc45 (Muramatsu et al., 2010, Sengupta et al., 2013). Of the firing factors, only Mcm10 is dispensable for CMG assembly (Yeeles et al., 2015). However, Mcm10 is essential for activation of the helicase, as indicated by impaired recruitment of the single-stranded DNA binding protein RPA in the absence of this firing factor (Yeeles et al., 2015, Heller et al., 2011, van Deursen et al., 2012, Kanke et al., 2012).

In this section, I describe the roles of the origin firing factors in CMG assembly and activation in more detail. Further, I provide an overview of the current understanding of the molecular mechanism of Cdc45 and GINS recruitment, the role of Polε in this process and the models for how Mcm10 activates origins.

1.4.1 DDK

DDK (Dbf4-dependent kinase) is a complex of the Cdc7 kinase associated to its regulatory subunit Dbf4 (Dumbbell forming 4), which alleviates an autoinhibitory mechanism in Cdc7 (Dowell et al., 1994, Jackson et al., 1993). The *CDC7* and *DBF4* genes are essential for transition into S-phase (Kitada et al., 1992) and the *CDC7* product is a serine/threonine kinase (Sclafani, 2000). The first line of evidence that DDK functions at origins stemmed from a one-hybrid screen in budding yeast where *DBF4* was isolated in an ARS1-dependent manner (Dowell et al., 1994). Parallel

studies subsequently identified that the essential role of Cdc7 extends to firing of both early and late replication origins throughout S-phase (Bousset and Diffley, 1998, Donaldson et al., 1998) while during the G1 phase Dbf4 levels are kept low by APC/C mediated proteasome degradation to prevent pre-mature kinase activity (Weinreich and Stillman, 1999).

Since then, a substantial number of studies have shown that the target of Cdc7 phosphorylation is the Mcm2-7 complex specifically bound to origins (Sheu and Stillman, 2006, Randell et al., 2011, Francis et al., 2009, Masai et al., 2006, Sheu and Stillman, 2010). The current consensus is that DDK poly-phosphorylates the N-terminal extensions of Mcm4 and Mcm6, which in eukaryotes contain long unstructured tails of more than 100 amino acids (Labib, 2010, Randell et al., 2010). For example, the Mcm4 N-terminal tail comprises a highly conserved serine/threonine rich sequence amongst eukaryotes (contributes roughly a fifth of the 174 and 148 amino acids in the Mcm4 N-terminal tail in yeast and humans, respectively) (Sheu and Stillman, 2006, Sheu and Stillman, 2010). Importantly, phosphorylation of Mcm4 and Mcm6 N-termini are a strict pre-requisite for the recruitment of the first helicase co-activator Cdc45 via Sld3/7 both *in vivo* and *in vitro* (see *section 1.4.6.1*)(Sheu and Stillman, 2006, Deegan et al., 2016).

However, the mechanism of Mcm2-7 activation by DDK remains unclear. One view is that DDK causes a conformational change in the Mcm2-7 complex that is important for its downstream activation (Labib, 2010). The principal demonstration of this came from a *bypass of block 1* mutation in budding yeast, which overcomes loss of cell viability due to *CDC7* and *DBF4* deletion, and is a point mutation of the highly conserved proline 83 to a leucine in the Mcm5 A-domain region (bob1 mutation)(Hardy et al., 1997). When an analogous *mcm5-bob1* mutation was introduced into the archaeal homo-hexamer, a slight shift of the A-domain was observed with respect to the C-domain, termed the ‘domain-push’ effect (Fletcher et al., 2003). The higher flexibility of the A-domain observed in this mutant predicts a destabilisation of the double hexamer and may represent the mechanism by which DDK phosphorylation activates the MCM. Further supporting that DDK may induce an activating MCM conformational change, deletion of the central stretch of residues in the Mcm4 tail relieves an inhibitory mechanism of this region and similarly

bypasses requirement for DDK (Sheu and Stillman, 2010). Nonetheless, low-resolution structures of the DDK phosphorylated double hexamer showed no major conformational differences compared to unmodified MCM (On et al., 2014, Sun et al., 2014). The structural effect of DDK function might therefore be subtler, which I will attempt to address in this study by characterising a DDK phosphorylated MCM double hexamer structure by cryo-EM (Chapter 3).

1.4.2 Sld3/7

SLD3 was initially identified in a budding yeast screen for factors that are synthetically lethal with *dpb11-1* (Kamimura et al., 2001) and was later shown to encode a conserved, essential replication factor amongst eukaryotes (called Treslin/TICRR in humans) (Kumagai et al., 2010, Boos et al., 2011). Budding yeast Sld3 and the human orthologue Treslin/TICRR are regulated by S-phase kinase signalling and play integral roles in the recruitment of both Cdc45 and GINS to licensed origins (Kanemaki and Labib, 2006, Kamimura et al., 2001, Boos et al., 2013). Specifically, Sld3 binds to Cdc45 and the recruitment of this complex to origins is regulated by DDK phosphorylation (Kamimura et al., 2001, Kanemaki and Labib, 2006, Deegan et al., 2016). Additionally, CDK directly phosphorylates Sld3, which enables binding of Sld3 to the BRCT (BRCA1 C-terminal) repeats of Dpb11 (Tanaka et al., 2007, Zegerman and Diffley, 2007) and subsequent recruitment of GINS to MCM.

While the Sld3 C-terminus is a target of CDK phosphorylation, the N-terminus of Sld3 constitutes both a Cdc45 and Sld7 binding domain (Itou et al., 2014, Tanaka et al., 2011b, Itou et al., 2015). Although not essential for yeast cell viability, Sld7 depletion leads to compromised cell growth, reduction of Sld3 levels and delayed progression through S-phase (Tanaka et al., 2011b). Furthermore, Sld7 interacts with origins in an Sld3 dependent manner before dissociating upon DNA replication initiation. Thus, it is thought that Sld3/7 form an inter-dependent stable dimer that work in concert during origin firing (Tanaka et al., 2011a).

1.4.3 Cdc45

First discovered in *S. cerevisiae* genetic screens for factors involved in cell cycle progression (Moir et al., 1982), *CDC45* encodes an indispensable and highly conserved DNA replication factor that forms a *bona fide* component of the eukaryotic replicative helicase (Gambus et al., 2006, Pacek et al., 2006, Moyer et al., 2006). The recruitment of Cdc45 to Mcm2-7 is a key event towards activation of the replicative helicase (Aparicio et al., 1997). As mentioned above, this process depends on the phosphorylation of licensed origins by DDK to promote binding of the Cdc45 chaperone Sld3/7 (a proposed mechanism is outlined in *section 1.4.6.1*) and occurs in a CDK independent manner (Yeeles et al., 2015, Heller et al., 2011). In the fully assembled CMG, Cdc45 binds the A-domains of Mcm2 and 5 to keep the Mcm2-5 gate firmly shut and also binds the Psf1 B-domain of GINS to stabilise holo-helicase assembly (as previously discussed in *section 1.2.7*).

Beyond this important observation, mechanistic insights into the role of Cdc45 have been gained through its demonstrated evolutionary relationship with the bacterial DNA repair 3'-5' single-stranded RecJ exonuclease (Sanchez-Pulido and Ponting, 2011, Krastanova et al., 2012), the archaeal GINS-associated nuclease (GAN) (Oyama et al., 2016) and a recent crystal structure of human Cdc45 (Simon et al., 2016), which has confirmed previous predictions of a similarity with RecJ (Pellegrini, 2017).

The core conserved features of Cdc45 include an N-terminal RecJ-like DHH and C-terminal DHHA1 domains bridged by a three-helix motif (3HM), which confers flexibility of the two DDH domains via the C-terminal alpha helix in 3HM (**Figure 1.12**) (Simon et al., 2016). In RecJ, the active site is juxtaposed between the two DHH domains, forming a groove that binds single-stranded DNA and hydrolyses phosphodiester bonds in the presence of two Mn^{2+} ions (Cheng et al., 2016). In Cdc45, however, the DHH and DHHA1 domains are intra-molecularly linked through hydrophobic interactions to effectively prevent such a DNA binding channel (Simon et al., 2016). Together with the observation that residues involved in coordination of Mn^{2+} ions at the N-terminal DHH domain are mutated, this explains why Cdc45 has lost its ancestral catalytic activity (Pellegrini, 2017).

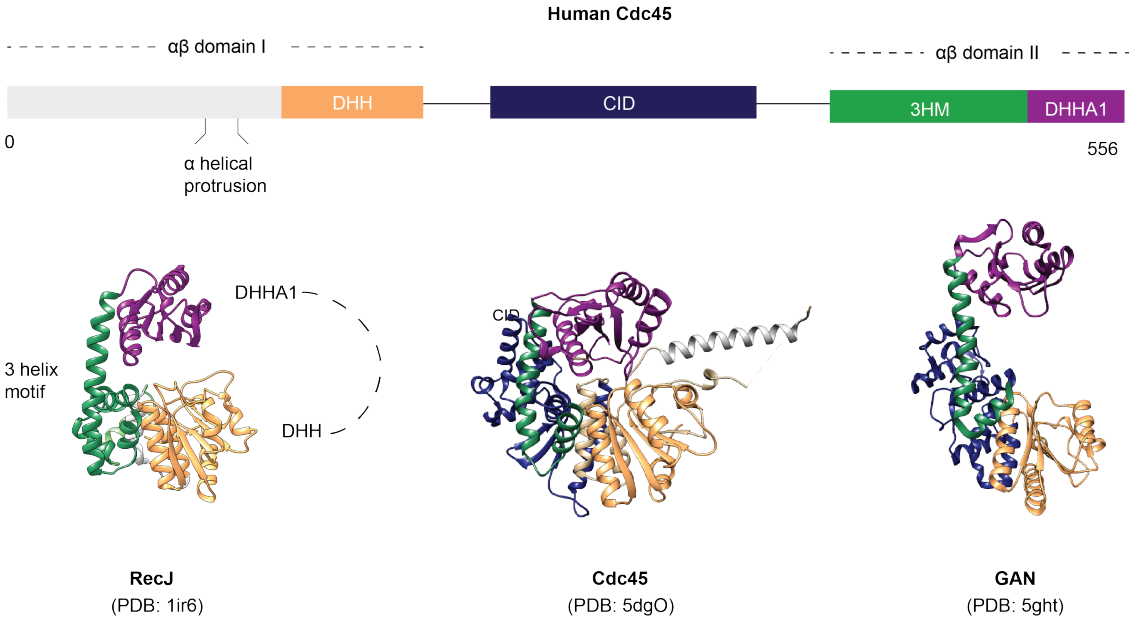


Figure 1.12: Domain organisation of human Cdc45 and structural comparison to RecJ and GAN
All three proteins share significant structural homology at the DDH and DHHA1 although in catalytically active RecJ and GAN their open configuration permits binding of DNA whereas intramolecular interactions in Cdc45 close this interface, partially explaining why Cdc45 is catalytically inactive.

In addition to the conserved RecJ fold, Cdc45 also contains a CMG interacting domain (CID) that binds Mcm2-5. However, the interaction surface with MCM is deemed insufficient to stabilise an MCM-Cdc45 complex, rationalising the need for the Sld3 chaperone to keep Cdc45 tethered ahead of GINS incorporation (Simon et al., 2016).

Archaeal GAN forms a processive 5'-3' single-stranded exonuclease that binds to Gins51 of the GINS heterodimer (Li et al., 2011) and shares high structural similarity with Cdc45 and RecJ (Oyama et al., 2016). In fact, GAN also contains a CID and the GAN-Gins51 interaction shows a resemblance to the binding mode of Cdc45 to the GINS Psf1 subunit (Oyama et al., 2016). A key difference, however, is that the GAN C-terminal 3HM alpha helix bends backwards to fully uncover its active site (right panel, **Figure 1.12**). Taken together, this suggests that GAN activates the archaeal MCM in a similar manner to the eukaryotic counterpart but the distinct open conformation of GAN allows additional activity of the exonuclease in archaea.

Despite the inhibitory conformation of the DHH and DHHA1 domains in Cdc45, an alternative DNA binding site cannot be ruled out. Indeed, several studies have shown that Cdc45 can bind DNA (Krastanova et al., 2012, Szambowska et al., 2014,

Petojevic et al., 2015). In particular, upon opening of the CMG at the Mcm2-5 gate, the leading strand template has been shown to interact with Cdc45 (Petojevic et al., 2015) as depicted in **Figure 1.5c**.

1.4.4 CDK and the pre-LC complex

At the onset of S-phase, a rise in expression of specific cyclins (Clb5 and Clb6) leads to an increased activity of CDK (Siddiqui et al., 2013). Although CDK has been reported to phosphorylate Mcm4 (Devault et al., 2008), Sld2 and Sld3 represent the minimum targets required for S-phase progression (Tanaka et al., 2007, Zegerman and Diffley, 2007). In turn, phosphorylated Sld2 and Sld3 interact with the key origin firing factor Dpb11 to promote GINS and Polε recruitment (Tanaka and Araki, 2013). Sld3 is mainly phosphorylated at two Thr600 and Ser622 sites whereas Sld2 has 11 phosphorylation motifs, of which Thr84 phosphorylation appears to be essential for binding to Dpb11 (Tak et al., 2006, Masumoto et al., 2002, Zegerman and Diffley, 2007, Tanaka et al., 2007). In the case of Sld2, it is thought that prior CDK phosphorylation of all other sites is required for Thr84 to become accessible to CDK and that this sets a threshold for kinase activity to prevent premature association of Sld2 to Dpb11 (Tak et al., 2006).

DPB11 was initially identified as an essential gene in *S. cerevisiae*, which in high copy numbers suppresses mutations in the C-terminal *POL2* and *DPB2* genes of Polε (Araki et al., 1995). It has since become clear that Dpb11 does not form a constitutive complex with Polε but is rather part of a bigger and more transient pre-LC complex that contains Polε, GINS and Sld2 (Muramatsu et al., 2010). Formation of the pre-LC depends on CDK activity and is stabilised by Sld2 but is independent of DDK and can occur in solution away from origins (Muramatsu et al., 2010).

The domain architecture of Dpb11 is fundamental to pre-LC assembly. Dpb11 consists of phosphoprotein binding domains called the BRCT repeats (BRCT I-IV) (Zegerman and Diffley, 2007, Tanaka et al., 2007). CDK phosphorylated Sld3 and Sld2 bind to the Dpb11 N- (BRCT I/II) and C-terminal (BRCT III/IV) repeats, respectively, and GINS binds Dpb11 in between those two sites (Tanaka et al.,

2013). This enables Dpb11 to function as a scaffolding factor to control helicase assembly (discussed further in *section 1.4.6.2*).

1.4.4.1 GINS

The first identification of the GINS (Go, Ichi, Nii and San, Japanese for 5, 1, 2 and 3) hetero-tetramer all came in 2003 from three separate studies that employed different experimental approaches (Labib and Gambus, 2007). The first used a combination of *S. cerevisiae* genetic multi-copy suppressor screens, immunoprecipitation and co-purification techniques to identify that Psf (partner of Sld5) 1, 2, 3 form a stable complex with Sld5 (Takayama et al., 2003). This study showed that each GINS subunit is required for cell viability and that strains carrying *SLD5* or *PSF1* thermosensitive mutations showed impaired origin firing efficiency. Furthermore, chromatin immunoprecipitation experiments showed that GINS and Dpb11 association at origins is interdependent.

A second independent study used a heat inducible degron systematic screening approach in budding yeast to identify novel cell cycle factors that have essential but unclear functions (Kanemaki et al., 2003). This resulted in the co-discovery of Sld5 (Cdc105)-Psf1 (Cdc101)-Psf2 (Cdc102), with purification of Psf2 from yeast extracts further revealing the presence of Psf3 in the GINS complex. Additionally, this study showed that GINS associates with origins in early S-phase and that inactivation of Psf2 or Sld5 caused severely impaired progression of DNA replication, demonstrating that GINS is a key constituent at the replication fork. Later studies confirmed that GINS is required for unwinding, establishing that in the absence of this complex, MCM-Cdc45 remains in an inactive origin bound state and that GINS is required for RPA recruitment to single-stranded DNA (Kanemaki and Labib, 2006).

A third study in *Xenopus* egg extracts confirmed the GINS complex stoichiometry and showed that it forms a ring like architecture as shown by rotary shadowing EM (Kubota et al., 2003). Moreover, the authors demonstrated that GINS binds to chromatin in a CDK-dependent manner and interacts with Cdc45 and MCM on replicating chromatin. Only the full Sld5-Psf1-3 complex is sufficient for DNA replication in this system as demonstrated by depleting GINS from extracts and

adding all subunits back *in trans*. In contrast, a Sld5-Psf1-Psf2 subcomplex was insufficient to rescue DNA replication defects (Kubota et al., 2003).

The crystal structure of human GINS was solved by three different groups in 2007 and revealed the near full-length architecture of the complex apart from the C-terminus of Psf1, which was either omitted from the construct (Choi et al., 2007, Kamada et al., 2007) or invisible due to flexibility (Chang et al., 2007). Each GINS subunit forms two structural domains: an alpha helix bundle called the A-domain, containing 5 alpha helices (except Psf2 which has a β strand instead of the 4th alpha helix) and a β sheet rich B-domain, composed of an alpha helix and two antiparallel β strands (**Figure 1.13**). The Sld5 and Psf1 proteins follow an A→B domain order whereas the reverse B→A order is true for Psf2 and Psf3. Although not solved in the crystal structures, Psf1 B-domain by analogy to the other GINS subunits is expected to carry a β sheet rich fold. The stability of the complex is thought to be mediated in part by two completely conserved (from archaea to human) arginine and glutamate residues in the A-domain of each GINS subunit, which form a pair through hydrogen bonding within the internal, hydrophobic surface of the tetramer.

The structure of the GINS tetramer has been described as a 'trapezoid' where each subunit makes up a quarter of the complex (**Figure 1.13**)(Kamada et al., 2007). In this context, Sld5-Psf1 form the top level and Psf2-Psf3 the bottom level although each GINS subunit makes extensive contacts both horizontally and diagonally. Collectively, this arrangement yields a central, highly acidic constricted pore and gives the complex a pseudo 2-fold symmetry (**Figure 1.13**).

Although predicted to be structurally similar, the Sld5 and Psf1 B-domains fulfil different roles within the human GINS complex. Whereas Psf1 B-domain is dispensable for GINS assembly, Sld5 B-domain is completely buried within the tetramer, situated between the Sld5 and Psf2 A-domains. This interaction is key to the structural integrity of the complex as deletion of the Sld5 B-domain diminishes interactions between the top and bottom levels of GINS (Kamada et al., 2007). On the other hand, Psf1 B-domain is critical for CMG assembly as revealed by negative stain EM and biochemical studies of the full-length recombinant CMG complex (Costa et al., 2011, Costa et al., 2014), confirming previous observations that Psf1

lacking the B-domain do not replicate sperm chromatin even when added back *in trans* (Kamada et al., 2007).

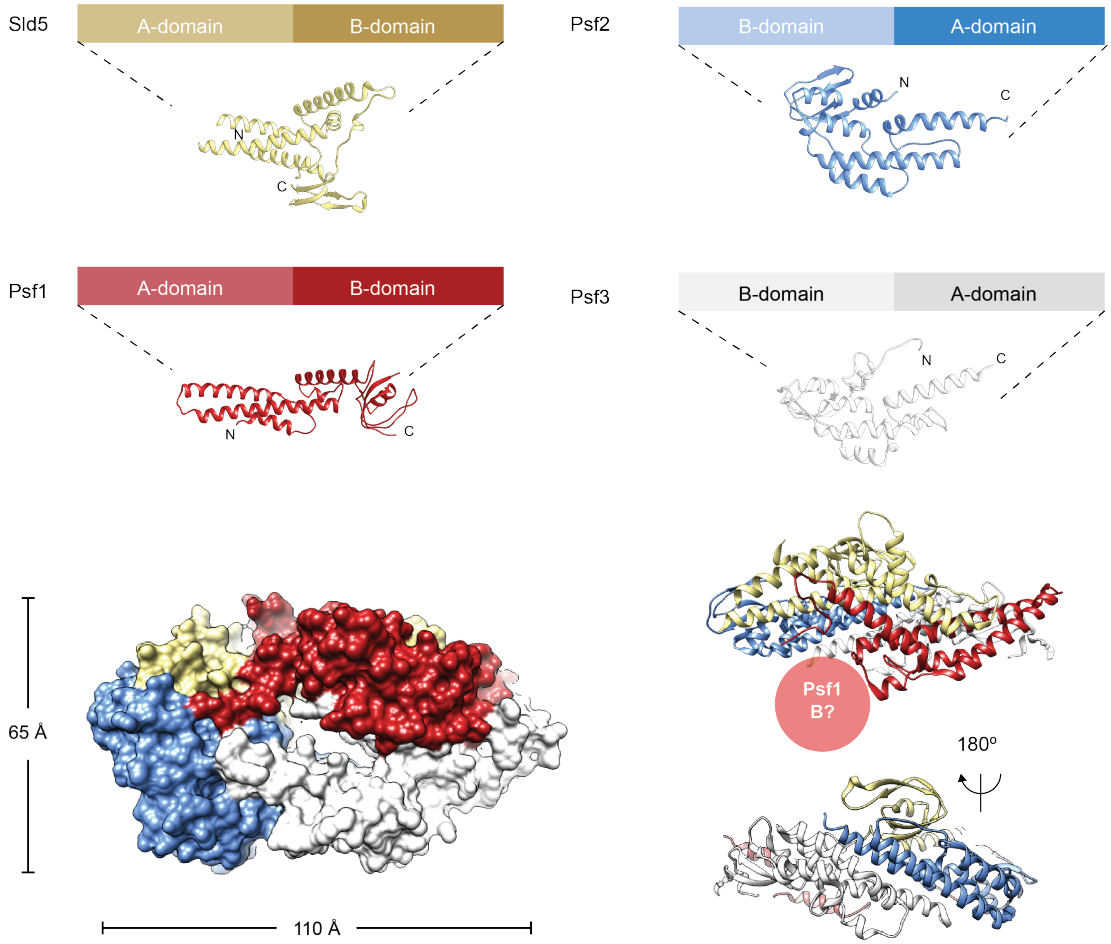


Figure 1.13: Crystal structure of human GINS
Domain architecture of each GINS subunit with the corresponding structure taken from the full GINS complex (PDB: 2Q9Q) except for Psf1 B-domain which is modelled based of the *Thermococcus* kodakarensis GINS structure (PDB: 3ANW). Lower panel left shows a surface view of GINS with a narrow pore in the middle. Right panel shows two views of the GINS complex with the Psf1 B-domain predicted position highlighted.

Aside from the essential binding of Psf1 B-domain to Cdc45 (Costa et al., 2014), Psf1 B-domain also interacts with the N-terminal domain of Dpb2 to enable GINS incorporation onto MCM-Cdc45 (Sengupta et al., 2013) although the mechanism of GINS recruitment to the helicase is poorly understood at the structural level.

The GINS proteins cluster into a highly diverse family of paralogous proteins that share common ancestry (Labib and Gambus, 2007). Most archaeal genomes encode a single copy of the GINS gene, either Gins15 (distantly related to Psf2-3) or Gins23

(distantly related to Sld5-Psf1). In some archaea (e.g. *P. furiosus* and *T. kodakaraenisis*) both Gins23 and Gins15 genes are present and the encoded heterodimers can associate to form a tetrameric architecture akin to the eukaryotic variant (Oyama et al., 2011, Yoshimochi et al., 2008). Each archaeal heterodimer complex seems to have been derived from each other by circular permutation of their A and B domains (Makarova et al., 2005). Although the early stages of GINS evolution are unknown, one possibility is that the four eukaryotic GINS genes evolved from one lone ancestral gene that underwent two rounds of gene duplication and an A→B domain swap (permutation) after the first duplication round (Makarova et al., 2005, Kamada, 2012).

GINS also interacts with the wider replisome to couple helicase unwinding with DNA synthesis (Gambus et al., 2009, Gambus et al., 2006). For example, Sld5 has been shown to bind the replisome trimer Ctf4, which acts as a hub that can bind multiple client proteins concomitantly, including Polα-primase (Simon et al., 2014, Gambus et al., 2009). GINS has also been reported to mediate helicase interactions with key factors at nascent forks, including Csm3/Tof1, FACT, Mrc1, Mcm10 and Top1 (Gambus et al., 2006).

A major aim of this study is to better delineate the roles of the Cdc45 and GINS subunits in initiation and origin firing. This will mainly be addressed through the structural characterisation of various helicase assembly intermediates and the fully formed CMG with or without bound Polε (Chapters 4 and 5).

1.4.4.2 Polymerase ε

Polε belongs to the B family of DNA polymerases and in both yeast and humans comprises four subunits: Pol2 (POLE1), Dpb2 (POLE2), Dpb3 (POLE3) and Dpb4 (POLE4) (Hogg and Johansson, 2012) (**Figure 1.14a**). Pol2 is made up of a catalytic N-terminal half (N-Pol2), where the 5'-3' polymerase and 3'-5' exonuclease activities reside (Hogg et al., 2014); and a non-catalytic C-terminal domain (C-Pol2), which is predicted to contain an inactive polymerase-exonuclease fold (Tahirov et al., 2009). Although C-Pol2 shares overall poor sequence identity with other homologues, it contains two highly conserved cysteine rich regions at the extreme C-terminus, which

coordinate binding to zinc and are integral to interactions with Dpb2 (ZF1 and 2) (Dua et al., 1998, Dua et al., 1999, Baranovskiy et al., 2017).

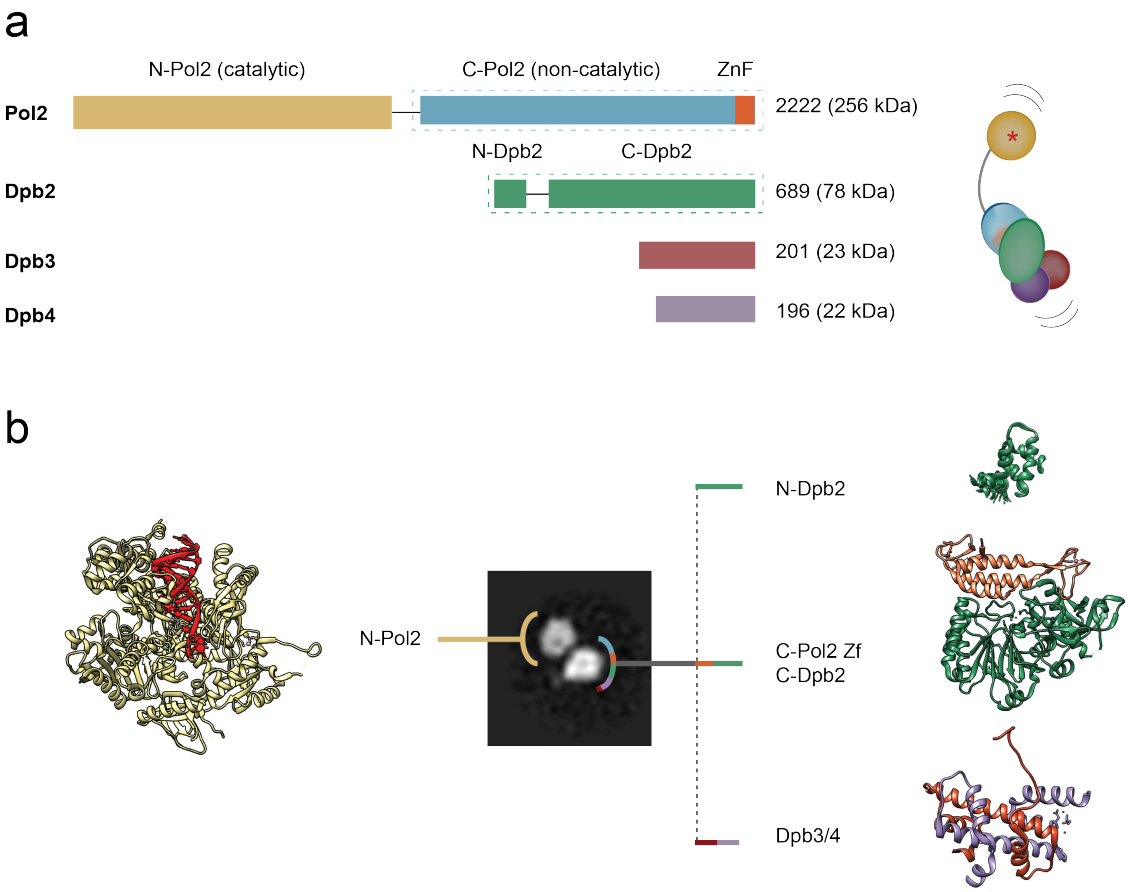


Figure 1.14: Domain architecture and structural characterisation of individual Polε domains
(a) Domain architecture of Polε. Dashed box represents indispensability. Cartoon representation depicts flexibility of the two Polε lobes as indicated by negative stain EM studies. **(b)** Negative stain EM 2D class average of wild-type Polε (Zhou et al., 2017) and crystal structures of N-Pol2 (PDB: 4m8o), N-Dpb2 (PDB: 2v6z), C-Pol2 Zf/C-Dpb2 (PDB: 5vbn) and Dpb3-Dpb4 (PDB: 5y26).

Dpb2 is a non-enzymatic protein that is thought to play a structural role in the complex (Isoz et al., 2012) and contains three domains: an N-terminal module (N-Dpb2), which comprises a fold reminiscent of the AAA+ lid domain (Nuutinen et al., 2008), a central OB-fold (Koonin et al., 2000, Baranovskiy et al., 2017), and a C-terminal (C-Dpb2) calcineurin-like phosphoesterase domain (Aravind and Koonin, 1998) (**Figure 1.14b**). A crystal structure of human Dpb2 bound to a region of C-Pol2 revealed a strong interaction between C-Dpb2 and zinc bound ZF2 (Baranovskiy et al., 2017) whereas N-Dpb2 is flexible and cannot be resolved as part of this Dpb2/C-Pol2 complex.

Dpb3-4 both contain a histone fold motif and these Polε subunits act together with Pol2/Dpb2 to confer a double stranded DNA binding ability (Tsubota et al., 2006) but do not stimulate the catalytic activity of Pol2 (Aksenova et al., 2010, Ganai et al., 2015). More recently it has been shown that Dpb3-4 form a H2A-H2B like dimer and play an important role in the inheritance of epigenetic marks (He et al., 2017, Yu et al., 2018).

Together, full-length monomeric Polε forms a ~379 kDa flexible structure as described by cryo-EM (Asturias et al., 2006). However, a succeeding crystal structure of N-Pol2 (Hogg et al., 2014)(**Figure 1.14b**, left) could not readily fit into the low-resolution cryo-EM density (Asturias et al., 2006). Therefore, the architecture of Polε was re-visited using dropout mutants and negative stain EM, revealing a bi-lobed density composed of N-Pol2 connected via thin linker to the non-catalytic C-Pol2, Dpb2 and Dpb3-4 subunits (Zhou et al., 2017) (**Figure 1.14b**, central panel).

Notably, only the non-catalytic C-Pol2 and Dpb2 modules are essential for viability (Isoz et al., 2012, Dua et al., 1999, Araki et al., 1991a) whereas N-Pol2 and Dpb3-4 can be deleted and still support DNA replication and cell growth in *S. cerevisiae, S. pombe* and *D. melanogaster* (Araki et al., 1991b, Ohya et al., 2000, Hogg et al., 2014, Feng and D'Urso, 2001, Suyari et al., 2012). Emerging biochemical and structural evidence are beginning to explain these conserved Polε phenotypes. First, and as mentioned previously, an interaction between N-Dpb2 and Psf1 B-domain has been identified (Sengupta et al., 2013), confirming observations from previous yeast two-hybrid assays (Takayama et al., 2003), and demonstrating that this non-catalytic domain of Polε is integral for GINS incorporation into the helicase. This is consistent with Polε-GINS recruitment as part of the same pre-LC complex to origins (Muramatsu et al., 2010) and the first demonstration that CMG and Polε directly interact (Langston et al., 2014). Second, negative stain EM structures of the CMG bound to Polε have shown that the non-catalytic subunits stably bind the AAA+ MCM tier while N-Pol2 is flexibly tethered and switches between two different conformations (Zhou et al., 2017) (discussed further in *section 1.5*). However, the precise contributions of each Polε subunit in CMG assembly and fork progression are unknown. In this study, I will attempt to address the molecular mechanism of

CMG activation by Polε by characterising a CMGε complex bound to a model replication fork using cryo-EM (Chapter 5).

1.4.5 Mcm10

MCM10 was discovered by independent *S. pombe* (referred to as *Cdc23*) and *S. cerevisiae* genetic screens (referred to as *DNA43* and *MCM10*) as a gene encoding a DNA replication factor that is essential for cell division (Nasmyth and Nurse, 1981, Dumas et al., 1982). A temperature sensitive allele of *DNA43* caused defective entry and progression through S-phase (Dumas et al., 1982) and other screens identified *mcm10* mutants that impair maintenance of minichromosomes (Maine et al., 1984) and initiation of DNA replication (Merchant et al., 1997). Since then, homologues have been found in all eukaryotic model organisms investigated from yeast to human (Wohlschlegel et al., 2002, Aves et al., 1998, Christensen and Tye, 2003) but not in archaea or bacteria.

Mcm10 is a non-enzymatic protein comprising an N-terminal domain, an internal domain (ID) and a C-terminal domain (Robertson et al., 2008, Thu and Bielinsky, 2013). The NTD contains a conserved coiled-coil domain, which mediates self-oligomerisation of Mcm10 into dimers and trimers (Du et al., 2013). The ID domain is a highly conserved region containing an OB-fold and ZF, which promote strong Mcm10 binding to DNA (Warren et al., 2009, Warren et al., 2008), with a preference for single-stranded DNA (Fien et al., 2004, Di Perna et al., 2013, Eisenberg et al., 2009). The ID domain also mediates protein-protein interactions to ORC, Mcm2-7, Cdc45, PCNA, Polα and Ctf4, thereby placing Mcm10 both at origins and advancing replication forks (Christensen and Tye, 2003, Di Perna et al., 2013, Ricke and Bielinsky, 2004, Das-Bradoo et al., 2006, Zhu et al., 2007). Finally, the CTD (conserved amongst metazoan species) contains a WHD, ZF and zinc ribbon module, which are collectively involved in DNA binding and localisation of Mcm10 to the nucleus (Sharma et al., 2010, Thu and Bielinsky, 2013).

Despite its discovery over 30 years ago, the timing and mechanism of Mcm10 function during DNA replication is still unclear (Thu and Bielinsky, 2013). Early evidence for Mcm10 as a key player in DNA replication initiation stemmed from the

observation of a dramatic decrease in firing of origins in an *mcm10-1* temperature sensitive strain (Merchant et al., 1997). Through cellular fractionation of *S. cerevisiae* extracts, a recent study subsequently showed that endogenous Mcm10 co-purifies with chromatin bound double hexamers (Quan et al., 2015). Using an *mcm10* mutant deficient in MCM binding, the authors observed delayed double hexamer separation and thus predicted that Mcm10 is required to remodel the two tethered MCM rings into single, replicative helicases. Consistent with such a role, direct binding of Mcm10 to Mcm2 and Mcm6 on licensed origins has been demonstrated using purified budding yeast proteins (Douglas and Diffley, 2016). This is further in keeping with earlier genetic and biochemical evidence that Mcm10 interacts with pre-RCs in *S. cerevisiae*, *D. melanogaster* and *X. laevis* (van Deursen et al., 2012, Christensen and Tye, 2003, Ricke and Bielinsky, 2004, Wohlschlegel et al., 2002).

The role of Mcm10 in CMG assembly is controversial. On the one hand, studies using *Xenopus* egg extracts reported that depletion of Mcm10 leads to a reduction in Cdc45 association to sperm chromatin that could be rescued by addition of Mcm10 *in trans* (Wohlschlegel et al., 2002). This is consistent with reports in *S. cerevisiae* of reduced Cdc45 binding to ARS1 origins in an *mcm10-1* mutant and another report showing an Mcm10 (alongside RecQL4 and Ctf4) dependent role for stable CMG assembly in human HeLa cells (Im et al., 2009). On the other hand, *in vitro* reconstitution of origin firing using purified budding yeast proteins (Yeeles et al., 2017, Yeeles et al., 2015) and yeast extract experiments (Heller et al., 2011) have demonstrated that Mcm10 is dispensable for CMG assembly, in line with reports showing that depletion of budding yeast Mcm10 using an inducible degradation system does not impair Cdc45 or GINS recruitment *in vivo* (van Deursen et al., 2012, Watase et al., 2012).

A number of studies have shown that Mcm10 is essential for helicase unwinding and replication fork establishment by the CMG. For example, Mcm10 depletion from *Xenopus* egg extracts leads to impaired recruitment of the single-stranded DNA binding protein RPA, which is used as a proxy for defective DNA unwinding (Wohlschlegel et al., 2002). This observation is consistent with omission of Mcm10 decreasing RPA recruitment and abolishing DNA replication *in vitro* (Yeeles et al.,

2015) and degradation of Mcm10 abrogating chromatin immunoprecipitation of RPA with Sld5 (van Deursen et al., 2012).

Recently, a model was suggested for Mcm10 mediated CMG activation. Looke *et al.* showed that Mcm10 binds to a conserved N-terminal motif of Mcm2; a mutation was identified within this region of Mcm2 that restored growth in cells depleted of Mcm10 (Looke et al., 2017). This suggests that Mcm10's essential role in origin firing is to remodel the CMG complex but the mechanism is poorly understood (roles of Mcm10 in origin firing are summarised in *section 1.4.6.3*). This study further connected Mcm10 to replication fork elongation, identifying an *mcm10* separation of function mutant that was capable of activating DNA unwinding by CMG but was defective in promoting replication elongation, consistent with previous data showing that Mcm10 is a constituent of the replisome (Gambus et al., 2006, Pacek et al., 2006, Ricke and Bielinsky, 2004).

1.4.6 Mechanism of origin firing

1.4.6.1 Recruitment of the first helicase co-activator: Cdc45

After origin licensing (step 1, **Figure 1.15**), the first step in pre-IC formation is the recruitment of the Sld3/7-Cdc45 complex in a DDK dependent but CDK independent manner (step 2, **Figure 1.15**) (Walter, 2000, Yeeles et al., 2015, Deegan et al., 2016, Heller et al., 2011, Deegan and Diffley, 2016, Kanemaki and Labib, 2006, Kamimura et al., 2001). Sld3 is a reader and binder of phospho-Mcm4,6, which by association allows the recruitment of Cdc45 to origins (Deegan et al., 2016). Specifically, DDK poly-phosphorylation of Mcm4 and Mcm6 yields a negatively charged patch that promotes binding of Sld3 via its numerous positively charged lysines. In this context, Mcm4 is more important for Sld3 recruitment but Mcm4 is not adjacent to the Cdc45-binding Mcm2-5 sites within the same hexamer. Therefore, it is suggested that Sld3 chaperones Cdc45 to Mcm2 of the opposite hexamer, which could act as a safeguarding mechanism to ensure that Cdc45 binding only occurs on licensed origins (Deegan et al., 2016).

Notably, a combination of two crystal structures (N-Sld3/N-Sld7 and C-Sld7) indicate that Sld3/7 form a dimer of hetero-dimers where two copies of Sld3 are bridged by

an Sld7 dimer (Itou et al., 2015). This stoichiometry may have important implications for the concomitant recruitment of two Cdc45 molecules to the two Mcm2-5 sites of the pre-RC. In this context, the two N-terminal Sld3 domains form convex shaped molecules that are predicted to bind acidic residues on the concave surface of two Cdc45 copies (Itou et al., 2014).

1.4.6.2 Recruitment of the second helicase co-activator: GINS

At the G1/S transition, CDK activity increases and this is required for phosphorylation of Sld2 and Sld3 to promote their interaction with the scaffold Dpb11 (Zegerman and Diffley, 2007, Tanaka et al., 2007). A pre-LC complex composed of Dpb11, Sld2, GINS and Polε is subsequently formed (step 3, **Figure 1.15**) (Muramatsu et al., 2010). In turn, CDK phosphorylated Sld3, bound to MCM-Cdc45, brings the pre-LC to origins via Dpb11 to stably recruit GINS and Polε at the replisome (Tanaka and Araki, 2013). As such, Sld3 integrates both DDK and CDK signals during CMG assembly.

Several pre-LC interactions are key to ensure GINS recruitment to MCM. GINS binds in between the two pairs of BRCT repeats in Dpb11 and is therefore sandwiched between Sld2 and Sld3 at origins. Deletion of the GINS binding region in Dpb11 significantly impairs DNA replication initiation (Tanaka et al., 2013). Second, C-Pol2 within the pre-LC interacts with Dpb11-Sld2 (Muramatsu et al., 2010) whereas Psf1 has been shown to bind to Dpb2 (Sengupta et al., 2013, Takayama et al., 2003). Therefore, it is thought that Polε links the Sld2-Dpb11 complex to GINS (Muramatsu et al., 2010).

While Sld2, Sld3 and Dpb11 are released ahead of DNA replication initiation (Kanemaki and Labib, 2006, Masumoto et al., 2000, Gambus et al., 2006), GINS and Polε maintain stable association throughout replication fork progression (Sengupta et al., 2013). The presence of Polε and GINS within the same pre-LC complex could ensure an inherent coupling mechanism of unwinding and synthesis functions at the onset of origin firing and could also be a means to pre-determine the architecture of the replisome (discussed further in *section 1.5.4.2*).

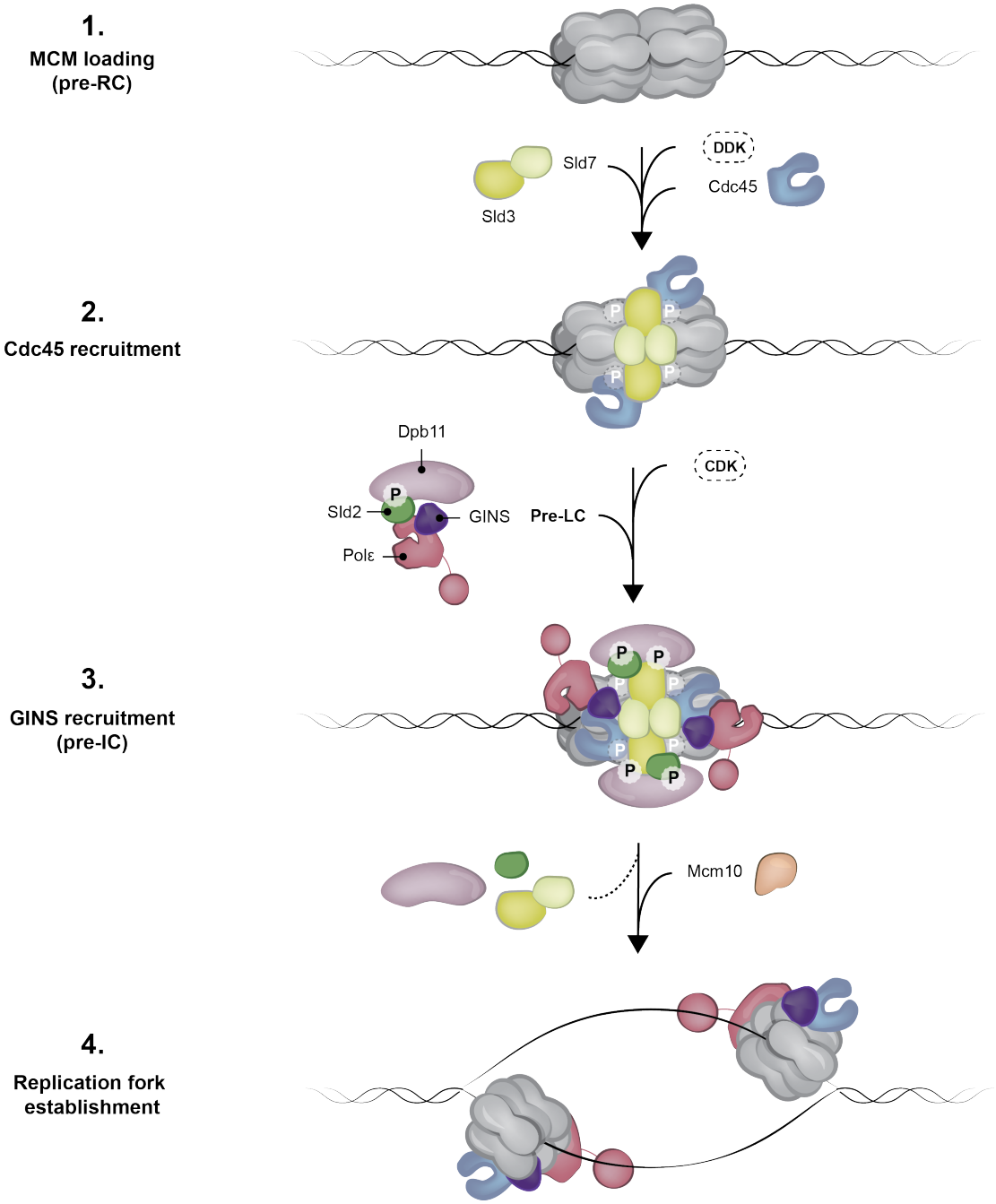


Figure 1.15: A model for step-wise assembly and activation of the eukaryotic replicative helicase
DDK functions ahead of CDK to recruit the first helicase co-activator Cdc45. CDK phosphorylates Sld2 and Sld3, thereby promoting recruitment of the pre-LC to origins. This enables GINS and Polε helicase incorporation. Mcm10 is required for CMG activation and functions downstream of helicase assembly. See section 1.4.6 for full description of model.

1.4.6.3 Helicase activation by Mcm10

The current view is that Mcm10 functions after CMG assembly to promote ATP-dependent fork unwinding although its precise mechanism of action remains

enigmatic (step 4, **Figure 1.5**) (Yeeles et al., 2015, van Deursen et al., 2012). Encompassing the main genetic and biochemical data, three non-mutually exclusive models have been suggested for the basis of Mcm10-mediated CMG activation. Mcm10 (1) splits the CMG double hexamers (Quan et al., 2015) and/or (2) reconfigures the CMG towards an active conformation (Looke et al., 2017) and/or (3) stabilises single-stranded DNA binding by the CMG (Thu and Bielinsky, 2013, Yeeles et al., 2015).

Despite a growing understanding of the step-wise pathway that leads to CMG assembly, the molecular mechanism of helicase activation is unclear. In particular, the individual roles the origin firing factors play in double hexamer separation, origin melting and stabilisation of replication fork progression are poorly understood. It is also unclear whether DDK phosphorylation elicits symmetrical docking sites for Sld3 to enable the concerted recruitment of Cdc45 to the double hexamer and what the function of Polε is with respect to both helicase assembly and stabilisation of the replisome. These themes will be explored further in the results sections of this thesis.

1.5 The CMG helicase at the replication fork

Once activated, the CMG helicase unwinds DNA with a 3'-5' directionality in an ATP dependent manner and drives replisome progression (Bell and Botchan, 2013). However, the molecular basis for CMG strand separation remains poorly understood, partially because a structural understanding of CMG replication fork engagement is lacking. Furthermore, although the architecture of the minimal eukaryotic replisome is beginning to emerge from low-resolution EM studies (Sun et al., 2015, Zhou et al., 2017) a detailed understanding of the spatial relationship between CMG and the replicative polymerases is elusive. In this section, I will introduce the proposed mechanisms of CMG unwinding by drawing comparison to homo-hexameric helicases co-crystalised with single-stranded nucleic acids. Subsequently, I will provide an overview of progress made towards structural characterization of the eukaryotic replisome and outstanding questions relating to helicase-polymerase coupling.

1.5.1 Mechanistic models of helicase unwinding

Currently, there are three main mechanisms to explain how a hexameric molecular machine coordinates nucleotide turnover with substrate translocation: probabilistic, concerted or rotary cycling (Lyubimov et al., 2011)(**Figure 1.16**).

The probabilistic model predicts a stochastic mechanism with no specific sequence to active site firing. This model is primarily based on studies of the hexameric ClpX unfoldase, which does not depend on all six subunits for activity (Martin et al., 2005). Instead, the power stroke is produced by a single asynchronous ATP hydrolysis active site.

The concerted model, supported by the crystal structure of L-Tag helicase, suggests an 'all-or-none' firing mechanism where the six active sites bind, hydrolyse and release nucleotide concomitantly (Gai et al., 2004).

The rotary cycling mechanism is supported by several homo-hexameric helicases including T7 gp4 (Singleton et al., 2000), Rho (Thomsen and Berger, 2009), E1 (Enemark and Joshua-Tor, 2006) and DnaB (Carney et al., 2017) and argues that nucleotide hydrolysis occurs in an ordered, sequential manner from one active site to the next around the ring. The E1 and Rho structures both show a radial arrangement of different active site nucleotide states that appears to represent the entire hydrolysis cycle (nucleotide exchange state, ATP bound state, hydrolysis competent state) (Thomsen and Berger, 2009, Enemark and Joshua-Tor, 2006).

In addition, E1 and Rho exhibit a right-handed spiral staircase of ATPase pore loops which matches the chirality of the bound single-stranded nucleic acids (**Figure 1.17**). In the E1 helicase, the top, middle and bottom DNA binding loops correlate with sequential ATP, ADP and apo nucleotide states, respectively (Enemark and Joshua-Tor, 2006). This led to the suggestion of a staircase model for unwinding whereby successive clockwise ATP hydrolysis events around the ring drive conformational changes of the loops, which in turn 'escort' one DNA base at a time through the channel. In this mechanism, each DNA-binding loop maintains contact to one DNA phosphate and upon ATP hydrolysis the loop moves downwards. ADP release

causes the loop to reach the bottom of the staircase and disengage with the DNA. Upon re-binding of ATP, the loop resets its position to the top of the staircase to repeat this process. Each subunit binds, hydrolyses and releases nucleotide meaning that a full cycle around the ring requires 6 ATP hydrolysis and 6 ADP release events to translocate 6 nucleotides.

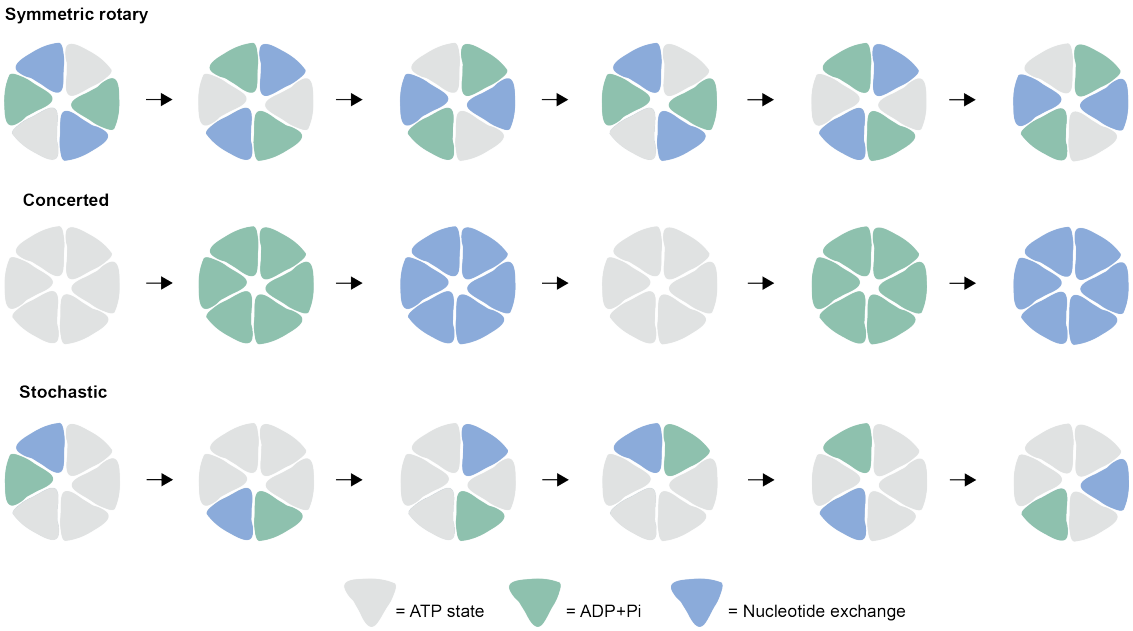


Figure 1.16: Models for ATP hydrolysis by hexameric helicases.
Symmetric rotary: ordered binding, hydrolysis and release of ATP around the ring. Concerted: simultaneous binding, hydrolysis and release of ATP. Stochastic: no specific order for ATP binding, hydrolysis and release. Based on summary in (Lyubimov et al., 2011)

Features of this mechanism is shared by the Rho and DnaB helicases where their 5'-3' translocation directionality is thought to be propagated by counter-clockwise sequential ATPase firing (Thomsen and Berger, 2009, Itsathitphaisarn et al., 2012). However, DnaB differs in that the bound DNA has a helical pitch that is much more extended than the nucleic acids co-crystalised with E1 and Rho (**Figure 1.17**) (Itsathitphaisarn et al., 2012). Indeed, the DnaB single-stranded DNA approximates the conformation of A-form DNA and two-phosphodiester bonds are bound compared to E1 and Rho's step-size of one nucleotide per subunit.

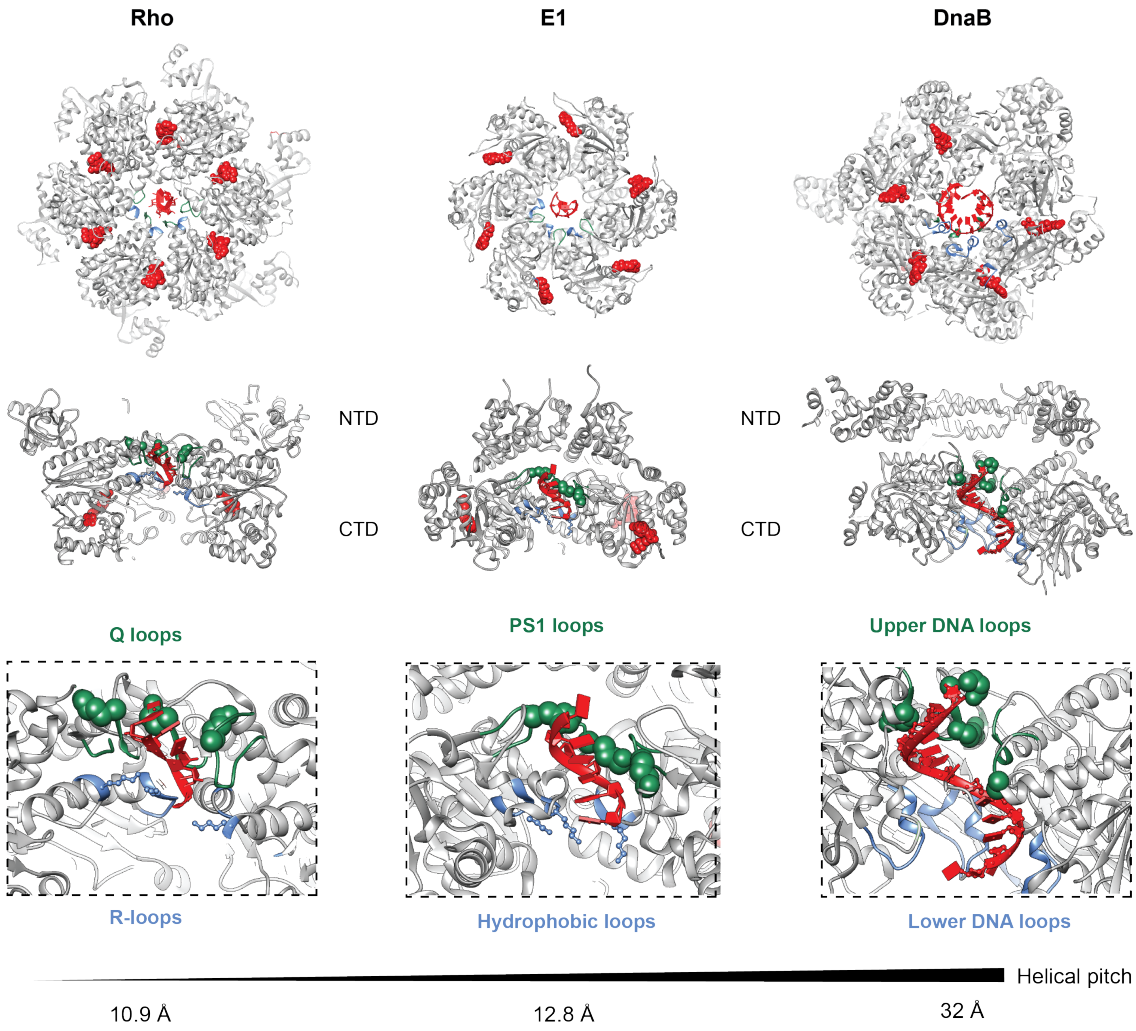


Figure 1.17: Co-crystal structures of hexameric helicases bound to single-stranded DNA or RNA. Comparison of top, side and central channel views between bacterial Rho transcription terminator (SF5) (PDB: 3ICE), E1 (SF3) (PDB: 2GXA) and DnaB (SF4) (PDB: 4ESV). Red densities in top views highlight nucleotide occupancy. In the lower panels, two types of DNA binding elements are observed in all cases, shown as 'spheres' (green) or 'sticks' (blue). All nucleic acids also form a right-handed spiral which match the upper DNA binding loops of Rho and E1.

Like other replicative helicases, Mcm2-7 forms a toroidal ring, which could enable coupling of sequential waves of ATP hydrolysis with propagation of conformational changes around the hexamer. However, unlike the rotary cycling homo-hexameric helicases discussed above, each subunit in the Mcm2-7 hetero-hexamer contribute unevenly to ATPase activity (Ilves et al., 2010). Indeed, mutagenesis experiments targeting each MCM ATPase active site revealed that only Mcm2-5 and Mcm5-3 are indispensable whereas the rest of the sites can be mutated with only a 20-50% reduction in helicase activity (Ilves et al., 2010). Whether this functional asymmetry contributes to a different translocation mechanism by Mcm2-7 compared to traditional homo-hexamers remains to be studied.

1.5.2 The mechanism of helicase strand engagement

One key unresolved question early on in my PhD is how the CMG engages a replication fork. Multiple biochemical, structural and biophysical experiments converged to propose three main models for how the CMG unwinds DNA: steric exclusion, strand extrusion and ploughshare (**Figure 1.18**)(Takahashi et al., 2005, Bochman and Schwacha, 2009).

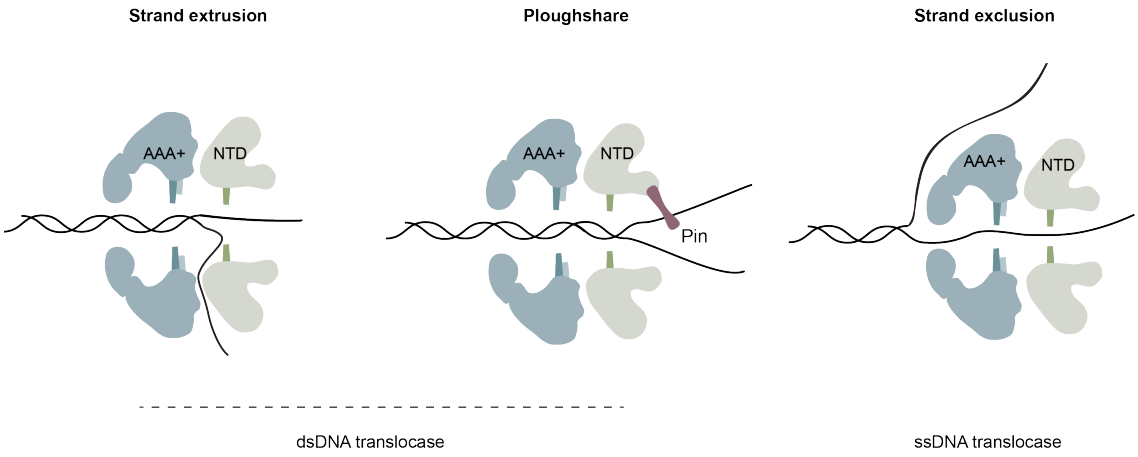


Figure 1.18: Models for CMG DNA fork engagement
Strand extrusion: AAA+ motor translocates on duplex DNA and unwinding occurs internally. Lagging strand template exits via a side channel formed between the two tiers whereas the lagging strand template comes out of the NTD pore. Ploughshare: helicase translocates on duplex DNA and a molecular wedge splits the two strands behind the helicase. Strand exclusion: CMG only translocates on single-stranded DNA and displaces the lagging strand template to the outside of the helicase.

1.5.2.1 Steric exclusion

The steric exclusion model posits that the helicase strictly translocates along single-stranded DNA while excluding the complementary strand from entering the channel (right, **Figure 1.18**). This is the prevailing view of how hexameric helicases function as evidenced by co-crystal structures of DnaB, E1 and Rho in **Figure 1.17** and biochemical studies of L-Tag (Yardimci et al., 2012, Itsathitphaisarn et al., 2012, Thomsen and Berger, 2009, Enemark and Joshua-Tor, 2008). This model for CMG unwinding is also supported by biochemical studies in *Xenopus* egg extracts which showed that the eukaryotic replisome can readily bypass a roadblock on the lagging strand template but not on the leading strand template (Fu et al., 2011). Furthermore, crosslinking studies of the CMG bound to DNA in the presence of a slowly

hydrolysable ATP analogue indicates a topological interaction with the leading strand template whereas the lagging strand template shows much weaker crosslinks on the external surface of the helicase (Petojevic et al., 2015).

1.5.2.2 Plough share

In contrast to the steric exclusion model, the ploughshare model envisions translocation along duplex DNA to induce the torsional strain required for DNA unwinding (Takahashi et al., 2005)(centre, **Figure 1.18**). The two strands are subsequently separated by a proteinaceous wedge at the point where DNA exits from the helicase channel akin to a plough pulled by a tractor. In line with this model, several crystal structures of bacterial DNA helicases, including the RecBCD repair helicase suggest a requirement for a 'pin' to unwind the double helix (Singleton et al., 2004). In RecBCD, two motor domains (RecB and RecD) pull duplex DNA over towards RecC, which contains a β hairpin that splits the two strands of DNA at the entry point to the RecBCD complex. However, in Mcm2-7 it is envisioned that the pin is towed behind the helicase (Takahashi et al., 2005). The negative-stain structure of CMG argues against Cdc45 or GINS forming the splitting pin, as these factors are situated to the side of the MCM ring away from the helicase pore (Costa et al., 2011). Whether an element within the MCM hexamer itself splits the two strands awaits structural characterisation of a CMG-DNA complex.

1.5.2.3 Strand extrusion

According to the strand extrusion mechanism for unwinding, the Mcm2-7 ring partially translocates on duplex DNA and only splits the two strands inside the channel (Li et al., 2015) (left, **Figure 1.18**). In this arrangement, the leading strand template passes through the entire length of the helicase channel whereas the lagging strand template egresses through a lateral opening formed between the two MCM tiers. Consistent with this model, the yeast double hexamer structure shows an opening between Mcm6 and Mcm2 at the neck between the NTD and AAA+ modules that could be used for extrusion of the lagging strand template (Li et al., 2015). This model is further supported by *Xenopus* egg extract DNA replication termination experiments which indicate that the CMG can translocate on both single and double-stranded DNA (Dewar et al., 2015), and *in vitro* unwinding experiments

using yeast CMG which show that the helicase is blocked by roadblocks on both the leading and lagging strand template (Langston and O'Donnell, 2017).

Irrespective of the mechanism of unwinding by the CMG, the displacement of the lagging strand template does not mean that it is insignificant to helicase function. Indeed, interactions of the 5' tail of the DNA with the external surface of archaeal MCM is essential to helicase activity (Graham et al., 2011, Graham et al., 2016, Rothenberg et al., 2007). In this context, it is thought that large-scale reconfigurations of the protruding A-domains expose a positively charged DNA binding site that enable the lagging strand template to wrap around the external surface of the helicase (the so-called 'strand-exclusion with wrapping' mechanism). It is unclear whether these A-domain conformational changes extend to the eukaryotic helicase. However, mutagenesis of external β hairpins from all six Mcm2-7 subunits as part of the *Drosophila* CMG complex led to a 95% reduction of DNA unwinding compared to wild-type (Petojevic et al., 2015).

1.5.3 Polarity of the helicase at the replication fork

While it is clear that bacterial and eukaryotic-based helicases track with 5'-3' and 3'-5' polarities, respectively, the orientation of the two helicase tiers with respect to the replication fork vary across helicase SFs (**Figure 1.19**). An understanding of helicase directionality is important to resolve how origin unwinding is initiated and how the replisome is organised for optimal coordination of helicase-polymerase activities. In the eukaryotic-based E1 helicase, the 3' end of the single-stranded DNA is proximal to the CTD, implying that the NTD tier tracks ahead of the motor domains upon translocation (**Figure 1.19**) (Enemark and Joshua-Tor, 2006). In contrast, bacterial helicases Rho and DnaB translocate with their CTD tier first with respect to the fork nexus, as implied by co-crystal structures of these helicases (Thomsen and Berger, 2009, Itsathitphaisarn et al., 2012). This orientation of CTD first directionality is shared by Mcm2-7 as shown by CMG binding to a 3' tailed substrate and FRET studies of archaeal MCM on a DNA fork substrate loaded using ADP (McGeoch et al., 2005, Costa et al., 2014). However, this has recently been challenged by helicase-fork DNA labelling experiments, showing that the CMG is positioned with its NTD tier towards the streptavidin-tagged duplex DNA end, albeit at a distance from

the fork nexus (Georgescu et al., 2017). It is possible that variations of model systems and non-productive fork binding lead to these discrepancies. One way to unambiguously address Mcm2-7 orientation on the replication fork is to activate the helicase step-wise from its duplex DNA origin loaded position to its unwinding mode, in the presence of ATP and covalently linked DNA roadblocks. The recent *in vitro* reconstitution of origin firing has now enabled this issue to be studied in this manner as will be shown in Chapter 4 (Yeeles et al., 2015).

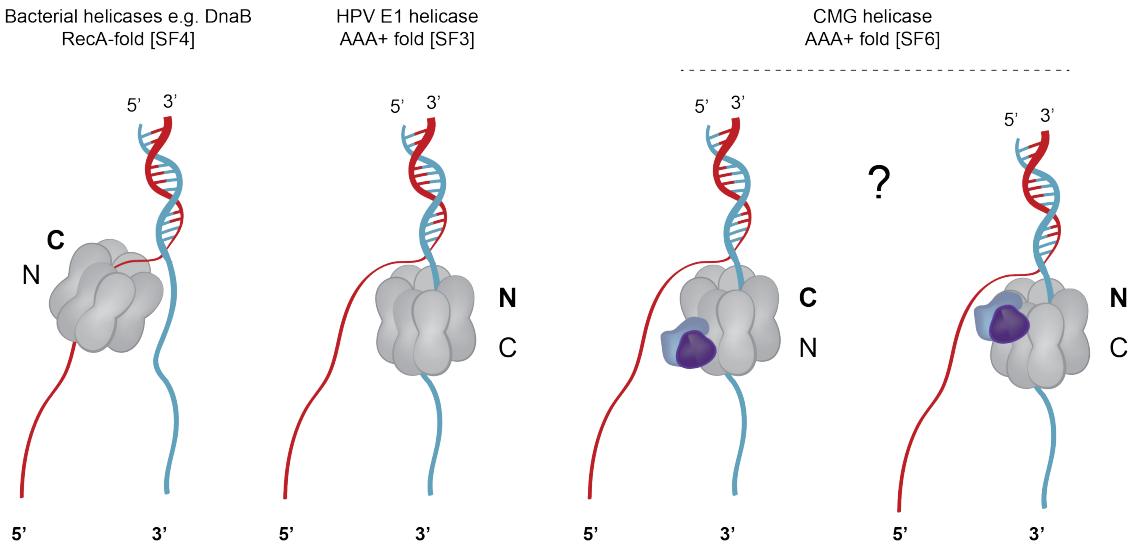


Figure 1.19: DNA binding polarity of hexameric helicases
Bacterial helicases translocate on the lagging strand template (5'-3') with their motor domain in front (CTD tier). All eukaryotic based AAA+ helicases studied so far translocate with a 3'-5' polarity but the orientation of the helicase relative to the fork nexus is a matter of debate. E1 translocates with its NTD tier first whereas the CMG orientation is still unclear. CMG CTD tier first is based on (Costa et al., 2014, McGeoch et al., 2005) and NTD tier first is based on (Georgescu et al., 2017).

1.5.4 The CMG helicase forms the core of the replisome

Replicative helicases function within the context of the replisome, a dynamic multi-protein complex that includes enzymes, scaffolding proteins and co-factors (Bell and Botchan, 2013, Burgers and Kunkel, 2017, Yao and O'Donnell, 2016). At a minimum, eukaryotic replication *in vitro* requires all factors that load, assemble and activate the helicase (CMG), a single-stranded DNA binding protein (RPA), a lagging strand polymerase-primase (Polα-primase), a leading strand polymerase (Polε) and topoisomerase II for removal of DNA supercoils (Yeeles et al., 2015). However, for *in vivo* rates of replication fork progression, replisomes further require a dedicated lagging strand polymerase (Polδ); sliding clamps (PCNA), which form hexameric assemblies that anchor the polymerases to DNA; clamp loaders (RFC), which load

PCNA molecules onto DNA; fork stabilization factors (Mrc1 and Csm3/Tof1), which stimulate helicase activity during unperturbed replication and promote fork pausing during perturbed replisome progression; and histone chaperones (e.g. FACT), which are important for reassembly of nucleosomes on replicated DNA (Yeeles et al., 2017, Kurat et al., 2017).

Whereas bacterial replisomes contain two identical polymerases for synthesis of daughter strands, the eukaryotic replisome uses two distinct polymerases (Polε and Polδ) for bulk chromosome duplication (Burgers and Kunkel, 2017). However, polymerase association to target strands is not strict in eukaryotes, with recent demonstration that Polδ can also extend the leading strand template ahead of Polε take-over (Yeeles et al., 2017, Johnson et al., 2015). This polymerase switching mechanism necessitates a high level of plasticity, and therefore, the replisome is not only compositionally complex but also highly flexible. As such, structural studies of the replisome so far have focused on the components that are tethered to the CMG helicase: Polε and Polα via Ctf4 (**Figure 1.20**) (Sun et al., 2015, Zhou et al., 2017).

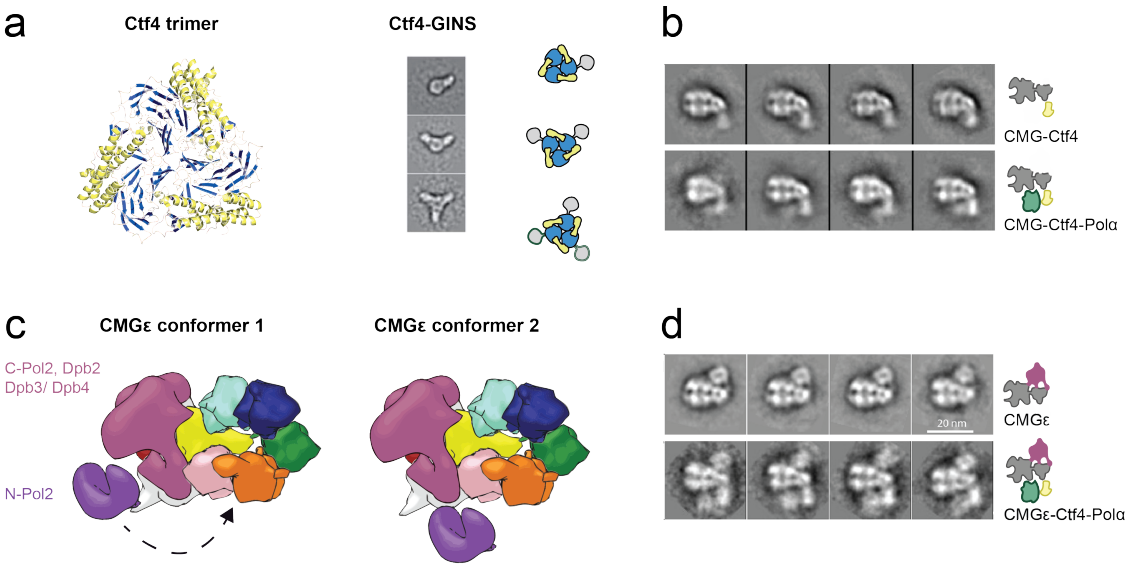


Figure 1.20: Structural characterisation of a minimal replisome
(a) Panel on the left: crystal structure of *S. cerevisiae* Ctf4 (construct lacking the WD40 domains) revealing homo-trimerisation of the protein. Panel on the right: negative stain EM 2D class averages of Ctf4 bound to 1, 2 or 3 GINS molecules simultaneously. This panel is adapted from (Simon et al., 2014). **(b)** Negative stain EM 2D class averages of CMG bound to Ctf4 (khaki) and flexibly linked Polα (green) via MCM NTD tier. Panel adapted from (Sun et al., 2015). **(c)** Cartoon representation of negative stain EM 3D reconstruction of CMGε in two different conformations. Based on (Zhou et al., 2017). **(d)** Negative stain EM 2D classes of CMG bound to Polε (pink density, via MCM AAA+) and Polα-Ctf4 (via MCM NTD). Panel adapted from (Sun et al., 2015).

1.5.4.1 CMG-Polα architecture

After activation of CMG unwinding, Polα-primase synthesizes *de novo* RNA-DNA primers that are extended by the leading and lagging strand polymerases (Frick and Richardson, 2001, Pellegrini, 2012). The hetero-dimeric primase initially synthesises a short RNA primer and the catalytic core of Polα mediates the initial extension of this primer prior to Polε and Polδ bulk synthesis.

Polα-primase remains associated to the replisome throughout fork progression (Gambus et al., 2006) as the semi-discontinuous mode of DNA replication dictates regular priming of DNA synthesis on the lagging strand template (Frick and Richardson, 2001). Studies in the past decade have established that Polα's association with the helicase requires the Ctf4 factor (Gambus et al., 2009, Simon et al., 2014). In this role, Ctf4 constitutively associates with the Sld5 subunit of GINS (Samora et al., 2016) and recruits the N-terminal catalytic subunit of Polα to mediate stable helicase-polymerase coupling (Simon et al., 2014). Since the identification of the important role of Ctf4 in this process, structural studies have provided the molecular basis for Polα recruitment to the helicase (Simon et al., 2014, Sun et al., 2015). Crystallographic studies showed that Ctf4 forms a β-propeller domain that can constitutively oligomerise into a homo-trimer (**Figure 1.20a**) (Simon et al., 2014). A common peptide motif was found C-terminal to the trimerisation domain, which is present in both Sld5 and Pol1 subunit of Polα. EM studies subsequently showed that Ctf4 can link up to three client proteins at one time, explaining how Ctf4 could act as a replisome bridge (**Figure 1.20a**) (Simon et al., 2014). Thereafter, EM studies of the full CMG-Ctf4-Polα assembly revealed that Ctf4 associates at the NTD tier of the CMG (via GINS) whereas a highly mobile Polα appears directly below the MCM-NTD pore, suitably positioned to capture the lagging strand template (**Figure 1.20b**) (Sun et al., 2015).

1.5.4.2 CMG-Polε architecture

In contrast to Polα, and as mentioned previously, CMG engages Polε directly. Multiple lines of evidence point to this conclusion, including the Polε-GINS simultaneous recruitment as part of the pre-LC complex (Muramatsu et al., 2010), yeast two hybrid experiments inferring a Dpb2-Psf1 interaction (Takayama et al.,

2003), the direct demonstration of N-Dpb2-Psf1 B-domain binding (Sengupta et al., 2013) and the reconstitution of the full 15-subunit CMGε holo-enzyme complex (Langston et al., 2014). Subsequent negative stain EM studies have shown that Polε indeed anchors to the MCM AAA+ tier of helicase, decorating the same Mcm2-5-3 side of the ring as Cdc45 and GINS (**Figure 1.20c**) (Zhou et al., 2017). This mode of binding allows both Polα and Polε to concomitantly engage the helicase as shown by 2D classes of the 20 subunit CMG-Polε-Polα-Ctf4 assembly (Sun et al., 2015)(**Figure 1.20d**). Interestingly, the opposite location of polymerases on either side of the CMG contrasts the classical view of the replisome, where the helicase forms the leading edge of fork progression and the polymerases follow behind (O'Donnell et al., 2013).

In a separate study by our laboratory, mild cross-linking conditions allowed reconstitution and structural characterisation of the full CMGε complex (**Figure 1.20c**) (Zhou et al., 2017). Drop out mutants were further used to establish that Polε associates with the helicase via a fixed lobe that comprises the non-catalytic domains (C-Pol2, Dpb2, Dpb3/4) whereas the catalytic N-Pol2 switches between a compact (Cdc45 binding) and extended conformation (periphery of the MCM ring) (**Figure 1.20c**). Assuming the CMG is situated at the fork nexus with its AAA+ tier first, then these structural data indicate that Polε is tethered in front of the helicase toroid, with N-Pol2 able to extend to capture the leading strand template, whereas Polα is situated at the back of the helicase to initiate daughter strand synthesis. The flexibility of N-Pol2 observed may also enable a polymerase switching mechanism suggested by a recent biochemical study to make room for Polδ to initially extend the leading strand, and may also provide the space for initial PCNA loading onto the DNA (Yeeles et al., 2017). All the while, the fixed binding of the non-catalytic, essential Polε domains to CMG ensures consistent helicase-polymerase coupling for maintenance of genome stability throughout replication fork progression.

Several mechanistic questions remain with regards to the essential CMG-Polε association that I will attempt to address in Chapter 5: (1) how does Polε structurally support CMG assembly? (2) what are the contributions of the individual Polε subunits to the stability of the CMGε holoenzyme? (3) how does the CMGε complex engage a DNA fork and does Polε remain associated during fork unwinding?

1.6 Aims

In summary, this thesis aims to address the mechanism of Mcm2-7 activation and the molecular basis for helicase fork engagement. To this end, I used a combined biochemical, cryo and negative stain EM approach using both *S. cerevisiae* and *Drosophila* model systems to structurally and functionally characterise various helicase intermediates. Chapter 3 provides an overview of the inactive licensed origin of replication, which acts as a base to then understand how DDK phosphorylation and subsequent origin firing factor recruitment structurally reconfigures the helicase for activation. Chapter 4 builds on these observations and entails a description of an active CMG helicase bound to a model replication fork. This provides insights into the roles of Cdc45 and GINS in stabilizing the helicase and the mode of DNA binding by the CMG. By comparing a CMG structure to the double hexamer structure from Chapter 3, a model for origin melting is further proposed. The final section of this chapter establishes the orientation of the active helicase at a replication fork. In Chapter 5, the spatial relationship between CMG and the essential Polε subunits is structurally characterised, revealing the multi-faceted roles of C-Pol2 and Dpb2 in both promoting productive replication fork engagement by the CMG and structurally bridging GINS and MCM to form a stable holo-helicase complex.

Chapter 2. Materials & Methods

2.1 Reagents

2.1.1 *S. cerevisiae* strains

Table 2.1: Strains used in this study.

Name	Genotype	Reference
Polymerase epsilon Pol2 exo- yAE99	*MATa ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100 bar1::Hyg pep4::KanMX ura3::URA3-Gal1-10 Dpb2 + Dpb3 Pol2::Pol2-FLAG-NAT trp1::TRP1-Gal1-10 Pol2 exo- + Dpb4-CBP*	This study
Polymerase epsilon wild-type yAJ2	*MATa pep4::KanMx4 bar1::Hph-NT1 trp1-1::TRP1pRS304/Pol2-GAL1-10-DPB4-TEV-CBP ura3-1::URA3pRS306/DPB2-GAL1-10-DPB3*	(Zhou et al., 2017)
Polymerase epsilon Pol2Δ1262 yAJ25 (Δcat)	*MATa ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100 bar1::Hyg pep4::KanMX ura3::URA3pRS306/Dpb2, Dpb3 trp1::TRP1pRS304/Pol2Δ1262, Dpb4-Tev-CBP*	(Zhou et al., 2017)

CMG (double FLAG tag) yJCZ2	*MATa pep4::KanMx4 bar1::Hph-NT1 trp1-1::TRP1pJF3 (MCM5-GAL1-10-MCM4) leu2-3::LEU2pJF4 (MCM7-GAL1-10-MCM6) ura3-1::URA3pJF6 (MCM2-GAL1-10–3xFLAG-MCM3) CDT1::CDT1-TAPTCP(HIS3)*/yJCZ1: *MATα* pep4*::KanMx4 bar1::Hph-NT1 his3-11::HIS3pJY13 (CDC45^{iFLAG2}-GAL1-10-GAL4) trp1-1::TRP1pJCZ1 (PSF1-GAL1-10-SLD5) leu2-3::LEU2pJCZ2 (PSF2-GAL1-10-PSF3) ade2-1::pJCZ3 (ADE2)*.	(Zhou et al., 2017)
CMG (CBP, FLAG tag) yJCZ3	*MATa pep4::KanMx4 bar1::Hph-NT1 trp1-1::TRP1pJF3 (MCM5-GAL1-10-MCM4) leu2-3::LEU2pJF4 (MCM7-GAL1-10-MCM6) ura3-1::URA3pRS306/MCM2-GAL1-10-CBP-TEV-MCM3*/yJCZ1: *MATα* pep4*::KanMx4 bar1::Hph-NT1 his3-11::HIS3pJY13 (CDC45^{iFLAG2}-GAL1-10-GAL4) trp1-1::TRP1pJCZ1 (PSF1-GAL1-10-SLD5) leu2-3::LEU2pJCZ2 (PSF2-GAL1-10-PSF3) ade2-1::pJCZ3 (ADE2)*	(Zhou et al., 2017)

2.1.2 Oligonucleotides

Table 2.2: Oligonucleotides used in this study

Name	Sequence	Purpose
*Dm*CMG leading strand template	5′-CACTCGGGCTCGTTTTACAACGTCGTGACTGGGCACTTGATCGGCCAACCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3′	Binding to DmCMG
*Dm*CMG lagging strand template	5′-CTGGCGTCGGGTCGGCGGTTGGCCGATCAAGTGCCCAGTCACGACGTTGTAAAACGAGCCCGAGTG-3′	Binding to DmCMG
*Sc*CMGε leading strand template	5’-/5deSBioTEG/GCAGCCacgctGGCCGTTTTACAACGTCGTGACTGGGCACTTGATCGGCCAACCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3’	Binding to ScCMGε
*Sc*CMGε lagging strand template	5' – GGCAGGCAGGCAGGCAGGCAGGCCGTGCGCGTGGTCGTGCGGTTGGCCGATCAAGTGCCCAGTCACGACGTTGTAAAACGGCCagcgtGGCTGC-3’	Binding to ScCMGe
CMG helicase unwinding lagging strand template	5’-GGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGTTGGCGAATTCCCATTGCCTCAGCAGATATCACCTCAGCACGACGTTGTAAAACGAG-3’	Helicase unwinding assay
CMG helicase unwinding leading strand template	5’-GGCTCGTTTTACAACGTCGTGCTGAGGTGATATCTGCTGAGGCAATGGGAATTCGCCAACCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3’	Helicase unwinding assay

2.1.3 Purification and reconstitution buffers

Table 2.3: Buffers used in this study

Buffer Name	Components	Use
Buffer C	25 mM Hepes pH 7.6 10% Glycerol 0.02% Tween-20 1 mM EDTA 1 mM EGTA	Base for anti-FLAG and first MonoQ CMG purification steps
Buffer C Resuspension	Buffer C 15 mM KCl 0.4 mM PMSF 2 mM $MgCl_2$ 2 mM 2-mercaptoethanol Protease Inhibitor tablets (Roche Diagnostics)	Resuspension of cell pellets before snap freezing
Buffer C NGNT (No Glycerol, No Tween)	25 mM Hepes pH 7.6 1 mM EDTA 1 mM EGTA	Base for second MonoQ CMG purification step
Buffer C Dialysis (*Dm*)	25 mM Hepes 7.6 50 mM Sodium acetate 10 mM Magnesium acetate 1 mM DTT	*Dm*CMG Dialysis buffer
Buffer C Dialysis (*Sc*)	25 mM Hepes 7.6 100 mM Potassium acetate 10 mM Magnesium acetate 1 mM DTT	*Sc*CMG Dialysis buffer
Buffer E	25 mM Hepes pH 7.6 10% Glycerol 1 mM DTT	Base for all Polε purification steps

Buffer MCM	25 mM Hepes pH 7.6, 5 mM Magnesium acetate 100 mM Potassium glutamate 0.02% NP-40 substitute 10% glycerol	MCM-Cdt1 purification
Buffer ORC	25 mM Hepes pH 7.6, 300 mM KCl 0.05% NP-40 substitute 10% glycerol	ORC purification
Buffer A	50 mM K_2HPO_4/KH_2PO_4 pH 7.5 5 mM $MgCl_2$ 1% Triton X-100 1 mM DTT	Cdc6 purification
Buffer DDK	25 mM Hepes pH 7.6 400 mM NaCl 0.05% NP-40 substitute 10% glycerol	DDK purification
Buffer CD	40 mM Hepes-KOH pH 7.6 10% glycerol 0.02% NP-40-S	CDK purification
Buffer H	25 mM Hepes pH 7.6 10% glycerol 0.02% NP-40S 1 mM EDTA 1 mM DTT	Dpb11, Sld2, Sld3/7, Cdc45 purification
Buffer D	25 mM Tris-HCl pH 7.2 10% glycerol 1 mM DTT 0.02% NP40S	GINS purification

Buffer M	25 mM Tris-HCl pH 7.6 10% glycerol 0.01% NP40S	Mcm10 purification
DNA binding buffer	25 mM Hepes 7.6 1M NaCl 10% glycerol 0.01% NP-40 1 mM EDTA	CMGε-DNA reconstitution
Protein binding buffer	25 mM Hepes 7.6 100 mM K acetate 10 mM Mg acetate 10% glycerol 0.01% NP-40 1 mM DTT 2 mM ATPγS	CMGε-DNA reconstitution
CMGε elution buffer	25 mM Hepes 7.6 100 mM K acetate 10 mM Mg acetate 0.01% NP-40 1 mM DTT 2 mM ATPγS, 400 nM biotin	CMGε-DNA reconstitution
Buffer 1	10 mM Tris pH 7.2 1 mM EDTA 2 M NaCl	CMG assembly buffer
Buffer 2	10 mM Hepes pH 7.6 1 mM EDTA 1 M K-acetate	CMG assembly buffer
Buffer 3	10 mM Hepes pH 7.6 1 mM EDTA	CMG assembly buffer
Buffer X	25 mM Hepes pH 7.6 50 mM sodium acetate	Single molecule buffer

	10 mM magnesium acetate 10% glycerol 1 mM DTT 2 mM Trolox	

2.1.4 Media for budding yeast cells

S. cerevisiae cells were added to Yeast Peptone (YP) medium (1% yeast extract, 2% peptone) with added 2% raffinose during cell growth and 2% galactose for expression.

2.1.5 Media for insect cells

Sf9 and Hi5 cells were grown in Graces medium supplemented with 10% foetal calf serum.

2.2 Protein expression

2.2.1 Expression using the baculovirus-insect cell system

To prepare infected cell extracts for subsequent expression of *Drosophila* CMG, bacmids were made from pFastBac1 plasmids encoding each of the 11 helicase subunits using the Bac-to-Bac Vector Kit (ThermoFisher Scientific) (pFastBac1 plasmids kindly gifted by Dr. Michael R Botchan). To obtain the first viral passage (P1), bacmids at 1 µg/ml were transfected via the Mirus transfection reagent (Mirus) into 3 ml Sf9 cells at $1x10^6$ cells/ml in insect cell media (*section 2.1.5*) then plated on 6-well plates and incubated for 3 days at 27ºC. 500 µl of each harvested baculovirus was amplified in a suspension containing $1x10^6$ cells/ml Sf9 and insect cell media for 3 days at 27ºC. To make P3 baculoviruses of each of the subunits, P2 stocks were amplified in 100 ml Sf9 at $5x10^5$ cells/ml for 5 days grown in insect cell media. For protein expression, 200 ml of each of the P3 stocks were added into a total of 4L Hi5 cells at $1x10^6$ cells/ml, giving a multiplicity of 5 with the virus stock at a titre of 108 pfu/ml. The infected cells incubated at 27ºC for 72h before harvesting cells by centrifugation and washing the collected pellet with PBS buffer (137 mM NaCl, 2.7

mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) supplemented with 5 mM $MgCl_2$. Cells were resuspended in 200 ml Buffer C-resuspension (**Table 2.3**) and snap frozen before storage at -80ºC.

2.2.2 Expression of yeast strains

Yeast Extract-Peptone (YEP)-Dextrose plates were streaked with a loop-full of flash-frozen yeast glycerol stock and incubated for 48 hours at 30ºC to form colonies. To prepare the starter culture, a single colony was picked and inoculated into 50 ml YEP+2% raffinose media and left to incubate for 24 hours at 30ºC and 180 rpm. Cells were counted and $2x10^8$ cells/ml were added to a 2L flask containing 800 ml YEP media+100 ml 20% raffinose. Cells were left to grow overnight at 30ºC and 180 rpm. For the Polε strains, cells were arrested in G1 with alpha-factor the next day to a concentration of 100 ng/ml for 3 hours at 1-$3x10^7$ cells/ml. Cells were induced at 1-$3x10^7$ cells/ml with 100 ml 20% Galactose (2% final) for 3 hours at 30ºC and 180 rpm. The cells were harvested by centrifugation at 4000 rpm for 40 min at 4ºC in a Beckman JS-4.2 rotor (Beckman Coulter J6-MC Centrifuge). Each pellet was washed with 50 ml resuspension Buffer C for CMG or Buffer E supplemented with 400 mM K-acetate for Polε (**Table 2.3**) and spun again for 4000 rpm for 20 min at 4ºC in a Beckman Coulter Allegra® X-15R Centrifuge. Pellets were resuspended with 0.5 column volumes resuspension buffer and frozen dropwise into liquid nitrogen. Frozen droplets were crushed in a 6875D Freezer/Mill® Dual Chamber Cryogenic Grinderfreezer mill (SPEX SamplePrep) using 6 cycles, intensity 15 (Precool 1 min, run 2 min, break 1 min), operating at -80ºC.

2.3 Protein purification

2.3.1 *D. melanogaster* CMG purification

Resuspended cells were thawed and broken up with a dounce homogeniser (20-40 strokes). The salt concentration was brought up to 100 mM KCl and lysate was cleared by centrifugation at 15,000 rpm for 20 min in a JA 25.50 rotor placed in an Avanti J-26S XP centrifuge. Cleared extract was incubated with 2 ml anti-FLAG M2-agarose beads (Sigma Aldrich) (pre-equilibrated in Buffer C (**Table 2.3**) supplemented with 100 mM KCl (BufferC-100)) for 2.5 hours at 4ºC with end-over-end mixing. Flow-through was collected and beads were washed with 100 ml Buffer-

C 100 in a 20 ml Poly-Prep disposable column (BioRad). Beads were transferred to a 10 ml column (BioRad) where bound complexes were eluted in Buffer C-100 supplemented with 200 µg/ml flag peptide and one tablet protease inhibitor cocktail (Roche). The first elution step was performed with 5 ml elution buffer for 15 min at room temperature, elution was collected, and beads were incubated with another 5 ml for 10 min. Fractions were pooled and pumped through a Mono S HiTrap SP FF column equilibrated in Buffer C-100. The flow through and a further 4 ml of Buffer C-100 wash was collected and injected into a MonoQ 5/50. Bound complexes were eluted over a 20 column volume gradient in BufferC from 100 mM to 550 mM KCl supplemented with 1 mM DTT. 0.5 ml fractions were collected and the CMG fractions (eluted between 410-440 mM KCl) were pooled and diluted to 150 mM in buffer C that does not contain glycerol or Tween-20. Sample was injected into a MonoQ PC 1.6/5 column equilibrated in Buffer C NGNT (25 mM Hepes pH 7.6, 1 mM EDTA, 1 mM EGTA, 1 mM DTT) supplemented with 150 mM KCl. Elution was performed over 20 column volumes in a Buffer C NGNT gradient from 150 mM to 550 mM KCl and 0.075 ml fractions were collected. Peak fractions were pooled and dialysed overnight in Buffer C Dialysis (*Dm*) for 16h (25 mM Hepes pH 7.6, 50 mM sodium acetate, 10 mM magnesium acetate, 1 mM DTT). Protein concentration was determined using the predicted molecular weight and extinction coefficient of the CMG complex on a Nanodrop (ThermoFisher).

2.3.2 *S. cerevisiae* CMG purification

Powder from 14 L expression was resuspended with 200 ml Buffer C supplemented with 100 mM KCl (Buffer C-100) and four protease inhibitor tablets (Roche). The lysate was left to stir for 30 min at 4ºC to achieve a homogenous resuspension. Lysate was cleared by ultracentrifugation at 45,000 rpm for 1 hour at 4ºC (Ti45 rotor, Beckman Coulter Optima L-100 XP Ultracentrifuge). Cleared lysate was added to 4ml anti-FLAG M2 agarose beads pre-equilibrated in Buffer C-100. The bead-lysate mixture was incubated with end over rotation for 2.5 hours at 4ºC. Flow-through was collected by gravity flow from a 20ml poly-prep disposable column and the beads were subsequently washed with 50 column volumes Buffer C-100. CMG was eluted in two steps using Buffer C-100 supplemented with 500 µg/ml FLAG peptide and one protease inhibitor tablet (Roche). First, 5 ml of elution buffer was incubated with the

beads with end over rotation for 30 min at room temperature followed by collection of the elution. In a second step, 4 ml of elution buffer was incubated with the beads with end over rotation for 20 min. The elutions were pooled and ~9 ml was injected into SP Sepharose Fast Flow 1 ml column (GE Healthcare) attached to a MonoQ 5/50 GL 1 ml column (GE Healthcare) in an Äkta purifier (GE Healthcare) pre-equilibrated with Buffer C-100. The columns were washed with 15 ml Buffer C-100 and CMG was eluted after removal of SP Sepharose Fast Flow column over a 20 CV gradient from 100-550 mM KCl (Buffer C100-BufferC500). 500 µl fractions were collected and checked by SDS-PAGE stained with Coomassie InstantBlue™ (Expedeon). Peak fractions were pooled and diluted 1:3 with Buffer C and this sample was injected onto a MonoQ PC 1.6/5 connected to an Äkta Micro pre-equilibrated in Buffer C 150 (NGNT). The column was washed with 2 ml Buffer C150 prior to CMG elution over 20 CV from 150-550 mM KCl (BufferC150-BufferC550 NGNT). 75 µl fractions were collected and analysed by SDS-PAGE. Peak fractions were pooled and dialysed into 2L Buffer C-dialysis (Sc) (**Table 2.1.3**) using a 10,000 MWCO Slide-A-Lyzer® MINI Dialysis Unit (Thermo Scientific) overnight at 4ºC.

2.3.3 Polymerase epsilon (wild-type, ∆cat and exo-) purification

Powder from 10-15 L expression was resuspended with 250 ml Buffer E supplemented with 400 mM K-acetate (Buffer E-400) and protease inhibitors (Roche). Lysate was cleared by ultracentrifugation at 45,000 rpm for 1 hour at 4ºC (Ti45 rotor, Beckman Coulter Optima L-100 XP Ultracentrifuge). Cleared lysate was added to 3 ml calmodulin beads (GE Healthcare) pre-equilibrated in Buffer E-400 supplemented with 2 mM $CaCl_2$ and left to incubate with end over rotation for 2 hours at 4ºC. Flow through was collected and beads were washed with 100 column volumes of Buffer E-400 supplemented with 2 mM $CaCl_2$. Protein was eluted using elution Buffer E-400 supplemented with 2 mM EGTA and 2 mM EDTA. In successive steps 3, 2 and 1 ml of elution buffer were incubated with the beads at 4ºC for 30, 20 and 10 minutes, respectively. The elutions were pooled and ~6 ml was injected into SP Sepharose Fast Flow 1 ml column (GE Healthcare) attached to a MonoQ 5/50 GL 1 ml column (GE Healthcare) in an Äkta purifier (GE Healthcare) pre-equilibrated with Buffer E-400. Protein was washed with 20 column volumes Buffer E 400 and eluted after removal of SP Sepharose Fast Flow column over a 15 column volume gradient

(Buffer E 400-1000 mM KOAc) and 0.5 ml fractions were collected. The purest fractions were pooled and dialysed against Buffer E-400 using a 10,000 MWCO Slide-A-Lyzer® MINI Dialysis Unit (Thermo Scientific). For further purification, the dialysed samples were applied to either Superdex 200 Increase 3.2/300 or Superose 6 Increase 3.2/300 gel filtration columns on an Äkta Micro system (both GE Healthcare), collecting 75 µl fractions. Samples were assessed by SDS-PAGE analysis and the purest fractions were pooled and snap frozen before further use in reconstitution assays.

2.3.4 Purification of replication factors for *in vitro* reconstitution

All the experiments described in this section were carried out by Dr. Max Douglas. Strain genotypes are listed in the Frigola et al., 2013 and On et al., 2014 papers.

2.3.4.1 Cdt1-Mcm2-7

Budding yeast strain yAM33 grown in YEP media and 2% raffinose was arrested during G1 for 3 hours using alpha factor (100 ng/ml) (Frigola et al., 2013). Protein expression was induced with the addition of galactose (2% final at 30ºC) for 3 hours. Cells were harvested and lysed by freezer milling in buffer MCM (25 mM Hepes pH 7.6, 5 mM MgOAc, 100 mM K-glutamate, 0.02% NP-40 substitute and 10% glycerol) supplemented with protease inhibitors (Roche). Cell powder thawed at room temperature was supplemented with an equal volume buffer MCM. Cell lysate was cleared by ultracentrifugation (235,000xg for one hour at 4ºC). 2 mM $CaCl_2$ was added to cleared lysate and applied to calmodulin resin for affinity purification. Elution was performed by addition of 5 mM EDTA to buffer MCM and fractions were pooled and concentrated. Concentrated fraction was loaded onto a pre-equilibrated (Buffer MCM) superdex 200 gel filtration column. Fractions containing MCM-Cdt1 were concentrated and snap frozen in liquid nitrogen.

2.3.4.2 ORC

ORC was expressed using yeast strain ySDORC (Frigola et al., 2013) and cells grown in YEP media + 2% raffinose were arrested in G1 phase with the addition of alpha factor (100 ng/ml). 2% galactose was added for expression at 30ºC for 3 hours. Harvested cells were resuspended with buffer ORC (25 mM Hepes pH 7.6, 300 mM

KCl, 0.05% NP-40 substitute and 10% glycerol) supplemented with protease inhibitor tablets (Roche). Cell powder and buffer MCM were mixed in a 1:1 volume ratio. Cleared lysate was obtained by ultracentrifugation (235,000xg, 1 hour, 4ºC). Cleared lysate was supplemented with 2 mM $CaCl_2$ and calmodulin resin was used to affinity purify ORC. Elutions were performed with 5 mM EDTA supplemented to buffer ORC and fractions containing ORC were pooled, concentrated and injected into a pre-equilibrated (buffer ORC) superdex 200 gel filtration column. Peak fractions were pooled, concentrated and dialysed against buffer ORC containing 200 mM KCl prior to snap freezing fraction in liquid nitrogen for future use.

2.3.4.3 GST-Cdc6

Cdc6 was expressed in BL21 DE3 Codon1 RIL bacterial cells (Stratgene) that had been transformed with pAM1 plasmid (Frigola et al., 2013). Expression was induced at an OD600 of 0.4-0.6 with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 hours at 25ºC. Cells were harvested and resuspended in buffer A (50 mM K_2HPO_4/KH_2PO_4 pH 7.5, 5 mM $MgCl_2$, 1% Triton X-100 and 1 mM DTT) supplemented with 2 mM ATP, 150 mM KOAc, protease inhibitor tablets (Roche) and 100 mg/ml lysozyme. Resuspension was incubated for 30 min on ice and sonicated. Cleared lysate was obtained by centrifugation at 20,000xg for 30 min. Cleared lysate was added to glutathione Sepharose resin (GE Healthcare) for 1 h at 4ºC. Column was washed with 20 column volumes of buffer A (with added 150 mM KOAc+2 mM ATP) and Cdc6 eluted by cleaving off from beads using the wash buffer supplemented with preScission protease (GE Healthcare) (2 hour incubation of protease with beads at 4ºC). The flow-through was collected and diluted down to reduce KOAc down to 75 mM concentration. This flow-through was subsequently added to pre-equilibrated hydroxyapatite resin (buffer A supplemented with 75 mM KOAc and 2 mM ATP). Column was first washed with buffer A + 2 mM ATP and then with buffer A + 150 mM KOAC + 15% glycerol. Cdc6 fractions were pooled and snap frozen in liquid nitrogen for future use.

2.3.4.4 DDK

DDK was expressed and purified from yeast strain ySDDK8 (On et al., 2014). To induce expression galactose was added to 2% at 30ºC for 5-8 hours. Cells were

harvested and lysed by freezer milling in buffer DDK (25 mM Hepes pH 7.6, 400 mM NaCl, 0.05% NP-40 substitute and 10% glycerol) supplemented with protease inhibitor tablets (Roche). Cell powder resulting from freezer milling was mixed in a 1:1 ratio with buffer DDK and the cleared lysate was obtained by ultracentrifugation at 235,000xg at 4ºC for 1 hour. The cleared lysate was supplemented with 2 mM $CaCl_2$ and added to calmodulin resin (Roche). DDK was eluted using 5 mM EDTA and fractions containing DDK were incubated with lambda phosphatase (NEB) for 1 hour at 4ºC. DDK fractions were concentrated and injected onto a Superdex 200 gel filtration column pre-equilibrated in buffer DDK. Peak fractions were combined and dialysed against buffer DDK comprising 200 mM K-glutamate. Protein fractions were concentrated and snap frozen in liquid nitrogen for future use.

2.3.4.5 CDK

The following sections (*2.3.4.5-12*) are adapted from, and based on, purification schemes and strain genotypes as outlined in Yeeles et al., 2015.

CDK was expressed and purified from yeast strain yAE37 (Yeeles et al., 2015). Cell powder (approximately 40 g) was resuspended with 80 ml buffer CD (40 mM Hepes-KOH pH 7.6, 10% glycerol, 0.02% NP-40-S) supplemented with 300 mM KOAc and protease inhibitor tablets (Roche). Cleared lysate was obtained by ultracentrifugation (235,000xg for 1 hour at 4ºC). $CaCl_2$ was added to cleared lysate and this mixture was incubated with calmodulin affinity resin (Agilent). The bead-lysate mixture was incubated for 1 hour at 4ºC. Resin was washed with buffer CD supplemented with 300 mM KOAc and 2 mM $CaCl_2$ extensively prior to elution overnight with TEV protease (100 µg/ml) in buffer CD supplemented with 2 mM $CaCl_2$. To remove TEV protease from the preparation, the elution was loaded onto a Talon resin column and the resulting flow-through was collected. The flow-through was then loaded onto a Superose 6 column (GE Healthcare) pre-equilibrated in buffer CD + 300 mM KOAc. Peak fractions were combined, concentrated and snap frozen for storage.

2.3.4.6 Dpb11

Dpb11-FLAG was expressed and purified from strain yJY26 (Yeeles et al., 2015). Cell powder was mixed in a 1:2 ratio with buffer H (25 mM Hepes pH 7.6, 10%

glycerol, 0.02% NP-40S, 1 mM EDTA, 1 mM DTT) supplemented with 500 mM KCl. Cleared lysate was obtained by ultracentrifugation at 235,000xg for 1 hour at 4ºC. Dpb11 was incubated with anti-Flag M2 affinity gel (Sigma) for 1.5 h at 4ºC. The resin was washed with 75-100 column volumes Buffer H supplemented with 500 mM KCl. Proteins were eluted by supplementing Buffer H+500 mM KCl with 0.5 mg/ml 3xFLAG peptide. Elutions were dialysed against buffer H + 150 mM KCl. After dialysis, the sample was loaded onto a MonoS column (1 ml, GE Healthcare) and Dpb11 was eluted using a 30 column volume gradient from 0.15-1 M KCl (buffer H). Peak fractions were pooled and dialysed against buffer H supplemented with 300 mM KOAc.

2.3.4.7 Polymerase ε

As described in *section 2.3.3*.

2.3.4.8 Sld2

FLAG-Sld2 was expressed and purified from strain yTD8 (Yeeles et al., 2015). Cell powder was thawed and mixed with buffer H supplemented with 500 mM KCl and protease inhibitor tablets (Roche) in a 1:3 ratio. Cleared lysate was obtained by ultracentrifugation at 235,000xg for 1 hour at 4ºC. Solid ammonium sulphate was supplemented to the supernatant to 32%. The mixture was stirred gently for 10 min at 4ºC followed by clearance of insoluble debris by centrifugation at 27,000xg for 20 min at 4ºC. At this stage the ammonium sulphate concentration was increased to 48% prior to further gentle stirring for 10 min at 4ºC. To collect the precipitated protein, the sample was spun down at 27,000xg for 20 min at 4ºC. The precipitated protein was resuspended with 1/3 of the volume buffer H supplemented with 500 mM KCl and protease inhibitor tablets. Sld2 was bound to anti-Flag M2 affinity gel resin for 30 min at 4ºC. Beads were extensively washed in buffer H supplemented with 500 mM KCl and protease inhibitor tablets (Roche). Sld2 was eluted in buffer H supplemented with 500 mM KCl and 0.5 mg/ml 3xFLAG peptide. Fractions containing Sld2 were combined and dialysed against buffer H supplemented with 280 mM KCl. Dialysed Sld2 was loaded onto a 1 ml HiTrap SP FF column (GE Healthcare), pre-equilibrated in buffer H supplemented with 250 mM KCl. An 18 column volume gradient from 0.25-1 M KCl in buffer H was used to elute Sld2 from

the column. Fractions were pooled and dialysed against buffer H supplemented with 350 mM KCl and 40% glycerol (to concentrate the protein). Fractions were snap frozen in liquid nitrogen for storage.

2.3.4.9 Sld3/7

TCP tagged Sld3/7 was expressed and purified using strain yTD6 (Yeeles et al., 2015). Powder was resuspended in a 1:1 ratio with buffer H supplemented with 500 mM KCl and protease inhibitors. Cell debris was pelleted by ultracentrifugation at 235,000xg for 1 hour at 4ºC. Sld3/7 was purified by addition of cleared lysate with IgG Sepharose 6 Fast Flow (GE Healthcare) resin for 40 min at 4ºC. Beads were washed with buffer H supplemented with 0.5 mM EDTA and 500 mM KCl. To elute Sld3/7, tobacco etch virus (TEV) protease was added to 50 µg/ml in buffer H supplemented with 0.5 mM EDTA and 500 mM KCl for 2 hours at 4ºC. TEV protease was removed by applying the sample onto Talon beads and collecting the flow-through. Flow-through was concentrated prior to loading on a pre-equilibrated Superdex 200 gel filtration column (buffer H supplemented with 500 mM KCl).

2.3.4.10 Cdc45

Cdc45 was expressed and purified using strain yJY13. Cell powder was resuspended with buffer H supplemented with 500 mM KOAc and protease inhibitor tablets (Roche) (no NP40). Cleared lysate was obtained by ultracentrifugation at 235,000xg for 1 hour at 4ºC and incubated with 8 ml anti-Flag M2 affinity resin (Sigma). Lysate and bead mixture was incubated for 4 hours at 4ºC. Beads were washed thoroughly in buffer H supplemented with 500 mM KOAc followed by buffer H supplemented with 300 mM KOAc. Cdc45 was eluted from the beads by adding 1 column volume buffer H supplemented with 300 mM KOAc and 0.5 mg/ml 3xFLAG peptide. The elution was dialysed against buffer P (150 mM KOAc, 0.5 mM DTT, 10% glycerol) supplemented with 20 mM K-phosphate pH 7.4 and loaded to a pre-equilibrated 1.5 ml hydroxyapatite column (Bio-rad) in the same buffer. Buffer P supplemented with 80 mM K-phosphate was used to was the column and Cdc45 was subsequently released with buffer P supplemented with 250 mM K-phosphate. A final dialysis of the elution was performed against buffer H with no NP40 supplemented with 300 mM KOAc prior to snap freezing in liquid nitrogen for storage.

2.3.4.11 GINS

GINS was expressed and purified from BL21 (DE3) Rosetta pLysS cells transformed with plasmid pFJD5 (Yeeles et al., 2015) (Psf3 contains an N-terminal His-tag). Cells were grown in LB media at 180 rpm and 37ºC to an OD600 of 0.5. To induce protein expression, IPTG was added to 1 mM for 2 hours. Cells were harvested by centrifugation and resuspended in buffer D (25 mM Tris-HCl pH 7.2, 10% glycerol, 1 mM DTT, 0.02% NP40S) supplemented with 400 mM NaCl, 10 mM imidazole and protease inhibitor tablets. To lyse, cells were sonicated and cleared lysate was obtained by centrifugation at 257,000xg for 30 min at 4ºC. Cleared lysate was added to Ni-NTA resin for 1 hour at 4ºC. Resin was washed with buffer D supplemented with 400 mM NaCl and 10 mM imidazole followed by buffer D supplemented with 100 mM NaCl and 200 mM imidazole. GINS containing fractions were pooled and dialysed against buffer D supplemented with 100 mM NaCl. Dialysed GINS was loaded onto a 1 ml MonoQ column equilibrated in buffer D supplemented with 100 mM NaCl and GINS was eluted using a 25 column volume gradient from 100-500 mM NaCl in buffer D. GINS fractions were combined, concentrated and loaded onto a Superdex 200 gel filtration column pre-equilibrated in buffer D supplemented with 150 mM NaCl. A final dialysis against buffer H supplemented with 200 mM KOAc was performed prior to snap freezing in liquid nitrogen.

2.3.4.12 Mcm10

Mcm10 with an N-terminal His-tag was expressed in BL21 (DE3) cells transformed with plasmid pET28a-Mcm10 (Yeeles et al, 2015, Douglas and Diffley, 2016). Cells were grown to an OD 600 of 0.6 at 37ºC in LB media. To express protein, IPTG was added to 1 mM and the temperature was reduced to 25ºC, and growth was continued for 3 hours at this temperature. Cells were collected and resuspended in buffer M (25 mM Tris-HCl pH 7.6, 10% glycerol, 0.01% NP40S) supplemented with 500 mM NaCl and protease inhibitors. Cells were sonicated and cleared lysate was obtained by centrifugation (257,000g for 30 min at 4ºC). Cleared lysate was applied to Ni-NTA agarose beads and resin was washed with 15 column volumes buffer M supplemented with 500 mM NaCl followed by 10 column volumes buffer M supplemented with 200 mM NaCl and 20 mM imidazole. To elute Mcm10, 200 mM NaCl and 200 mM imidazole was supplemented to buffer M and incubated with the

beads. The elution was loaded onto a 1 ml MonoS column (pre-equilibrated in buffer M supplemented with 200 mM NaCl). Mcm10 was eluted using a 20 CV gradient from 0.2-1M NaCl in Buffer M. Fractions from the resulting peak were pooled and diluted to 200 mM NaCl using buffer M and subsequently reapplied to a 1 ml MonoS column. Mcm10 was eluted as described and fractions were pooled prior to a final dialysis against buffer H supplemented with 0.01% NP40S and 200 mM K-glutamate.

2.4 Reconstitution reactions

2.4.1 DmCMG-DNA-ATPγS reconstitution

The oligonucleotides used to create a model replication fork were synthesized by Integrated DNA Technology (IDT). The two strands making up the fork (as listed in **Table 2.2**) were mixed in an equimolar manner to a final 50 µM stock concentration in a 50 µl volume. Annealing of the two strands was performed by heating the mixed oligos at 95°C for 3 min and slow cooling to room temperature for 60 min. This fork is designed to have a 50 bp duplex region, a 16 nt CG rich lagging strand template (5' end) and a 40 nt poly-T leading strand template (3' end).

To reconstitute the *Dm*CMG-DNA-ATPγS complex, forked DNA substrate was added to dialysed (buffer A, **table 2.1.3**) 300 nM CMG in a 2:1 molar ratio in the presence of 0.1 mM ATPγS (Sigma). To promote binding, the CMG–DNA–ATPγS mixture was incubated at 30 °C for two hours prior to cryo-grid preparation.

2.4.2 ScCMG-Polε-DNA-ATPγS reconstitution

The two oligo nucleotides that formed the model replication fork are listed in **Table 2.2** and were purchased from IDT. The two oligos were annealed as described in *section 2.4.1* except the desthiobiotin labelled oligonucleotide was added in a 1.2 fold molar excess with respect to the other strand.

To reconstitute the ScCMG-Polε-DNA-ATPγS complex, a bead-DNA affinity strategy was used. For each reaction, 7 µl M-280 Streptavidin Dynabeads™(Thermo Fisher) slurry was used and all incubations occurred at 30ºC with 1250 rpm shaking in a thermomixer. Beads were washed twice with 20 µl DNA binding buffer (buffers in this section listed in **Table 2.1.3**). 20 µl 250 nM fork DNA was added and incubated with

the beads for 30 min to promote attachment. Supernatant was collected and the beads were washed once with 20 µl DNA binding buffer and once with 20 µl protein binding buffer. 20 µl 250 nM CMG was added and incubated for 30 min. Supernatant containing unbound CMG was collected and 20 µl 80 nM Polε variant (either wild-type, ∆cat or exo-) was added to the beads and incubated for 20 min at 30ºC and 1250 rpm. Beads were washed once with 20 µl protein binding buffer and once with 20 µl protein binding buffer without glycerol. Complexes were eluted by addition of 10 µl elution buffer (containing biotin) and incubating for 30 min. Elutions were pooled and used for cryo-grid preparation.

2.4.3 Helicase assays

Helicase assays were performed by Dr. Jin Chuan Zhou using *D. melanogaster* or *S. cerevisiae* CMG purified by me. Purified CMG (123 to 410 fmol) was added to 20 µl reactions containing 25 mM Hepes, pH 7.6, 50 mM NaCl, 0.5 mM ATP, 10 mM MgOAc, 1 mM DTT, 0.1 µg/µL BSA, 27 µM ATPγS, and [γ-32P]-ATP–labeled fork. This mixture was incubated for 3 hours at 30ºC. 7 mM ATP was subsequently added to initiate DNA unwinding for 5 min at 30ºC. Reactions were quenched with buffer containing 150 mM Tris, pH 8.0, 3% SDS, 120 mM EDTA. The unwinding products were checked on an 8% non-denaturing PAGE gel running in 0.1% SDS and 1XTBE. The fork DNA construct used in this assay is listed in **Table 2.2**.

2.4.4 Single molecule DmCMG-DNA experiments

All single molecule experiments were carried out by Drs. Hailey Gahlon and David Rueda as described (Abid Ali et al., 2016). Briefly, oligonucleotides carrying fluorescent Cy3 and Cy5 dyes were annealed. Reaction chambers formed by quartz slides and coverslips, were incubated with 0.2 mg/ml BSA (Sigma) in buffer T50 containing 10 mM Tris-HCl pH 7.0 and 50 mM NaCl for 10 min. BSA was washed off with T50 supplemented with 0.2 mg/ml neutravidin and incubated for 10 min. Unbound neutravidin was washed off in buffer X containing 25 mM Hepes pH 7.6, 50 mM sodium acetate, 10 mM magnesium acetate, 10% glycerol, 1 mM DTT and 2 mM Trolox. To immobilize DNA, biotinylated, Cy3-Cy5 labelled template was incubated (25 pM) in the reaction chamber. Unbound DNA was washed off in Buffer X supplemented with 50 µM ATPγS or ATP, 7 nM CMG, and oxygen scavenging

reagents (10 mM 3,4-dihydroxybenzoic acid and 10 mM 3,4-dihydroxybenzoic acid). DNA molecules were analysed on a home-built prism-based total internal reflection fluorescence microscope. Measurements were recorded at room temperature via a green excitation operating at 1 mW and 30 ms time resolution. FRET efficiency was determined using the following equation: $FRET=I_A/(I_D+I_A)$, where A stands for acceptor intensity and D stands for donor intensity.

2.4.5 *In vitro* reconstitution of MCM double hexamers

Experiments described in this section were carried out by Dr. Max Douglas from Dr. John Diffley's group at the Francis Crick institute (Douglas et al., 2018). Protein purification of all factors used for the *in vitro* reconstitution of origin licensing and origin firing were carried out by Dr. Max Douglas as described in *section 2.3.4* (Yeeles et al., 2015).

The DNA used to load the MCM double hexamers was a 5' biotinylated 2.8 kb linear duplex DNA fragment containing the ARS1 origin. M280 streptavidin magnetic beads (ThermoFisher Scientific) were pre-washed twice with buffer 1 (10 mM Tris pH 7.2, 1 mM EDTA, 2 M NaCl). 50 ng DNA was added per 1 µl of bead slurry together with an equal volume of buffer 1. After incubation at room temperature, beads were washed twice in buffer 2 (10 mM Hepes pH 7.6, 1 mM EDTA, 1 M K-acetate) and twice with buffer 3 (10 mM Hepes pH 7.6, 1 mM EDTA). The beads were resuspended in buffer 3 at half the starting bead slurry volume.

Each reaction contained 5 µl bead-DNA, 7.5 nM ORC, 50 nM Cdc6, 100 nM Mcm2-7:Cdt1 and 5 mM ATP in a final volume of 80 µl containing buffer 4 (25 mM Hepes pH 7.6, 10% glycerol, 10 mM Mg-acetate, 0.2% NP-40 substitute). After incubation for 20 min at 30ºC, reactions were either mock treated or DDK was added to a concentration of 80 nM. Reactions were left to incubate for 30 min at 30ºC before two washes with high salt buffer (400 µl 25 mM Hepes 7.6, 5 mM Mg-acetate, 0.2% NP-40 substitute, 10% glycerol and 0.3 M KCl) and one wash with EM buffer (100 µl 25 mM Hepes 7.6, 5 mM Mg-acetate and 250 mM K-glutamate). 6 reactions per cryo-grid prepared were pooled and elutions were obtained after addition of 8 µl EM buffer

supplemented with 0.4 U/µl MseI restriction enzyme (New England Biolabs) at 30ºC for 10 min.

2.4.6 *In vitro* reconstitution of CMG assembly

Experiments described in this section were carried out by Dr. Max Douglas from Dr. John Diffley's group at the Francis Crick institute (Douglas et al., 2018). 60 ng linear DNA (as described above or with additional covalent roadblocks on each end) was immobilized on M-280 streptavidin magnetic beads (Invitrogen). ORC (37.5 nM), Cdc6 (50 nM) and 100 nM Mcm2-7:Cdt1 were added and incubated for 20 min at 30ºC. 50-100 nM DDK was subsequently added and the reaction was incubated for another 30 min at 30ºC. Supernatant was collected and the beads were washed with 250 mM K-glutamate, 25 mM Hepes pH 7.6, 10 mM Mg-acetate, 0.02% NP-40-S, 8% glycerol, 400 µg/ml BSA, 5 mM ATP, 1 mM DTT. The firing factors were pre-mixed to give a final concentration of 30 nM Sld3/7, 50 nM Cdc45, 20 nM Clb5-Cdc28, 50 nM Dpb11, 55 nM Sld2, 30 nM Polε, 200 nM GINS, 2.5 nM Mcm10 and added to the bead reactions for 8 min. Reactions were subsequently washed twice with 25 mM Hepes 7.6, 5 mM Mg-acetate, 10% (vol/vol) glycerol and 0.02% NP-40-S and once with 25 mM Hepes, 5 mM Mg-acetate, 250 mM K-glutamate. DNA bound complexes were released from beads by restriction enzyme digest with 0.1 U/µl MseI (NEB) in 5-10 µl buffer EM after incubation for 10 min at 30ºC. Cleaved fragment sizes were in the range of 1.5-2 kb.

2.4.7 *In vitro* replication assays

All *in vitro* replication reactions were performed by Dr. Max Douglas. Purification of additional replication factors listed in this section (PCNA, RFC, Polδ, Polα, RPA, TopI and Topo II) were performed exactly as described in (Yeeles et al., 2017). Replication reactions were also based on (Yeeles et al., 2017) apart from the DNA template used and incubation times. All reactions were performed at 30ºC. To initially load the MCMs on DNA, reactions were set up in buffer comprising 25 mM Hepes-KOH pH 7.6, 100 K-glutamate, 10 mM Mg-acetate, 100 µg/ml BSA, 1 mM DTT, 0.01% NP40S, 5 mM ATP, 45 nM Cdc6, 22.5 nM ORC, 100 nM Cdt1-Mcm2-7 and 4 nM circular DNA template (8.2 kb origin-containing DNA template) with a 20-minute incubation period. DDK was subsequently added to 25 nM for 20 minutes. The reaction volumes

were doubled by adding nucleotide to 200 µM CTP, GTP and UTP as well as 80 µM dCTP, dATP and dTTP and 33 nM α^{32}P-dCTP. To initiate replication, a master mix of replication proteins was added to final concentrations of 25 nM Sld3/7, 20 nM S-CDK, 50 nM Sld2, 30 nM Dpb11, 210 nM GINS, 40 nM Cdc45, 20 nM Polε (either wild-type, drop out variant or mutant Polε), 5 nM Mcm10, 20 nM RFC, 10 nM PCNA, 10 nM Topo I, 20 nM Topo II, 20 nM Ctf4, 100 nM RPA, 20 nM Polα, 10 nM Mrc1 and 10 nM Polδ. Following a 40 minute incubation, reactions were quenched using EDTA and the products were analysed using a 0.6% denaturing alkaline agarose gel.

2.5 Electron microscopy sample preparation

2.5.1 Negative stain grid preparation

Negative staining was performed on 400-mesh copper grids (Agar Scientific) with an additional film of carbon added to the grids. This was achieved by first evaporating carbon onto cleaved mica for 2 min at 46 mA using a ‘ramped carbon evaporation’ setting (Quorom Technologies) and floating the carbon from the cleaved mica onto the grids using a grid-coating trough. Grids were either stained in 2% uranyl acetate or 2% uranyl formate. Uranyl acetate stain was prepared by weighing out 100 mg uranyl acetate (U001 from Taab) and mixing with 5 ml MiliQ water. After a vortex mix and spinning down the solution in a micro centrifuge (20,000 x g for 5 min) the stain was filtered and wrapped in aluminum foil before use. Uranyl formate stain was prepared by weighing out 100 mg uranyl formate (U002 from Taab) and adding 5 ml of hot MiliQ water. The solution was left to mix on a vortex for 10 min. The pH of the uranyl formate solution was adjusted by addition of 20 µl 5 M NaOH (dropwise, 5 µl added four times and shaking in between). After another round of vortexing for 10 min, the uranyl formate solution was spun down and filtered. 3-4 µl drops of the sample was added to the grid and incubated for 1 min. Meanwhile, four 75 µl drops of stain were placed on parafilm. The sample was blotted almost to dryness prior to stirring the grid (held by tweezers) on top of the stain four stain droplets for 5, 10, 15 and 20s, respectively. Excess stain was blotted away and grids were left to dry before imaging.

2.5.2 Positive stain grid preparation

To positively stain the sample, two-week old 400-mesh copper grids covered with a layer of carbon were used (Agar Scientific). These grids were glow-discharged minimally for 10s at 45 mA (Electron Microscopy Sciences). 3 µl sample was added on top of the grids and incubated for 30s. Half the solution was blotted away prior to staining on one 75 µl uranyl acetate drop (stirring for 30s). To wash away the stain and promote positive staining, the grids were stirred on four sequential 75 µl milliQ water drops for 5s. Grids were blotted to dryness and stored before usage.

2.5.3 Cryo-EM grid preparation

Grids used for cryo-EM preparation were either open holes (Quantifoil R 1.2/1.3 and C-flat 1/1) or lacey grids with an ultrathin layer of carbon support (Agar Scientific). Grid preparation was performed in collaboration with Dr. Ludovic Renault and Dr. Julia Locke. An automated plunge freezer (FEI ThermoFisher Vitrobot mark IV) or Cp3 (Gatan) was used to plunge grids into liquid ethane at all times. 4 µl sample was always applied and grids were stored in liquid nitrogen prior to imaging. For all datasets, a humidity of 100% temperature of 22ºC was set. For the MCM double hexamer and CMGε datasets, a double-application process was adopted with a 0.5s blot in between the two applications. See **Table 2.4** for grid type, blotting and incubation parameters used for each dataset.

Table 2.4: Cryo-grid preparation parameters

	MCM DH (all datasets)	**CMG dataset 1**	**CMG dataset 2**	**CMGε (all datasets)**
Plunge freezer	Vitrobot (IV)	Vitrobot (IV)	Cp3	Vitrobot
Grid type	Lacey with carbon layer	QF 1.2/1.3	C-flat 1/1	Lacey with carbon layer
Glow discharge	1 min, 45 mA	30s, 45 mA	30s, 45 mA	1 min, 45 mA
1st application wait	2 min	30 s	2 min	2 min
2nd application wait	2 min	n/a	n/a	2 min
Final blot time	2 s	4 s	3 s	3 s

2.6 Electron microscopy image acquisition

2.6.1 Negative-stain data collection and cryo-grid screening

Negative-stain and cryo-grids were screened and imaged on a Tecnai Lab6 G2 Spirit transmission electron microscope operating at 120 keV (FEI ThermoFisher, EM-STP, The Francis Crick Institute). Micrographs were acquired using a 2K x 2K GATAN Ultrascan 100 camera at two different magnifications: 21,000 (4.92 Å pixel size) or 30,000x (3.45 Å pixel size). The defocus was kept to within a -0.5 to -2.5 µm range during collection and movements were tracked to prevent double exposures. Images were acquired manually using the digital micrograph 3 software (DM3, Gatan). Cryo-grids were inserted into the microscope using a Gatan 626 single tilt liquid nitrogen transfer holder (Gatan). Specimens were imaged with the low-dose settings, where the 'Focus' mode was used to search for ice thickness at 2100x mag and the exposure mode at 30,000x magnification was used to adjust focus to -4 µm. Images were acquired at 30,000x magnification and 1.0 s using the 'Expose' mode on the FEI user interface software.

2.6.2 Cryo-EM data collection

All cryo-grids were imaged on 300 keV electron microscopes, either FEI ThermoFisher Polara or Titan Krios equipped with a K2 summit direct electron detector operating in counting mode (Gatan) and an energy-filter in zero-loss mode (GIF Quantum, Gatan). Aside from the CMG cryo-datasets, which were collected manually using the Serial EM software, all other datasets were acquired using automatic data collection set up in EPU (ThermoFisher). See **Table 2.5** for image acquisition parameters for all datasets collected in this study.

Table 2.5: Movie acquisition parameters

	DH DDK (1&2)		**DH -DDK**	**CMG (1&2)**	**CMGε (1&2)**
Microscope	Krios	Krios	Krios	Polara	Krios
Institution	NeCEN	Diamond	Crick	STRUBI	Crick
Detector	K2 summit				
Frames	29	28	30	25	30

Exposure (s)	5.8	14	12	10	12
Dose ($e^-/Å^2$)	50	50	50	48	50
Movies	3,996	1,974	8,331	2,098	8,782
Pixel size (Å)	1.4	1.36	1.38	1.35	1.38

2.7 Electron microscopy image processing

2.7.1 A general pipeline for negative-stain image processing

All negative-stain EM datasets were processed in a similar manner. First, .dm3 micrographs were converted to .mrc using Bsoft bconvert option ('bconvert -output mrc *.dm3') (Heymann, 2018). Second, all particles were picked using EMAN2 (v2.07) in a semi-automatic manner (e2boxer.py 'Swarm' option) (Tang et al., 2007). The boxsize was set to 128 pixels for micrographs collected with a 3.45 Å/pix pixel size and 98 pixels for micrographs collected with a 4.92 Å/pix pixel size. Third, micrographs were imported into either RELION-1.4 (Scheres, 2015)(DmCMG negative stain dataset) or RELION-2 (Kimanius et al., 2016)(the rest of the negative stain data shown in this thesis) and all subsequent steps were performed using this image processing package. Within RELION, contrast transfer function (CTF) parameters were estimated for all micrographs using a Gctf wrapper (Zhang, 2016), setting the acceleration voltage to 120 keV, astigmatism amount between 200-500 Å, spherical aberration to 6.2 mm, the defocus range to -0.5 to -2.5 μm and the resolution limits to 75-15 Å, while leaving all other parameters as default. The Thon-ring images were inspected and the best CTF model fits compared to the experimental Thon rings were selected. Particles were extracted with the same box size as the picking step and subjected to reference free 2D classification. In general, 15-20 classes were used for every 10,000 particles. The regularisation parameter T was set to 4 and the additional argument '--only_flip_phases' was used to correct for CTF effects of the phases (phase flip). A circular mask of 320 Å was set for all CMG particles and 350 Å for all double hexamer datasets. After 2D classification, particles baring the same dimensions and features in different orientations were selected while discarding unrecognisable or 'junk' classes, which did not average. For 3D reconstruction, a previously published structure of CMG was sized to the correct pixel size and box size using the EMAN2's e2proc3d.py programme ('e2proc2d.py in.mrc out.mrc --clip=[pixel size division factor] --scale=128,128,128'). The map was

imported to RELION, filtered to 50 Å and used as a reference model against the best set of particles obtained from the 2D classification step. Resolution was judged by the 0.143 gold-standard Fourier shell correlation criterion.

2.7.2 Cryo-EM image processing and analysis

A generalized scheme was adopted for all cryo-EM datasets. Collected movies were corrected for beam induced motion and integrated using Motioncorr (CMG) or MotionCor2 (double hexamer and CMGε datasets). In the case of MotionCor2, 5x5 patch alignment was utilised. Non-dose-weighted micrographs were used for particle picking and CTF estimation (to enhance signal) and dose-weighted micrographs were used for particle extraction, 2D classification and all the 3D steps. EMAN2's e2boxer.py programme was used for all particle picking in a semi-automated manner (version 2.07). All CTF correction was either performed using CTFFIND4 (CMG dataset) or Gctf (all other datasets) within RELION-1.4 or 2.0 (whole frame correction). Particle extraction and 2D classification were also performed in RELION using the 'ignore CTFs until the first peak' switched on and --maxsig 5 added in the additional arguments option. 3D reconstruction was either performed in RELION or cryoSPARC (v0.5.6). The next three sections will provide a description of the image processing pipeline for the individual cryo-EM structures that are presented in this study. See **Table 2.6** for a summary of the main image processing parameters.

Table 2.6: Cryo-EM image processing parameters

	DH+DDK	**DH -DDK (1&2)**	**CMG (1&2)**	**CMGε (1&2)**
Motioncor version	MotionCor2		MotionCorr	MotionCor2
CTF correction	Gctf v1.06		CTFFIND3	Gctf v1.06
Particle picking	EMAN 2 v2.07			
2D and 3D classification	Relion 2.0		Relion 1.4	Relion 2.1
3D refinement	Relion 2.0 (initial) and cryoSPARC 1.0 (final)		Relion 1.4	Relion 2.1 (initial) and cryoSPARC 2.0 (final)
Raw particles	299,604	580,423	340,573	404,676

Final particles ensemble structure	90,918	52,319	60,287	n/a
Final particles DNA-bound structure	8,343	15,529	38,792 (compact ATPase), 13,692 (relaxed ATPase)	89,809
B-factor applied ensemble structure (Å)	-92.7	-162.246	-100	n/a
Resolution ensemble structure (Å)	4.3	4.8	7.2 (poor ATPase)	n/a
B-factor applied DNA-bound (Å)	-480.3	-163.3	-150	-100
Resolution DNA-bound (Å)	7.5	6.9	7.4 & 9.8	5.0
EMDB submission	3833	3834, 3960 (DNA-bound)	3318 (compact ATPase) 3319 (relaxed ATPase)	0288

Table 2.7: Model refinement parameters

Refinement (Phenix)	**Unmodified MCM (PDB: 6F0L)**	**CMGε-DNA (PDB: 6HV9)**
Initial model used (PDB code)	3ja8	5u8s
Model resolution (Å)	4.77	4.9
FSC threshold	0.143	0.143
R.m.s. deviations		
Bond lengths (Å)	0.34	0.012
Bond angles (°)	0.62	1.870
Validation		
MolProbity score	2.09	2.08
Clashscore	10.91	9.26
Poor rotamers (%)	0.21	0.48
Ramachandran plot		
Favored (%)	90.64	88.65
Allowed (%)	99.92	11.05
Disallowed (%)	0.08	0.29

2.7.2.1 Unmodified and DDK phosphorylated double hexamer datasets

MotionCor2 was used for correction of beam induced motion for the number of movies listed in **Table 2.5**. The first two frames were removed and 5x5 patch alignment was applied for the remaining frames prior to integration of movies (herein called micrographs). After semi-automatic picking of all particles from non-dose-weighted micrographs using EMAN2 (v2.07), whole-frame CTF correction was performed using Gctf (v1.06). The best quality micrographs were inspected and selected for further processing in RELION-2. Particles were first extracted from dose-weighted micrographs using bin by 2 parameters (256 pixel box size downscaled to 128 pix) giving a pixel size ranging from 2.7 to 2.76 Å for all double hexamer datasets. These particles were subjected to rounds of 2D classification, 3D classification (with and without alignment) and 3D refinement as shown in the flowcharts in **Figures 3.2** and **3.3**. For 3D refinement and 3D classification steps in RELION, the yeast MCM double hexamer cryo-EM structure was used as an initial model (filtered to 60 Å). The initial model was resized using EMAN2 e2proc3d.py ('e2proc3d.py in.mrc out.mrc --clip=[pixel size division factor] --scale=128,128,128'). The best subset of particles, defined in this case by isotropy of the 3D classes, DNA occupancy and high-resolution features, were reverted to unbinned parameters in RELION, yielding a 1.36-1.42 Å pixel range for all double hexamer datasets. Over-representation of views was mitigated by rounds of 3D classification and discarding the clearly anisotropic classes. To obtain the final structures, cryoSPARC was used for one round of ab initio 3D classification (3 classes, no symmetry imposition) and homogenous 3D refinement (with applied C2 symmetry) (Punjani et al., 2017). Final post-processing was performed in RELION-2, where a soft mask was applied and the map was sharpened.

To visualize the subset of DNA bound particles in the double hexamer datasets, a signal subtraction approach was adopted in RELION-2 based on previously published methods (Bai et al., 2015). First, the AAA+ tiers were segmented out of the 3D structure obtained from cryoSPARC using a combination of color zone and segmentation tools implemented in UCSF Chimera. This isolated density was used to create a mask in RELION-2 (mask filtered to 9 Å, 5 pixel drop off). The threshold for the mask was decided upon inspection of the map in Chimera; using the volume

viewer slider, the lowest possible threshold before the appearance of noise density was used as the binarization threshold input value in the Mask Create tab in RELION-2. The re-projected mask density was used to subtract the AAA+ tiers from the aligned experimental particles (using the signal subtraction function). This generated a new particle set, containing only the two NTD tiers and any density that occupies the channel. These particles were subjected to 2D classification, showing helical protrusion from a subset of particles that corresponds to DNA. These particles were selected, reverted back to their full original density in the 'signal subtraction' tab of RELION-2.1 and subjected to *ab initio* 3D classification and refinement in cryoSPARC. All structures were post-processed in RELION with B-factor values as stated in **Table 2.6**. Map local resolution was estimated using ResMap (Kucukelbir et al., 2014).

DNA modelling and real space refinement was performed by Dr. Valerie Pye (Francis Crick Institute). To create a template model for double-stranded DNA, the online make-na server was used where a random sequence was used as input. The generated model was manually adjusted in Coot against the cryo-EM density (Emsley et al., 2010). The Phenix.geometry_minimization tool was used to regularize model geometry of the DNA (Adams et al., 2010). The MCM model (PDB: 6F0L) was generated by docking in the previously published yeast MCM atomic coordinates (PDB: 3JA8) into the unphosphorylated MCM map, manually adjusting it in Coot (Emsley et al., 2010) followed by real-space refinement in Phenix (Adams et al., 2010) (**Table 2.7**).

2.7.2.2 CMG-DNA-ATPγS datasets

Movies were corrected for beam induced drift by whole frame alignment and subsequent integration using MotionCorr (first two frames were discarded) (Li et al., 2013). After semi-automatic picking of particles using EMAN2 (Tang et al., 2007), CTF parameters were estimated using CTFFIND4 (Rohou and Grigorieff, 2015). Low quality micrographs were discarded prior to further image processing in RELION-1.4 (Scheres, 2015). Particles were extracted with binned by 2 parameters, giving a pixel size of 2.7 Å/pixel and a box size of 112x112 pixels. After three rounds of 2D classification from both datasets, a set of highly recognisable CMG particles in a

range of orientations were selected whereas isolated Mcm2-7 subcomplexes were discarded (see **Table 2.6** for particle numbers). CMG particles from both datasets (60,287 and 44,665, respectively) were pooled (giving 104,952 particles) and subjected to series of 2D classification steps. 57,465 CMG particles from classes showing projections of secondary structure elements were selected and subjected to 3D refinement with the CMG-ATP cryo-EM structure as initial model (EMD-3321, filtered to 50 Å) to give a 7 Å structure. However, the ATPase tier was poorly resolved and as such CMG particles were subjected to 3D classification without alignment to separate out different conformational states in the particle population (10 classes) (the 7.0 Å structure from the previous step was used as initial model). After inspection and discarding poorly resolved/featureless classes, two sets of particles were selected for further processing: compact ATPase (43,773 particles) and relaxed ATPase (13,692 particles). Class 3 (38,792 particles) of the compact ATPase classes yielded a well-resolved DNA density and ATPase tier (7.4 Å resolution structure) after map-sharpening (B-factor applied -100 Å). All relaxed ATPase particles (13,692) were pooled and yielded a final structure at 9.8 Å resolution after map sharpening (B-factor applied -100 Å). Resolution was estimated using the gold-standard FSC method (0.143 criterion). Local resolution was estimated using ResMap and the maps were inspected in UCSF Chimera. See **Table 2.6** and **Figure 4.5** for the image processing parameters and work-flow.

2.7.2.3 CMGε-DNA-ATPγS datasets

Movie stacks were corrected for beam induced motion and integrated using MotionCor2, keeping all the frames and using a patch alignment of 5x5. Each non-dose-weighted micrograph was subjected to CTF estimation using Gctf. Quality of micrographs were inspected manually within the RELION-2 graphical user interface. High quality micrographs were selected and subjected to bin by 2 particle extraction giving 252,705 particles from a first dataset. A second dataset was collected, where a total of 151,971 particles were extracted from 5,687 dose weighted micrographs. The particles from the first dataset were subjected to 3D classification using the negative stain CMGε structure as an initial model (EMD-6465). This initial model was filtered to 60 Å and 5 classes were used with the spherical mask set at 350 Å and the ‘ignore CTFs until the first peak’ option switched on. The resulting two high-

resolution classes, comprising 115,068 particles, were selected and re-extracted back to unbinned parameters (256 pixel box size, 1.38 Å/pix). These unbinned particles were used for rounds of 3D refinement and classification without alignment. The particles from the second dataset were subjected to rounds of 2D classification and the highest quality particles (5,671 particles) were unbinned and combined with the first 89,809 particles from the first dataset. The new merged dataset contained 95,480 particles, which were then subjected to 3D classification with alignment. The best two classes from that step were combined (78,556 particles) and subjected to a final homogenous 3D refinement in cryoSPARC. Post processing was performed within the LocRes function in RELION-2.1, where two-half maps from the cryoSPARC reconstruction were used as input and sharpening the maps with a B-factor of -300 Å.

Molecular models for C-Pol2 and Dpb2 were determined by Drs. Panchali Goswami and Andrew Purkiss and CMG real-space refinement was performed by Dr. Andrew Purkiss. First, HHpred (Soding et al., 2005) was used to obtain a homology model based on the atomic structure of yeast Pol1 (PDB:4FYD) and the co-crystal structure of human Pole2/Pole1 (PDB: 5VBN). Rigid body docking of these models was first carried out using UCSF chimera and manually adjusted in coot. Real space refinement was performed in Phenix. See **Table 2.6** for a summary of image processing parameters, **Table 2.7** for model validation and **Figure 5.7** for the image processing flow chart.

2.7.2.4 Cryo-electron tomography of MCM trains

The protocols described in this section were performed by Dr. Andrea Nans (The Francis Crick Institute). The software Tomography 4.0 was used to collect 20 tilt series on a Titan Krios (Thermo Fisher) with a K2 summit direct electron detector and an additional GIF quantum energy filter (operated at slit width 20 eV) (Gatan). The tilt series was collected at a pixel size of 2.71 Å/pixels in a range from ±54º using a 3º increment at a -5 to -7 µm defocus range. The total dose from four movie frames was 111 e-/Å^2 (each exposure received 3 e-/Å^2). IMOD was used to align movie frames (via the alignframes tool) and the tilt series (via the fiducial-less patch tool). IMOD was also used for 2D CTF correction (via ctfphaseflip). To reconstruct

tomograms, weighted back-projection was utilised and given a SIRT-like filter (for one iteration).

For the template matching approach, MOLMATCH was used (Forster et al., 2010). Cryo-EM maps of MCM (EMD-6338) and CMG (EMD-8519) were filtered to 20 Å and used as templates for the MOLMATCH, which involves systematic rotation and translation prior to cross-correlation calculation within a cropped region of the aligned tomogram. The scanning parameters utilised was 0-360 phi, 0-360 psi and 0-180 theta, imposing a 10º increment. To account for the missing wedge in the tomogram, a missing wedge compensation was used for the template. The AV3 toolbox in Matlab (specifically av3_createmotl) was used to extract the coordinates of the highest correlation coefficients. To display the highest MCM and CMG matches (as determined by MOLMATCH) against the tomogram, the software Amira version 5.3 was used (Thermo Fisher).

Chapter 3. The Structure of the MCM double hexamer and insights into CMG Activation

3.1 Introduction

DNA replication begins at loci across the genome termed origins (Mechali et al., 2013). In eukaryotes, ORC and Cdc6 recruit the motor of the replicative helicase Mcm2-7, which is initially bound by Cdt1 (Remus and Diffley, 2009). The final step of this reaction is the loading of catalytically inactive head-to-head Mcm2-7 double-hexamers encircling duplex DNA, a stage referred to as origin licencing (Blow and Laskey, 1986, Remus et al., 2009, Evrin et al., 2009). Despite advances in the structural characterisation of the double hexamer, it is still unknown how the MCM binds DNA and whether the MCM melts the double helix at this early stage in replication as has been previously predicted (Li et al., 2015).

Several structures of both the archaeal and eukaryotic Mcm2-7 complexes have revealed the helicase motor organisation and the elements involved in ATP binding and DNA translocation (Li et al., 2015, Froelich et al., 2014). All six MCM subunits share a similar domain organisation: an N-terminal domain, divided into three A, B, C sub-domains; a AAA+ domain (comprising ATPase sites, DNA binding hairpins [pre-sensor 1 and helix 2 insert] and a winged helix appendix).

Upon entry into S-phase, the Mcm2-7 rings are activated in a step-wise manner (Yeeles et al., 2015, Labib, 2010). First, DDK phosphorylates MCM (Sheu and Stillman, 2006) and promotes the association of Sld3/7 and the helicase co-activator Cdc45 (Deegan et al., 2016). This step is essential as omission of DDK prevents CMG assembly (Yeeles et al., 2015, Deegan and Diffley, 2016) although how DDK structurally changes MCM to enable downstream recruitment of activating factors is unclear (On et al., 2014).

In a second stage of helicase activation, CDK mediates the recruitment of Polε, Sld2, Dpb11 and the second helicase co-activator GINS (a tetramer of Psf1-3, Sld5) (Muramatsu et al., 2010, Zegerman and Diffley, 2007). Together, Cdc45, Mcm2-7

and GINS form the 11-subunit CMG replicative holo-helicase (Moyer et al., 2006, Ilves et al., 2010). Full DNA unwinding by the CMG depends on Mcm10, resulting in 'origin firing' and replication fork progression (Yeeles et al., 2015). However, the molecular mechanism and spatiotemporal order of the Mcm10 activating switch is unclear. It is also unknown when the two hexamers separate ahead of bi-directional unwinding.

Recent breakthroughs in both the biochemistry of DNA replication (Yeeles et al., 2015, Yeeles et al., 2017, Georgescu et al., 2014, Georgescu et al., 2015) and cryo-EM (Kuhlbrandt, 2014) have provided a powerful approach to study the mechanism of helicase assembly and activation. In this chapter, the *in vitro* reconstitution assay of origin firing will be used (in collaboration with Dr. Max Douglas and Dr. John Diffley), combined with negative stain and cryo-EM, to address the following questions:

1. How does the inactive MCM contact DNA? Is the DNA melted during origin licensing?
2. How does DDK structurally change MCM for downstream activation?
3. At what stage during origin firing do double hexamers dissociate?

3.2 Structure of the MCM Double Hexamer

A central event in replisome assembly is the loading of the MCM double hexamer onto origin DNA during the G1 phase of the cell cycle. At this stage, the MCM remains catalytically inactive until G1/S-phase transition, when the action of DDK and CDK promote the binding of origin firing factors for CMG assembly and activation (Yeeles et al., 2015). Several structural studies have provided extensive mechanistic insight into the overall architecture of the double hexamer (Remus et al., 2009, Sun et al., 2014, On et al., 2014, Li et al., 2015) but how MCM binds DNA remains elusive. In addition, how DDK structurally changes the MCM to enable downstream CMG assembly is unclear (On et al., 2014, Labib, 2010).

To determine the structure of the double hexamer and the subsequent effect of DDK phosphorylation on the MCM, we employed an *in vitro* reconstitution assay using

purified budding yeast proteins (Remus et al., 2009, Yeeles et al., 2015, On et al., 2014). Here, double hexamers were loaded onto bead-immobilised DNA and either mock-treated or phosphorylated with DDK (outlined in **Figure 3.1a**). A high salt wash was carried out to remove weakly bound proteins, including DDK in the relevant sample (**Figure 3.1b**, lane 2). DNA bound complexes were released from beads by restriction enzyme digest to give fragment sizes ranging from 1.5-2 kb (helicase loading experiments performed by Dr. Max Douglas, see methods *section 2.4.5*). Samples were subsequently applied to lacey grids containing a thin layer of carbon and cryo grids were imaged on a Titan Krios equipped with a K2 summit camera (grid preparation by Dr. Julia Locke)(**Figure 3.2** and **Figure 3.3**, step 1).

3.2.1 Image processing of the MCM double hexamer

Both the unmodified and DDK phosphorylated double hexamer datasets were processed in a similar manner. After several rounds of 2D classification, high resolution classes that depicted the double hexamer were selected (**Figure 3.2** and **Figure 3.3**, step 2) and subjected to two rounds of 3D classification in RELION-2 (Kimanius et al., 2016) (**Figure 3.2** and **Figure 3.3**, steps 3-4). The density maps were inspected and particles from the highest resolution, most isotropic classes were taken forward for *ab initio* 3D classification and homogenous 3D refinement in cryoSPARC (Punjani et al., 2017) (**Figure 3.2** and **Figure 3.3**, steps 5-6). After map sharpening in RELION, I obtained structures of the unmodified and DDK phosphorylated MCM double hexamers at 4.8 Å and 4.3 Å resolution, respectively (**Figure 3.1c**; **Figure3.2** and **Figure3.3**, step 7; **Figure 3.4** and **Figure 3.5**).

3.2.2 Overall architecture of the MCM double hexamer

The first step towards addressing how MCM binds DNA is to describe the double hexamer on its own. To this end, the atomic coordinates from the endogenous MCM cryo-EM structure (PDB: 3JA8) (Li et al., 2015) were re-adjusted to the correct pixel size and individual MCM subunits were docked in as rigid bodies to the cryo-EM maps. The docked coordinates were then used to refine an atomic model against the highest resolution cryo-EM map in this study (DDK phosphorylated MCM) using Phenix (modelling by Dr. Valerie Pye at the Francis Crick Institute)(Afonine et al., 2018).

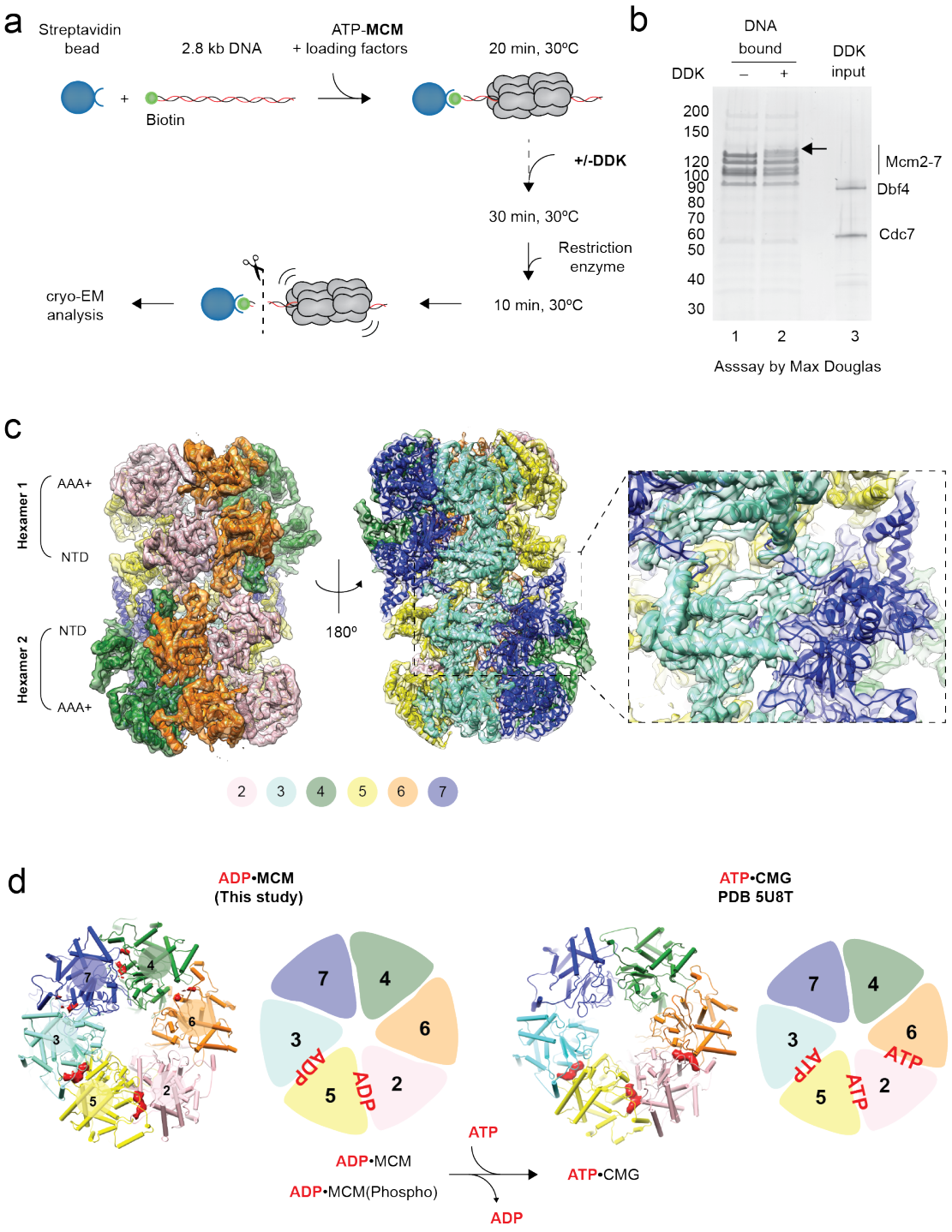


Figure 3.1: Sample preparation and cryo-EM structure of the MCM double hexamer
(a) Scheme for the loading of MCM double hexamers on DNA. **(b)** A silver stain gel by Dr. Max Douglas of the unmodified (lane 1) and DDK treated (lane 2) double hexamer. Arrow points to a shifted Mcm6 band upon DDK phosphorylation. **(c)** The sharpened DDK phosphorylated map at 4.3 Å resolution with docked atomic coordinates of the real space refined structure (based on PDB: 3JA8) **(d)** Comparison of nucleotide occupancy between the inactive and active MCM in the CMG (PDB: 5u8t).

Docking in the real space refined model to the unmodified double hexamer gave an excellent fit, indicating that DDK phosphorylation does not induce a significant conformational change at the core MCM level (**Figure 3.6**) (discussed further in *section 3.4*). As such, the DDK phosphorylated structure at 4.3 Å resolution will be used to describe the double hexamer hereafter in this section.

The cryo-EM map reveals the characteristic four tier architecture of the double hexamer, where the dimerization interface is formed by the NTD tiers while the AAA+ tiers face outward (**Figure 3.1c**). The structure here confirms all the main previous observations from the DNA-free double hexamer (**Figure 1.4** and **1.11**) (Li et al., 2015), including the off-axis tilt (14º) and lateral offset alignment (~20 Å) of the two hexamers, and the shifted NTD and AAA+ tiers within the single hexamer. This arrangement is largely due to the staggered B-domain Zf interlocking contacts across the two NTD tiers (Li et al., 2015). Further inter hexamer interactions are stabilised through the previously observed interaction between Mcm7 and Mcm5 (**Figure 3.1c,** inset). Compared to archaea, eukaryotic Mcm7 has an additional 70 amino acids at its N-terminal alpha helical domain (A-domain) (Li et al., 2015). This folds into a helix-turn-helix, which projects towards and binds Mcm5 from the second hexamer at the N-terminal region to provide stable contact.

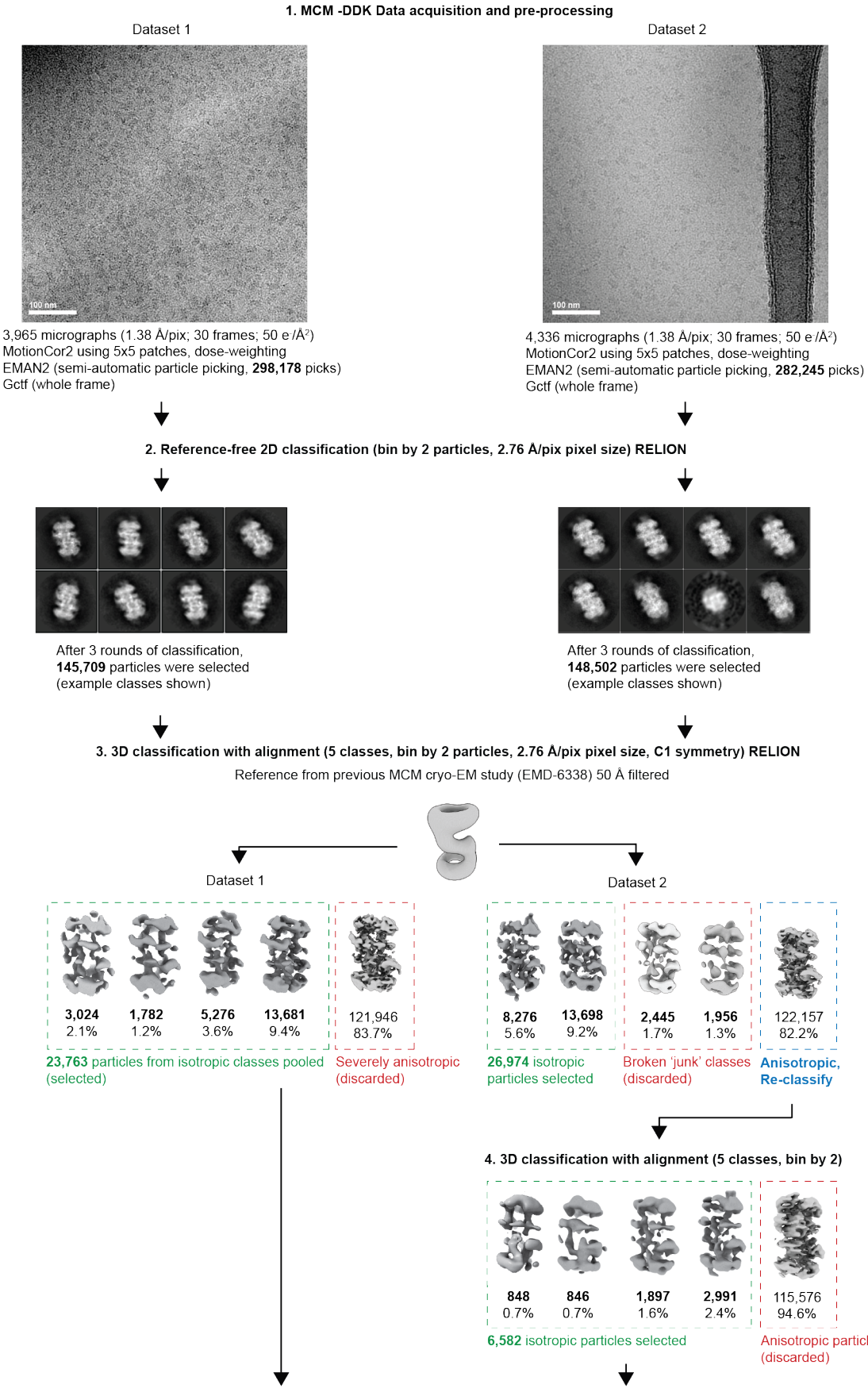
3.2.3 ATPase state of the MCM double hexamer

The resolution achieved here allows a description of the nucleotide occupancy in the MCM double hexamer. Although ATP density cannot be discerned from ADP with confidence, previous studies have shown that the MCM double hexamer is a post-catalytic structure and remains stably bound to ADP until CMG assembly (Li et al., 2015, Douglas et al., 2018).

Inspection of a top view of the ring from the AAA+ side reveals strong nucleotide density between the Mcm2-5 and Mcm5-3 subunits (**Figure 3.1d**, left) and much weaker density between the Mcm3-7 and Mcm7-4 interfaces. The Mcm6-2 interface is open and nucleotide-free, indicating nucleotide release at this interface. In line with this nucleotide occupancy pattern, ATP hydrolysis by Mcm2-5 and Mcm5-3 have been shown to be essential for double hexamer loading (Coster et al., 2014).

Furthermore, this pattern is identical in the unmodified double hexamer and thus DDK treatment does not alter the ATPase function of MCM. In contrast, the endogenous yeast MCM double hexamer cryo-EM structure had ADP built into nucleotide density that was present at every interface around the ring (Li et al., 2015). A reason for the discrepancy in nucleotide pattern between double hexamer structures could be that Li et al. imaged the double hexamers in buffer containing 3 mM ATP whereas we performed high salt washes in the absence of nucleotide.

In comparison, a cryo-EM structure of yeast CMG in the presence of non-hydrolysable ATP analogue AMP-PNP showed nucleotide at the Mcm6-2, Mcm2-5 and Mcm5-3 interfaces (Georgescu et al., 2017) (**Figure 3.1d**, right). This is in line with the finding that a nucleotide exchange has to occur during CMG assembly where the MCM is induced to release ADP by an unknown origin firing factor before ATP reuptake (Douglas et al., 2018).


1. MCM -DDK Data acquisition and pre-processing
Dataset 1
Dataset 2
100 nm
100 nm
3,965 micrographs (1.38 Å/pix; 30 frames; 50 e /Å²)
MotionCor2 using 5x5 patches, dose-weighting
EMAN2 (semi-automatic particle picking, 298,178 picks)
Gctf (whole frame)
4,336 micrographs (1.38 Å/pix; 30 frames; 50 e /Å²)
MotionCor2 using 5x5 patches, dose-weighting
EMAN2 (semi-automatic particle picking, 282,245 picks)
Gctf (whole frame)
2. Reference-free 2D classification (bin by 2 particles, 2.76 Å/pix pixel size) RELION
After 3 rounds of classification,
145,709 particles were selected
(example classes shown)
After 3 rounds of classification,
148,502 particles were selected
(example classes shown)
3. 3D classification with alignment (5 classes, bin by 2 particles, 2.76 Å/pix pixel size, C1 symmetry) RELION
Reference from previous MCM cryo-EM study (EMD-6338) 50 Å filtered
Dataset 1
Dataset 2
3,024 2.1%
1,782 1.2%
5,276 3.6%
13,681 9.4%
121,946 83.7%
8,276 5.6%
13,698 9.2%
2,445 1.7%
1,956 1.3%
122,157 82.2%
23,763 particles from isotropic classes pooled (selected)
Severely anisotropic (discarded)
26,974 isotropic particles selected
Broken 'junk' classes (discarded)
Anisotropic, Re-classify
4. 3D classification with alignment (5 classes, bin by 2)
848 0.7%
846 0.7%
1,897 1.6%
2,991 2.4%
115,576 94.6%
6,582 isotropic particles selected
Anisotropic particles (discarded)

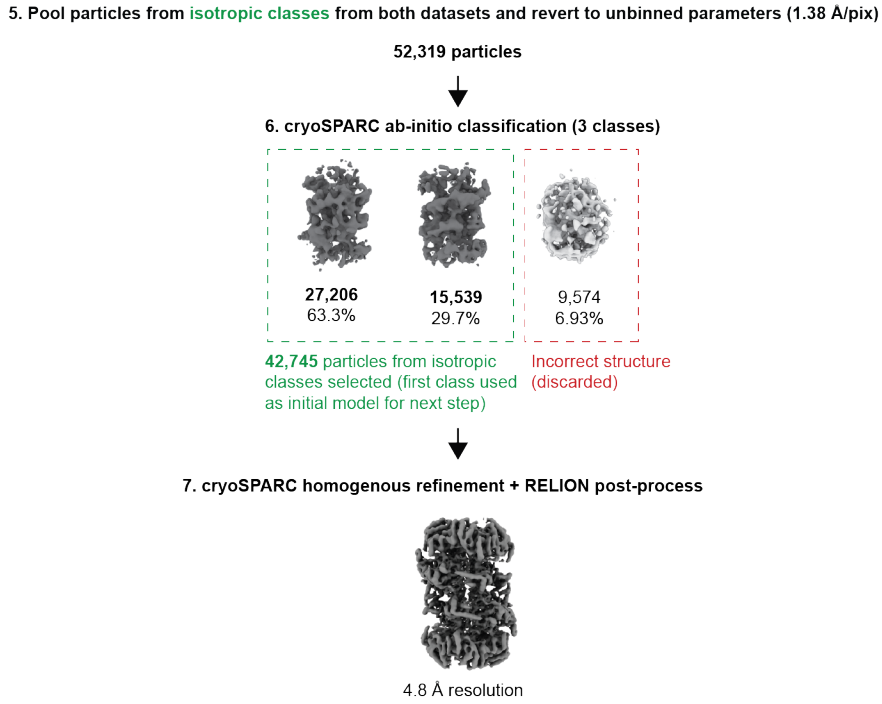
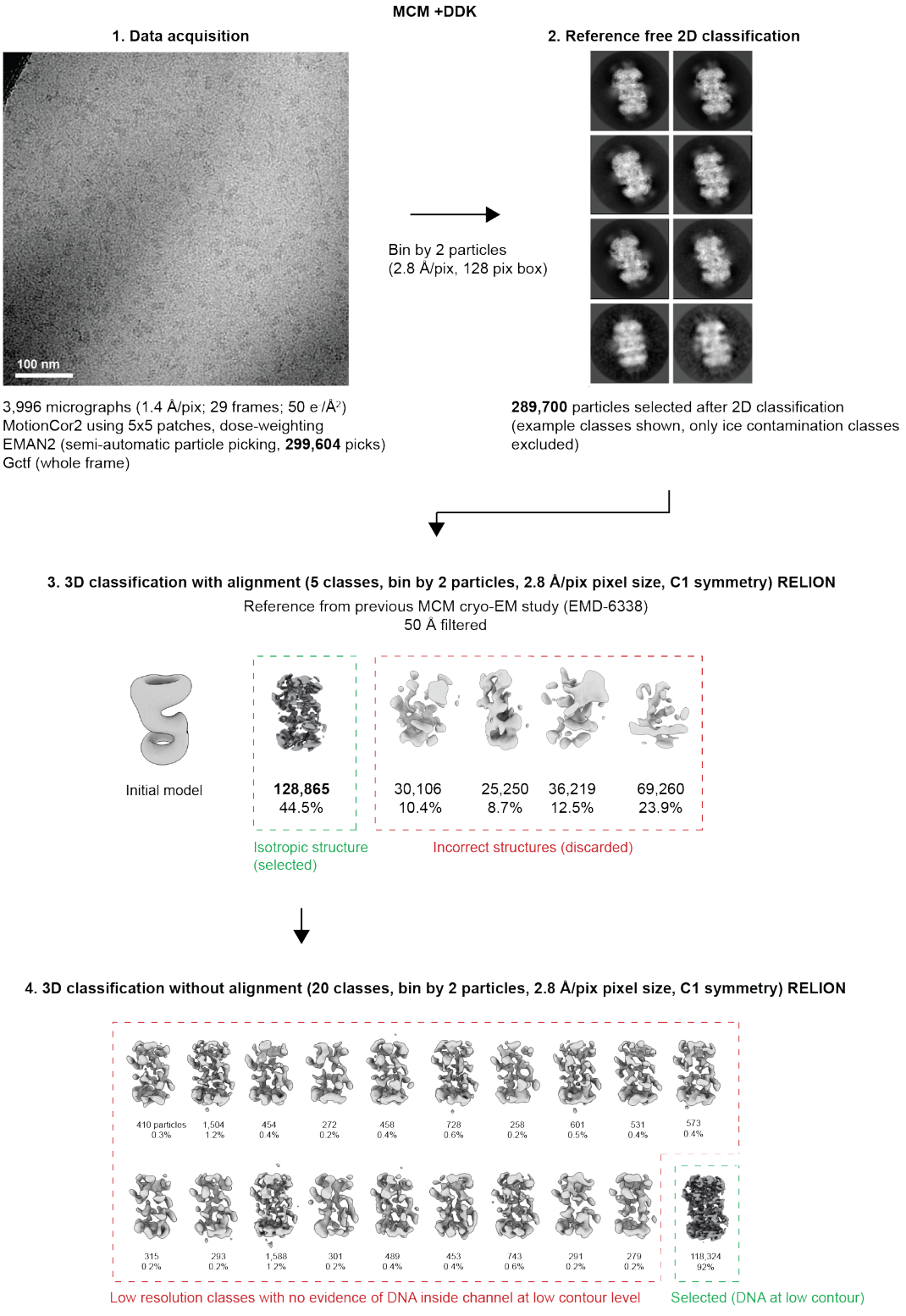


Figure 3.2: Image processing workflow of the unmodified MCM double hexamer
Two datasets were collected on a Titan Krios equipped with a K2 summit direct electron detector. All pre-processing (2D classification and 3D classification) was performed in RELION-2 (steps 1-4), apart from particle picking, which was performed in EMAN2 (e2boxer.py). Particle picking and CTF correction was performed on non-dose weighted integrated movies. The rest of the steps were performed on dose-weighted integrated movies. Steps 2-4 were performed on binned particles to speed up computation and increase signal-to-noise ratio for alignment of particles. The two datasets were processed separately to select the particles from isotropic 3D classes that resembled previously published structures of the double hexamer (green box). Classes not resembling double hexamers or exhibiting strong preferential orientation (anisotropy) were either discarded (red box) or re-classified (blue box). The selected particles (green) mostly belonged to lower resolution classes, thus to increase resolution of the final structure, particles from both datasets were pooled (step 5). These particles were subjected to ab-initio classification in cryoSPARC (step 6) for a final round of cleaning (junk particles removed) and the particles contributing to double hexamer classes from this step were selected. These particles were finally subjected to homogenous refinement with dynamic masking in cryoSPARC prior to sharpening in RELION-2 (auto-sharpening -163 Å B-factor, step 7).


MCM +DDK
1. Data acquisition
100 nm
Bin by 2 particles
(2.8 Å/pix, 128 pix box)
2. Reference free 2D classification
3,996 micrographs (1.4 Å/pix; 29 frames; 50 e⁻/Å²)
MotionCor2 using 5x5 patches, dose-weighting
EMAN2 (semi-automatic particle picking, 299,604 picks)
Gctf (whole frame)
289,700 particles selected after 2D classification
(example classes shown, only ice contamination classes
excluded)
3. 3D classification with alignment (5 classes, bin by 2 particles, 2.8 Å/pix pixel size, C1 symmetry) RELION
Reference from previous MCM cryo-EM study (EMD-6338)
50 Å filtered
Initial model
128,865
44.5%
30,106
10.4%
25,250
8.7%
36,219
12.5%
69,260
23.9%
Isotropic structure
(selected)
Incorrect structures (discarded)
4. 3D classification without alignment (20 classes, bin by 2 particles, 2.8 Å/pix pixel size, C1 symmetry) RELION
410 particles
0.3%
1,504
1.2%
454
0.4%
272
0.2%
458
0.4%
728
0.6%
258
0.2%
601
0.5%
531
0.4%
573
0.4%
315
0.2%
293
0.2%
1,588
1.2%
301
0.2%
489
0.4%
453
0.4%
743
0.6%
291
0.2%
279
0.2%
118,324
92%
Low resolution classes with no evidence of DNA inside channel at low contour level
Selected (DNA at low contour)

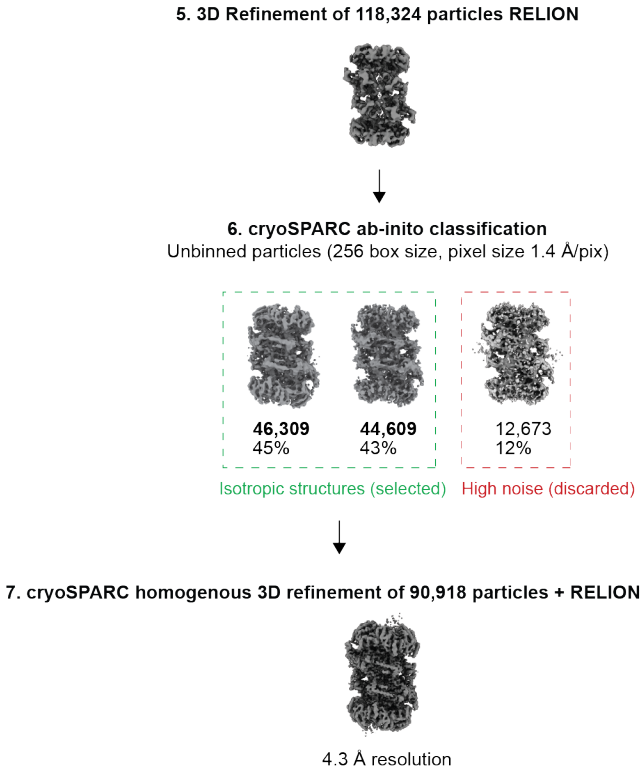


Figure 3.3: Image processing workflow for the DDK phosphorylated MCM double hexamer
Data were collected on a Titan Krios equipped with a K2 summit direct electron detector (pixel size 1.4 Å/pixel). After motion correction, particles were picked on filtered, non-dose weighted micrographs in a semi-automatic manner using EMAN2 e2boxer.py (Step 1). 2D, 3D classification and 3D refinement was performed in RELION on CTF (Gctf) corrected bin by 2 particles from dose-weighted micrographs (Steps 2-5). Binning was performed to speed up computation and increase the signal-to-noise ratio for initial alignment of particles. For 3D classification (Step 3) the initial model (MCM double hexamer, EMD-6338) was re-sized to a 128-pixel box size, a pixel size of 2.8 Å/pixel and filtered to 50 Å to prevent initial model bias. All 3D classification steps were performed with no symmetry imposed (3D refinements performed with C2 symmetry). Criteria for selection of 'good' particles (green boxes) were based on resemblance of 3D classes to previously published MCM double hexamer (EMD-6338), evidence of DNA inside the central channel at low contour level and isotropic density (inspection of euler angle plot and maps in Chimera). Anisotropic, noisy and 'junk' classes that did not resemble double hexamer (EMD-6338) were discarded (red boxes). 3D classification without alignment was performed in RELION to separate out different conformers in the particle population (Step 4). Vast majority of particles went into an isotropic class (green box, step 4), indicating that the dataset is homogenous and comprises a good distribution or particle orientations. This class was further refined (step 5) while all other low-resolution classes from step 4 were discarded after observation of no DNA in the central channel at low contour level. All particles from step 5 were selected and final ab-initio 3D classification + homogenous 3D refinement was performed in cryoSPARC using unbinned particle parameters (box size 256 pixels, pixel size of 1.4 Å/pixel) to further remove 'junk' particles and improve the resolution of the structure (Steps 6-7). Half maps from cryoSPARC 3D refinement was used as input for RELION-2 post-processing (masking and sharpening with a B-factor of -480 Å applied) and the map resulting from this step was used for real-space refinement of an atomic model based on previous coordinated (PDB: 3ja8). The final structure resolved to 4.3 Å as determined using gold-standard Fourier shell correlation, 0.143 criterion, step 7.

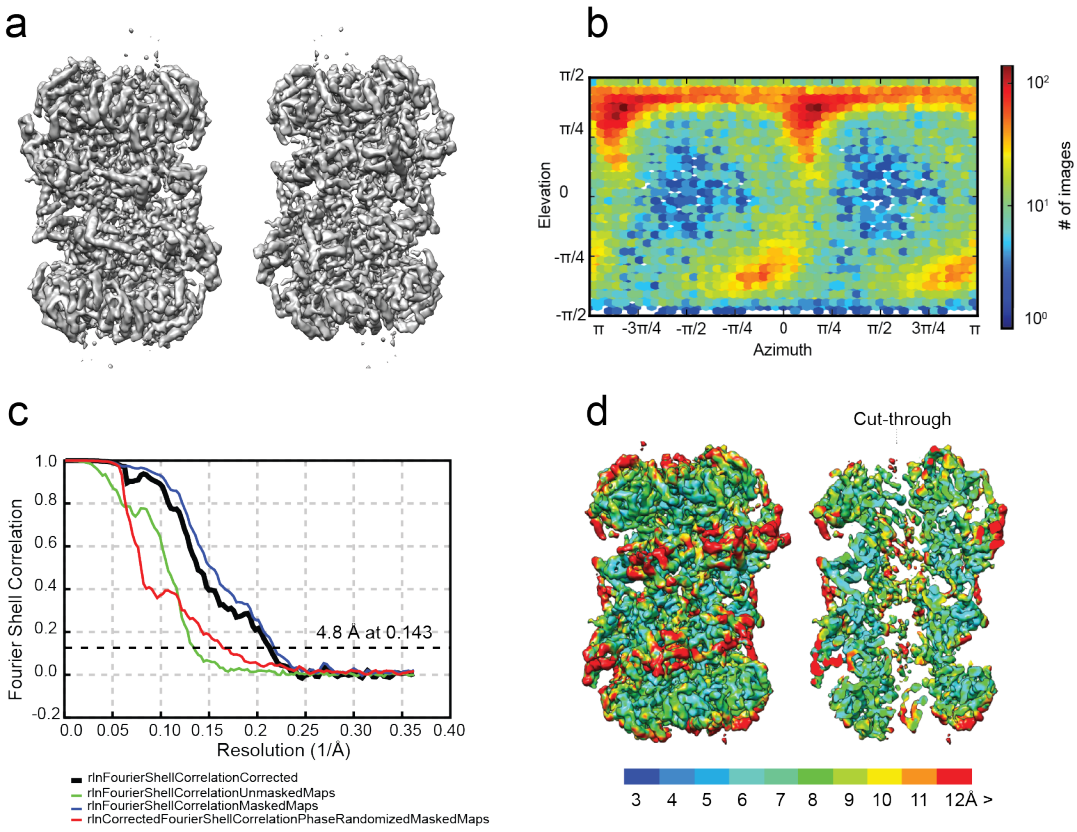


Figure 3.4: Cryo-EM of the unmodified MCM double hexamer

(a) Representative micrograph of the double hexamers imaged with a K2 summit camera in counting mode. **(b)** 2D class averages of the highest resolution particles. **(c)** 3D reconstruction shown in two views. **(d)** Angular distribution. **(e)** Fourier Shell Correlation (Gold Standard, using the 0.143 criterion). **(f)** Local resolution as determined by ResMap.

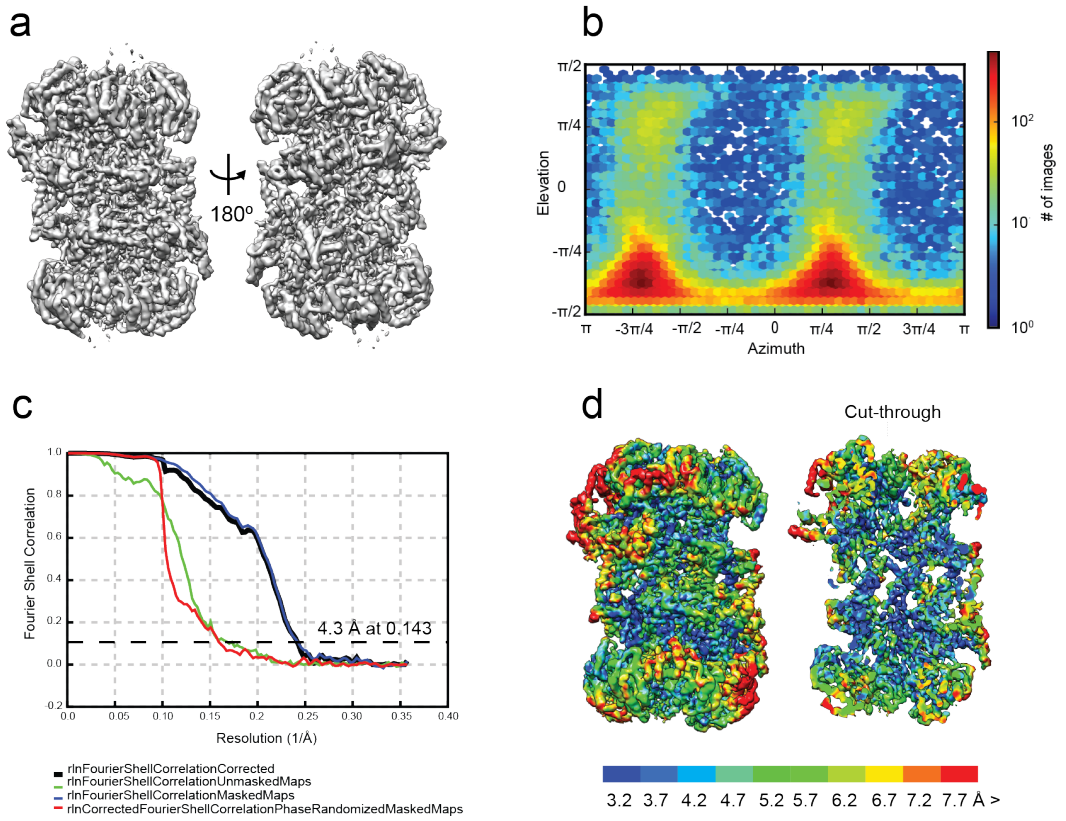


Figure 3.5: Cryo-EM of the DDK phosphorylated MCM double hexamer
(a) 3D reconstruction shown in two views. **(b)** Angular distribution. **(c)** Fourier Shell Correlation (Gold Standard, using the 0.143 criterion). **(d)** Local resolution as determined by ResMap.

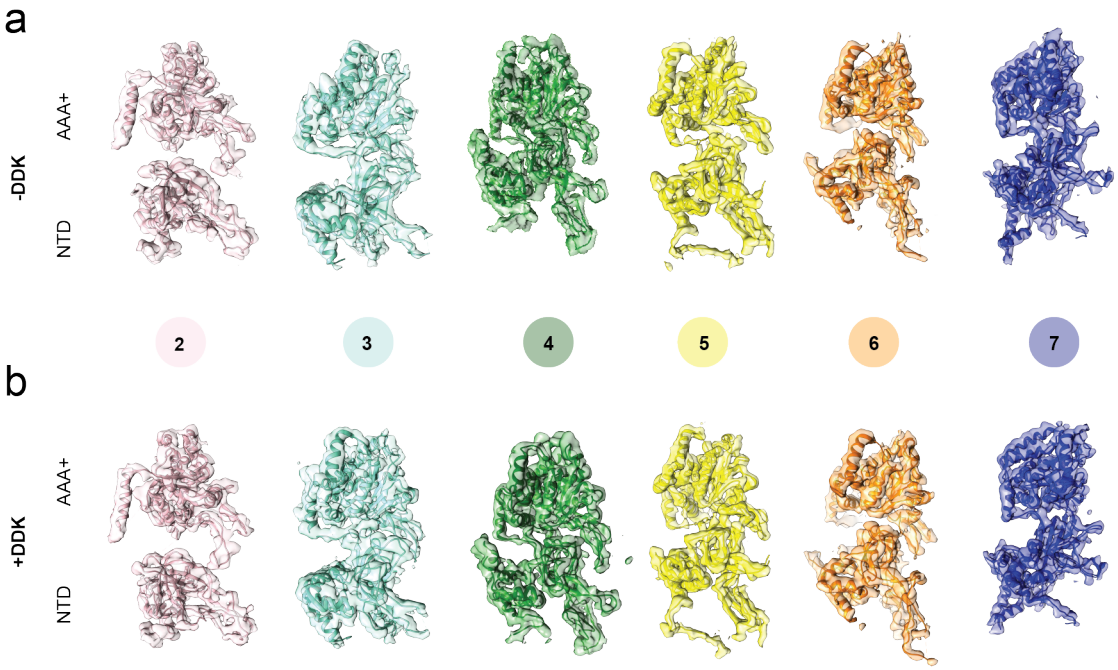


Figure 3.6: Subunit comparison of unmodified and DDK phosphorylated MCM double hexamer
(a) Unmodified and **(b)** DDK phosphorylated double hexamer density segmentations. Docked coordinates are the real-space refined model based on the PDB: 3ja8 atomic structure.

3.3 Structure of MCM double hexamer bound to DNA

Understanding the molecular basis of helicase activation requires a detailed description of DNA contacts in the inactive MCM. Rotary shadowing experiments of MCM bound to circular double stranded DNA, which failed to detect single-stranded DNA and the supercoiling effects that would be expected upon DNA twisting, suggest that the MCM double hexamer binds duplex DNA (Evrin et al., 2009, Remus et al., 2009). However, this notion has recently been contended by the cryo-EM study of the DNA-free endogenous yeast MCM double hexamer, which predict deformation of DNA due to the constricted central channel (Li et al., 2015) and atomic force microscopy experiments of the DNA loaded double hexamer showing egression of DNA through a side gate (Hizume et al., 2017). Numerous MCM loops have been suggested to bind DNA, including the ZF loops and OB hairpins (NTD tier) and the highly conserved PS1 and MCM-specific H2i loops (AAA+ tier) (Li et al., 2015, Froelich et al., 2014, McGeoch et al., 2005, Jenkinson et al., 2009, Barry et al., 2009). However, it is unknown how MCM contacts DNA and whether it melts the double helix at this early stage. To address these questions, I adopted a new approach to identify DNA bound particles.

3.3.1 Using signal subtraction to retrieve DNA bound particles

It is well established that DNA binding is required for double hexamer formation but once loaded MCMs can passively slide off linear DNA in high salt conditions (Remus et al., 2009, Evrin et al., 2009). The half-life of the double hexamer on 1 kb linear DNA in 0.5 M NaCl is 10 min (Remus et al., 2009), which is similar to the range of conditions used in the present study (1.5-2 kb, 0.3 M KCl). Indeed, despite extensive 2D and 3D classification attempts, I could only visualise poorly resolved discontinuous DNA density inside the channel (**Figure 3.7a**). This indicates that a fraction of DNA bound double hexamers in the sample were averaged out by a much larger subset of DNA-free MCM particles. This notion was confirmed by a positive stain experiment, which showed a combination of unbound and DNA-bound MCM particles in our samples (**Figure 3.7b**, white arrows).

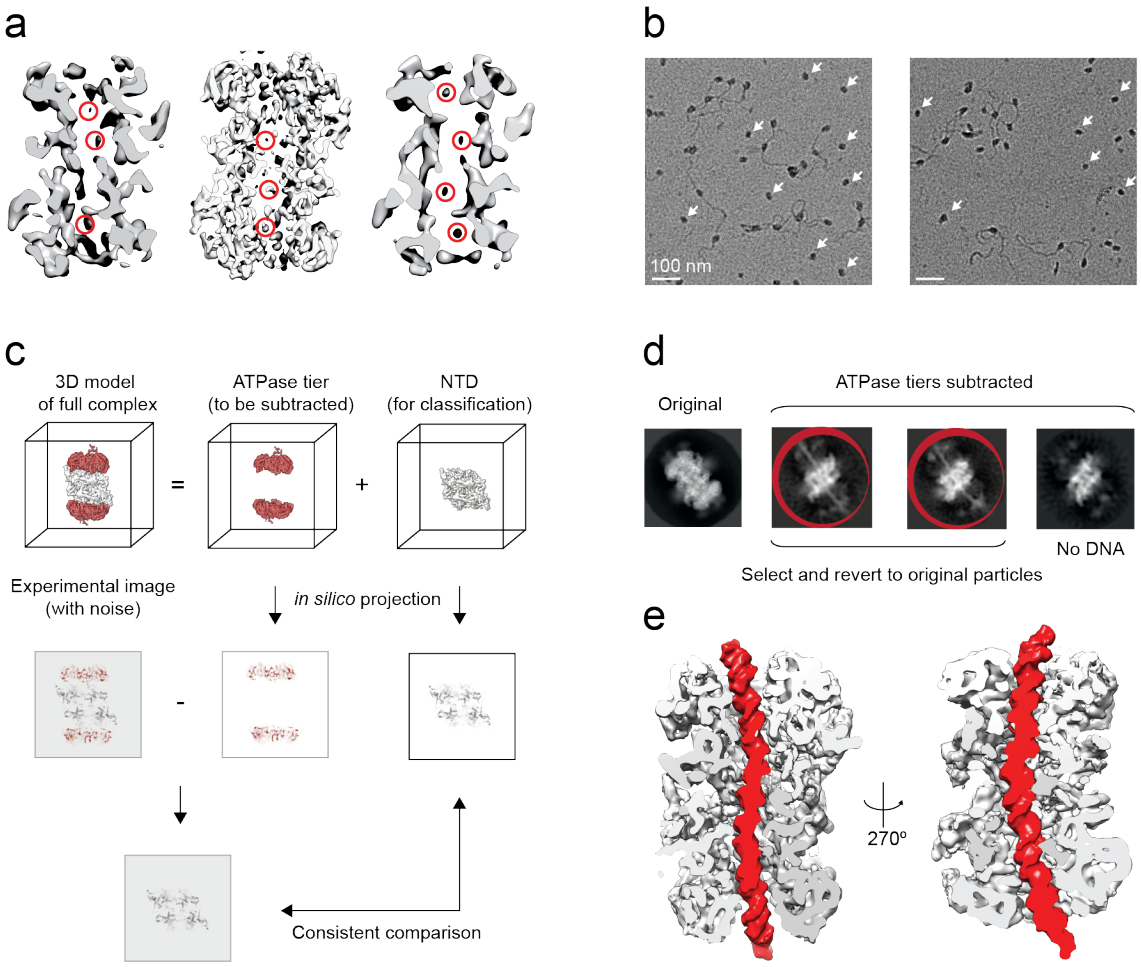


Figure 3.7: Signal subtraction of AAA+ motor allows visualisation of DNA bound particles
(a) Canonical 3D classification in RELION-2 fails to identify the fully occupied DNA bound MCM particles. **(b)** Representative positively stained reactions with arrows pointing at double hexamers that have slid off the DNA, acquired at 50,000x maginifcation on a Tecnai G2 F20 TWIN electron microscope operating at 200 keV (FEI; Electron Microscopy Centre, Imperial College London) equipped with a Falcon II direct electron detector (FEI) **(c)** Signal subtraction scheme based on (Bai et al., 2015). A mask is generated for the AAA+ tier which is used to subtract the AAA+ tier from all aligned experimental particles. This leaves behind only the two NTDs and any DNA that would emanate from the channel. **(d)** Subtracted particles are subjected to 2D classification, where the DNA bound particles can be confidently distinguished from the unbound particles. **(e)** DNA bound subtracted particles are selected and then reverted to their original density before 3D refinement in cryoSPARC. Shown here is the unmodified DNA bound MCM double hexamer at 6.9 Å resolution.

As canonical classification techniques failed to identify the DNA-bound particles, I employed a signal subtraction approach building on methods developed in the Scheres lab (Bai et al., 2015) that exposed the double hexamer channel for inspection (**Figure 3.7c**). In this protocol, the 2D projections of the AAA+ tiers were first subtracted from the aligned, raw subset of particles that contributed to the high-resolution structures. This left behind only the two NTD tiers (used for re-alignment

of the particles) and any DNA that would emanate from the channel. The subtracted particles were then subjected to 2D classification, allowing the unbound and DNA-bound MCM particles to be confidently distinguished (**Figure 3.7d**). This step showed that although most particles had diffused off the DNA, 7% DDK phosphorylated and 20% unmodified double hexamer particles contained a double helical density symmetrically protruding from the NTD core.

The DNA projecting out of the NTD tiers comes out at only one angle and is long enough to also be entrapped by the AAA+ tiers, suggesting that DNA could span the entire double hexamer channel in contrast to observations by atomic force microscopy (Hizume et al., 2017) (**Figure 3.7d**). To confirm this observation, I selected only the DNA-bound signal subtracted particles and reverted back to their full original density in RELION. Finally, I performed a homogenous 3D refinement in cryoSPARC to obtain a fully occupied DNA-bound MCM double hexamer at 6.9 Å (unmodified, **Figure 3.7e, Figure 3.8**) and 7.5 Å resolution (DDK phosphorylated, **Figure 3.8**). The structures show a continuous density that indeed covers the entire length of the central channel (~200 Å), with the minor and major grooves that are characteristic of the DNA double helix (**Figure 3.7e**). No conformational differences were observed between the DDK and unmodified DNA bound double hexamers, thus the highest resolution structure will be described hereafter (unmodified DNA bound MCM).

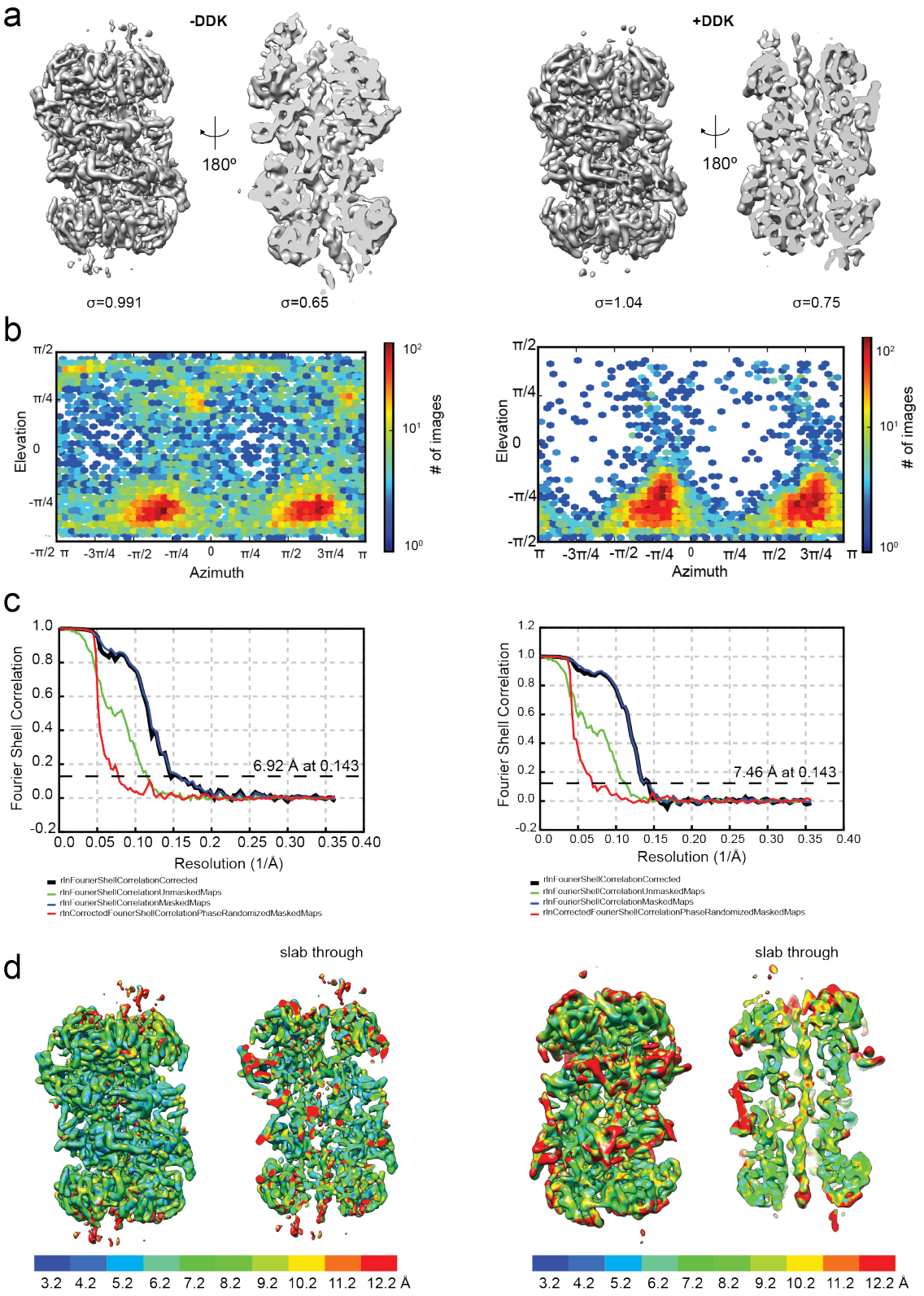


Figure 3.8: Cryo-EM of the unmodified and DDK phosphorylated DNA bound double hexamers
(a) 3D reconstruction of full surface view and a slab view at lower contour level. **(b)** Angular distribution. **(c)** Gold-standard Fourier Shell Correlation (0.143 criterion). **(d)** ResMap local resolution of full surface view and slab through.

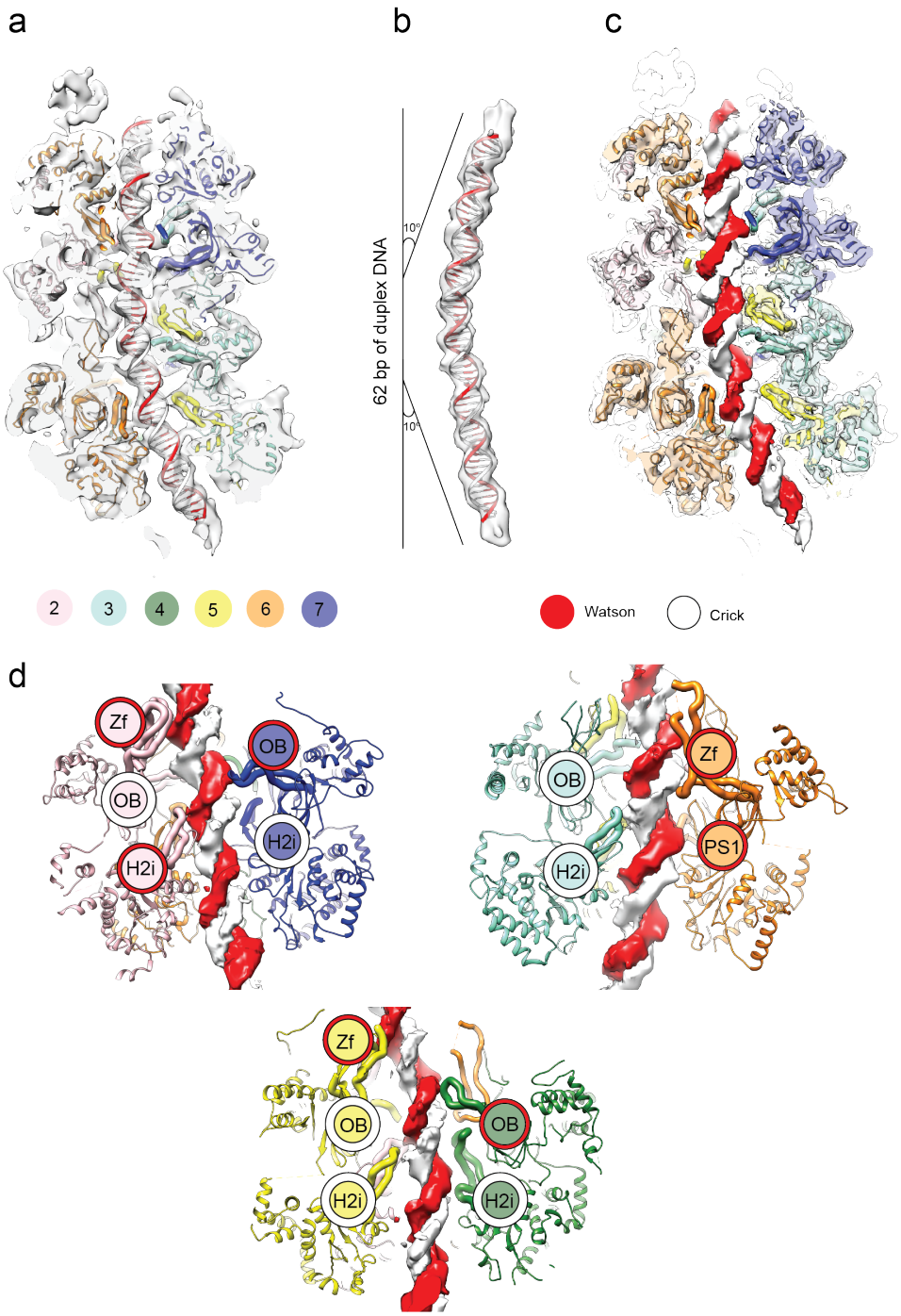


Figure 3.9: The MCM double hexamer bound to double stranded DNA.
(a) A cut-through view of the unmodified MCM double hexamer with the refined coordinates base on PDB: 3JA8 and manually built DNA (Dr. Valerie Pye). **(b)** 62 base pairs of duplex DNA fits inside the channel and bends as it enters the double hexamer interface (NTD-NTD). **(c)** A composite resolution structure of the 4.8 Å ensemble map and the 6.9 Å map for the DNA density. **(d)** Analysis of all the MCM loop contacts with the Watson (red) and Crick (white) strands. OB, oligosaccharide/oligonucleotide binding; H2i, helix-2 insert; Zf, zinc finger; PS1, pre-sensor 1.

3.3.2 The double hexamer binds duplex DNA

To analyse how MCM binds DNA, a model for the central channel density in the unmodified double hexamer was built by Dr. Valerie Pye (Francis Crick Institute). To create this model, a random sequence of DNA was used instead of the original DNA sequence from the 2.8 kb template in our experiments as MCM does not have any sequence specificity and nitrogen base identity cannot be determined at 6.9 Å resolution.

As major and minor grooves can clearly be resolved from our structure, 62 base pairs of continuous, duplex DNA can be modelled inside the axial channel of the double hexamer, consistent with previous measurements (Remus et al., 2009) (**Figure 3.9a**). Encircled DNA takes two 10º turns, leading to bending, but not melting of duplex DNA (**Figure 3.9b**). The DNA bend is propagated by the kinked central channel and occurs in the NTD region between the B (ZF) and C-domains (OB) (**Figure 3.9c-d**). The observation of a bend is in agreement with the findings recently reported by Noguchi et al., which were published at the same time as the present study (Noguchi et al., 2017). Therefore, melting must occur at a subsequent step in CMG activation (presented in *section 4.2.9*).

3.3.3 Mapping the DNA-MCM contacts in the double hexamer

To describe the DNA-MCM contacts, I docked in the higher resolution ensemble structure (4.8 Å) to the DNA bound structure (6.9 Å). This resulted in an unambiguous fit, revealing that DNA binding does not appear to alter the pore loop configurations. Therefore, a composite hybrid resolution structure where the protein model was based on the higher resolution 4.8 Å structure and the DNA model was based on the 6.9 Å structure was used for further analysis (**Figure 3.9d**).

Inside the channel, each hexamer is in contact with two full turns of DNA and an additional turn of DNA is situated at the double hexamer interface (**Figure 3.9d**). The channel is lined with highly conserved pore loops that project towards and make contact with the Watson and Crick strands of DNA. These include the ZF and OB loops of the NTD and the H2i and PS1 of the AAA+ tiers (**Figure 3.9c-d**). In our structure, 14 contacts are observed, including the ZF hairpin of Mcm2,5,6, OB-fold

of Mcm4,7, H2i of Mcm2 and PS1 of Mcm6 contact the Watson strand whereas the OB-fold of Mcm2,3,5 and H2i of Mcm3,4,5,7 contact the Crick strand (**Figure 3.9d**). Taken together, these contacts form a tight grip on DNA that may explain why a high salt wash is required to destabilise the DNA-protein contacts for sliding to occur (Remus et al., 2009, Evrin et al., 2009).

From this analysis, we observed that the ZF domains from only one half of the ring contact the DNA (Mcm6-2-5), positioning the DNA proximal to each Mcm2-5 gate, which is a site known to be required for helicase loading (Bochman et al., 2008, Samel et al., 2014) (**Figure 3.9d**). This specific Mcm2-5 ZF DNA interaction is mediated in part by the staggered configuration of the two rings and the DNA is therefore suitably positioned if extrusion of one strand occurs via this gate upon helicase activation. Furthermore, the position of the Mcm4 OB-fold loop (contacting the minor groove) is the point at which the DNA bend is seen in the MCM double hexamer and is also where the CMG contacts the duplex DNA junction in the forked DNA-CMG structure recently published (Georgescu et al., 2017) (**Figure 3.9d**). This suggests that the Mcm4 OB-fold could act as a fixed pivot point during CMG assembly and unwinding (discussed further in *section 4.2.9*). All H2i loops apart from Mcm6 contact DNA but do not follow a strict helical path. H2i loops only exist in SF6 helicases and are thought to release from DNA to enable movement during active translocation (Jenkinson and Chong, 2006, Enemark and Joshua-Tor, 2008). The firm DNA binding by H2i loops in this inactive MCM intermediate may therefore serve to block untimely helicase unwinding (**Figure 3.9d**). The PS1 of Mcm6 is the only MCM PS1 loop involved in DNA binding but is positioned 12 Å lower in the double hexamer channel compared to the yeast CMG. The movement of this loop may be involved in activating the helicase as will be explored in *section 4.2.9*.

In summary, the double hexamer forms a tilted assembly of two MCM rings, which can slide on DNA. The MCMs are inactive at this stage through stable binding of ADP at Mcm2-5 and Mcm5-3 interfaces and a locked configuration of the two rings. The double hexamer binds to double-stranded DNA through the entire span of the central channel and bends (instead of melting) at two inflection points near the dimerization interface. Both strands of DNA are gripped equally by pore loops and this may prevent premature directional translocation.

3.4 Effect of DDK phosphorylation on the double hexamer

DDK phosphorylates MCM (Sheu and Stillman, 2006, Randell et al., 2010), which is an essential step for the recruitment of Cdc45 via the Sld3/7 chaperone (Deegan et al., 2016, Yeeles et al., 2015). The majority of DDK phosphorylation sites are in the flexible N-terminal tails of Mcm4 and 6 at a region that is predicted to be proximal to the double hexamer interface (Randell et al., 2010). An *mcm5-bob1* mutant bypasses the requirement of DDK and causes a structural change in the A-domain of Mcm5 (Hardy et al., 1997, Fletcher et al., 2003). This has previously invited the hypothesis that DDK splits the double hexamer prior to CMG assembly (On et al., 2014). Data presented here and elsewhere (On et al., 2014, Sun et al., 2014), however, clearly show that the double hexamer is not destabilised upon DDK treatment and no single hexamers exist in the particle population despite efficient phosphorylation (**Figure 3.1b**). Furthermore, no other major conformational change is observed in the core double hexamer structure when compared to the unmodified MCM (**Figure 3.6**).

To investigate if phosphorylation of MCM might promote a subtler conformational change within the double hexamer, I focused in the comparison on the NTD region where DDK phosphorylation is known to locally take place (Randell et al., 2010, Deegan et al., 2016, Sheu and Stillman, 2006). Indeed, comparison of aligned 2D class averages from the unmodified and DDK phosphorylated samples revealed extra density present in the DDK phosphorylated particles but not the unmodified counterpart (**Figure 3.10a-b**). This is best recognised by subtracting the unmodified from the DDK phosphorylated particles, leaving behind only a symmetrical side density (**Figure 3.10a,** bottom row and **Figure 3.10b,** right panel). This density is unlikely to be DDK binding to the MCM as it is washed off after the phosphorylation reaction (**Figure 3.1b**).

In the ensemble DDK phosphorylated structure where C2 symmetry was applied, the extra density can be observed at low contour level as a flexibly tethered lobe at the NTD dimerization interface, mapping close to Mcm4 and 6 on both hexamers (**Figure 3.10c**). Asymmetric 2D classification revealed that this density still appears symmetrical and in two main configurations: compact and extended (**Figure 3.10d**).

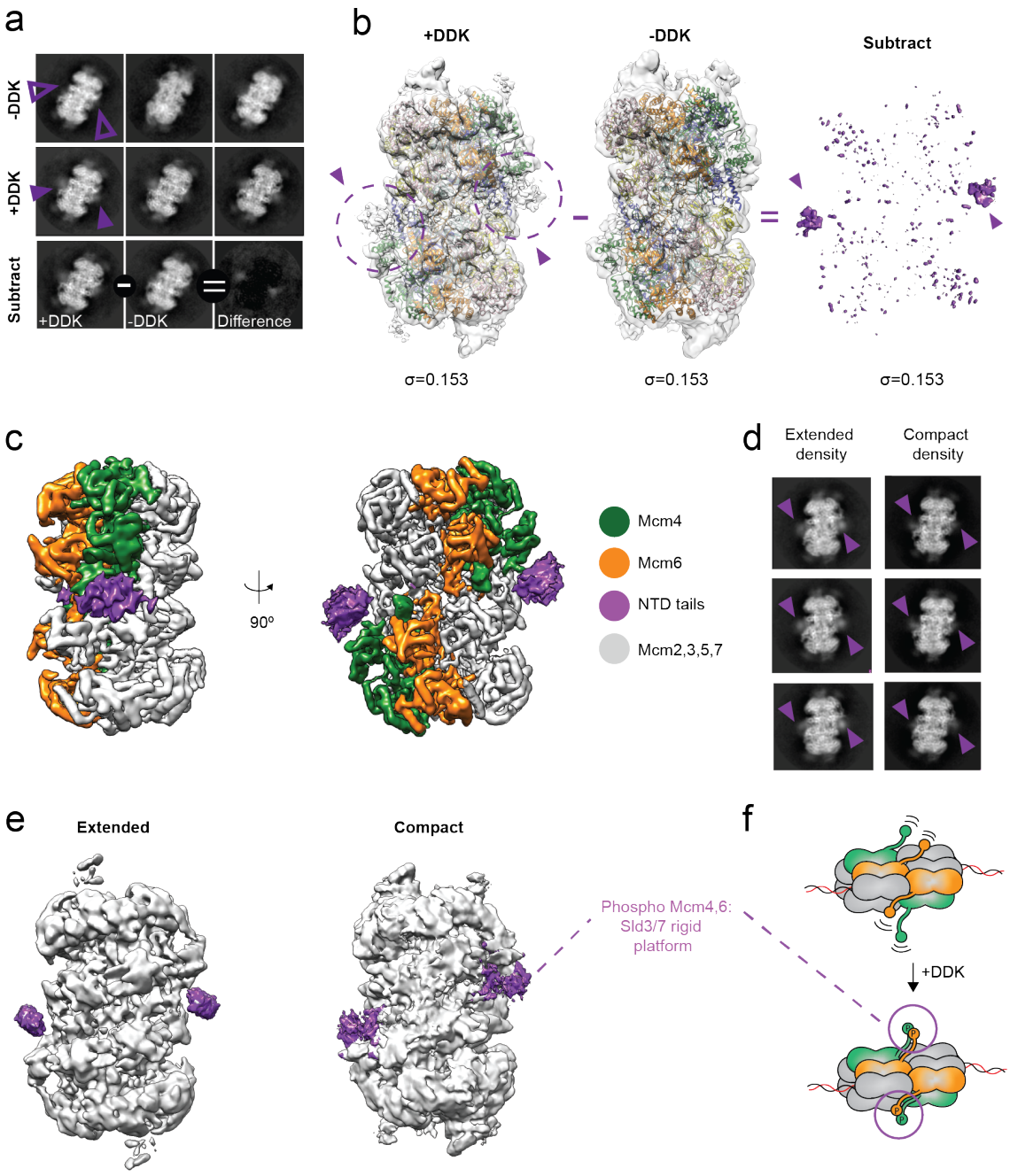


Figure 3.10: Effect of DDK phosphorylation on the double hexamer

(a) Comparison of 2D class averages between the unmodified (first row) and DDK phosphorylated (second row) MCM double hexamers. Third row shows a difference image (unmodified subtracted from the DDK phosphorylated class). **(b)** DDK phosphorylated and unmodified double hexamer unsharpened maps showed at low contour level. A difference map points to the extra density that is left behind when the unmodified is subtracted from the DDK phosphorylated double hexamer. **(c)** A hybrid contour level map of the DDK phosphorylated MCM double hexamer where the MCM core density is shown at σ=0.37 and the difference map at σ=0.08 (purple). **(d)** 2D classification reveals two types of additional density, extended and compact. **(e)** Refinement of two different DDK phosphorylated MCM double hexamer conformers. A mask was generated from the difference model in (b) and used for masked 3D classification. 38,971 particles (33% of total) belonging to the extended conformation and 78,067 particles (67% of total) belonging to the compact conformation were separately refined with C2 symmetry. Extended and compact structures were resolved to 7.2 Å and 5.5 Å and maps are shown at σ=0.0095 and σ=0.0054 contour level, respectively. **(f)** Summary of DDK effect on MCM double hexamer. DDK phosphorylation partially stabilises the N-terminal tails of Mcm4 and 6, making them visible upon averaging.

Therefore, particles contributing to each conformer were refined in C2 where 33% of particles went into the extended structure and 67% of particles went into the compact structures (**Figure 3.10d-e**). Although the significance of the two conformations remains to be established, these results indicate that phosphorylation of MCM by DDK promotes the relative stabilisation of the otherwise invisible tails of Mcm4 and 6. Phospho Mcm4 and 6 is read by Sld3 of the Sld3/Sld7 complex to promote binding of Cdc45 to the MCM (Deegan et al., 2016) and our structures suggest a model that underlies these events. Phospho-Mcm4 and 6 from both hexamers form a rigid, symmetrical density, which acts as a landing platform for the recruitment of two Sld3/7 dimers. In theory, the two-fold symmetrical features would in turn enable the docking of two copies of Cdc45 per double hexamer to start assembling the CMG concertedly (Itou et al., 2014, Itou et al., 2015, Deegan et al., 2016) (**Figure 3.10f**). However, reconstitution of an MCM-Sld3/7-Cdc45 intermediate is required to provide evidence for such a mechanism.

3.5 The transition from double hexamer to active CMG

CMG activation requires three distinct conformational changes. The MCM transitions from binding duplex DNA (*section 3.4*) to melting the DNA upon CMG assembly to retaining only one strand inside the channel (Georgescu et al., 2017, Douglas et al., 2018). At the same time the double hexamer has to separate prior to bi-directional unwinding as the active CMG exists as a monomer (Costa et al., 2011, Ilves et al., 2010, Yardimci et al., 2010) but the timing and structural changes that accompany this dissociation is unknown.

I showed earlier that DDK phosphorylation of MCM is insufficient to destabilise the double hexamer (see *section 3.4*). A key candidate for splitting the two helicases instead is Mcm10, which is not required for CMG assembly but is essential for full activation of unwinding (Yeeles et al., 2015). Mcm10 has previously been shown to have the capacity to bind both MCM and the CMG (Douglas and Diffley, 2016, Langston et al., 2017), split double hexamers (Quan et al., 2015) and induce full DNA extrusion (Douglas et al., 2018). While the precise architecture of the CMG helicases after step-wise assembly is unknown, a negative stain study of an overexpressed *Drosophila* CMG has shown that a CMG dimer intermediate can exist (Costa et al.,

2014). Therefore, a compelling hypothesis is that Mcm10 separates two tethered CMGs before strand extrusion. The reconstitution of eukaryotic DNA replication *in vitro* has provided a platform to address this hypothesis and identify what factors play a role in double hexamer separation (Yeeles et al., 2015).

3.5.1 Testing the suitability of the origin firing assay for structural studies

To first test whether the *in vitro* reconstitution reaction is an ideal target for EM studies, I collaborated with Dr. Max Douglas who performed the complete origin firing reaction with purified budding yeast proteins. First, the MCM loading factors were added (ORC, Cdc6, Mcm2-7-Cdt1) to 2.8 kb linear bead bound DNA, followed by DDK treatment. After washing off the supernatant, the remaining firing factors were added (CDK, Sld2, Sld3, Sld7, Dpb11, Polε, GINS, Cdc45, Mcm10). Another high salt wash was performed and the DNA was cut off before I negatively stained the sample for EM analysis (**Figure 3.11a**) (see *section 2.4.6* for methods). Both Mcm2-7 double hexamers and CMG are salt resistant complexes whereas the other firing factors should be washed away in high salt (Remus et al., 2009, Yeeles et al., 2015).

Negative stain EM analysis revealed that 58% of MCM containing particles were fully assembled CMGs, indicating that approximately 40% of the input double hexamers had been converted to activated helicases. This result is the first visual evidence of CMG formation using this assay and agrees well with the biochemistry (Yeeles et al., 2015). Notably, no evidence for single MCM rings nor double CMGs were found in this sample. Furthermore, the CMG-Mcm10 interaction is likely to be transient and sensitive to high salt washes as no additional density corresponding to Mcm10 was identified in the CMG classes (**Figure 3.11b**). The CMG particles in this reaction showed several different orientations, making them amenable to 3D reconstruction (**Figure 3.11c-e**). The 3D structure shows the characteristic two-tiered architecture of the CMG (Costa et al., 2011, Costa et al., 2014) and docking of the atomic coordinates of the CMG allowed mapping of the different subunits (Georgescu et al., 2017)(**Figure 3.11c**). This structure further agrees with the yeast CMG structures previously determined from over expression strains (Sun et al., 2015, Zhou et al., 2017). Thus, the CMG assembled step-wise in an *in vitro* origin firing reaction can be used for single particle analysis to inform helicase activation.

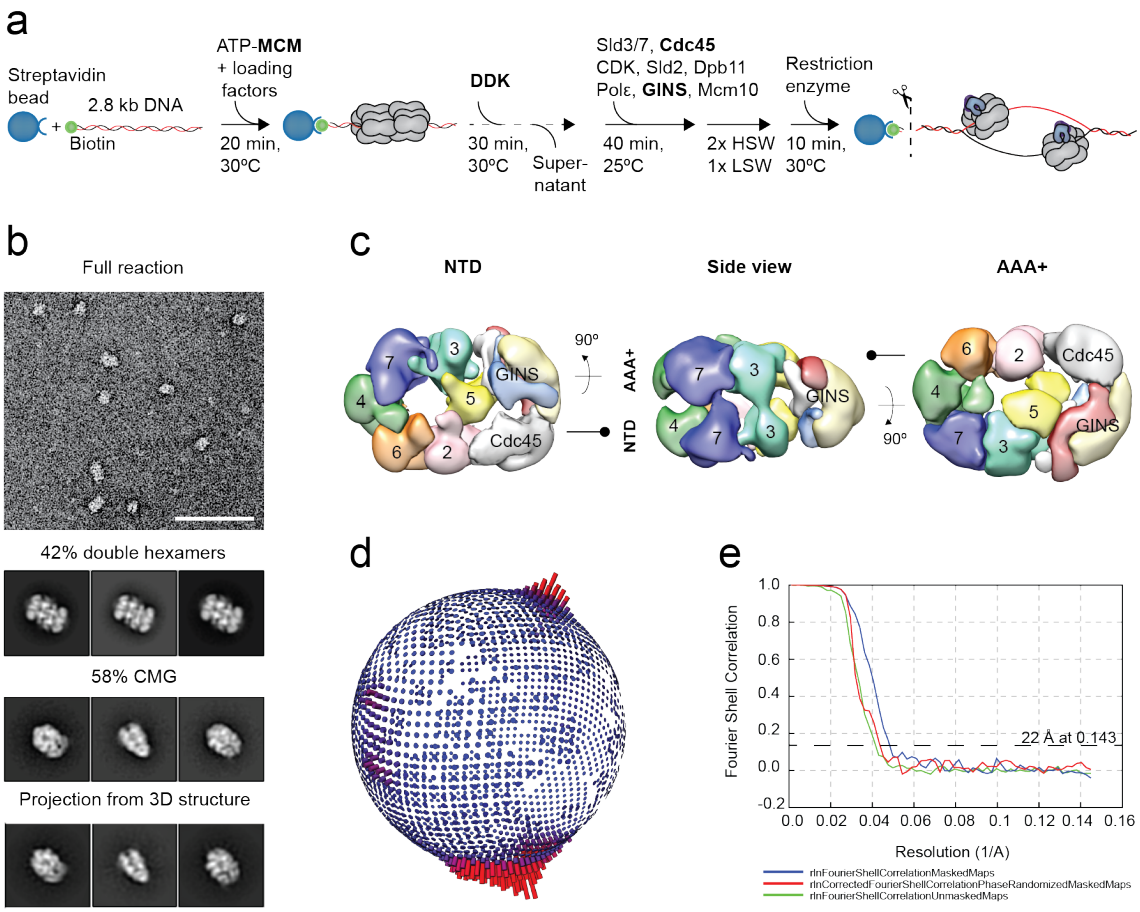


Figure 3.11: EM analysis of reconstituted origin firing and 3D reconstruction of active CMG
(a) Schematic overview of the reaction set-up to assemble and activate DNA bound MCM complexes using purified budding yeast proteins (carried out by Dr. Max Douglas). **(b)** Representative, cropped electron micrograph acquired at 30,000x collected using a Tecnai LaB6 G2 Spirit transmission electron microscope (FEI) operating at 120 keV. RELION-1.4 reference free class averages of helicase activation reactions washed in high salt buffer (128 pixel box size, 3.45 Å/pixel). 14,688 particles were CMG and 10,492 particles were double hexamers out of a total 43,320 particles. Scale bar is 100 nm. **(c)** 3D reconstruction of CMG (14,688 particles) using the negative stain DmCMG as a starting model (EMD: 2772) filtered to 60 Å in RELION-2. Atomic coordinates PDB: 5u8S were docked into the negative stain EM structure to map the subunits using the color zone function in Chimera (coordinates not shown for clear depiction of EM density). **(d)** Euler angle distribution. **(e)** Gold-standard Fourier Shell Correlation (0.143 criterion).

3.5.2 Establishing the temporal order of helicase activation

To identify when double hexamer separation occurs, we performed dropout experiments and analysed the contents of each reaction by EM. Omission of either of DDK, Sld3/7, Dpb11 or Sld2 showed 100% of the MCM-containing particles stuck in a double hexamer state (**Figure 3.12a**). This is in line with their essential role in CMG assembly (Yeeles et al., 2015). Furthermore, Cdc45 recruitment on its own (in samples where Sld2 of Dpb11 were omitted) is not sufficient to dissociate the two hexamers.

Next, I analysed a reaction where only Mcm10 was omitted (**Figure 3.12b**), a condition that allows CMG formation but not full CMG activation (Yeeles et al., 2015). Surprisingly, this sample showed a similar percentage (64%) of single CMGs compared to the complete reaction and no evidence of double CMGs (**Figure 3.12b**). To confirm that this result is not biased by a high salt wash, the reaction was repeated with a low salt wash and still exhibited a near identical proportion of single CMGs (63%) (**Figure 3.12b-c**). Therefore, separation of helicases occurs during CMG assembly and prior to full activation by Mcm10.

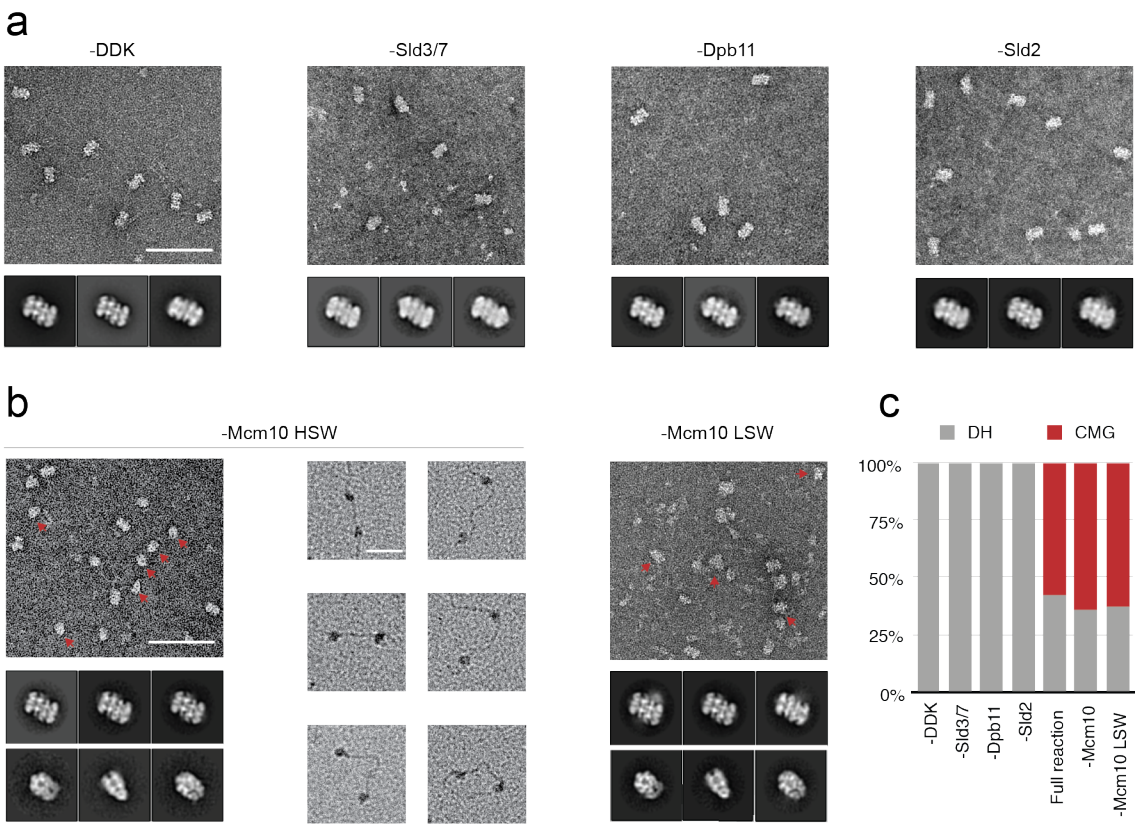


Figure 3.12: Structural characterisation of salt stable replicative helicase activation intermediates
(a) Representative, cropped electron micrographs of helicase activation reactions washed with high-salt wash buffer (see methods *section 2.4.6*) acquired at 30,000x magnification on a Tecnai LaB6 G2 Spirit transmission electron microscope (FEI) operating at 120 keV and using a 2K × 2K GATAN Ultrascan 100 camera. Lower panels show representative reference free 2D classes of MCM containing classes acquired in RELION-2. -DDK: 7,410 of 23,092 picked particles were double hexamers. -Sld3/7: 2,039 of 20,904 particles were double hexamers. -Dpb11: 5,032 of 6,815 particles were double hexamers. -Sld2: 2,866 of 10,670 particles were double hexamers. **(b)** Same as (a) apart from the right –Mcm10 HSW panel, which shows positively stained MCM containing complexes bound to DNA acquired on a Tecnai G2 F20 TWIN electron microscope operating at 200 keV (FEI; Electron Microscopy Centre, Imperial College London) equipped with a Falcon II direct electron detector (FEI). -Mcm10 HSW: 3,984 and 2,226 of 12,920 particles were CMG and double hexamers, respectively. -Mcm10 LSW: 8,859 and 5,120 of 38,885 particles were CMG and double hexamers, respectively. **(c)** Quantification of MCM double hexamer and CMG particles expressed as a percentage of the total number of MCM containing particles. N=1. Scale bar is 100 nm.

Positive stain analysis revealed that the separate, inactive CMG helicases can passively diffuse apart by up to 400 bp on the same DNA molecule *in vitro* (**Figure 3.12b**). However, *in vivo* the extent of separation may be restricted by nucleosome positioning.

In summary, MCMs remain tethered as double hexamers until the full CMG is assembled. This occurs concomitantly with untwisting of half a turn of the DNA double helix (Douglas et al., 2018) (See *section 4.2.9*). As such, Mcm10 likely acts later in origin firing, where it promotes replication fork establishment and ATP-hydrolysis powered DNA fork unwinding (Douglas et al., 2018).

3.6 Conclusions

In this study, we determined the structure of the MCM double hexamer bound to DNA both before and after DDK phosphorylation. Although sliding of MCM double hexamers off the DNA was observed, a subset of DNA bound MCM particles were identified using signal subtraction techniques. This allowed a fully occupied MCM-DNA structure to be reconstructed, revealing that duplex DNA spans the entire length of the MCM channel and bends at the double hexamer interface. Importantly, this observation is inconsistent with the prediction of DNA melting at the origin licensing stage (Li et al., 2015) or egression of the DNA via side gates of the double hexamer (Hizume et al., 2017).

DDK phosphorylation does not alter the conformation of the core MCM structure, the ATPase state nor the DNA inside the channel but instead stabilises the flexible N-terminal tails of phospho-Mcm4 and Mcm6. We suggest that this symmetrical structural feature forms a recruitment platform for the coordinated docking of Sld3/7 and Cdc45, required for the downstream activation of the helicase.

Taken together, this study suggests that double hexamers are inactive because they bind duplex DNA, ADP and are stably locked through extensive inter-hexamer interactions. The DDK phosphorylated double-hexamer structure indicates that the effect of phosphorylation is manifested locally on the Mcm4,6 tails. This goes against allosteric domain-push models previously proposed by *mcm* mutations that bypass

DDK requirement (Fletcher et al., 2003). Finally, GINS recruitment to MCM-Cdc45 is concomitant with double hexamer separation. This could ensure that the sequential Cdc45 and GINS incorporation occurs on the double hexamer for concerted assembly of two CMGs prior to bi-directional fork establishment.

Chapter 4. Architecture of the active CMG helicase

4.1 Introduction

In eukaryotes, a pre-requisite for DNA replication is the unwinding of the double helix by the CMG helicase (Cdc45, Mcm2-7, GINS) ahead of dedicated replicative polymerases that synthesise the daughter strands (Pol α-primase, Pol ε and Pol δ) (Burgers and Kunkel, 2017). These functions are contained within the replisome, a large protein assembly where the CMG acts as both an organizing centre and leading edge of fork progression. Despite identification of the essential properties of CMG, the molecular basis of helicase activation and replication fork establishment are poorly understood. Building on the results in Chapter 3, in this chapter I describe the cryo-EM structure of the pre-assembled CMG bound to DNA and analyse the reconstituted origin firing reaction using negative-stain EM. Collectively, the results presented here contribute to emerging principles on helicase activation and fork unwinding (Yuan et al., 2016, Yeeles et al., 2015, Georgescu et al., 2017, Douglas et al., 2018, Abid Ali et al., 2016). In the following section (*4.1.1*), I will summarise the key findings relating to helicase activation and the unanswered questions that motivated the aims of this study (*section 4.1.2*).

4.1.1 Background

The motor of the CMG helicase is the Mcm2-7 hetero-hexameric ring, comprising six ATPase sites at inter-subunit interfaces and a large interior channel that harbours the DNA substrate. As shown in Chapter 3, Mcm2-7 initially loads onto origin duplex DNA as an inactive double hexamer, which in turn acts as a platform for the recruitment of auxiliary factors Cdc45 and GINS in a phosphorylation dependent manner (Yeeles et al., 2015). Cdc45 and GINS bridge the Mcm2-5 site, which is a known gate in MCM that is used for DNA loading during origin licensing (Samel et al., 2014, Bochman and Schwacha, 2007, Costa et al., 2011). In binding to Mcm2-5, Cdc45 and GINS reconstitute the ATPase sites required for helicase activity (Ilves et al., 2010, Costa et al., 2011) although their precise roles in the helicase are unknown.

Notably, the recruitment of GINS during holo-helicase formation promotes a simultaneous dissociation of the double hexamer into two assembled 11-subunit

CMGs (as shown in *section 3.5.2*). This step is concomitant with untwisting of half a turn of duplex DNA per formed CMG (Douglas et al., 2018). However, the molecular basis for both double hexamer destabilisation and origin melting are unclear.

The two CMG helicases subsequently become activated by Mcm10 to trigger ATP hydrolysis and unwind two DNA forks bi-directionally (Douglas et al., 2018). Whether the CMG translocates on single or double-stranded DNA is unknown. Currently, there are three main models to explain how the CMG could drive replication fork progression: (1) strand extrusion (Li et al., 2015), (2) plough-share (Takahashi et al., 2005) and (3) strand exclusion (Fu et al., 2011). The strand extrusion model posits that the CMG initially harbours duplex DNA and the two strands separate only once inside the channel. This model envisions that the leading strand template exits from the main channel whereas the lagging strand template emerges from a side gate between the two tiers (Abid Ali and Costa, 2016, Li et al., 2015). In the plough-share model, the helicase binds duplex DNA and after origin melting separates the DNA using an associated proteinaceous 'pin' (Singleton et al., 2004, Saikrishnan et al., 2008). In contrast, the classical steric exclusion model suggests the Mcm2-7 ring in the CMG would only translocate on single-stranded DNA (leading strand template) while sterically excluding the lagging strand template from entering the main channel of the motor (Bochman and Schwacha, 2009, Fu et al., 2011). Thus, the mode of DNA translocation by the CMG is also an unresolved question in the field.

Although multiple lines of evidence indicate that the helicase unwinding occurs in a 3'-5' polarity (Kelman et al., 1999, Chong et al., 2000, Ilves et al., 2010, McGeoch et al., 2005), the orientation of the translocating CMG relative to a replication fork is still a matter of debate (Costa et al., 2014, Georgescu et al., 2017, Rothenberg et al., 2007, McGeoch et al., 2005, Froelich et al., 2014, Trakselis, 2016). As shown in Chapter 3, the two MCMs in the double hexamer associate via their NTDs whereas the AAA+ tiers face outward (**Figure 3.1c**). In a AAA+ tier first model, the two CMGs would be poised to immediately bi-directionally unwind upon activation. If, however, CMG translocates on DNA NTD first, then the two CMGs have to first cross their paths prior to extensive unwinding. Elucidating CMG polarity will therefore give insights into how unwinding is initiated and will also inform how the replisome is organised.

4.1.2 Aims

In this chapter, I will use biochemical approaches and single particle cryo-EM to structurally characterise a CMG-DNA complex. Combined with the DNA bound double hexamer from Chapter 3 and EM analysis of the *in vitro* reconstitution of origin firing (Douglas et al., 2018), this will help propose a model for helicase activation and replication fork establishment. Specifically, I aim to address the following questions:

1. How does the MCM in the CMG compare to the MCM in the double hexamer?
2. What is the architecture of the fully assembled CMG? What role do the Cdc45 and GINS activators play in the holo-helicase?
3. What are the CMG ATPase conformational states that could drive substrate translocation?
4. How does CMG assembly lead to DNA melting and double hexamer separation during this transition?
5. Does the CMG translocate on single or double-stranded DNA? What is the conformation of the DNA inside the channel?
6. What is the polarity of the active CMG helicase relative to a replication fork?

I will use two main methods to obtain a CMG complex: (*i*) pre-assembly of 11 *Drosophila* helicase subunits via overexpression and reconstitution with a model DNA replication fork (Costa et al., 2011; Georgescu et al., 2014; Ilves et al., 2010) and (*ii*) step-wise *in vitro* reconstitution of origin firing using budding yeast proteins (performed by Dr. Max Douglas, Yeeles et al., 2015). Results presented here will be analysed in the context of other CMG cryo-EM structures that were published at around the same time as this study (Georgescu et al., 2017, Yuan et al., 2016).

4.2 Results

4.2.1 Purification and negative stain EM analysis of CMG

The first step towards structural characterisation of the eukaryotic replicative helicase bound to DNA is to obtain a homogenous and pure CMG preparation. To this end, I expressed and purified recombinant *Drosophila melanogaster* CMG (hereafter simply CMG) from insect cells using a modified protocol from a previously published study (Ilves et al., 2010) (outlined in **Figure 4.1**). To express CMG, I first transfected plated *Spodoptera frugiperda* Sf9 insect cells with 11 different bacmids that each encode a CMG subunit (gifts from Dr. Michael Botchan). This step generated baculoviruses, which were harvested and amplified to get the second viral passage (P2). After repeating this step and obtaining the third viral passage (P3), I co-infected a four-litre suspension of Hi5 insect cells with all 11 baculoviruses to induce protein expression of the full CMG complex (detailed in **Figure 4.1a**). Hi5 cells were used as they give a 5-10-fold increase in expressed protein compared to expression in Sf9 cells (Wickham and Nemerow, 1993).

To isolate the full 711 kDa CMG complex from the cleared lysate, a four-stage purification scheme was employed (**Figure 4.1b**). First, CMG was pulled down via FLAG-Mcm3 (**Figure 4.2b**, lane 1) followed by a cation exchange chromatography step (SP Sepharose Fast Flow) to trap contaminating nucleases in the sample. The flow through from the SP Sepharose column was then applied to an anion exchange column (Mono Q 5/50 GL) to separate out MCM sub-complexes from the pure CMG fraction, the latter of which eluted as a sharp peak at ~40 mS/cm (**Figure 4.2a**). SDS-PAGE verification showed the presence of all 11 subunits at their expected molecular weight after this step and an improved purity relative to the pull-down elution (**Figure 4.2b**).

Traditionally, density gradients or gel filtration purification strategies are used as final protein sizing steps to obtain homogenous samples for structural studies. However, the yield of CMG was too low (<100 nM in a 75 µl volume) after the Mono Q step to attempt these methods and a relatively high concentration of CMG is required for cryo-EM sample preparation. To circumvent this problem, I applied the CMG to a

second but smaller Mono Q column (100 µl column, Mono Q PC 1.6/5) (**Figure 4.2c-d**). This concentrated the CMG 3-fold (~350 nM in a 75 µl volume) compared to the elution after the first Mono Q step and served to purify the complex further (**Figure 4.2c** compared to **Figure 4.2a**). Analysis of the peak fraction on a silver stained SDS-PAGE gel confirmed the presence of pure CMG after implementation of the complete purification strategy (**Figure 4.2d**).

To confirm that the CMG sample contained fully assembled complexes, the peak fraction from the second Mono Q stage was dialysed down to a low salt concentration (50 mM sodium acetate) and then bound to ATP for stabilisation, followed by negative-stain EM analysis (**Figure 4.3**). Under the electron microscope, the stained grids showed a homogenous distribution of monodisperse particles (**Figure 4.3a**), exhibiting dimensions and views that agree with previous CMG studies (Costa et al., 2011, Costa et al., 2014). Subsequent single particle image processing and reference-free 2D classification in RELION-1.4 (Scheres, 2012) revealed that the CMG was of one major molecular species and showed various orientations of the complex (**Figure 4.3b**), making the data suitable for 3D reconstruction. After 3D refinement in the image processing package SPARX (Hohn et al., 2007) using the apo CMG negative stain structure as an initial model (Costa et al., 2011), a map of CMG-ATP was obtained at 25 Å resolution (**Figure 4.3c**). For validation of the structure, the CMG 3D volume was re-projected via the image processing package IMAGIC (van Heel et al., 1996), and the resulting re-projections were compared to the reference-free 2D classes, showing a good agreement (**Figure 4.3b**). To further validate and interpret the EM density, the recent atomic structure of yeast CMG (Yuan et al., 2016) was docked into the structure, confirming subunit assignments of previous studies (Costa et al., 2011, Costa et al., 2014) (**Figure 4.3c**).

The CMG structure shows the distinguishing two-tiered rings of the helicase, where GINS and Cdc45 contact NTD MCM2-7 via the Mcm2-5-3 interface to restrain the Mcm2-5 gate at this tier (**Figure 4.3c**). In contrast, in this ATP bound state the Mcm2-5 gate is in an open configuration at the AAA+ tier. Aside from a conformational difference at Mcm2-5, the architecture of the CMG obtained by overexpression in this case is highly similar to the negative stain yeast CMG map acquired using the *in vitro* reconstitution strategy (*see section 3.5.1*). This suggests that both methods of

CMG complex assembly are suitable for EM studies and can be used interchangeably to gain mechanistic insights into helicase function. However, the concentration of *D. melanogaster* CMG achieved by co-overexpression here is much higher than the yeast CMG *in vitro* reconstitution strategy and thus the former will be the focus of the cryo-EM study.

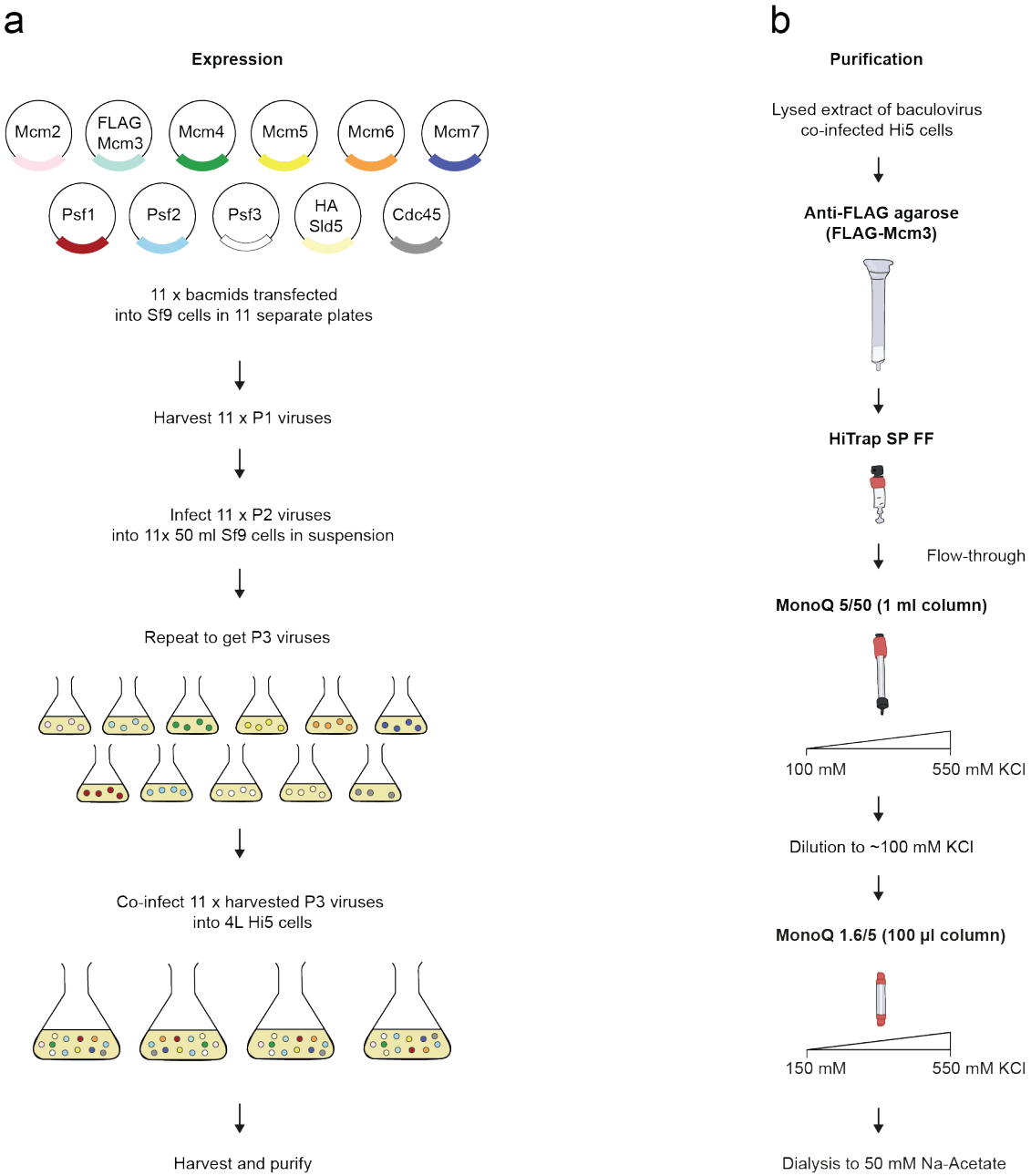


Figure 4.1: Overview of DmCMG expression and purification strategies
(a) Baculovirus-insect cell expression strategy. **(b)** Purification approach to obtain a homogenous CMG preparation.

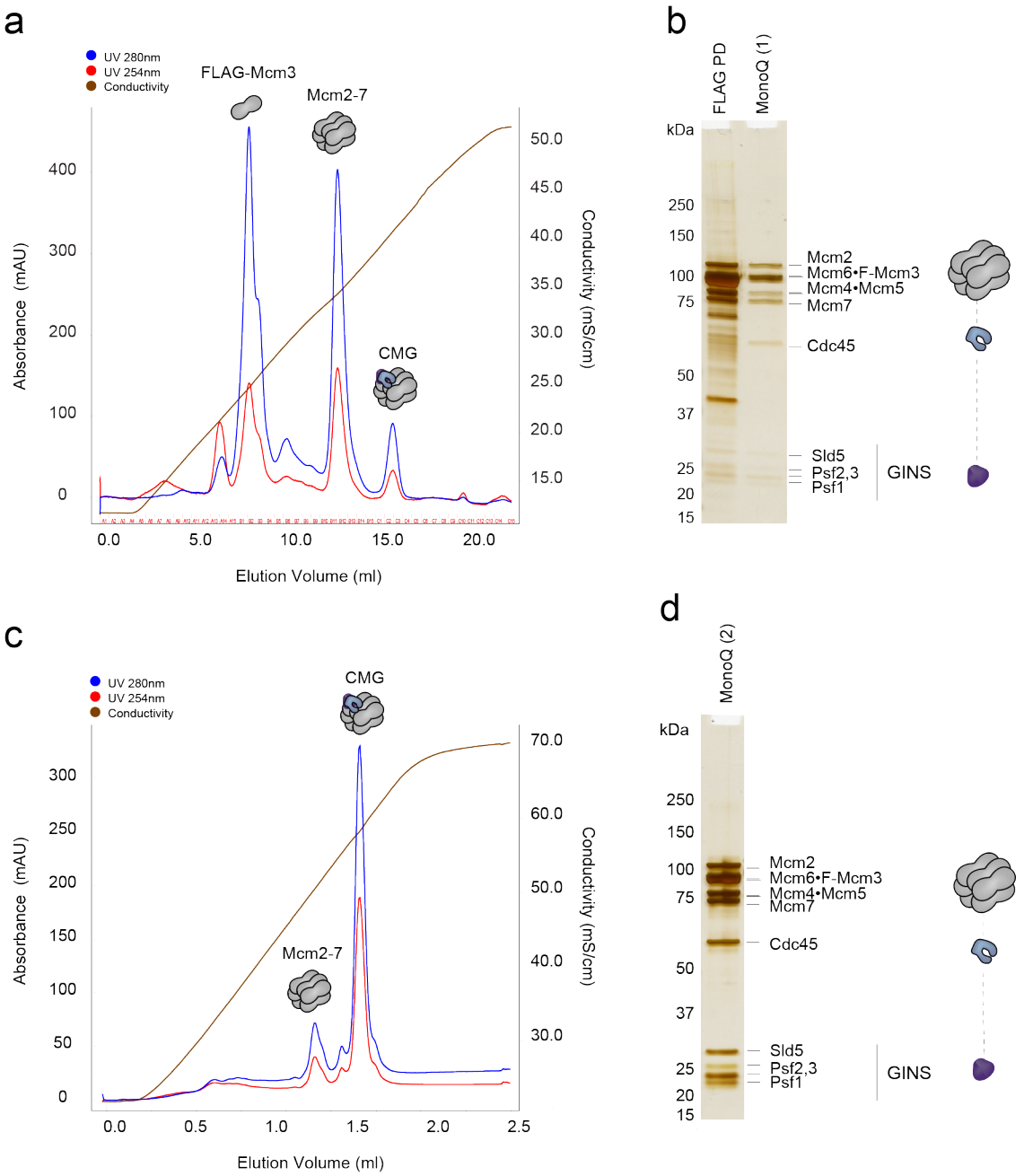


Figure 4.2: Purification of DmCMG
(a) Representative Mono Q 5/50 GL chromatogram shows the separation of the various MCM subcomplexes from the main CMG peak. **(b)** Silver stained SDS-PAGE gel of the FLAG pulldown elution and MonoQ 5/50 GL CMG peak elution. **(c)** Representative MonoQ PC 1.6/5 chromatogram shows the separation of MCM2-7 from CMG. **(d)** Silver stained SDS-PAGE gel of the Mono Q PC 1.6/5 CMG peak elution.

In summary, I have showed that the CMG sample prepared is pure and homogenous, and negative stain EM analysis indicates that the preparation is suitable for higher resolution structure determination of the helicase bound to DNA.

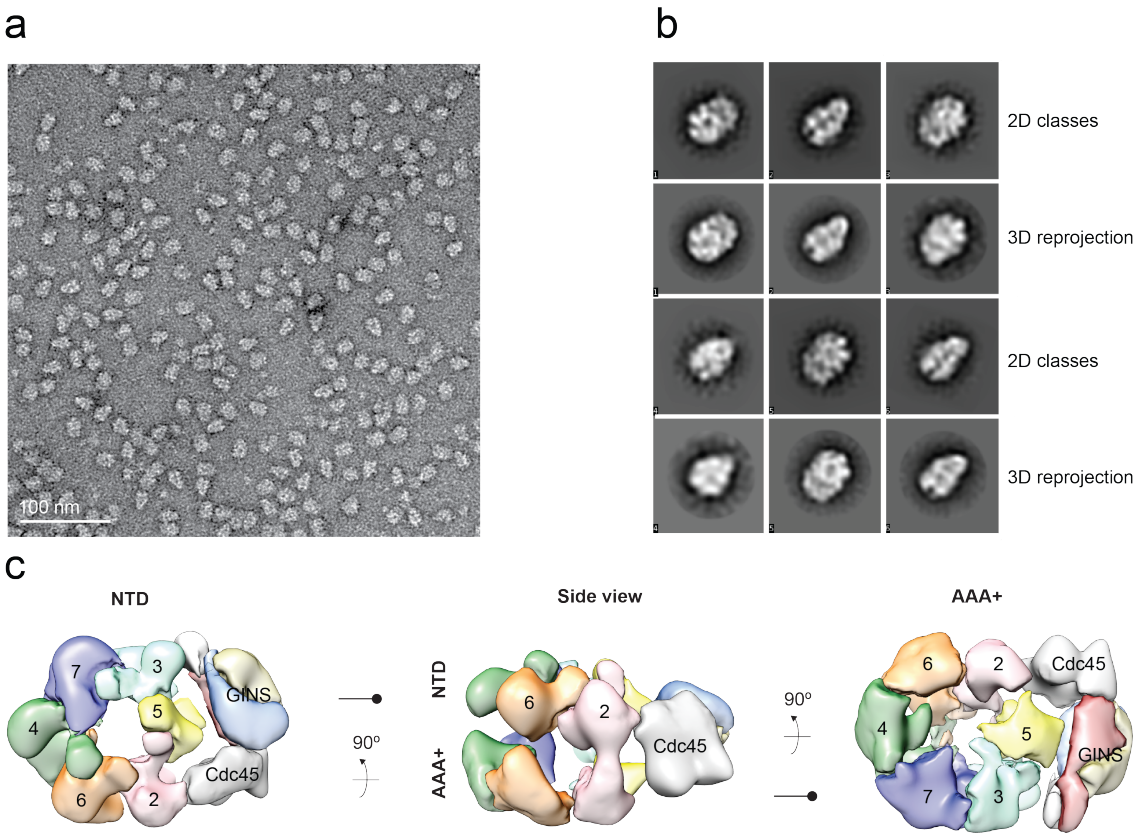


Figure 4.3: Negative stain EM analysis of DmCMG
(a) Representative electron micrograph of purified DmCMG imaged at 42,000 magnification (pixel size: 2.46 Å/pixel) on a Tecnai LaB6 G2 Spirit transmission electron microscope (FEI) operating at 120 keV and using a 2K × 2K GATAN Ultrascan 100 camera. 120 micrographs were collected in total. **(b)** 2D class averages of DmCMG (as performed in RELION) and 3D re-projections from the structure in (c) as implemented in IMAGIC (multi reference alignment). Box size 128 pix. **(c)** Negative stain 3D model of DmCMG-ATP at 25 Å obtained in SPARX from 9,287 particles. Atomic coordinates from a yeast CMG structure (PDB: 5u8s) were docked in and then "colour zoned" in Chimera to map and split all 11 subunits subunits.

4.2.2 Sub-nanometre structure of the CMG helicase

To improve our understanding of how ATP binding modulates DNA engagement by the CMG, I set out to reconstitute a CMG-DNA assembly and image it by cryo-EM (**Figure 4.4a**). In the *in vitro* reconstitution assay, double hexamers are first loaded onto duplex linear DNA and upon activation, the CMG transitions to replication fork DNA unwinding. In the present overexpression strategy, CMG is pre-assembled in the absence of DNA. To therefore capture and stabilise an active form of the replicative helicase, I supplemented the purified CMG sample with a pre-formed synthetic fork. The DNA fork contains a 50 bp duplex DNA region followed by two single-stranded overhangs: (*i*) the leading strand template (3' overhang) comprises 40 nt of a poly-T repeat, which has been shown to readily bind to CMG and (*ii*) the lagging strand template (5' overhang) comprises 16 nt of a CG repeat, which has

been shown to prevent non-productive binding (Petojevic et al., 2015). Following purification as described in *section 4.2.1*, CMG was incubated with ATPγS and two-fold molar excess fork DNA was added for two hours to form a ternary complex (**Figure 4.4a**). ATPγS nucleotide was used as it has been shown to be required for *in vitro* binding of CMG on synthetic DNA forks (Ilves et al., 2010, Petojevic et al., 2015) and this nucleotide will prevent the helicase from completely unwinding the substrate (as ATPγS is a slowly hydrolysable ATP analogue).

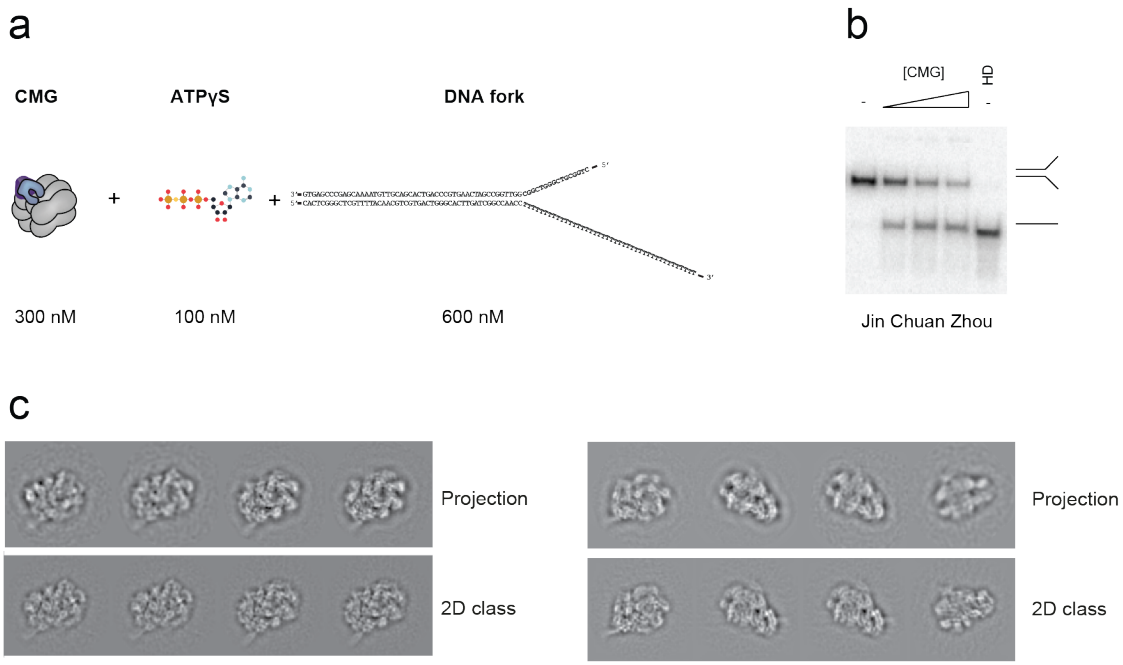


Figure 4.4: Sample preparation, data collection and initial image processing of CMG-DNA
(a) Schematic overview of the CMG-DNA reconstitution reaction. Reactions assembled in a 50 µl volume. **(b)** Unwinding assay of increasing amounts of CMG relative to a fixed amount of DNA fork. Each lane represents a 20 µl reaction comprising 25 mM Hepes pH 7.5, 75 mM NaCl, 0.5 mM ATP, 10 mM magnesium acetate, 1 mM DTT, 0.1 mg ml−1 BSA. Purified CMG ranging from 123 to 410 fmol in the three middle lanes was mixed with 27 µM ATPγS and 1.8 pmol of [γ-32P]-ATP-labelled fork DNA substrate. To enable fork loading, this reaction was incubated for 3 hours at 30 °C. To initiate unwinding, ATP was added to 7 mM for 5 min at 30 °C. Reactions were quenched with 6X stop buffer containing 150 mM Tris pH 8.0, 3% SDS, 120 mM EDTA. Reactions were loaded on 8% PAGE in 0.1% SDS and 1XTBE running buffer to separate unwinding products. HD: heat denatured. **(c)** Projection matching between 2D class averages and projections of the 3D map (multi-reference alignment as implemented in the software package IMAGIC).

To establish whether binding to a DNA fork successfully activates the CMG helicase I had purified, its unwinding activity was tested by a postdoc in the lab, Dr. Jin Chuan Zhou (**Figure 4.4b**). The CMG was first incubated with a radiolabelled fork in the presence of ATPγS for three hours to enable DNA binding and unwinding was initiated by adding excess of ATP for 5 min. Analysis of the unwinding products revealed that the CMG is indeed catalytically active, with accumulation of single-stranded DNA for increasing amounts of CMG (**Figure 4.4b**).

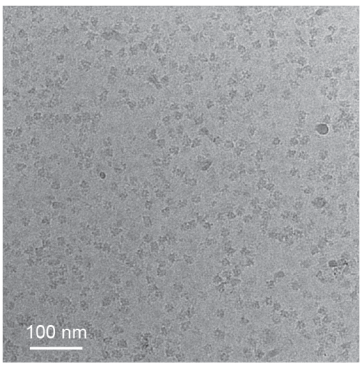
Given these results, I proceeded to structurally characterise CMG-DNA-ATPγS by cryo-EM to gain insights into the mode of CMG unwinding and conformational changes that occur during helicase translocation. To this end, particles were embedded in vitrified ice and imaged on a Thermo Fisher Polara electron microscope equipped with a K2 Summit camera operating in counting mode (grid preparation by Dr. Ludovic Renault, data collection by Dr. Abhay Kotecha). Approximately 2,000 movies were recorded across two datasets, allowing for correction of beam induced motion and recovery of high resolution information prior to summation of sub-frames using MorionCorr (Li et al., 2013) (**Figure 4.4c**, **Figure 4.5**, step 1).

After CTF estimation using CTFFIND4 (Rohou and Grigorieff, 2015) and semi-automatic particle picking using EMAN2 (Tang et al., 2007), 2D classification was performed in RELION-1.4 (Scheres, 2012), resulting in CMG averages in a range of orientations with recognisable projections of secondary structure elements (**Figure 4.4c, Figure 4.5**, step 2-3). These particles were selected and subjected to further 3D classification, 3D refinement and map sharpening in RELION-1.4 (**Figure 4.5,** step 4-6), leading to the reconstruction of two main CMG volumes: a predominant 'compact ATPase' form where the AAA+ tier is topologically closed, determined to 7.4 Å resolution (**Figure 4.6**) and a 'relaxed ATPase' form where the AAA+ tier is open (**Figure 4.7**), determined to 9.8 Å resolution. In the following sections, I will use the higher resolution 'compact ATPase' structure to first describe the MCM NTD and AAA+ tiers (*section 4.2.3*) and the role of Cdc45 and GINS in CMG (*section 4.2.4*). I will then compare the two CMG conformers to analyse ATPase reconfigurations in *section 4.2.5* and mode of DNA binding in *section 4.2.6*.

At the intermediate resolution range of 7.4 Å, alpha helices are resolved as tubular densities, allowing individual domain atomic coordinates to be fitted accurately. To map all of the subunits in the CMG density, I docked in the atomic coordinates that were available at the time of this study: yeast Mcm2-7 (PDB: 3JA8) (Li et al., 2015), human GINS (2Q9Q) (Chang et al., 2007) and human Cdc45 (5DGO) (Simon et al., 2016). The MCM subunits were individually separated into their subdomains (NTD [A, B and C subdomains] and AAA+) and docked in as rigid bodies giving an unambiguous fit (**Figure 4.6d, Figure 4.8**) (active MCM ring described in *section 4.2.3*).

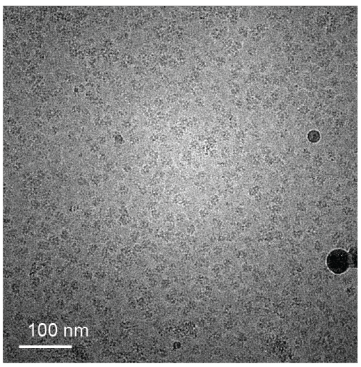
1. Data acquisition of CMG-DNA

Dataset 1



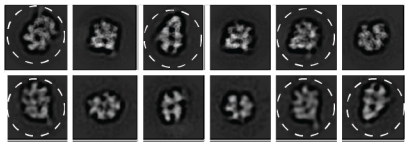
793 micrographs (1.35 Å/pix; 25 frames; 48 e /Å^2)
MotionCorr (bin first two frames)
EMAN2 (semi-automatic, 186,093 picked particles)
CTFFIND4 whole-frame

Dataset 2

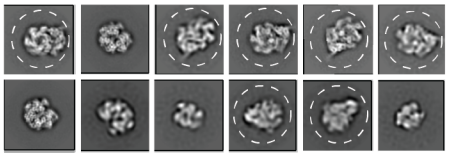


1,305 micrographs (1.35 Å/pix; 25 frames; 48 e /Å^2)
MotionCorr (bin first two frames)
EMAN2 (semi-automatic, 154,480 picked particles)
CTFFIND4 whole-frame

2. Reference-free 2D classification (bin by 2 particles, 2.7 Å/pix pixel size)



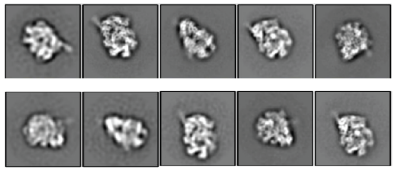
31,330 isolated Mcm2-7 particles (discarded)
60,287 CMG particles



22,892 isolated Mcm2-7 particles (discarded)
44,665 CMG particles

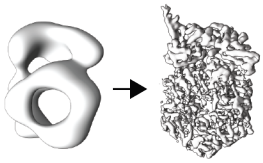
Combine CMG particles from two datasets
104,952 particles

3. Reference-free 2D classification of combined datasets (bin by 2 particles), 3 rounds



57,465 CMG particles selected, 'junk' particles discarded

4. 3D refinement RELION (bin by 2 particles)
Reference from negative stain (EMD-2772)
50 Å filtered



7.0 Å (after post-processing,
poorly resolved ATPase tier)

57,465 particles

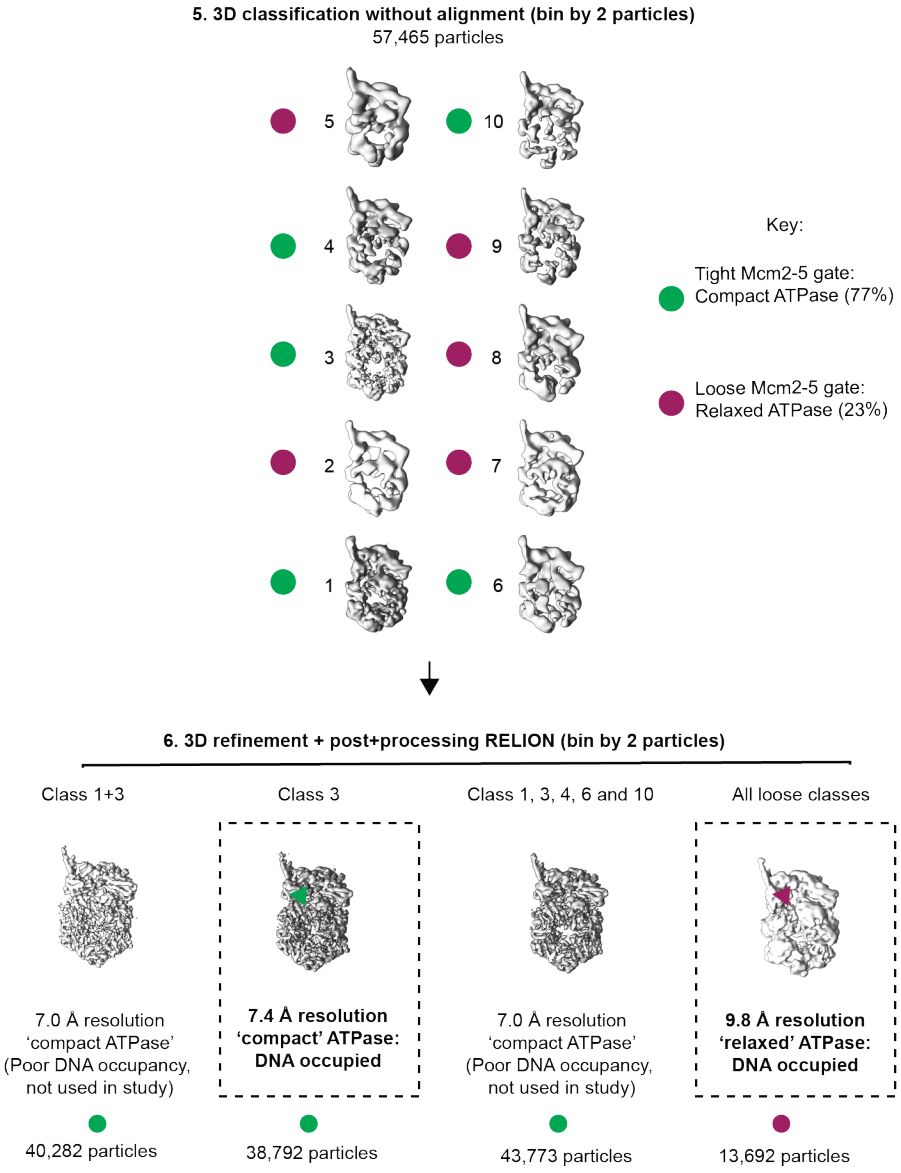


Figure 4.5: Single-particle CMG-DNA-ATPγS 3D reconstruction and classification overview.
Two datasets were collected with identical acquisition parameters on a Polara (ThermoFisher) equipped with a Gatan K2 summit direct electron detector in counting mode (step 1). All steps were performed in RELION version 1.4 with bin by 2 parameters (box size 112, pixel size 2.7 Å/pixel) to improve the signal-to-noise ratio. Three rounds of 2D classification were performed for each dataset (step 2) to remove junk particles, Mcm2-7 sub-complexes and ice contamination picks. CMG particles from both datasets (dashed circles) were pooled (step 3) and subjected to a further three rounds of 2D classification (step 3) and only the CMG classes exhibiting projections of secondary structure elements were selected for initial 3D refinement (step 4). This structure showed a poorly resolved ATPase tier, indicative of a mixture of conformational states. As such, the step 4 structure was used as an initial model for 3D classification without alignment (10 classes) to separate out conformers in the particle population (step 5). After inspection of classes, two main differences were observed in the ATPase tier between the Mcm2-5 interface, referred to as 'compact ATPase' (closed Mcm2-5, green circle) and 'relaxed ATPase' (open Mcm2-5 gate, purple circle). Different permutations of these classes were subjected to 3D refinement and only those structures comprising DNA and a well-resolved ATPase tier were selected for analysis in this study (dashed rectangles, step 6 [green and purple arrows point to Mcm2-5 gate]). After post-processing of the two half-maps from each structure, a 7.4 Å compact ATPase and 9.8 Å relaxed ATPase structure was obtained (using the gold-standard Fourier shell correlation at 0.143).

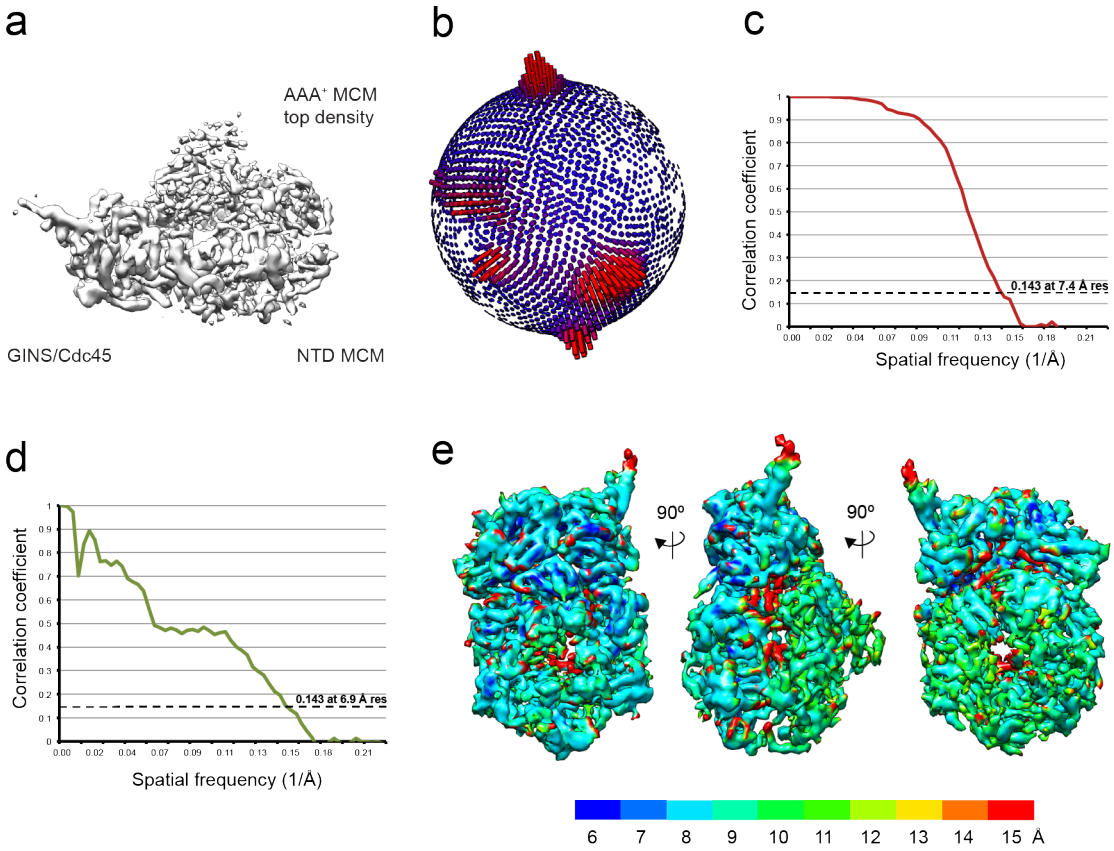


Figure 4.6: CMG-DNA-ATPγS Compact ATPase angular distribution and resolution estimation
(a) Side view of undocked cryo-EM map. **(b)** Euler plot. **(c)** Gold-standard Fourier Shell Correlation according to the 0.143 criterion. **(d)** cryo-EM map vs docked model Fourier Shell Correlation plot. **(e)** Local resolution as estimated by ResMap.

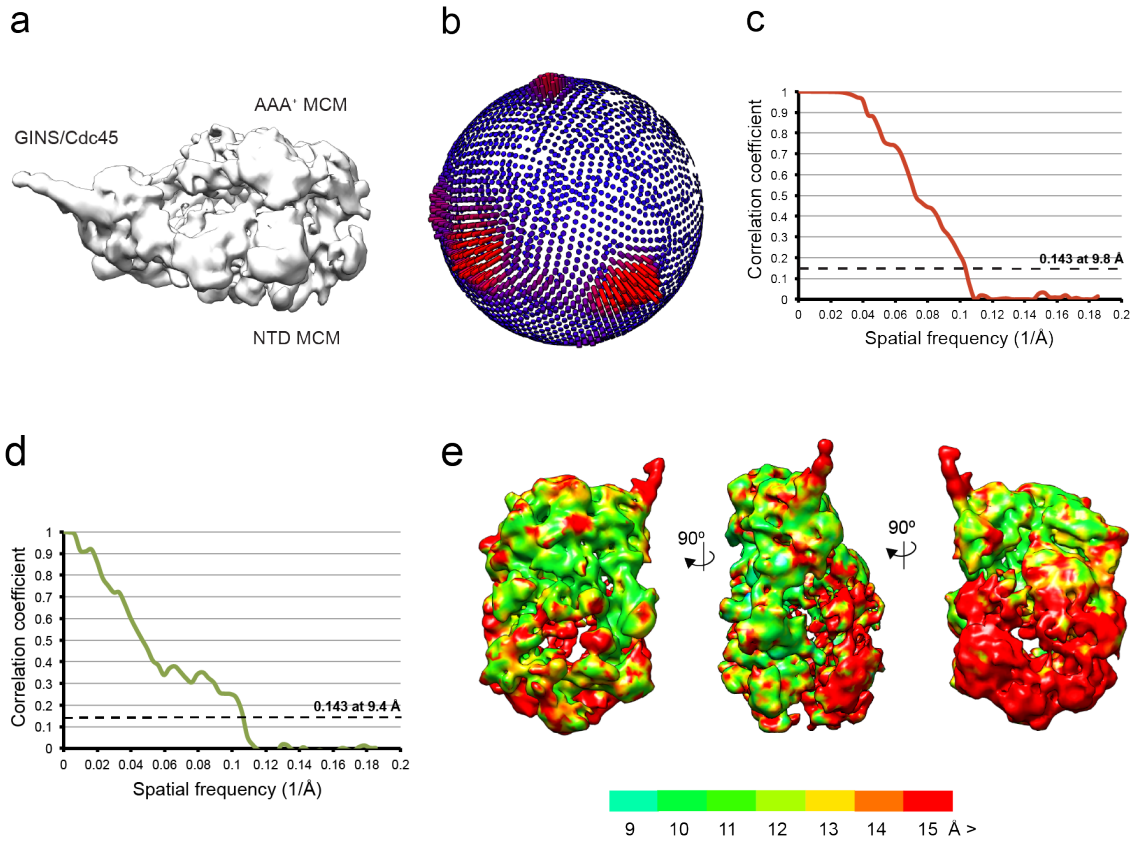


Figure 4.7: CMG-DNA-ATPγS Relaxed ATPase angular distribution and resolution estimation. **(a)** Side view of undocked cryo-EM map. **(b)** Euler plot. **(c)** Gold-standard Fourier Shell Correlation according to the 0.143 criterion. **(d)** cryo-EM map vs docked model Fourier Shell Correlation plot. **(e)** Local resolution as estimated by ResMap.

4.2.3 The active MCM complex in the CMG holo-helicase

Multiple lines of evidence suggest that the two tiers in the MCM ring preform different helicase functions (Costa et al., 2014). Whereas the NTD tier forms a non-catalytic DNA-interacting collar and is always found topologically closed, the AAA+ tier is dynamic and has been shown to reconfigure upon ATP and DNA binding (Froelich et al., 2014, Costa et al., 2014, Costa et al., 2011). The cryo-EM structure of the CMG here provides a much more detailed overview of how the active MCM complex is organised and remodelled upon Cdc45 and GINS binding (**Figure 4.8**).

As mentioned in the introduction, the NTD tier (**Figure 4.8a**) is arranged into a ring that structures the MCM complex. In the CMG, the central channel within the NTD ring is ~35 Å in diameter, which is wide enough to accommodate duplex DNA (20 Å in diameter). As with the inactive MCM, the A-domain projects peripherally from each subunit for protein-protein interactions (**Figure 4.8a**, lower panel).

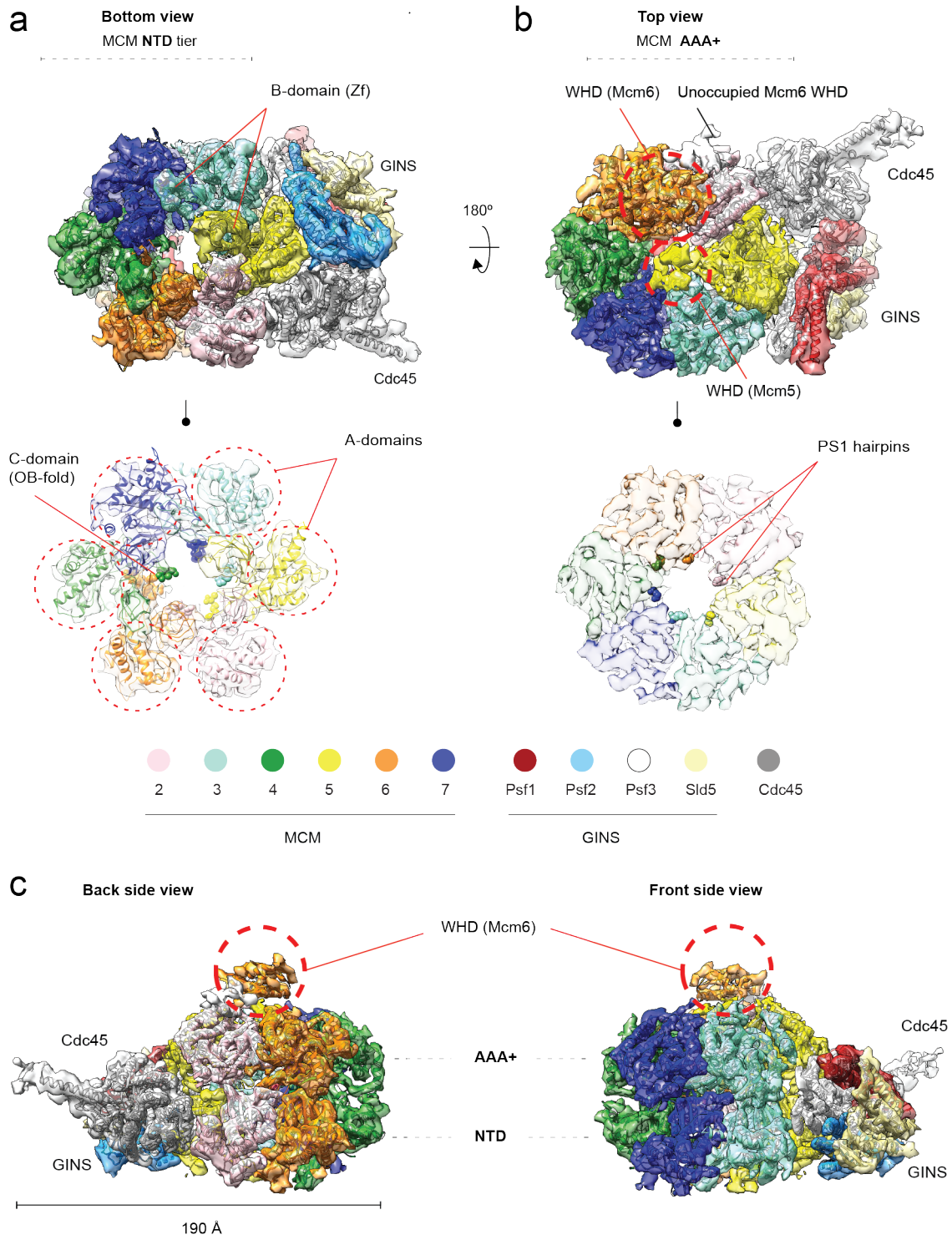


Figure 4.8: Cryo-EM structure of DmCMG at 7.4 Å resolution

(a) NTD tier of the MCM ring. Top panel highlights the B domain Zinc-finger crown. Inset shows the MCM ring, highlighting the peripheral A-domains and channel protruding C-domains. **(b)** The AAA+ tier of the MCM ring. WHD of Mcm5 partially occludes the channel. Mcm6 WHD rises above the AAA+ tier. Inset highlights the lysine and arginine residues of the PS1 ATPase pore loops that project towards the channel. **(c)** Side views of CMG complex highlight the Cdc45 and GINS interactions and the WHD of Mcm6. Docked coordinates: yeast Mcm2-7 (PDB: 3JA8), yeast WHD of Mcm5 and 6 (3JC7), human GINS (2Q9Q) and human Cdc45 (5DGO). NTD, N-terminal domain. AAA+, ATPases associated with various cellular activities. WHD, winged helix domain. PS1, presensor 1. OB-fold, oligonucleotide/oligosaccharide-binding fold.

In particular, the A-domains of Mcm2-5-3 interact with the co-factors, where Cdc45 binds Mcm2-5 while GINS contacts Mcm3-5. Side views of the CMG helicase reveal how Cdc45 interacts solely with the NTD tier whereas GINS wedges between the two MCM tiers (**Figure 4.8c**). Therefore, the NTD MCM ring is strengthened by multiple contacts from the helicase co-activators, which is also evidenced by the higher local resolution of this tier (6-8 Å) compared to the freely moving AAA+ ring (8-10 Å) (**Figure 4.6e**, left vs right panel).

The MCM NTD tier is significantly remodelled upon activation (**Figure 4.9**). Morphing between the double hexamer and CMG MCM-NTD structures revealed large inter-domain reconfigurations of the peripheral A-domains and a counter clockwise rotation of the B-C domains. During this transition, the B domains that form the internal interface simultaneously constrict and narrow the diameter of the MCM channel (movie is published online (Abid Ali et al., 2016)). Therefore, considerable protein rearrangements occur during the MCM to CMG transition (**Figure 4.9b**) that could conceivably also drive a reconfiguration of the encircled DNA (*see section 4.2.9*).

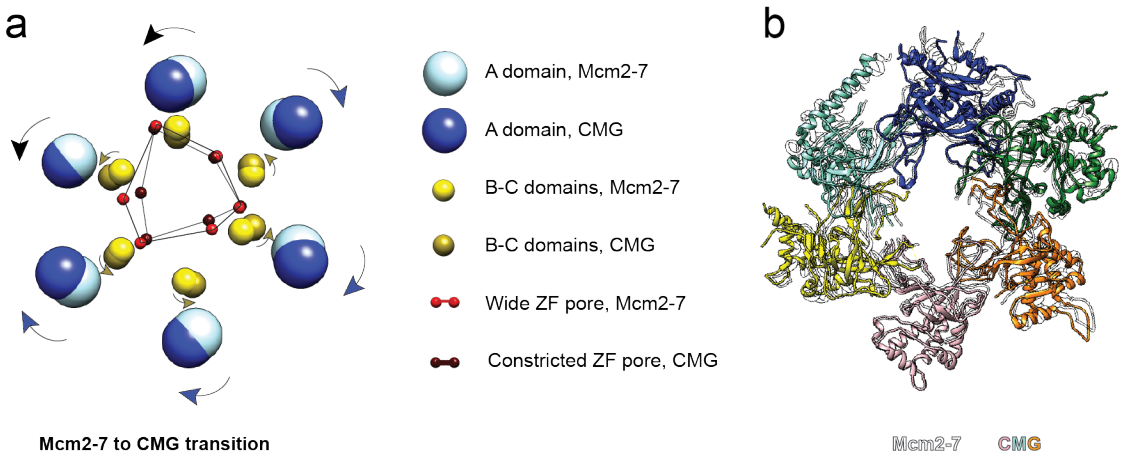


Figure 4.9: Transitions in the DNA interacting collar from MCM to CMG.
(a) Representation of N-terminal tier rearrangements during the MCM to CMG transition, where each domain is represented as a spherical centre of mass. The B-C domains (yellow) rotate as a rigid body in a counter-clockwise direction. Meanwhile, all ZF elements constrict to narrow the channel. **(b)** Cartoon depiction of conformational change at the N-terminal domain in the MCM (transparent) to CMG (coloured) transition.

The AAA+ tier of the compact ATPase structure sits on top of the NTD ring, with a slight left-handed offset of the NTD compared to the AAA+ tier. As mentioned previously, each AAA+ domain comprises two loops (PS1 and H2i) that project towards the centre of channel, which are highly conserved elements important for

DNA binding and unwinding (Jenkinson and Chong, 2006, Fletcher et al., 2003, Petojevic et al., 2015, McGeoch et al., 2005)(**Figure 4.8b**, lower panel [PS1 loops highlighted]). In the active CMG, the PS1 loops arrange into a spiral staircase akin to the E1 helicase (Enemark and Joshua-Tor, 2006). AAA+ ATPases comprise bipartite catalytic sites for ATP hydrolysis that are contributed by adjacent subunits (Erzberger and Berger, 2006). Although nucleotide density cannot be resolved at this resolution, the compact and planar arrangement of the MCM ATPase domains indicate that the ATPase sites are correctly aligned for productive nucleotide coordination. In agreement with this notion, a recently published yeast CMG structure bound to DNA and AMP-PNP at 4.9 Å closely matches the compact ATPase conformer in this study and revealed nucleotide binding at the consecutive Mcm2-6, Mcm2-5 and Mcm5-3 interfaces (Georgescu et al., 2017). Importantly, two of those nucleotide occupied ATPase sites (Mcm2-5 and Mcm5-3) are essential for helicase activity as determined by mutagenesis studies of *D. melanogaster* CMG (*see section 4.2.5*) (Ilves et al., 2010).

Extending away from the main AAA+ domain, two of the five MCM winged helix domains (WHD of Mcm5 and Mcm6) are well-defined in the CMG structure (**Figure 4.8b-c**) compared to the inactive MCM where these appendices are disordered (Li et al., 2015). These WHDs were also identified in a recent yeast apo CMG cryo-EM structure that was published at the same time as this study (Yuan et al., 2016), where an atomic model was built for most of the observed WHD density. Docking in the coordinates in the present CMG structure shows that Mcm5 WHD partially occludes and narrows the channel pore at the AAA+ tier down to ~21 Å (compared to ~35 Å at the NTD tier) (**Figure 4.8b**). In contrast, the Mcm6 WHD, best appreciated from a side view of the CMG complex, rises above the AAA+ tier and is exposed to solvent (**Figure 4.8c**). WHDs are generally involved in nucleic acid-protein interactions but interestingly, Mcm6 WHD does not interact with DNA (Wei et al., 2010). Instead, a number of biochemical studies have revealed that Mcm6 WHD binds to several different replication factors including Cdt1 (Wei et al., 2010), Mcm10 (Douglas and Diffley, 2016), Mrc1 (Komata et al., 2009) to ensure competent MCM loading, helicase activation and proficient translocation, respectively (Yeeles et al., 2017). Thus, the identified exposed configuration of Mcm6 WHD is compatible with those reported interactions (**Figure 4.8c**).

4.2.4 The role of Cdc45 and GINS in the CMG complex

To better understand the critical roles of the co-activators in the CMG, I analysed the docked crystal structures of Cdc45 and GINS in detail and mapped their interactions (**Figure 4.10** and **Figure 4.11**). The recently published crystal structure of human Cdc45 informs how this co-factor stabilises the CMG complex (Simon et al., 2016)(**Figure 4.10**). Cdc45 is evolutionarily linked to the bacterial RecJ nuclease, and its domain architecture reflects this, comprising two RecJ-like α/β domains (termed DHH and DHHA1) separated by a three-helix motif (3HM) and a Cdc45-specific CMG interacting domain (CID) (**Figure 4.10a**) (Sanchez-Pulido and Ponting, 2011, Krastanova et al., 2012, Szambowska et al., 2014). As introduced earlier (*section 1.4.3*), the RecJ active site is formed in between the two DHH domains. In both isolated Cdc45 and helicase incorporated Cdc45, the closed DHH and DHHA1 configuration blocks the DNA binding site (**Figure 4.10a**, right). However, a substitute DNA binding path cannot be ruled out as cross-linking studies show that Cdc45 can bind the leading strand template in the open CMG state (Simon et al., 2016, Petojevic et al., 2015).

Together, the Cdc45 DHH and CID domains fold into a disk-like architecture, which interact with GINS (via C-terminal Psf1 and Psf2) and the two adjacent Mcm2-5 A-domains, respectively (**Figure 4.10b-c**). These are the only Cdc45 interaction sites within the CMG, keeping a rigid brace at the NTD but leaving room for dynamic movement in the AAA+ tier as would be required for full helicase activity. The cryo-EM density shows a further helical protrusion from Cdc45 that projects away from the main MCM ring, a feature that was not previously resolved by negative stain CMG structures (**Figure 4.10a**, right) (Costa et al., 2014, Costa et al., 2011). A crosslinking and mass spectrometry study has shown that a region of the Polε catalytic subunit crosslinks to a lysine at the tip of this helix (Sun et al., 2015), which was subsequently verified by a low-resolution structure of CMG-Polε (Zhou et al., 2017). Other replisome components could also bind to both the top and bottom faces of Cdc45 as they are also not restrained by any GINS or MCM interactions (**Figure 4.8c**).

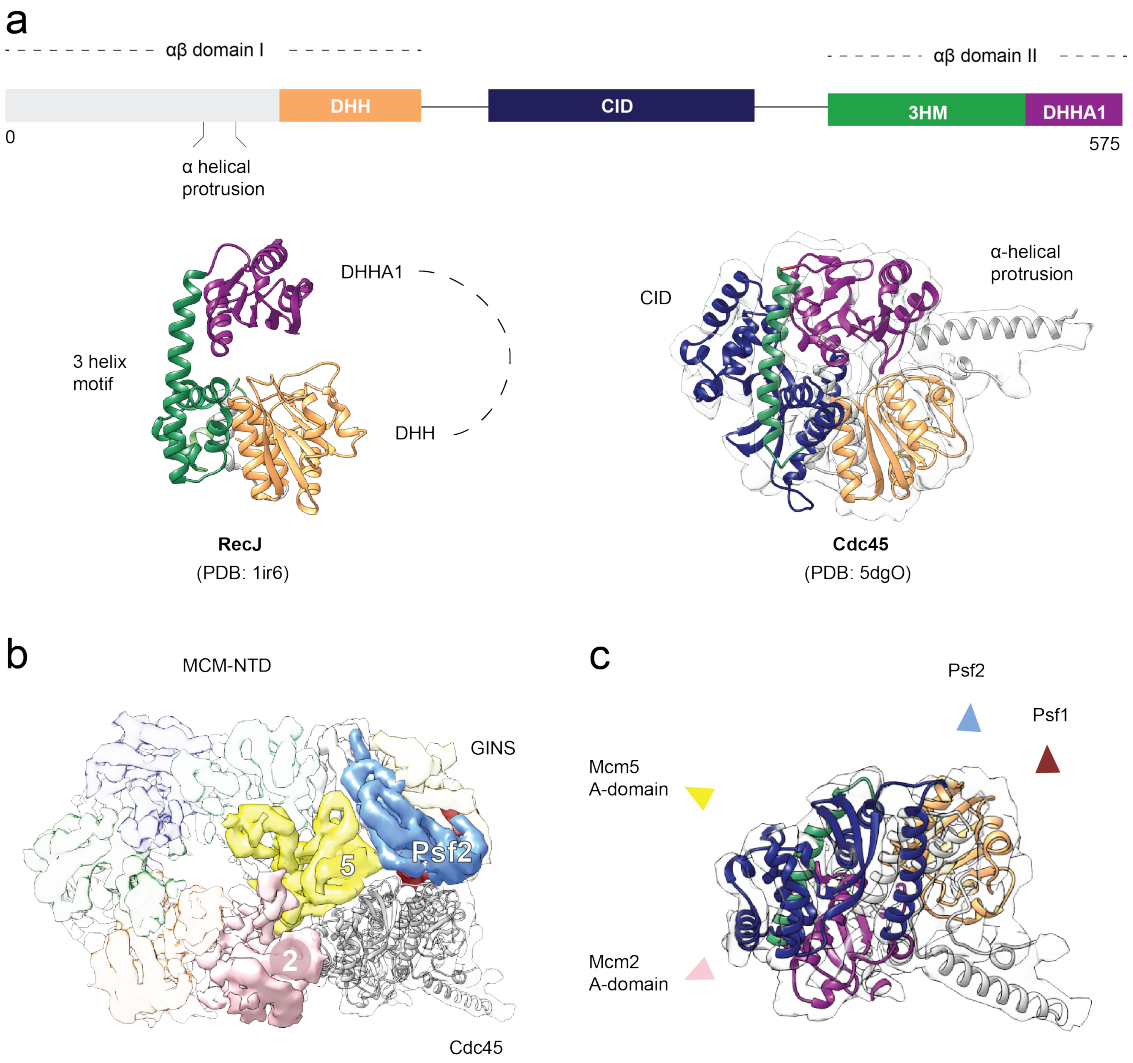


Figure 4.10: Domain architecture of Cdc45 and structural role in guarding the Mcm2-5 gate
(a) Predicted domain architecture of *Drosophila* Cdc45 based on the human Cdc45 structure. The crystal structure of bacterial RecJ and human Cdc45 (docked into the compact ATPase cryo-EM density) also shown. **(b)** Cdc45 interaction network in the CMG complex viewed from the MCM N-terminal tier. **(c)** Cdc45 subdomain interactions with MCM and GINS. CID, CMG interacting domain. 3HM, 3 helix motif. DHH and DHHA1 refer to the amino acid signature found in these motifs.

Like Cdc45, the GINS tetramer is essential for viability and plays a key scaffolding role within the CMG and wider replisome (Labib and Gambus, 2007). As introduced previously (*section 1.4.4.1*), each GINS subunit is formed of two domains: a 4-helix bundle (A-domain) and an αβ domain (B-domain) (**Figure 4.11a**). The crystal structure of human GINS (Chang et al., 2007) was docked into the density proximal to Cdc45, revealing the intra subunit interactions within the tetramer and a maintained pseudo two-fold symmetry of the complex within the helicase-incorporated GINS, where Sld5 and Psf1 form one asymmetric unit of the heterotetramer and Psf3-Psf2 the other (**Figure 4.11b**). An additional, unoccupied

density was observed after docking of all GINS subunits. This was confirmed to be the B domain of Psf1 after superposition of the full-length archaeal Psf1 crystal structure (Oyama et al., 2011). Although Psf1 B-domain is not required for GINS tetramerisation, it is essential for CMG assembly (Costa et al., 2014, Costa et al., 2011). The cryo-EM structure rationalizes this observation as it shows that the Psf1 β sheet core of the B-domain latches across and binds the DHH domain of Cdc45, confirming an interpretation from a low-resolution CMG study (Costa et al., 2014) (**Figure 4.10c**, right).

Two other key MCM interaction sites are mediated by GINS: Psf3 interacts with the Mcm3 A-domain whereas Psf2 α-helices 3 and 5 stabilise the Cdc45-bound Mcm5 A-domain at an adjacent site (**Figure 4.11c**). The only GINS subunit that does not interact with CMG is Sld5, which is situated towards the side of the complex away from the MCM ring (**Figure 4.11c**, middle). This is an ideal position for Sld5 to bind to its reported replication interactors (FACT, Ctf4, Mcm10, Csm3/Tof1 and Topoisomerase I) (Gambus et al., 2006, Gambus et al., 2009, Simon et al., 2014). Sld5 binding to the replisome Ctf4 trimer allows a number of factors to be linked to the replisome, including Polα (Gambus et al., 2009, Simon et al., 2014), the sister chromatid cohesion helicase Chl1 (Samora et al., 2016) and Dna2 (Villa et al., 2016). Therefore, GINS not only activates the CMG but also couples the helicase to DNA synthesis and sister chromatid cohesion.

In *section 3.5.2*, I showed that GINS recruitment during CMG assembly causes a concerted separation of the double hexamer but a molecular basis for this observation was lacking. By now being able to compare the inactive MCM to the active MCM from CMG a mechanism for this transition can be proposed. The region that Psf2 binds to on the Mcm5 A-domain of the CMG is initially bound by the yeast specific Mcm7 helical insert in the MCM double hexamer (as discussed in *section 3.2.2*). This suggests that GINS recruitment is sterically incompatible with the double hexamer (**Figure 4.12**). As shown in **Figure 4.9**, a large re-configuration of the Mcm5 A-domain occurs as the MCM transitions to the assembled CMG. This movement would cause release of Mcm7 to destabilise the double hexamer and promote Psf2 binding instead. However, the origin firing factor that drives the Mcm5 A-domain reconfiguration is still unknown. Previous studies have identified a mutation in Mcm5

that bypasses the requirement for DDK (Hardy et al., 1997). This mutation causes a conformational change in the Mcm5 A-domain, implying that DDK phosphorylation could lead to a similar structural change (Fletcher et al., 2003). However, our cryo-EM structure of the double hexamer shows no significant conformational changes in the core MCM upon DDK treatment. Whether subsequent recruitment of Sld3/7-Cdc45 or CDK mediated Sld2, Dpb11, Polε or GINS binding itself re-arranges Mcm5 remains to be investigated.

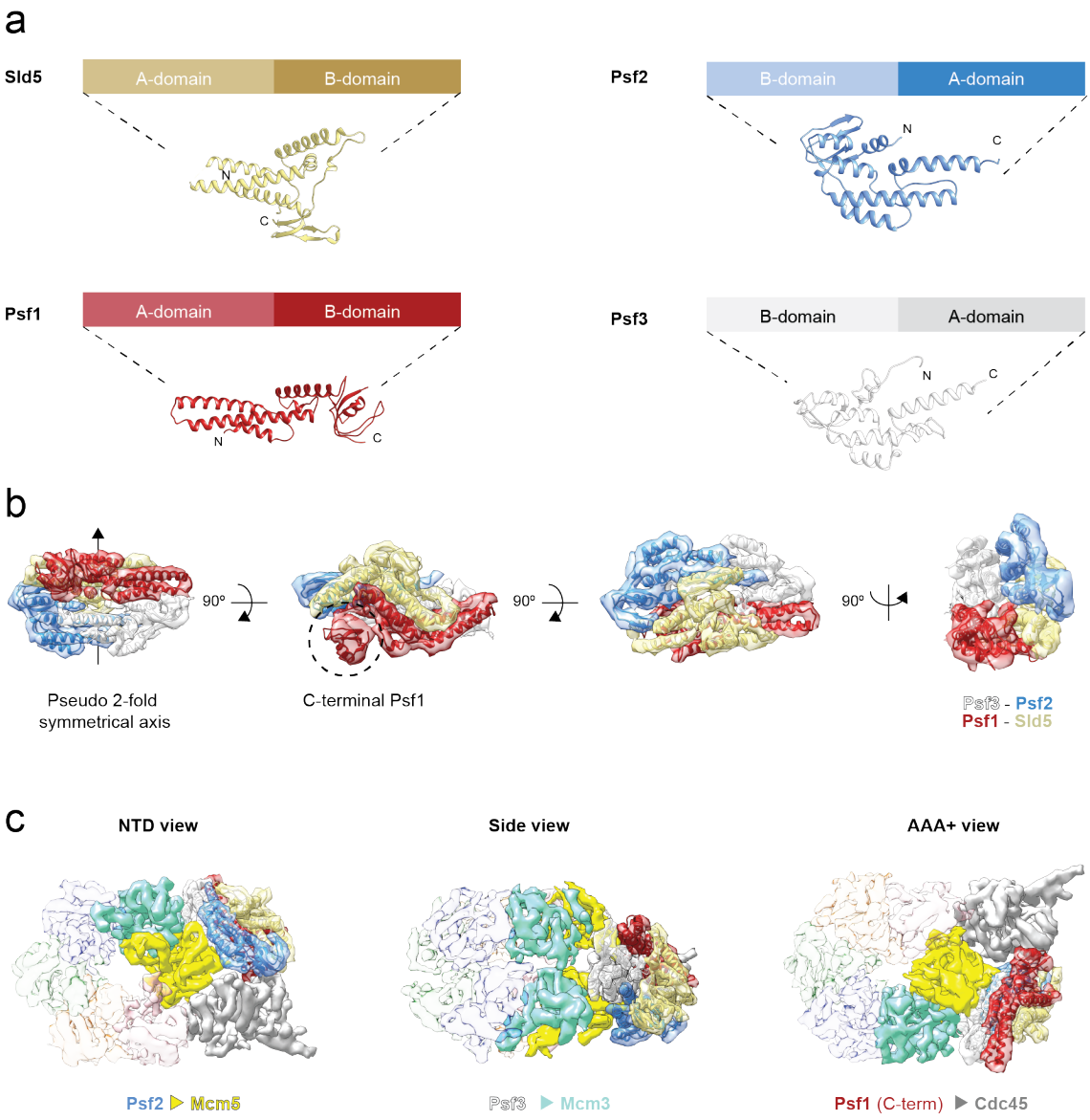


Figure 4.11: Structure and interactions of GINS with the MCM and Cdc45
(a) Domain architecture of GINS and the corresponding structures of each subunit from PDB 2Q9Q. B-domain of Psf1 modelled from PDB 3ANW. In Psf2 and Psf3 the domain architecture is inverted. **(b)** The full length *D. melanogaster* GINS structure from the CMG compact ATPase conformer. GINS shows a pseudo two fold symmetric architecture. C-terminal of Psf1 is stabilised in the cryo-EM structure (docked crystal structure: PDB 3ANW). **(c)** GINS interaction network in the CMG.

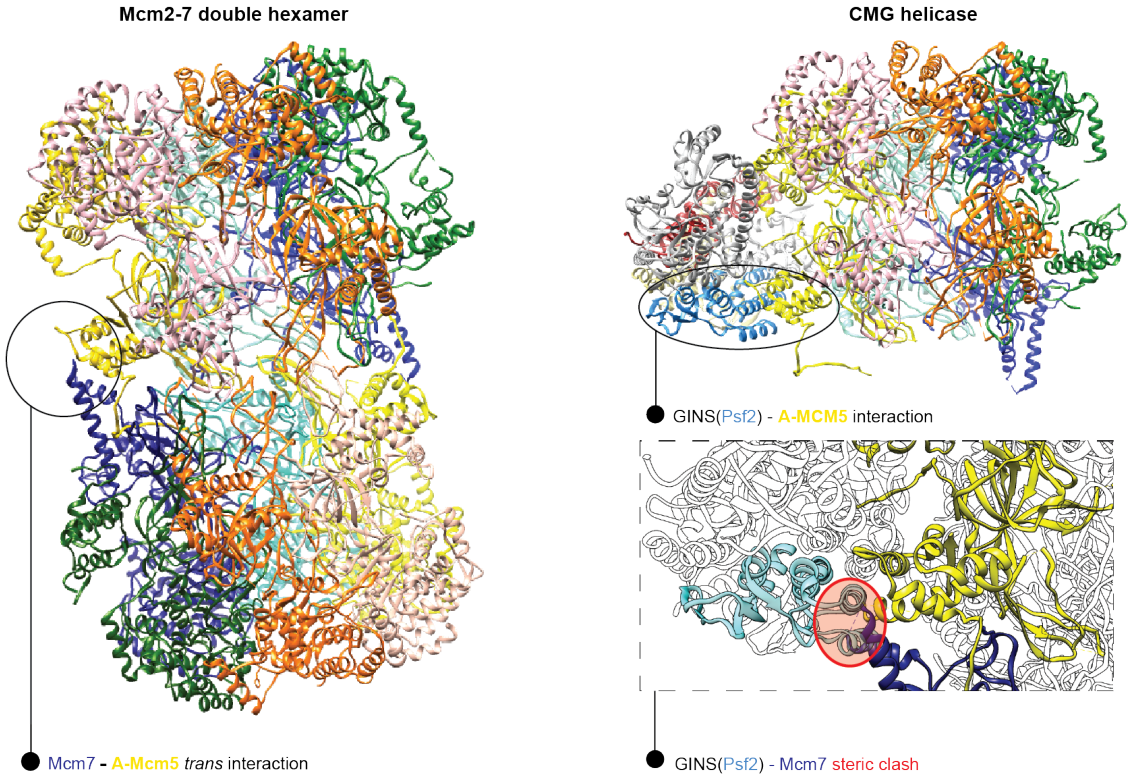


Figure 4.12: GINS binding may cause destabilisation of the MCM double hexamer
One of the key features in the double hexamer is the *in trans* binding of the yeast specific helix insertion of Mcm7 to the A-domain of Mcm5 in the opposing hexamer. In the CMG, Psf2 binds to the very same site in Mcm5, indicating that a rearrangement of the Mcm5-7 *trans* interaction has to occur for GINS to bind.

In summary, Cdc45 plugs the Mcm2-5 gate whereas GINS braces the Mcm5-3 interface, together closing the MCM NTD ring. This closed arrangement could prevent DNA from slipping out of the central channel (Petojevic et al., 2015) and explains why the replisome is obstructed by a bulky streptavidin roadblock on the translocating strand (Fu et al., 2011). GINS recruitment during helicase assembly causes dissociation of the two MCM rings likely because it competes for the same Mcm5 A-domain binding site that is initially bound by a *trans* acting Mcm7 α-helix in the double hexamer.

4.2.5 The AAA+ domain adopts relaxed and compact ATPase conformations

AAA+ proteins use the energy derived from ATP hydrolysis to perform mechanical work (Lyubimov et al., 2011). As mentioned in the previous section, two structures that showed a reconfigured AAA+ tier were identified: a compact ATPase (introduced earlier) and a relaxed ATPase at 9.8 Å resolution (**Figure 4.7**, **Figure 4.13**). The lower resolution of the relaxed ATPase structure is probably due to the low particle

numbers that contributed to this class (13,692 vs 38,792 in the compact ATPase structure) and the flexibility exhibited at the AAA+ ring (local resolution is ~15 Å at this tier) (**Figure 4.7e**). As evidenced by the earlier analysis of the CMG structure, Cdc45 plugs the gate at the NTD Mcm2-7 tier in both conformers (lower panel **Figure 4.13a-b**) but leaves room for movement in the AAA+ domain (upper panel **Figure 4.13a-b**). This gives rise to two major differences between the conformers. First, the AAA+ Mcm2-5 interface, a previously identified gate in the ring (Costa et al., 2011) is either notched at the ATPase face (relaxed ATPase) or closed (compact ATPase). Second, both the compact and relaxed ATPase structures contain extra density inside the central MCM channel in either the AAA+ or NTD tier, respectively (**Figure 4.13**, red densities). This density is assigned to DNA and is discussed further in *section 4.2.6*.

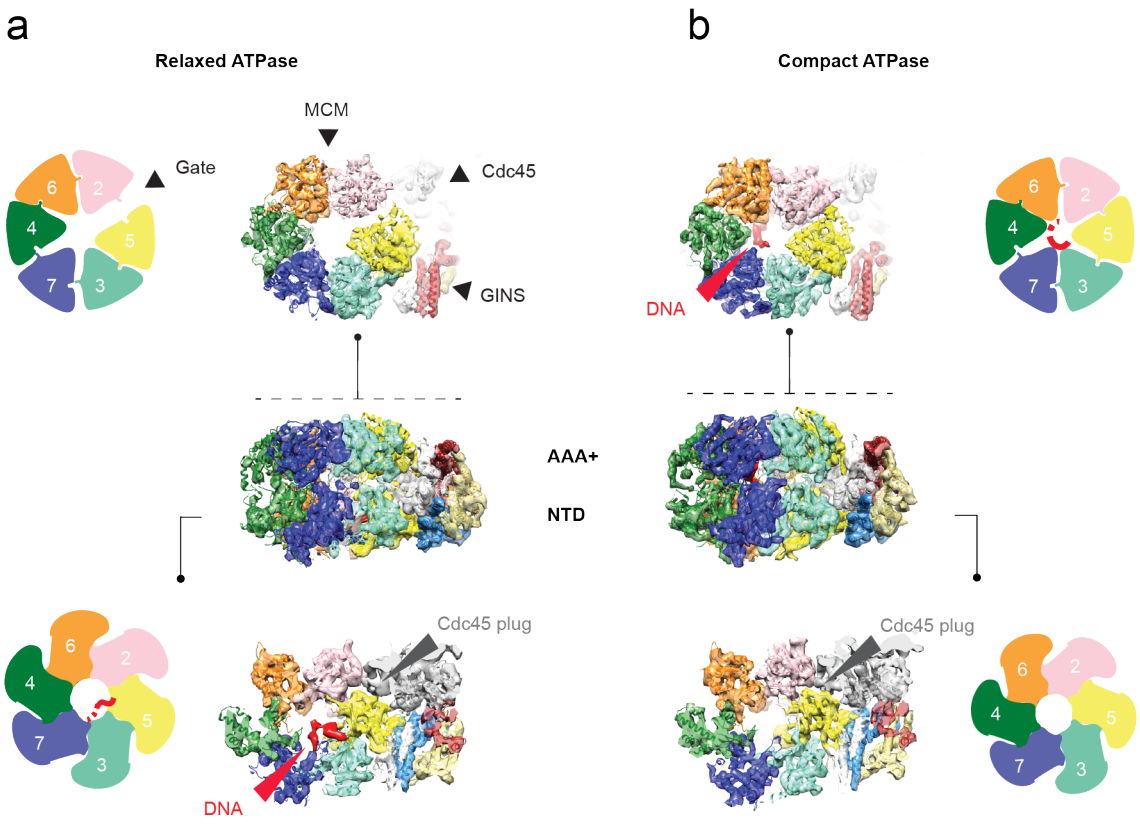


Figure 4.13: CMG bound to DNA and ATPγS exists in two main configurations
(a) Relaxed ATPase structure at 9.8 Å resolution reveals how the Mcm2-5 gate is notched at the AAA+ tier. A cut-through view of the N-terminal tier (bottom panel) shows additional density assigned to DNA. **(b)** Compact ATPase structure at 7.4 Å resolution shows a tight Mcm2-5 interface at the AAA+ tier. Additional density resides inside the channel is assigned to DNA. Cdc45 plugs the Mcm2-5 gate at the N-terminal tier to keep the complex topologically closed in both complexes.

Whereas the compact ATPase conformer comprises a planar MCM ring (**Figure 4.13b**), similar to the arrangement of MCM derived from the double hexamer structure (Li et al., 2015), in the relaxed ATPase form, the MCM subunits arrange into a right-handed, notched spiral that would match the chirality of DNA (**Figure 4.13a**, top). This configuration was previously observed at low resolution of a nucleotide free CMG structure (Costa et al., 2011). Comparison of segmented subunits from the two conformers in this study reveals significant movement of the AAA+ and NTD domains with respect to each other (**Figure 4.14a**). Such plasticity of the CMG helicase could be important during the unperturbed helicase translocation cycle or during fork stalling.

To specifically model an ATPase conformational transition from the relaxed to compact ATPase, the two CMG conformers were aligned at the NTD and compared at the AAA+ tier (**Figure 4.14b**). Atomic coordinates of ATPase domains bound to ADP from the yeast Mcm2-7 structure were docked in to the two conformers (Li et al., 2015). Although nucleotide density cannot be resolved at the resolution ranges of this study, ADP was used in this case as a marker to identify ATPase centres. Morphing between the two structures reveals that movement in the AAA+ ring is mainly dependent on the opening and closing of the Mcm2-5 gate (root mean square deviation, RMSD 8.2 Å), which is in turn mediated by the tightening of the adjacent Mcm6-2 (RMSD 8.3 Å) and Mcm5-3 interfaces (RMSD 5.3 Å). On the other hand, the movement of the Mcm4-7 and Mcm7-3 interfaces are more restrained (RMSD 3.1 and 3.3 Å, respectively). This points to a static half in the AAA+ tier of Mcm2-7 and a more dynamic interface at the Mcm2-5 site and its nearest subunits. In line with these observations, mutating the ATPase sites at the dynamic side of the ring abolishes DNA unwinding activity whereas mutations of the static Mcm3-7, Mcm7-4, Mcm4-6 hinge have minimal effects on unwinding (Ilves et al., 2010) (**Figure 4.14b**, lower left panel). Thus, the active MCM ring is both structurally and functionally asymmetric.

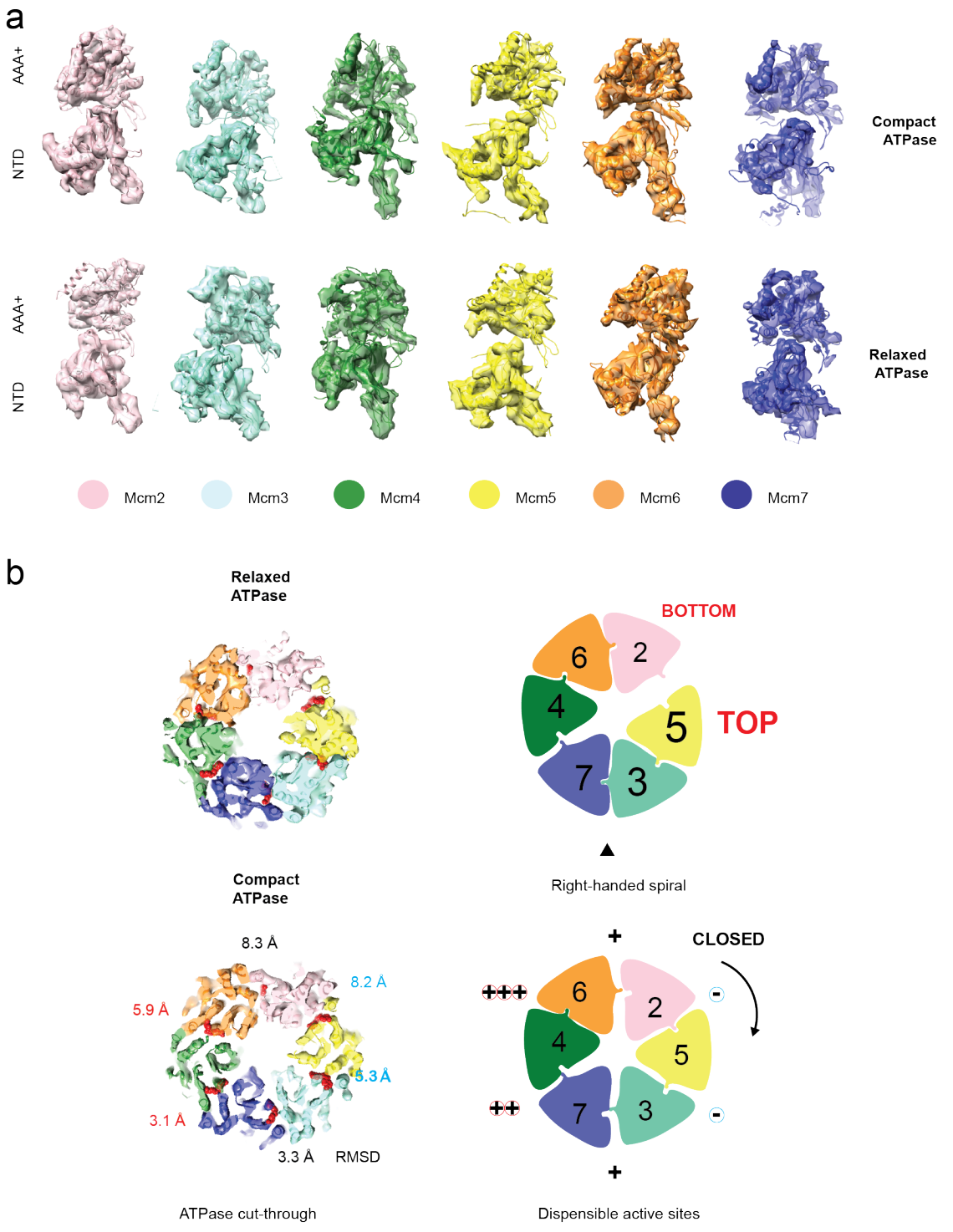


Figure 4.14: Remodelling of the CMG ATPase motor.
(a) Segmented densities of the individual MCM subunits with docked coordinates (PDB: 3JA8) from the two structures. **(b)** ATPase sites in the relaxed and compact structures, signposted by docked in ADP from PDB entry 3JA8 (although at this resolution nucleotide occupancy cannot be determined). Cartoon depictions of the relaxed ATPase structure highlights the open, right handed spiral configuration of this structure whereas the compact ATPase forms a more planar configuration. Comparison of the two AAA+ tiers and calculations of root mean square deviation (left panel) reveals that one half of the ring is near static (Mcm3-7, Mcm7-4, Mcm4-6). This corresponds to the active sites that have been shown to be dispensable for viability, referring to a Walker A mutagenesis experiment to test the contribution of different ATPase sites in DNA unwinding (Ilves et al., 2010). Dispensable sites are marked by "+" whereas indispensable sites are marked by "-".

As mentioned before, the compact ATPase conformer, which is contributed by the vast majority of the particles in the dataset (77%), bears high resemblance to a recently published cryo-EM structure of a CMG bound to single-stranded DNA and AMP-PNP at three MCM active sites (Georgescu et al., 2017). This supports the notion that the compact ATPase in the presence of ATPγS and DNA in this study represents a translocation intermediate. The relaxed ATPase conformer, contributed by the minority of the particles in the dataset (23%) could represent a second intermediate in the translocation cycle. Alternatively, this conformation could be a paused state where the Mcm2-5 gates opens upon during fork stalling (Petojevic et al., 2015). An open Mcm2-5 gate has previously been observed as the predominant CMG conformation in both nucleotide-free (Costa et al., 2011) and ATP conditions (Abid Ali et al., 2016). In CMG-ATP, it is thought that ATP hydrolysis at the essential Mcm2-5 site drives the MCM to the relaxed ATPase configuration. As ATPγS was used in the current study, such hydrolysis is unlikely to have occurred. A higher resolution structure of the CMG-DNA in the presence of ATP is required to identify the functional relevance of the notched right-handed spiral CMG conformation.

4.2.6 The CMG helicase translocates on single-stranded DNA

Replicative helicases harness the energy derived from ATP hydrolysis to drive replication fork progression. To start to understand the molecular basis for DNA binding by the CMG, I docked in the helicase atomic coordinates and inspected the unoccupied density present in the channel of both conformers, either trapped in the NTD or AAA+ tier (**Figure 4.15**). These additional densities are assigned to DNA as they are absent in a DNA-free CMG-ATP cryo-EM structure (work performed by Dr. Julian Gannon and Dr. Ludovic Renault) (Abid Ali et al., 2016). However, the duplex DNA fork junction could not be visualized in either the compact or relaxed ATPase structures, probably because the ATPγS condition prevents CMG translocation up to the fork nexus (Georgescu et al., 2017). The lagging strand template is also invisible, probably due to flexibility of this filament (further discussed in *section 4.2.7*).

The nucleic acid density in the compact ATPase MCM channel spirals with a right-handed configuration and spans solely the AAA+ tier (**Figure 4.15a**). This conformation of DNA is reminiscent of B-form DNA, and indeed density encircled by

the compact ATPase can accommodate one single strand extracted from B-form double helix (**Figure4.15b**). Therefore, the AAA+ motor of the CMG translocates on single and not duplex DNA. Docking of the DNA (**Figure 4.15b**) indicates that the helicase would translocate with its N-terminal tier first, a notion which will be explored directly in *section 4.2.10*.Six nucleotides are modelled inside the DNA density and the phosphate backbone contacts positively charged residues on the PS1 hairpins of Mcm4, 6 and 7 (Mcm7 K471, Mcm4 K602 and Mcm6 R478) (**Figure 4.15b**). These PS1 residues are conserved from archaea to human, and their mutation in both the archaeal and *D. melanogaster* MCM affect DNA binding and abrogate unwinding (McGeoch et al., 2005, Petojevic et al., 2015). Moreover, the DNA density is also in close proximity to the PS1 at Mcm3 K430, a residue required for viability in yeast (Lam et al., 2013).

A recently published yeast CMG structure is nearly identical to *D. melanogaster* CMG at the protein level but the single-stranded DNA entrapped inside the channel is found at opposing sides of the ring across the two structures (**Figure 4.15e**). Three possibilities could explain this difference: two different DNA translocation states may have been captured (Mcm4,6,7 binds DNA in *D. melanogaster* CMG and Mcm3,4, 2 bind DNA in yeast CMG), the two species have inherent differences in how CMG binds DNA or the DNA is bound upside down in one of the two structures.

In the relaxed ATPase conformer, there does not seem to be any DNA inside the AAA+ domain (**Figure 4.15c**). One rationale for this could be that the open Mcm2-5 gate reduces *in vitro* fork DNA binding or that a larger degree of substrate movement prevents 3D averaging of the nucleic acid in this region. Instead, continuous density spans the topologically closed NTD around highly conserved C-domain (OB-fold) hairpins in the pore (**Figure 4.15d**). The importance of this DNA binding site is shown by studies in archaea where mutation of the arginine residues that project from the OB-fold abolish DNA binding and unwinding (Barry et al., 2009, Froelich et al., 2014, Fletcher et al., 2003). However, it should be noted that the observed DNA conformation in the relaxed conformer is distinct from the configuration of the nucleic acid co-crystalised in the N-terminal domain of archaeal MCM, where single-stranded DNA lines the inner perimeter of the ring perpendicular to the central channel (MSSB motif) (Froelich et al., 2014).

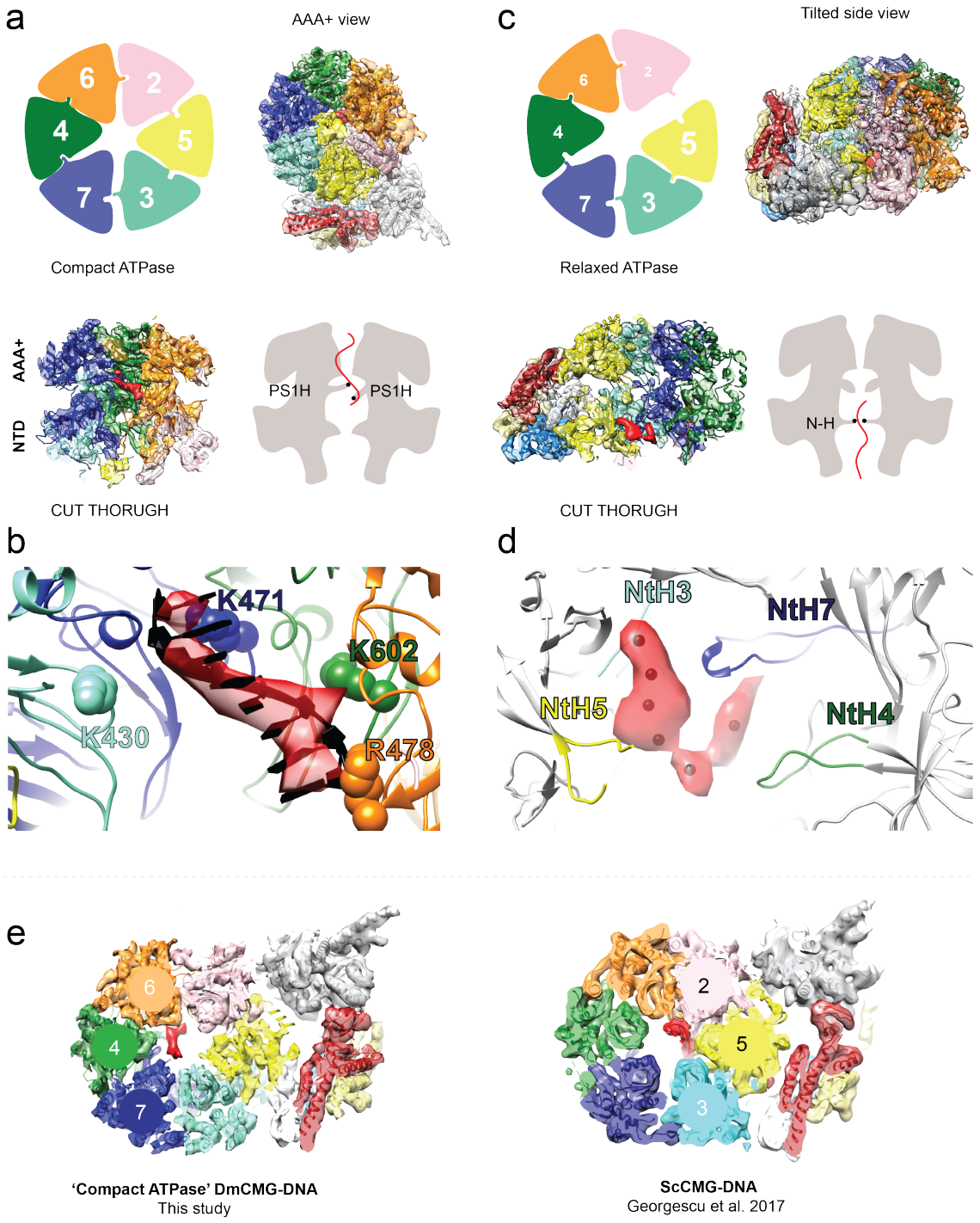


Figure 4.15: The CMG helicase binds to single-stranded DNA
(a) A cut through the AAA+ top, tilted view of the compact ATPase conformer reveals an elongated density inside the channel, assigned to DNA. **(b)** DNA in the compact ATPase spirals around Mcm7, Mcm4 and Mcm6 and is in close proximity to Mcm3 contacting positively charged residues on the PS1 hairpins of the AAA+ tier. **(c)** A cut-through the side, tilted view of the relaxed ATPase conformer reveals DNA density inside the NTD channel. **(d)** DNA contacts the N-terminal hairpins of Mcm5, Mcm3, Mcm7 and Mcm4. **(e)** DmCMG (this study) and yeast CMG (EMD-8518) structures are near identical at the protein level but differ in their DNA binding state. A cut-through the AAA+ top view is shown. PS1, pre-sensor 1 hairpin. NtH, N-terminal hairpin.

4.2.7 Locating the lagging strand template

Visualisation of only single-stranded DNA inside the MCM channel suggests a strand exclusion mode for CMG unwinding. In this translocation mode, the CMG sterically excludes the lagging strand template, which is not encircled by the helicase. As this filament is invisible in the cryo-EM structures, I used a fork that contains a biotinylated 5' end and added streptavidin to locate and visualise the lagging strand template on a CMG-DNA fork assembly. Streptavidin appears as an electron-dense feature under negative stain EM, allowing biotinylated fork substrate ends to be easily identified. To investigate how the lagging strand template interacts with the CMG, I incubated a purified CMG-ATPγS sample with a model replication DNA fork and subsequently added streptavidin (**Figure 4.16a**). EM analysis of CMG particles indicates that the tagged lagging strand template occupies distinct positions around the external surface of the CMG (**Figure 4.16b-c**). This experiment confirms that the lagging strand template is excluded from the main MCM channel, in agreement with CMG nucleoprotein/crosslinking studies (Petojevic et al., 2015). Furthermore, the lagging strand template is flexible relative to a fixed CMG, explaining why this strand is invisible in the CMG-DNA structure. Whether this flexibility is due to the CMG not engaging the fork nexus or an inherent plasticity of the excluded lagging strand template remains to be established.

Taken together with the cryo-EM structures, this study has shown two different leading strand template binding regions in the CMG and localised the lagging strand template outside the helicase, thereby providing evidence in support of the strand exclusion mechanism for CMG fork progression.

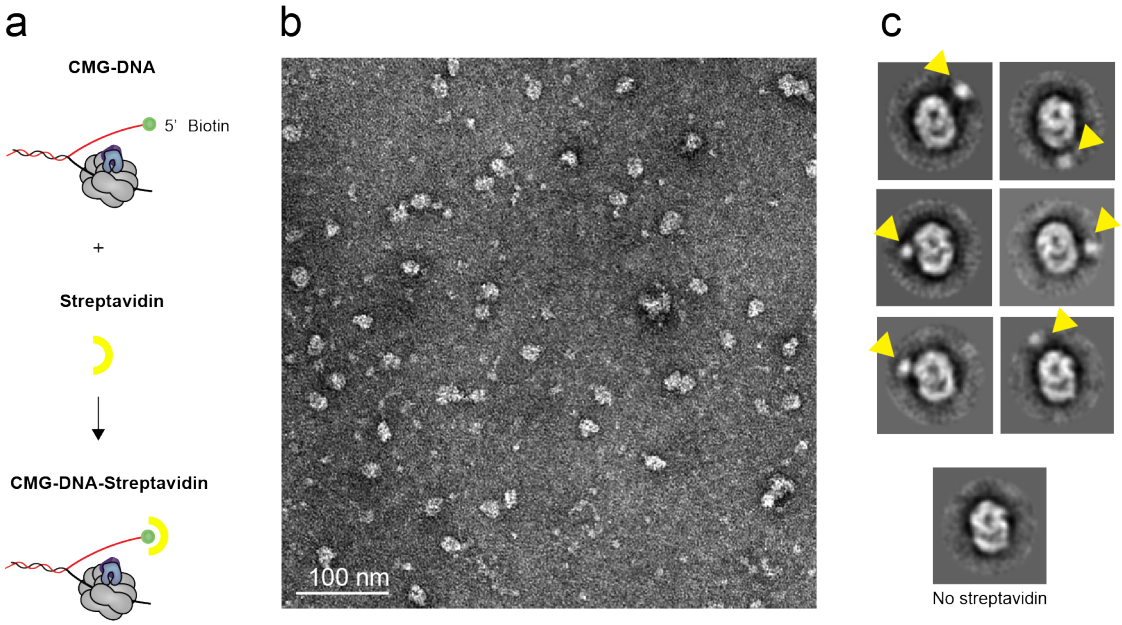


Figure 4.16: Locating the lagging strand template in a CMG-DNA complex
(a) Overview of the CMG-DNA reconstitution. This model replication fork is made from two oligonucleotides that are annealed such that the double stranded region spans 50-bp length and the two forked single-stranded overhangs comprise 40-nt with a biotin moiety engineered on the lagging strand template end. Annealing was performed at 90ºC for 3 min with subsequent cooling to room temperature. Purified CMG was first dialysed into Buffer A (methods secion 2.1.3) containing 0.1 mM ATPγS. 2-fold molar excess annealed fork substrate was added and after 30 min incubation 2-fold molar excess of streptavidin was added. **(b)** Representative micrograph of CMG-DNA-streptavidin complexes. 100 micrographs were collected using a Tecnai LaB6 G2 Spirit transmission electron microscope (FEI) operating at 120 keV at 42000x magnification. **(c)** Reference-free 2D classes of CMG-ATPγS-DNA was obtained in IMAGIC. Each class contains at least 50 particles. Only CMG top views shown and yellow arrows highlight streptavidin density.

4.2.8 Single-stranded DNA in the CMG is stabilised in an extended conformation

Comparison of the compact ATPase structure to other hexameric helicases can provide mechanistic insights into how CMG translocation occurs. Therefore, available high-resolution co-crystal structures of Rho (Skordalakes and Berger, 2003), DnaB (Itsathitphaisarn et al., 2012) and E1 (Enemark and Joshua-Tor, 2006) bound to single-stranded nucleic acids were aligned and compared to CMG-DNA (**Figure 4.17**). This revealed a common structural feature across the helicases of a right-handed spiralling of the ATPase pore loops, which matches the right-handed configuration of each bound nucleic acid. In this context, Rho and E1 bind and compress the RNA and DNA, respectively, contacting 1 nucleotide per subunit (**Figure 4.17**, top two structures) whereas DnaB contacts 2 nucleotides per subunit (**Figure 4.17**, third structure), a configuration similar to A-form DNA. In contrast, the

CMG binds to single-stranded DNA extracted from the more extended B-form DNA (**Figure 4.17**, bottom structure). This observation raised the question of whether CMG stretches DNA upon binding. To test this, we performed single-molecule FRET experiments of CMG-DNA with either ATPγS or ATP in the sample (in collaboration with Dr. Hailey Gahlon in Dr. David Rueda's laboratory, Imperial College London) (Ha et al., 1996).

Single molecule FRET (smFRET) is a widely used technique to probe DNA conformational changes induced by ATPase enzymes in real time, as has been done to assess how a Rad51 paralog ATPase remodels single-stranded DNA-Rad51 filaments (Taylor et al., 2015). In our experiments, a decrease in FRET efficiency would indicate stretching of DNA due to the increase in fluorophore inter-distance. Thus, smFRET acts as a 'molecular ruler' to probe the DNA-CMG association.

To prepare the sample for smFRET measurement, I first added biotinylated DNA containing a Cy3 and Cy5 donor-acceptor pair (separated by seven nucleotides) on a neutravidin coated reaction chamber. Next, I mixed CMG with the surface immobilised DNA in the presence of either ATPγS or ATP. DNA molecules were thereafter imaged on a home-built total internal reflection fluorescence microscope to monitor changes in FRET efficiency (by Dr. Hailey Gahlon). To calculate FRET efficiency for each molecule, the acceptor intensity was divided by the total donor and acceptor intensity. In all these measurements the acceptor intensity was corrected for 11% donor emission in the acceptor channel (**Figure 4.18d**). **Figure 4.18a-c** shows that CMG binding to DNA significantly reduces the FRET value from 0.94 (naked DNA, **Figure 4.18a**) to 0.44 in the ATPγS condition and to a similar 0.48 value in the ATP state (**Figure 4.18b** and **c**). This change in FRET efficiency is due to both a reduction in DNA molecular flexibility as well as stretching induced by CMG binding to the DNA. As revealed by comparing the histograms across the two CMG samples, binding events were less frequent in the ATP condition (**Figure 4.18b** and **c**), which is in line with previous work showing that ATPγS is required for CMG-DNA complex stabilisation *in vitro* (Ilves et al., 2010, Petojevic et al., 2015). Together with the structural data, these results support the notion that the CMG stabilises DNA in a stretched configuration. A comparison with hexameric helicases bound to nucleic acids (**Figure 4.17**) indicates that the CMG might contact two or more nucleotides of

DNA per ATP hydrolysed. Indeed, a recent yeast CMG-DNA structure determined at 4.9 Å resolution led to the estimation of two nucleotides per subunit step-size (Georgescu et al., 2017).

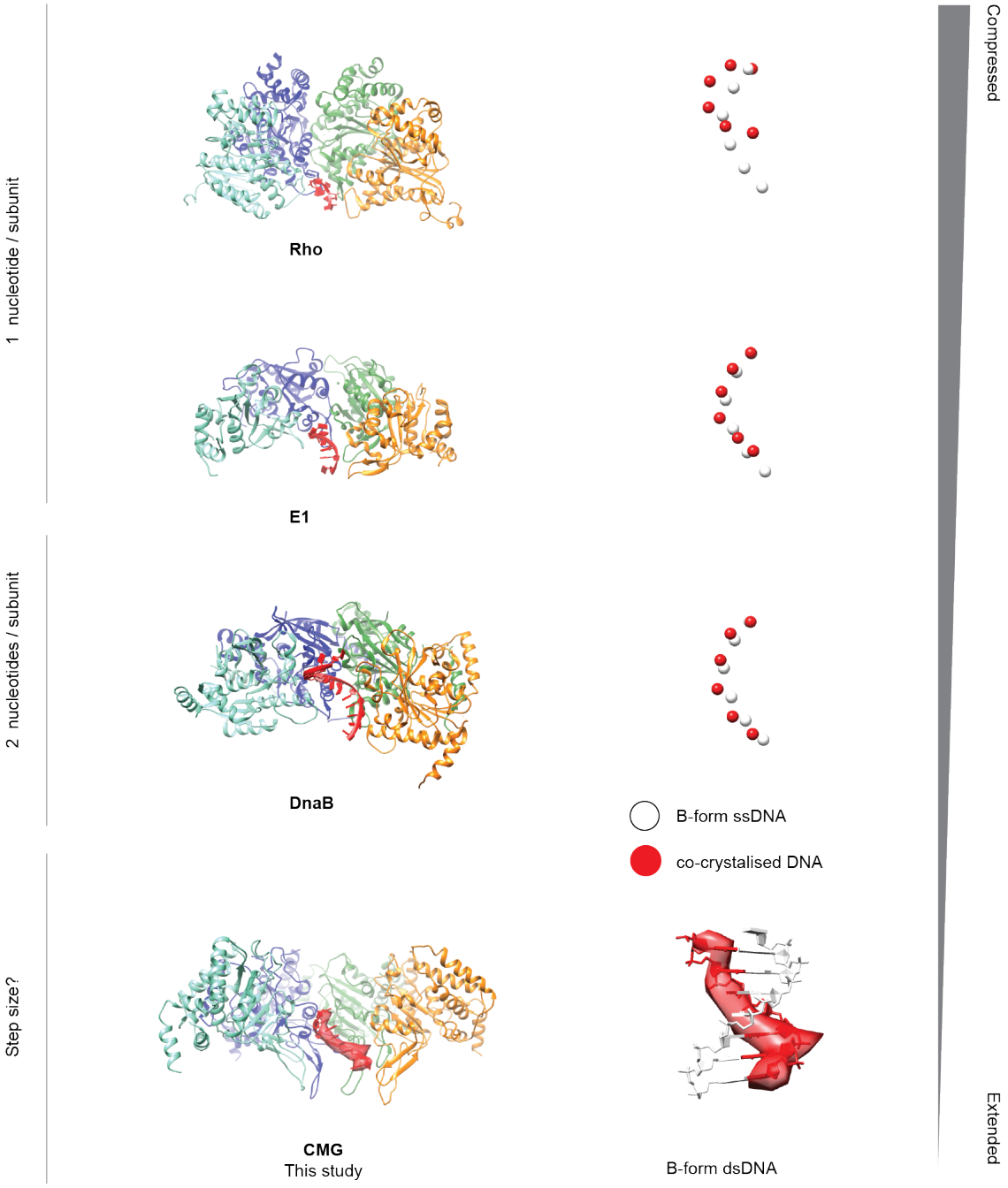


Figure 4.17: Comparison of CMG-DNA-ATPγS with helicases co-crystallised with nucleic acid
Helicases co-crystallised with nucleic acids are shown in red are compared to one single strand extracted from B-form duplex DNA. The spheres represent nucleotide centre of mass. Rho termination factor (PDB: 3ICE) and Papilloma E1 replicative helicase (PDB: 2GXA) bind and compress RNA and single-stranded DNA, respectively, with 1 nucleotide contacted per subunit. The DnaB replicative helicase binds slightly compressed DNA, approacting an A-form DNA configuration (PDB: 4ESV) where one subunit contacts two nucleotides. The CMG (this study) binds DNA in a more extended configuration. The DNA density matches the chirality and dimensions of one single strand in the B-form double helix.

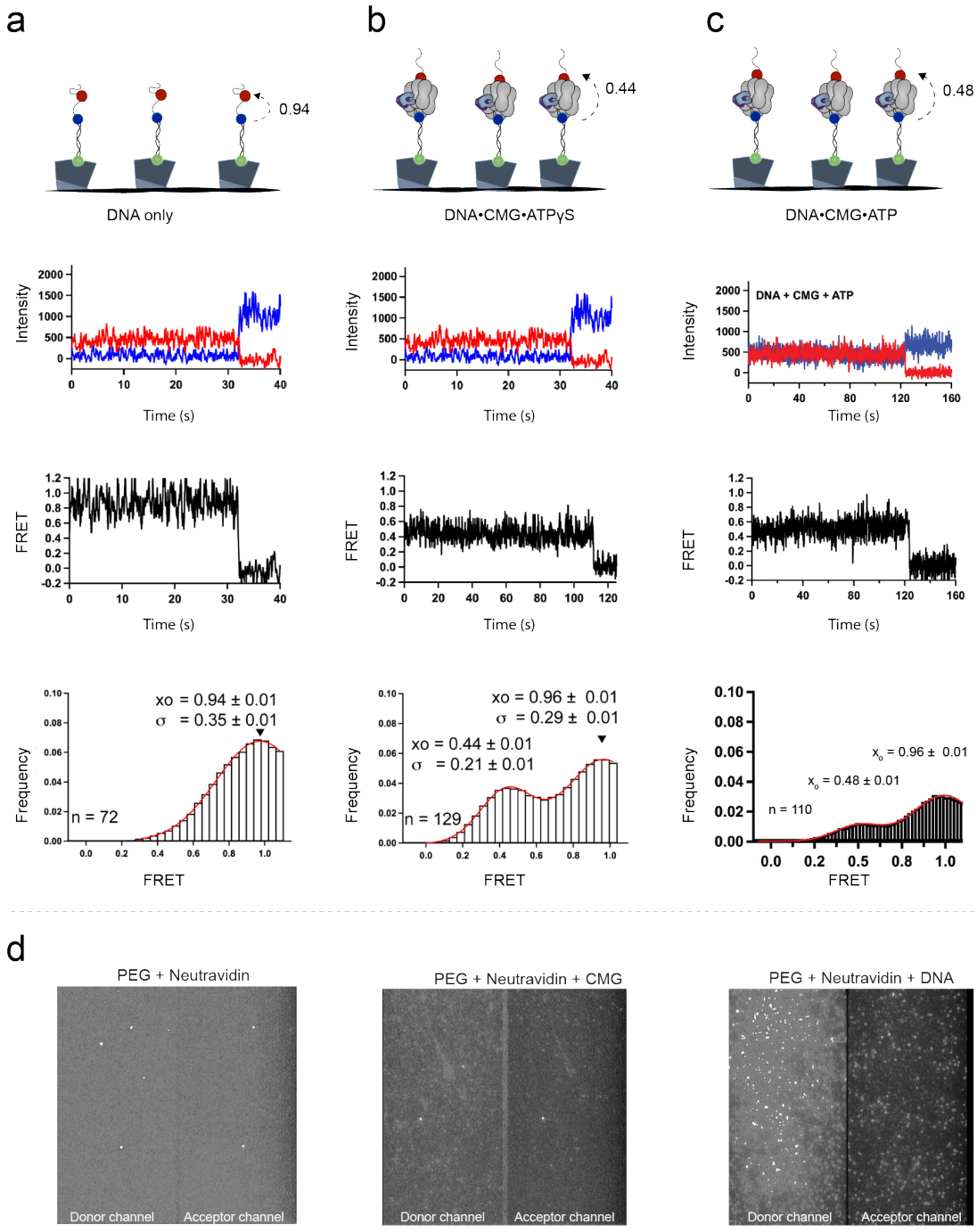


Figure 4.18: DNA deformation by the CMG helicase

Single molecule FRET analysis of **(a)** isolated DNA **(b)** CMG-ATPgS and **(c)** CMG-ATP. Top to bottom: representation of the FRET experiment. The blue sphere is Cy3 (donor) and the red sphere is Cy5 (acceptor), separated by 7-nt. DNA is immobilized through a biotin-neutravidin interaction on the PEG coated surface of the Quartz slide; donor (blue) and acceptor (red) intensity trajectories are anti-correlated until single-step photobleaching of the acceptor. FRET trajectories (black) between Cy3 and Cy5 exhibit a sharp drop to zero FRET when the acceptor photobleaches; histogram of FRET values collected from all molecules with a Gaussian fit are shown in red, where x0 is the mean FRET and σ is the distribution width. **(d)** FRET signals are only detected upon DNA immobilisation on quartz slide as shown by imaging of quartz slides with PEG+Neutravidin (left), PEG+Neutravidin+CMG (middle) and PEG+Neutravidin+DNA (right). See methods *section 2.4.4* for more details.

4.2.9 Comparison of MCM to CMG suggests a model origin DNA melting

In recent years, cryo-EM structures of the inactive MCM bound to DNA (Chapter 3) (Abid Ali et al., 2017, Noguchi et al., 2017) and the CMG in various nucleotide states (Georgescu et al., 2017, Yuan et al., 2016, Abid Ali et al., 2016) have provided an overview of origin licensing and initiation of DNA replication fork progression. So far in this chapter, I have shown that the CMG binds to single-stranded DNA and in Chapter 3, I showed that the MCM double hexamer binds to intact duplex DNA. A central unanswered question from those studies is how the MCM transitions from binding duplex DNA to single-stranded DNA upon replication fork establishment. What is the molecular basis for origin melting?

In bacteria, the AAA+ DnaA protein stretches the DNA in an ATP dependent manner before loading of the replicative helicase DnaB (Duderstadt et al., 2011). In eukaryotes, melting and unwinding are temporally separated but are thought to be carried out by the same MCM complex. Evidence in support of this notion has recently emerged, showing that the MCM within the assembled CMG untwists half a turn of DNA ahead of full activation by Mcm10 (Douglas et al., 2018). Thus, by observing the protein reconfigurations that promote DNA melting a mechanism for the initial strand separation can be proposed. To accurately model this transition, I compared the DNA-bound double hexamer MCM described in Chapter 3 to a CMG bound to forked DNA (Georgescu et al., 2017) from the same budding yeast species (**Figure 4.19a-b**).

One MCM ring from the double hexamer structure was first extracted and aligned at the N-terminal tiers with the CMG-DNA structure. The CMG structure contains well-resolved duplex DNA outside the N-terminal tier and single-stranded DNA inside the channel whereas the lagging strand template is invisible (Georgescu et al., 2017). Morphing between the MCM-DNA and CMG-DNA states reveal substantial conformational changes in the AAA+ tier that bring it into alignment with the NTD. As this occurs, the MCM channel narrows (as described in **Figure 4.9**) and the duplex DNA at the NTD rotates clockwise from binding Mcm2-5 in the MCM to binding the Mcm4-7 side in the CMG, in turn promoting DNA underwinding.

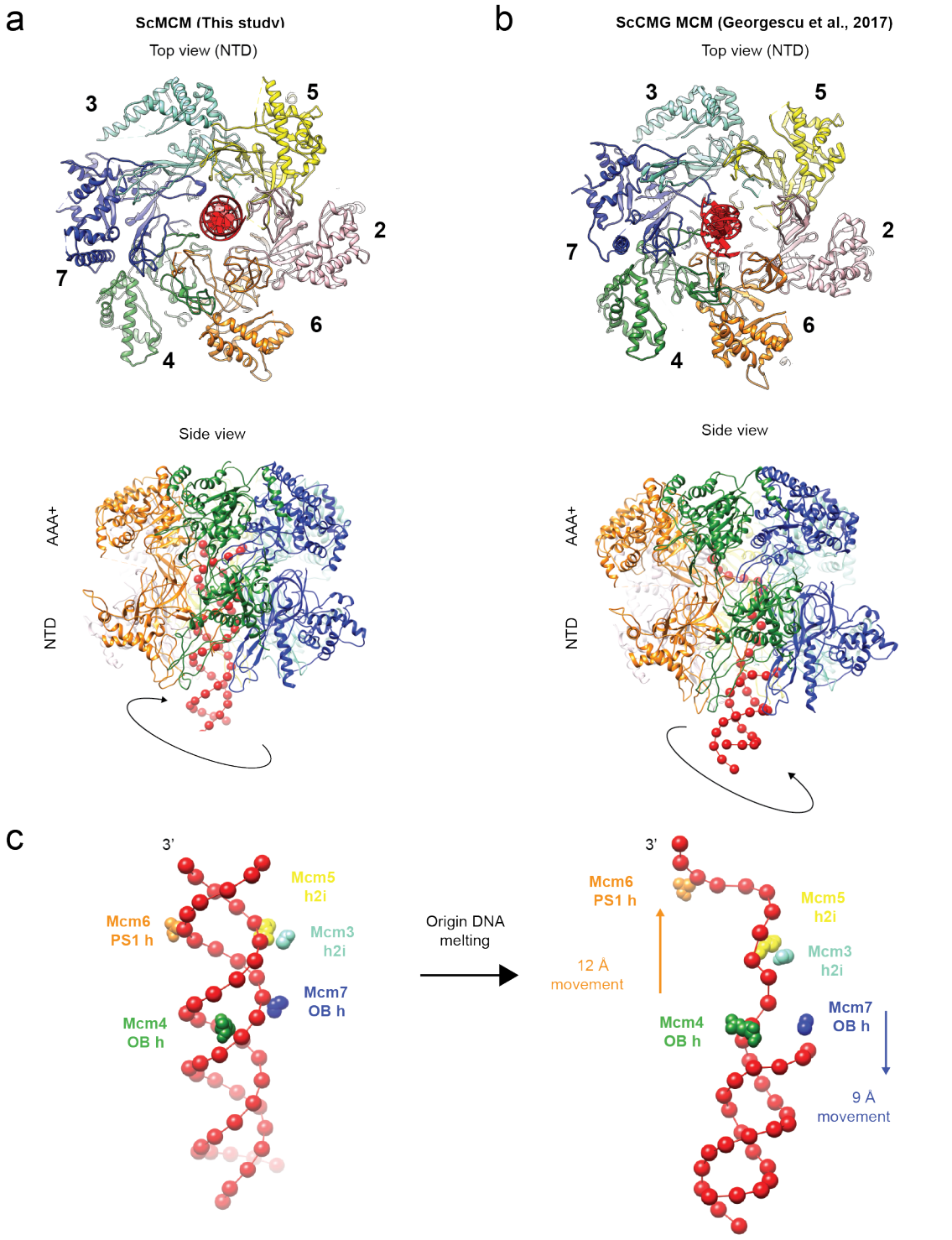


Figure 4.19: A model for DNA melting
Comparison of the **(a)** MCM and **(b)** CMG DNA bound complexes (PDB: 5u8s) reveal the protein and DNA movements **(c)** that lead to DNA untwisting. As the MCM transitions to CMG, duplex DNA moves from the Mcm2-5 side to the Mcm4-7 side in a clockwise direction, compatible with underwinding. Simultaneously, Mcm6 PS1 hairpin and Mcm7 OB hairpin move in opposite directions, while Mcm4 OB hairpin remains fixed at the duplex-single-stranded DNA junction. Altogether, this leads to a stretching motion of the DNA filament, which is ready to be captured by the Mcm5 and Mcm3 h2i loops.

Concomitant with DNA rotation, the Mcm6 PS1 loop moves upwards by 12 Å and the Mcm7 OB loop downwards by 9 Å while the OB-fold of Mcm4 remains fixed in both structures to stretch the DNA, which is a likely prerequisite for DNA untwisting (Duderstadt et al., 2011) (**Figure 4.19c**). In this model, the H2i of Mcm3 and Mcm5 are suitably positioned to bind to the stretched single strand in the channel and place the untwisted DNA at an interface between the AAA+ and NTD tiers. This model for DNA melting is further in agreement with our earlier single molecule FRET experiments, which showed that the CMG can stretch a single strand of DNA (*section 4.2.8*) and ascribes the specific MCM loops that mediate this DNA conformational change (**Figure 4.19c**).

The melted strand is not shown in this transition as it is absent from the solved CMG-DNA structure (Georgescu et al., 2017) but it is likely that this strand would be captured by the Mcm4-6-7 subunits of the NTD at the MSSB site during this rotation (Froelich et al., 2014). Furthermore, this melting model assumes that the MCM to CMG transition is not affected by the recruitment of other origin firing factors, for example Sld3/7 and the CDK dependent Sld2-Dpb11-Polε interaction.

In summary, a stretch-to-open DNA melting model is proposed where upon CMG assembly, the DNA inside rotates and stretches due to the opposing movements of MCM loops, thereby providing the force required for DNA melting. A similar mechanism has been previously described for the bacterial DnaA initiator, which sequesters and stretches one of the two strands of DNA in an ATP dependent mechanism prior to helicase recruitment (Duderstadt et al., 2011).

4.2.10 The orientation of the CMG helicase at the replication fork

The CMG helicase translocates on DNA with a 3’ to 5’ polarity (Bell and Labib, 2016). However, the fork nexus was not resolved in the cryo-EM CMG-DNA structure in this study and therefore the relative orientation of the CMG as it progresses is still a matter of debate. Knowing the polarity of the helicase is important to understand how the two helicases begin to unwind DNA and how the progressing replisome is organised.

To address helicase polarity, a previous study reconstituted *D. melanogaster* CMG on a primer template junction and imaged this complex by negative stain EM (Costa et al., 2014). In those experiments, the CMG was bound to DNA using ATPγS and its orientation was analysed relative to a biotin-streptavidin label on the duplex end of the substrate. This showed that the *D. melanogaster* CMG is positioned on the primer template junction AAA+ tier first, which is in contrast to the yeast CMG orientation of NTD first under the same experimental conditions (Georgescu et al., 2017). This discrepancy could be because the MCM can bind DNA in both directions as determined by our DNA bound double hexamer structure (*see section 3.3*).

To address which DNA-engagement mode supports productive unwinding, we returned to the *in vitro* reconstitution assay to assemble and activate CMGs (biochemical preparation by Dr. Max Douglas). In our strategy, the aim is to capture active CMG helicases that push dormant MCM double hexamers against a roadblock. In this way, CMG polarity can be assessed by its relative position against MCM double hexamers (depicted in the lower panel of **Figure 4.20a**). To this end, we introduced covalent protein roadblocks at the ends of 3-kb linear DNA substrate to prevent MCM sliding off the ends and loaded double hexamers on this bead-bound DNA substrate in excess. A subset of these double hexamers was activated with the complete set of origin firing factors and a high-salt wash was performed to remove the weakly bound proteins (**Figure 4.20a**) (Yeeles et al., 2015). The nucleo-protein complexes were subsequently released by a single-site restriction enzyme digest to preserve the roadblock positions.

Negative stain EM analysis of the reaction revealed the presence of multiple MCM ‘trains’ amongst a mixture of double hexamer and CMG single particles (**Figure 4.20b**). Each train observed in the raw micrographs was capped off by a CMG-like density on one side (**Figure 4.20b**, insets), suggesting the presence of a universal feature on each train. To confirm this notion, manual particle picking was performed along the lengths of all trains in EMAN2 (Tang et al., 2007) then averaged and classified in RELION-2 (Kimanius et al., 2016).

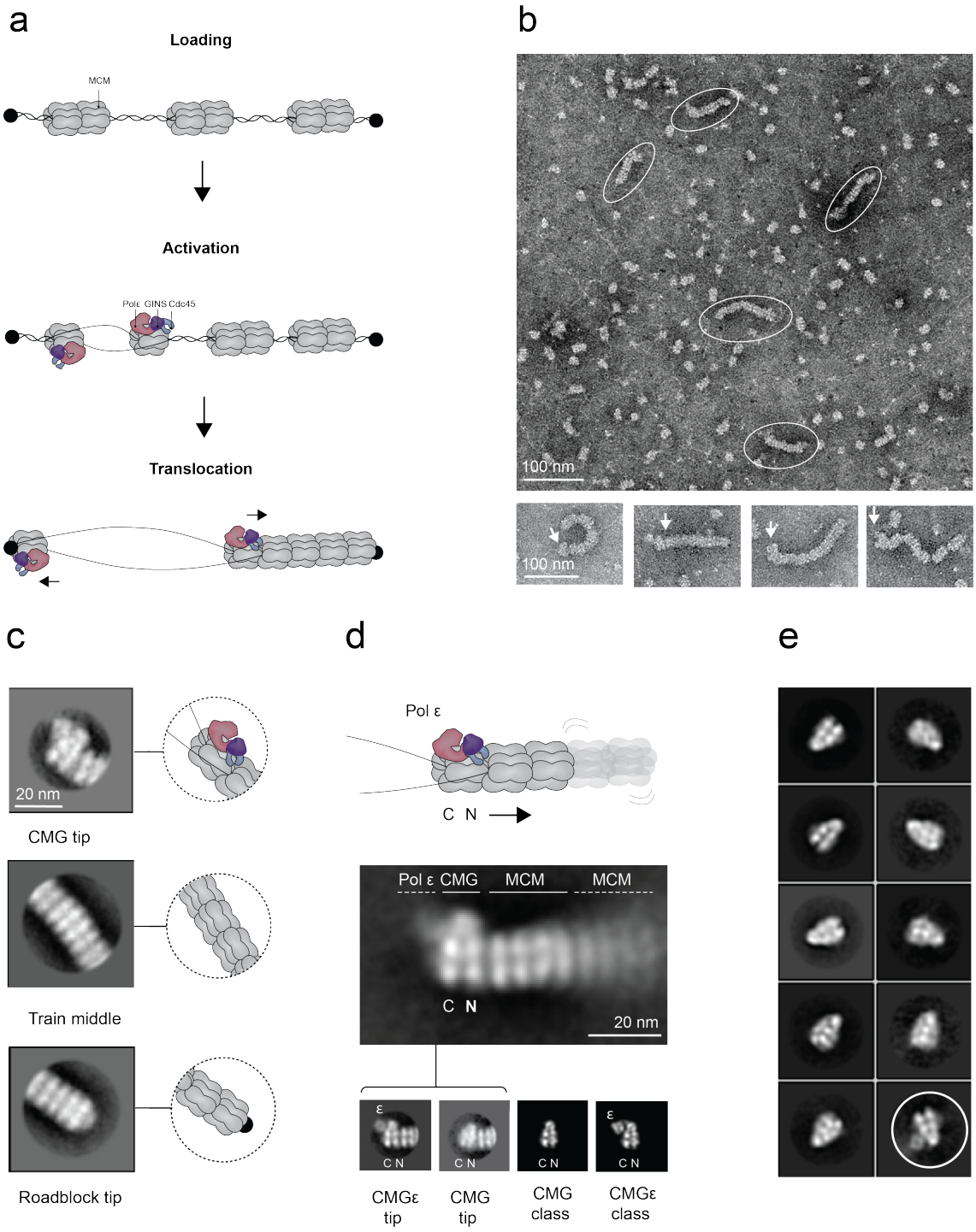


Figure 4.20: The CMG translocates with its N-terminal tier ahead of the AAA+ engine
(a) Schematic of assay to assemble and activate only a subset of CMG helicases on road blocked linear DNA. **(b)** Negative stain electron micrograph (out of 405 collected) of full origin firing reaction on road blocked DNA. White circles highlight the presence of MCM 'trains'. Individual examples of trains showing that a CMG particle always caps off one end are shown below the micrograph. **(c)** 2D class averages of various sections of the trains. Box size: 128 pixels. **(d)** Cropped 2D class average of train with a 250 pixel box size, showing the CMG N-terminal tier first against the MCMs. Trains are either capped off with a CMG bound to polymerase epsilon (80% of particles) or with isolated CMG. 2D class averages of isolated CMG and CMG bound to polymerase epsilon are shown for comparison purposes. 469 particles contributed to the big class. **(e)** Classification of CMG particles not part of the trains. 98% (7,928) of particles are isolated CMG whereas 2% (151 particles) are epsilon bound (Total particles: 8,090).

The resulting 2D classes give high signal-to-noise views of three sections of the train: (*i*) a side view of the CMG with additional density against the start of an MCM train ('CMG tip') (*ii*) the middle section of a train ('train middle') and (*iii*) the end of the MCM train pushed against the roadblock ('roadblock tip') (**Figure 4.20c**). The additional, doughnut shaped density in the CMG tip is characteristic of the Polε structure recently described, which binds to the C-terminal AAA+ tier of the active helicase (Zhou et al., 2017, Sun et al., 2015). Therefore, the position of Polε relative to the CMG tip can inform the polarity of the helicase at the fork junction with confidence.

To get a better overview of the CMG tip and analyse the orientation of the translocating helicase, the particles that contributed to the CMG tip class were re-extracted to a bigger box size (from 98 to 250 pixels) and the circular mask was expanded before 2D averaging. This revealed a CMG-Polε complex against a train of two double hexamers that gets increasingly more flexible (**Figure 4.20d**). Comparison of the CMG tip of the train to 2D projections of CMG and CMG-Pol epsilon (Costa et al., 2014, Zhou et al., 2017) reveals that the CMG translocates NTD tier first ahead of its motor AAA+ tier (**Figure 4.20d**, lower panel). This is in line with the orientation identified by the cryo-EM structure of recombinant yeast CMG bound to forked DNA (Georgescu et al., 2017), the single-stranded DNA in the DmCMG structure (*section 4.2.6*) and the E1 helicase (Enemark and Joshua-Tor, 2006). Therefore, eukaryotic-based helicases bind DNA with the same polarity.

Notably, over 75% of CMG tips are Polε bound (**Figure 4.20d**, lower panel). In contrast, 98% of CMG particles that have not translocated to the train tips from the same reaction are Polε free (**Figure 4.20e**). This indicates that Polε may have an active role in stabilising productive fork engagement by the CMG (see Chapter 5 for structural characterisation of CMG-Polε).

One caveat of this assay is that the trains could be the result of passive diffusion of the MCMs towards the roadblock rather than the result of CMG activity. To rule this out, we analysed five control experiments by negative stain EM analysis and single particle image processing (**Figure 4.21**). A reaction comprising the complete origin firing reaction but on a non-roadblocked substrate completely abolished train formation regardless of the number of DNA cleavage sites (**Figure 4.21a**, multisite

cut vs single site cut). Furthermore, these trains were dependent on a high number of MCM double hexamers being loaded in the first step of the reaction (**Figure 4.21b**). Finally, omitting Mcm10 from the reaction where CMG assembly but not activation can occur, is insufficient for train formation in both roadblocked and non-roadblocked substrates (**Figure 4.21c**). Collectively, this suggests that the CMGs observed in the trains (**Figure 4.20b**) are active and push the double hexamers towards the roadblock, where they are stopped from falling off the DNA.

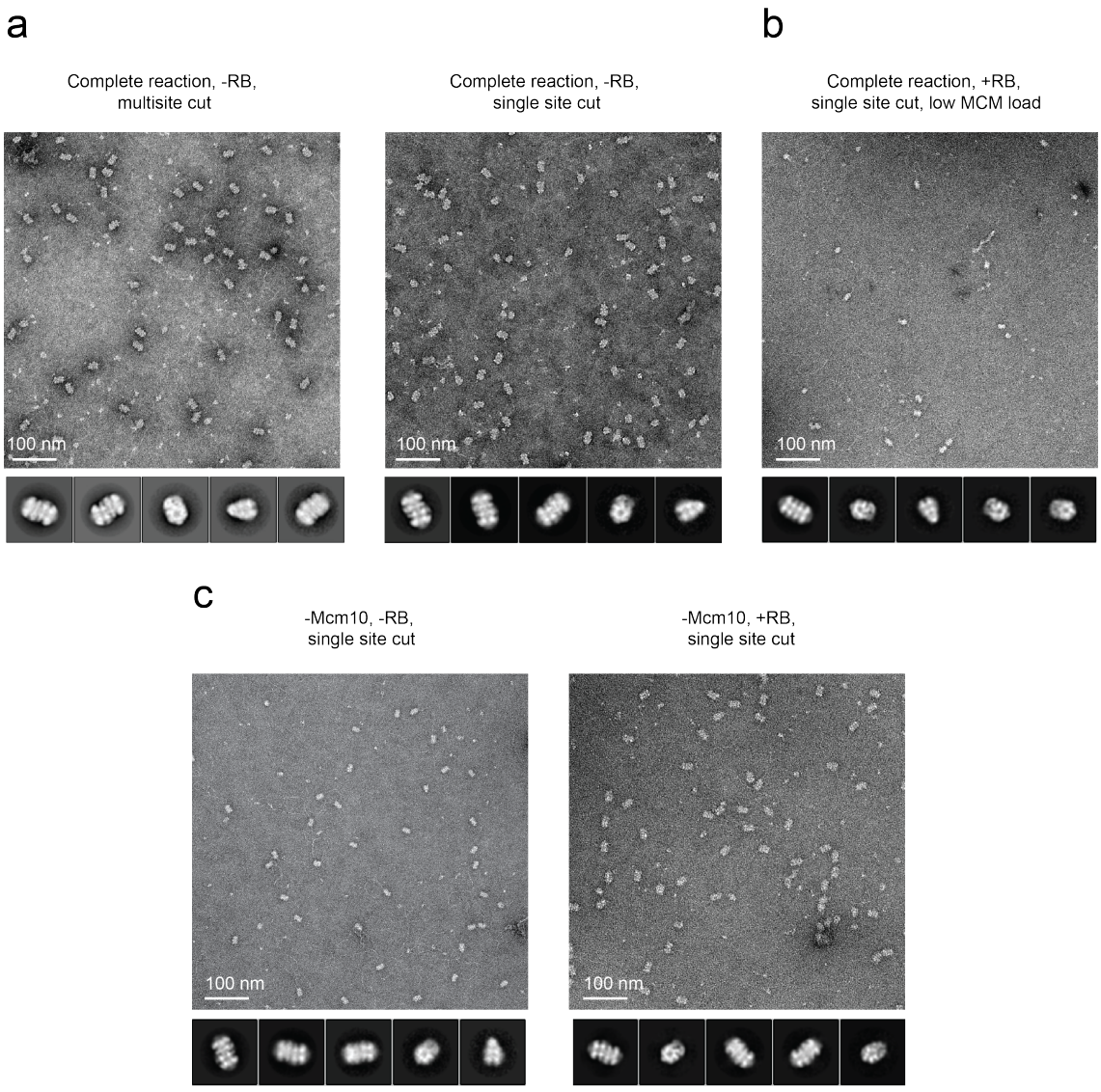


Figure 4.21: Trains only form in the presence of roadblocks and Mcm10
(a) Absence of roadblocks causes trains to fall off the DNA and disassemble and **(b)** trains cannot form if MCMs are loaded at a low concentration. **(c)** Omission of Mcm10 from reaction prevents train formation, indicating that only active CMGs push MCMs towards a roadblock. Complete reaction (-RB) multisite cut: 228 micrographs (9,302 particles); Complete reaction (-RB) single site cut: 112 micrographs (7,977 particles); Complete reaction (+RB) single site cut, low MCM load: 722 micrographs (13,698 particles); -Mcm10 (-RB) single site cut: 92 micrographs (4,614 particles); -Mcm10 (+RB) single site cut: 82 micrographs (8,013 particles). 98 pixel box size used throughout for 2D classification.

As the complete trains are large, flexible and variable in shape, they are refractory to 3D reconstruction by canonical single particle methods. This prompted us to further investigate the structure of the trains in the near-native conditions of cryo-electron tomography. In this technique, the sample is imaged as it is tilted to different angles to produce a series of 2D projections (tilt series) that can then be computationally reconstructed to a 3D volume (tomogram) (Wan and Briggs, 2016).

In collaboration with Dr. Andrea Nans from the Crick structural biology STP, we set out to collect several tomograms and used template matching approaches to find the positions of MCM and CMG-Pol epsilon in a representative train (**Figure 4.22**). To this end, the complete origin firing reaction on roadblocked DNA (as described in **Figure 4.20a**) was applied twice to lacey grids comprising an ultrathin layer of carbon prior to plunge freezing and imaging on a Titan Krios equipped with a K2 summit camera (grid preparation by Dr. Julia Locke).

Figure 4.22a shows the complete field of view of a 3D tomogram section and the presence of a train at the bottom. Several examples of different trains were observed across different tomograms, including straight and compressed (**Figure 4.22b**, top) and curved, looser trains (**Figure 4.22b**, bottom). To detect the position of CMG-Polε relative to the MCM trains we used the template matching software MOLMATCH (Forster et al., 2010). Using this software package, the known structures of MCM (Chapter 3) and CMGε (Chapter 5) are used as templates to calculate their cross-correlation function relative to a cropped experimental tomogram (also referred to as pattern recognition) (**Figure 4.22c-d**) (Forster et al., 2010). After cross-correlation peak determination, the highest matching scores were chosen and viewed as a macromolecular atlas in Amira (ThermoFisher) (see methods *section 2.7.2.4*). This showed a clear view of a CMGε complex translocating in an NTD tier first directionality against a train of ten MCM double hexamers, thereby confirming the negative stain experiments (**Figure 4.22d**).

The orientation of NTD-first at the fork nexus has important implications as it means that the two CMGs would cross paths upon activation and ensure that all DNA is unwound at the origin. Furthermore, the observation of CMGs actively pushing inactive double hexamers could be relevant *in vivo*, where more double hexamers

are loaded than are fired (Donovan et al., 1997). In this context, the CMG could drive the unfired double hexamers forward, which could act as a “back-up” source upon fork collapse (Remus et al., 2009, Gros et al., 2015).

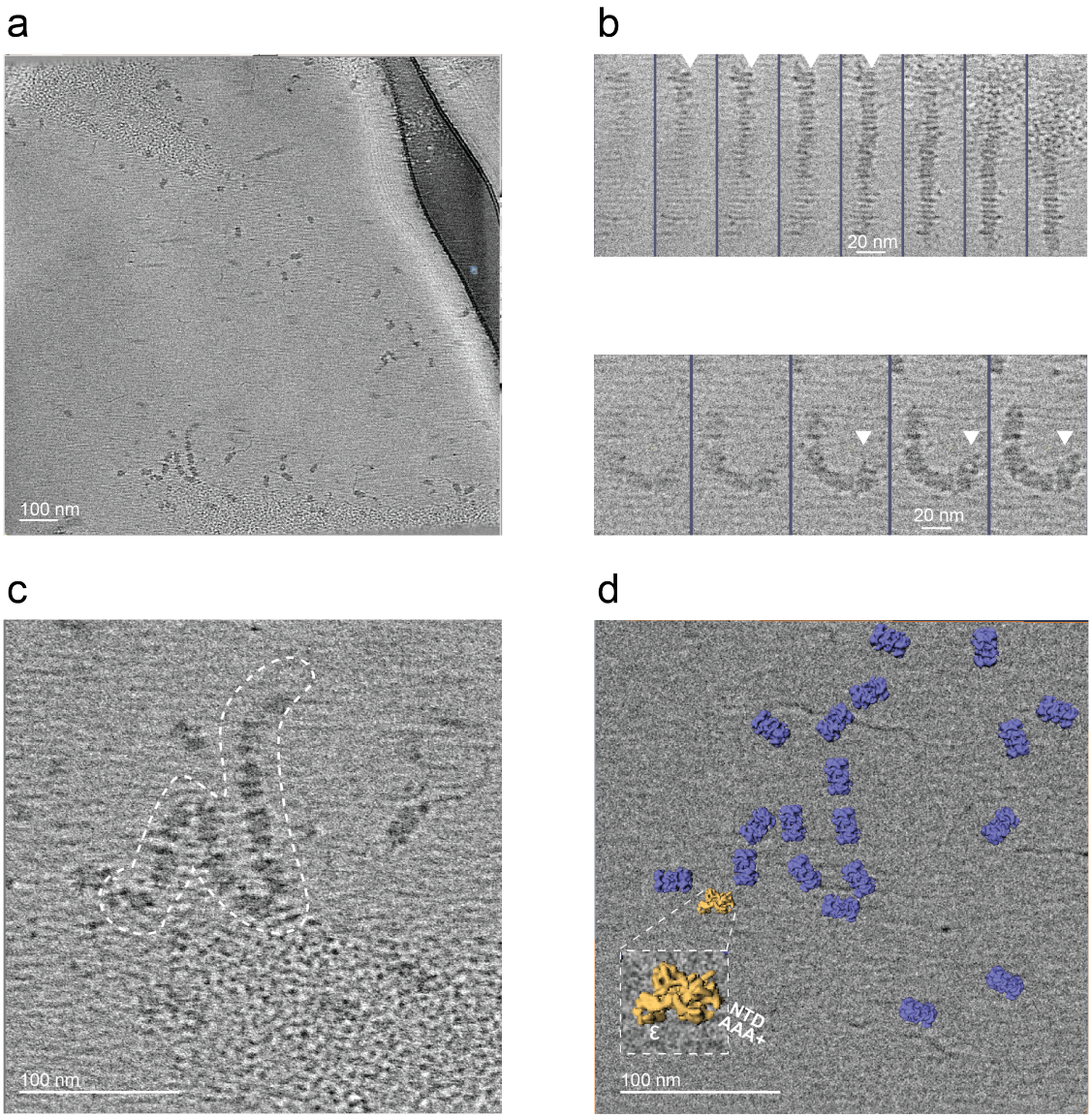


Figure 4.22: Cryo-ET of MCM trains confirms that the active CMG helicase translocates N-tier first
(a) An x-y slice of a 3D tomogram of the complete field of view, containing the full origin firing assay reconstituted *in vitro*. **(b)** Examples of trains from two different tomograms, visualised along the z axis (displayed using the multi Z slicer from IMOD). **(c)** Cropped view of MCM train obtained from the tomogram in (a). **(d)** MOLMACTH placed MCM double hexamers (blue) and CMG-ε shows that the CMG pushes towards the MCM N-terminal tier ahead of the ε-bound AAA+ tier (visualized in Amira).

4.3 Conclusions

In this chapter, I used both the overexpressed and step-wise assembled CMG to provide mechanistic insights into helicase mode of DNA translocation, ATPase state switch and polarity at the replication fork. In conjunction with MCM double hexamer results presented in Chapter 3, I further proposed how the MCM to CMG assembly transition could lead to double hexamer separation, untwisting of DNA and activation of the helicase.

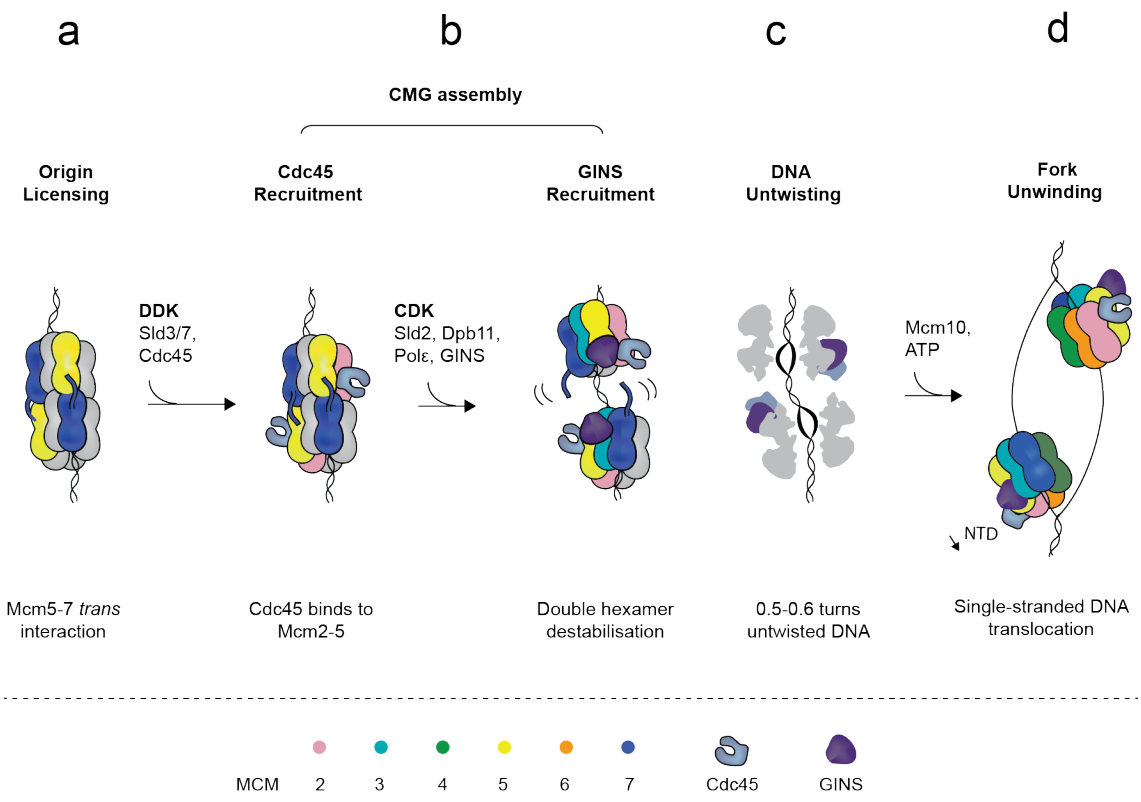


Figure 4.23: A model for CMG helicase activation and translocation.
(a) The loaded double hexamers are held together in part by the in trans interaction between Mcm5 and Mcm7. **(b)** The two helicases are still in double hexamer form as Cdc45 is recruited but GINS binding leads to destabilisation of the Mcm5-7 interaction of the double hexamer. **(c)** Concomitant with CMG assembly, each helicase begins to untwist the double helix, due to the narrowing of the channel and the opposing loop movements of Mcm6 and Mcm7 that stretch the DNA. **(d)** Upon origin firing and ATP hydrolysis, the CMG helicase translocates on single-stranded DNA, and may alternate between a closed and open configuration during unwinding. Replisome progression occurs with the CMG NTD ahead of the motor AAA+ tier, meaning that all DNA has to be unwound over the origin as the two CMGs pass each other.

Together, the data suggests a model (**Figure 4.23**) where the inactive double hexamer forms a tight association in part due to the Mcm5-7 in *trans* interaction (**Figure 4.23a**). Cdc45 and GINS are recruited in separate steps (**Figure 4.23b**). GINS binding to MCM requires the same Mcm5 site that is protected by Mcm7 in the

double hexamer. Mcm5 undergoes a conformational change to enable GINS binding and Mcm7 release, leading to dissociation of two fully formed CMG helicases. Concomitant with double hexamer separation, half a turn of DNA is untwisted per assembled CMG; in the MCM to CMG transition, the channel narrows while a rotation of the DNA coupled with the counteracting movements of Mcm6 and Mcm7 leads to the stretching and untwisting the double helix (**Figure 4.23c**). In line with CMG stretching DNA, single molecule FRET studies show that single-stranded DNA is stabilised in an extended configuration inside the MCM channel.

Upon origin firing, the CMG helicase translocates on single-stranded DNA while excluding the lagging strand template (**Figure 4.23d**). The CMG may switch from a translocation state (compact ATPase) during unperturbed replication (Georgescu et al., 2017) to a second state that either represents a paused intermediate (relaxed ATPase) during fork stalling (Petojevic et al., 2015) or a second translocation state. The change between the two conformers is driven by the opening and closing of the Mcm2-5 gate. In both conformations, Cdc45 and GINS keep a topologically closed N-terminal tier to prevent DNA from slipping out of the channel (Petojevic et al., 2015). During unwinding, the CMG translocates with its N-terminal tier ahead of the AAA+ motor, meaning that the two helicases have to pass each other after activation, which may act as a fail-safe mechanism to ensure that all origin DNA is unwound (Douglas et al., 2018).

Chapter 5. Cryo-EM structure of the CMG helicase bound to Polε

5.1 Introduction

The CMG helicase separates the two component strands of the double helix and acts as the replisome organising centre to ensure that unwinding and daughter strand synthesis are tightly coordinated (Johansson and Dixon, 2013, Sun et al., 2015, Gambus et al., 2009, Langston et al., 2014). As part of this large macromolecular replisome machinery, the CMG advances the replication fork and Polα-primase synthesises short primed RNA-DNA sequences for both template strands (Kuchta and Stengel, 2010, Ilves et al., 2010, Moyer et al., 2006). These primed sites are then extended by Polδ and Polε on the lagging and leading strands, respectively (Kunkel and Burgers, 2008, Pursell et al., 2007, Clausen et al., 2015). In agreement with a leading strand specific role of Polε, the CMG helicase, which encircles the leading strand template, binds directly to Polε (Sun et al., 2015, Zhou et al., 2017, Langston et al., 2014). In fact, GINS and Polε are recruited together during CMG assembly to form a stable CMGε complex (Yeeles et al., 2015, Muramatsu et al., 2010), ensuring that coupling of unwinding and synthesis occurs at the onset of origin firing. In contrast, Polα and Polδ are dispensable for CMG formation and activation of unwinding (Yeeles et al., 2015, Douglas et al., 2018). Therefore, characterisation of the association between CMG and the leading strand polymerase is paramount to understanding DNA replication initiation and replisome progression.

A negative-stain EM study of Polε by our laboratory has provided an architectural overview of this hetero-tetrameric complex (Zhou et al., 2017). Polε is organised into two lobes: *(i)* the Pol2 N-terminal catalytic domain (N-Pol2) (Hogg et al., 2014), which is dispensable for viability (Feng and D'Urso, 2001, Kesti et al., 1999), and *(ii)* the Pol2 C-terminal catalytically inactive polymerase domain (C-Pol2), which is indispensable for viability (Dua et al., 1999) and is bound to the essential Dpb2 catalytically inactive exonuclease subunit (Isoz et al., 2012, Araki et al., 1991a) and the ancillary Dpb3 and Dpb4 factors (He et al., 2017). An exhaustive description of

the low-resolution CMGε structure was subsequently provided (Zhou et al., 2017), showing that the catalytically inactive module of Polε interacts with the AAA+ tier of the CMG. In contrast, the catalytic N-Pol2 domain is mobile and switches between two configurations that could be important for its polymerase activity and enhancing replication fork progression rates (Yeeles et al., 2017, Zhou et al., 2017).

Despite these advances, it remains unknown how the CMGε complex binds DNA, what the function of the essential, inactivated Polε domains are and what specific interactions keep Polε directly associated with the replicative helicase. To address this and to build on our understanding of helicase activation and replisome progression from Chapter 3 and 4, I set out to obtain a high-resolution structure of a CMGε complex bound to a model replication fork. In this chapter, I describe the purification of yeast CMG from a recently developed diploid strain in the lab and give an overview of the biochemical reconstitution of CMGε using a fork DNA affinity approach. I then present the subnanometre cryo-EM structure of CMG-DNA bound to the essential Polε subunits and detail the CMGε interaction network. Finally, I propose a model for the role of Polε in helicase activation and replication fork establishment.

5.2 Results

5.2.1 Purification and EM analysis of recombinant yeast CMG

To reconstitute a CMGε complex bound to DNA, I first had to obtain pure and concentrated CMG. Previously in Chapter 4, I detailed the expression and purification of *D. melanogaster melanogaster* CMG from baculovirus-infected insect cells. Since that project, we have benefited from investigating DNA replication using the yeast model system in collaboration with the Diffley laboratory (see Chapter 3 and *section 4.2.10* of Chapter 4) (Yeeles et al., 2015, Douglas et al., 2018). However, several attempts to get large quantities of yeast CMG for high-resolution studies using the *in vitro* reconstitution approach were unsuccessful. Therefore, I turned to overexpression and purification of recombinant yeast CMG using a recently developed diploid strain by our laboratory (yJCZ3 made by Dr. Jin Chuan Zhou) (**Figure 5.1**).

To create the diploid strain, a *MATa* Mcm2-7 strain was mated with a *MATα* GINS-Cdc45 strain, thereby providing sufficient selection markers for integration of plasmids containing the codon optimised CMG genes (Zhou et al., 2017). The diploid strain comprises an internal FLAG (iFLAG) tag on Cdc45 and an additional N-terminal CBP-tag on Mcm3. To purify the CMG complex, I used the same protocol outlined in *section 4.2.1* apart from modifications in cell lysis and pull-down strategy (iFLAG-Cdc45 was used instead of pulling down on Mcm3, see *section 2.3.2* for full description) (**Figure 5.1a**).

Figure 5.1b-c shows how the first anion exchange step separates the full CMG complex from Cdc45-subcomplexes after the FLAG pulldown. To further purify out excess Cdc45 and enrich for the CMG complex, the elution was applied to a second anion exchange step (**Figure 5.1d**). On a Coomassie stained SDS-PAGE gel the resulting peak fraction showed the presence of near-stoichiometric CMG (elutes at ~2 ml), which was separated from a CMG fraction containing contaminating bands (~1.9 ml) (**Figure 5.1e**). At this stage, the concentration of CMG is in the low micro molar range, which is a sufficient starting point for further purification or reconstitution.

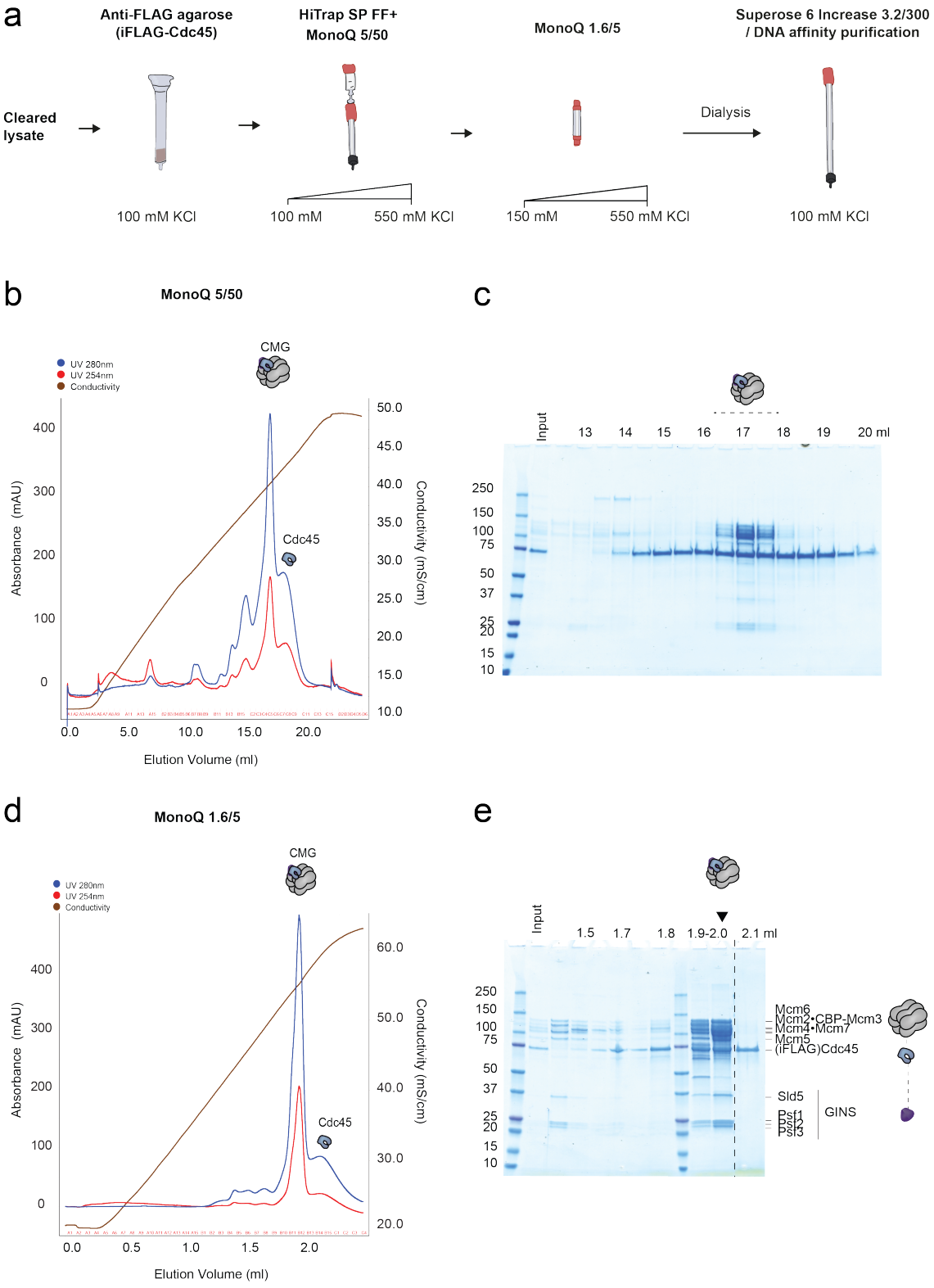


Figure 5.1: Purification of recombinant *S. cerevisiae* CMG
(a) Purification scheme. **(b)** Mono Q 5/50 chromatogram showing the separation of CMG from Cdc45. **(c)** Coomassie SDS PAGE verification of Mono Q 5/50 purification. Dashed line indicates pooled fractions. **(d)** Mono Q 1.6/5 chromatogram showing the further removal of excess Cdc45 from CMG peak. **(e)** Coomassie SDS PAGE verification of Mono Q 1.5/5 purification with labelled bands.

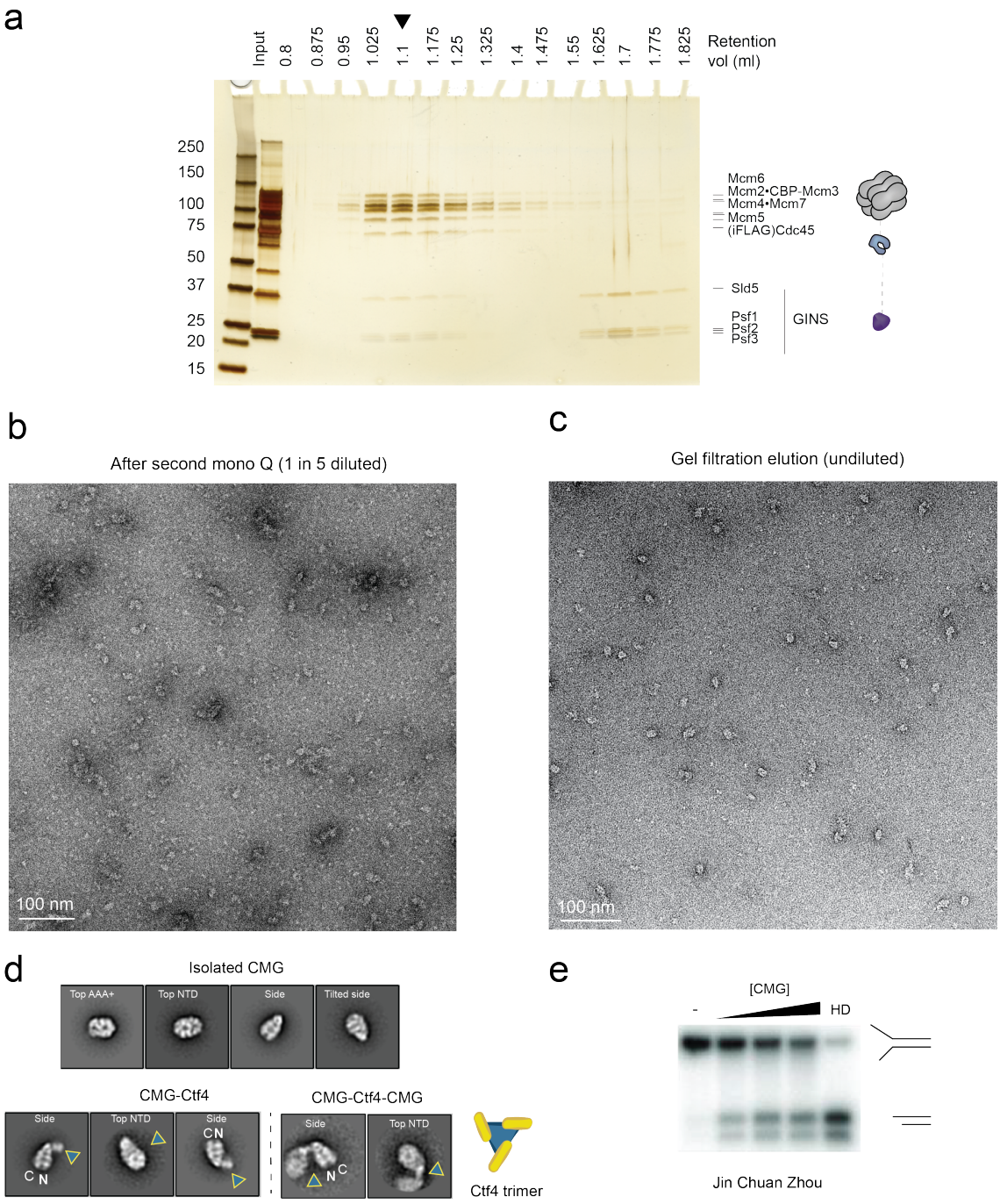


Figure 5.2: EM analysis of yeast CMG and co-purification of endogenous Ctf4

(a) Silver stain SDS-PAGE of a CMG gel filtration run on a Superose 6 Increase 3.2/300 column. **(b)** Representative micrograph of a diluted fraction from the Mono Q 1.6/5 peak reveals heterogeneity in the sample. Collected on a ThermoFisher FEI F20 TEM with a Falcon II detector at 50,000x magnification (2.05 Å/pixel). 98 micrographs collected. **(c)** Representative micrograph of gel filtration elution (Superose 6 Increase 3.2/300 column) showing improved homogeneity compared to (b). Collected on a FEI Spirit (TWIN) TEM at 30,000x magnification (3.45 Å/pix) (120 micrographs collected). **(d)** Reference free 2D class averages of CMG using Relion-2 (Kimanius et al., 2016). The first 7 classes are from (c) (total 8,731 particles) whereas the last two classes, showing the double CMG are from (b) (total 528 particles), 156 pixel box size for all classes. A cartoon representation of Ctf4 is shown alongside arrows pointing to where Ctf4 is placed on the CMG. **(e)** Unwinding assay performed by Dr. Jin Chuan Zhou using the purified CMG fraction from Figure5.1e. From the second lane to the third lane the 159, 397, and 793 fmol CMG was incubated with 20 µl buffer containing 25 mM Hepes, pH 7.6, 50 mM NaCl, 0.5 mM ATP, 10 mM Mg-acetate, 1 mM

DTT, 0.1 µg/µL BSA, 27 µM ATPγS, and 1.8 pmol [γ-32P]-ATP–labeled fork for 3 hours at 30ºC. Unwinding was initiated by incubation with ATP to 7 mM for 5 min at 30ºC. Reactions were quenched 150 mM Tris, pH 8.0, 3% SDS, 120 mM EDTA. Products were separated on 8% PAGE gel with 0.1% SDS running in 1X TBE.

To further assess sample purity and homogeneity after the two anion exchange steps, I dialysed the CMG down to low salt (100 mM KCl) and applied it to a gel filtration column (Superose 6 Increase 3.2/300). A silver stain gel of the fractions from the size exclusion step reveals a relatively broad distribution of the CMG peak, indicating that the CMG exists in multiple conformational and/or compositional states (**Figure 5.2a**). Moreover, comparison of the anion exchange step elution to a fraction from the gel filtration peak revealed that the size exclusion step gave a dramatic improvement in homogeneity under the electron microscope (compare **Figure 5.2b** to **Figure 5.2c**).

Subsequent single particle image processing of the fully purified CMG in RELION-2 (Kimanius et al., 2016) produced reference-free 2D class averages of the helicase in several different orientations, showing the characteristic side, tilted and top views (**Figure 5.2d**). Notably, a large subset of particles was bound by an additional factor at the NTD tier of the CMG (**Figure 5.2d**, bottom row). These classes were highly similar to previously published CMG-Ctf4 classes (Sun et al., 2015)(**Figure 1.20**) and gel band mass spectrometry confirmed the presence of Ctf4 in the preparation (in collaboration with Dr. Len Packman, Cambridge Biochemistry, data not shown). From a separate dataset of the sample prior to gel filtration, one CMG is seen to connect to another CMG to form a double-CMG complex, where the Ctf4 trimer acts as the linking bridge (**Figure 5.2d**, bottom row). The observation of this assembly is in line with several studies reporting that Ctf4 can bind multiple client proteins at the same time, including Sld5 (Simon et al., 2014, Gambus et al., 2009) and that the Ctf4-Sld5 interaction is stable throughout the cell cycle (Samora et al., 2016, Gambus et al., 2009). Importantly, the results presented in **Figure 5.2** indicate that the diploid yeast strain can be used to co-purify recombinant CMG with stably bound endogenous replication factors. Future work will reveal whether more transient replisome complexes can also co-purify with the CMG from its native source.

With verification of complex formation, we next assessed whether the purified yeast CMG is catalytically active. In collaboration with Dr. Jin Chuan Zhou, we tested

helicase activity on a radiolabelled fork as described before in *section 4.2.1*. As shown in **Figure 5.2e**, increasing the amount of yeast CMG in the reaction increases the unwound DNA product detected, thereby confirming that the purified helicase is active.

In summary, I have showed that the recently generated diploid strain in our lab can be used to purify CMG to a high concentration. Negative stain EM analysis shows that the CMG particles average and a subset are bound by the replisome factor Ctf4. Together, this makes the yeast CMG strain an ideal tool to use for reconstitution with Polε for high-resolution structural determination.

5.2.2 Purification of yeast polymerase epsilon variants

The next step towards CMGε reconstitution is to purify Polε. To this end, I expressed and purified a recently developed strain that expresses full-length Polε but with two inactivating point mutations (D290A and E292A) in the exonuclease domain of N-Pol2 (yAE99 strain by Dr. Anne Early). This construct (hereafter termed Polε-exo-) prevents the degradation of 3' termini of DNA and as such is a good candidate to capture an intact CMGε-fork DNA complex (Sun et al., 2015, Ganai et al., 2015).

The purification of Polε-exo- entailed a CBP pull-down followed by anion exchange and gel filtration steps (**Figure 5.3a**). **Figure 5.3b-c** show that after anion exchange chromatography a highly pure Polε-exo- complex was obtained compared to the earlier pull-down stage. The purest fractions of the peak (eluting at ~9-9.5 ml) were subsequently pooled, dialysed and applied to a gel filtration column for separation of sub-complexes. As shown in **Figure 5.3d**, a single peak elutes that corresponds to all four subunits of Polε-exo- (**Figure 5.3e**).

The structure of wild-type Polε has been previously structurally characterised via cryo-EM (Asturias et al., 2006). However, a published crystal structure of the catalytic domain (Hogg et al., 2014) does not readily fit into the cryo-EM density. To revisit the architecture of Polε, our laboratory has extensively studied the architecture of the leading strand polymerase using several different dropout mutants (Zhou et al., 2017). These studies established that a single Polε monomer forms two distinct but

flexibly tethered lobes (comprising N-Pol2 and non-catalytic Polε [C-Pol2, Dpb2, Dpb3, Dpb4], respectively).

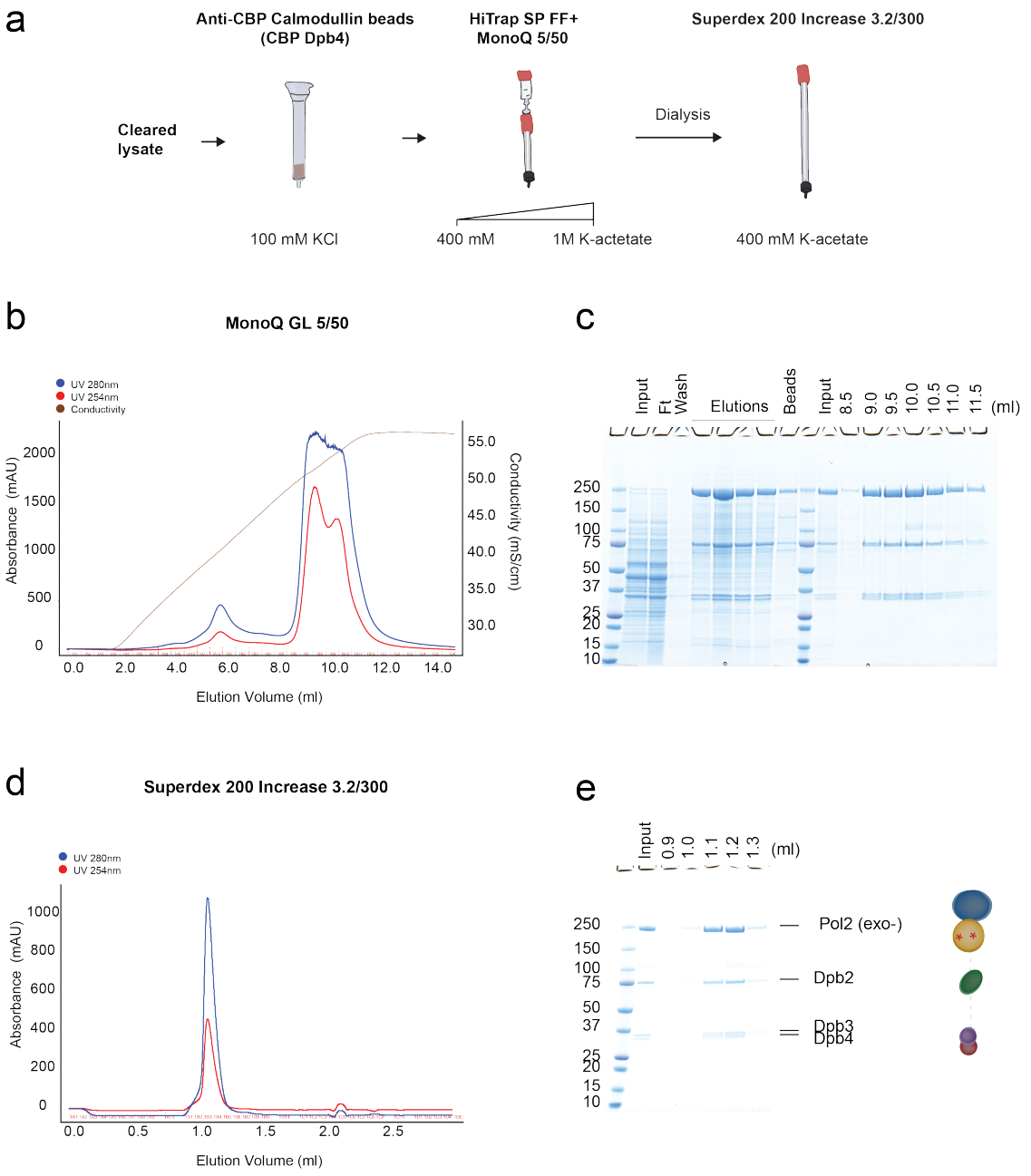


Figure 5.3: Purification of *S. cerevisiae* Polε-exo-
(a) Purification strategy. **(b)** Mono Q 5/50 step after pull-down. **(c)** Coomassie stained SDS-PAGE gel shows the improvement in purity from the cleared lysate to the anion stage of the purification. **(d)** Gel filtration step after Mono Q 5/50 gives a sharp peak that corresponds to **(e)** all subunits of Polε-exo- (Coomassie stained SDS-PAGE gel).

To assess whether the Polε-exo- mutant matches previous structural observations of wild-type Polε, I stained and analysed a purified Polε-exo- fraction by EM. The representative micrograph in **Figure 5.4a** shows monodisperse particles that vary in

shape and conformation. 2D classification in RELION-2 (Kimanius et al., 2016) revealed the presence of two distinct configurations in the particle population (**Figure 5.4b**). The first conformation falls into a 'compact' category, where the two lobes are closely tethered to each other (**Figure 5.4b**, top and middle row). In the well-resolved classes (**Figure 5.4b**, class 1 and 3, top row), a triangular shaped lobe that highly resembles the non-catalytic domains of Polε (Zhou et al., 2017) sits above a circular lobe, which by process of elimination is assigned to N-Pol2 (**Figure 5.4b**, cartoons). The second 'extended' conformation shows two less well resolved spatially separated lobes that are probably connected by a thin linker (Sun et al., 2015) (**Figure 5.4b**, bottom). The presence of two conformational states underscore the dynamic nature of Polε and may explain why a full-length version of this protein has been refractory to crystallographic studies.

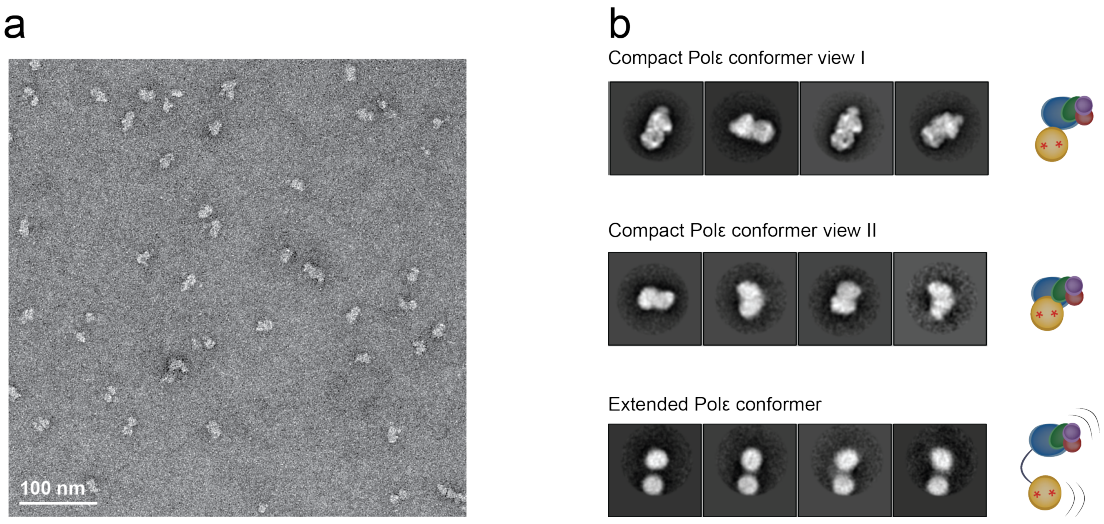


Figure 5.4: Negative-stain EM analysis of Polε exo- shows that the leading strand polymerase can switch between compact and extended conformations
(a) Representative micrograph (730 collected) of a 1 in 10 diluted fraction from the gel filtration elution in Figure 5.3. Collected on a FEI Spirit (TWIN) TEM at 42,000 x magnification (2.46 Å/pix). **(b)** Reference 2D class averages from RELION-2 (Kimanius et al., 2016) reveal the conformational flexibility of Polε-exo-. 14,572 and 13,228 out of 27,560 particles are compact and extended conformers, respectively. Box size 156 pixels (mask 320 Å).

In summary, I obtained pure Polε-exo- and showed that its architecture approximates that of wild-type Polε. As such, it is a suitable candidate for reconstitution with CMG and a model replication fork.

5.2.3 Optimisation and reconstitution of a CMGε complex

To understand how CMG and Polε coordinate for replication fork establishment, I reconstituted a CMGε-DNA complex by fork affinity purification (**Figure 5.5a**).

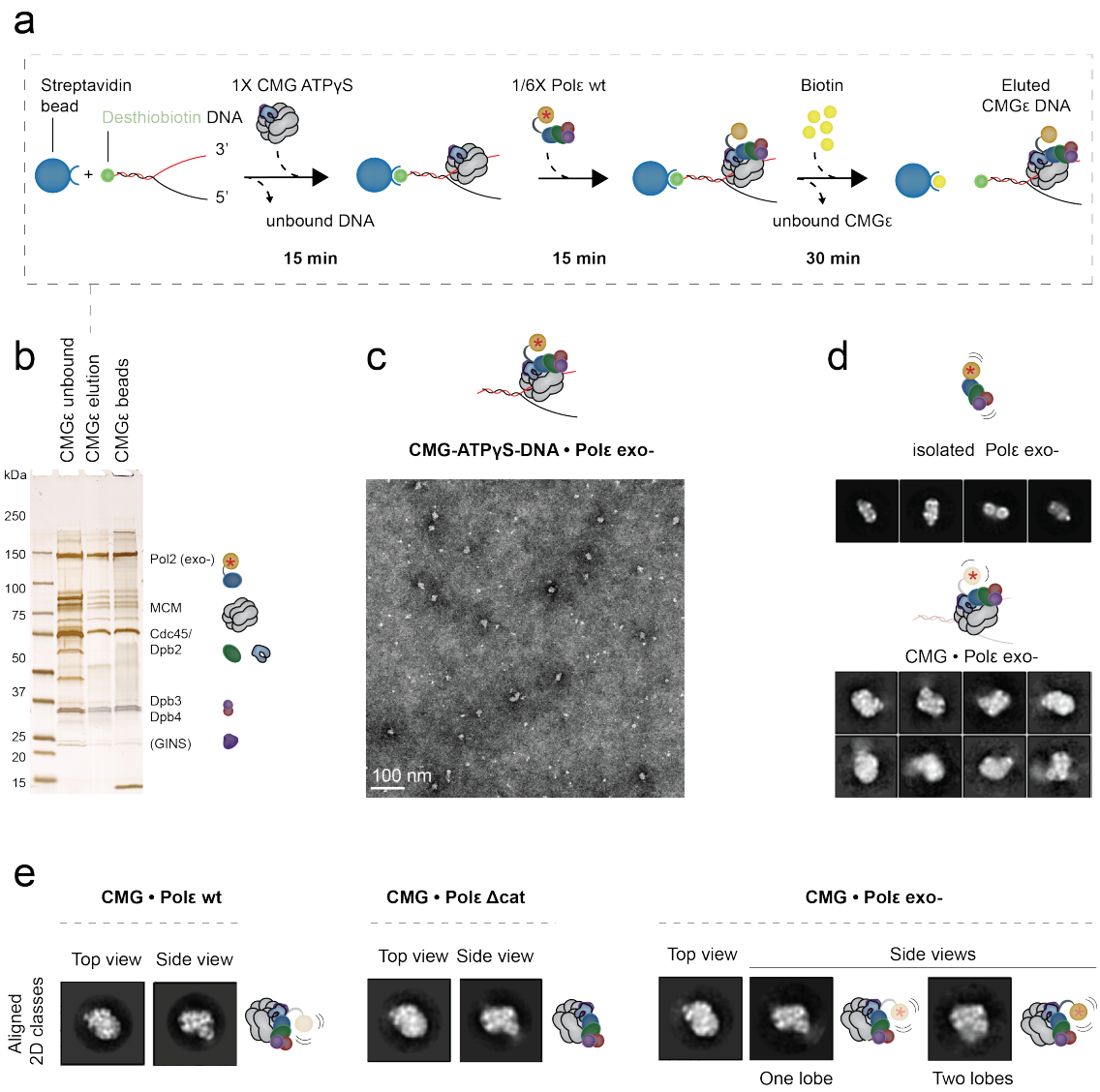


Figure 5.5: Reconstitution and EM analysis of CMG-Polε variant complexes on fork DNA
(a) Optimised strategy for CMG-Polε DNA reconstitution. **(b)** Silver stained SDS-PAGE gel showing the unbound, eluted and bead-bound fractions from the fork affinity assay. **(c)** Representative micrograph of a CMGε-DNA complex (510 micrographs collected). Collected on a FEI Spirit (TWIN) TEM at 21,000 x magnification (4.92 Å/pix). **(d)** Reference free 2D class averages from RELION-2 (Kimanius et al., 2016) of isolated Polε-exo- (18,729 particles) and CMG-Polε-exo- (5,982 particles). Total particles picked: 30,982; box size: 98 pix. **(e)** Alignment of top and side views of CMG-Polε wt (particles shown 8,208), CMG-Polε Δcat (3,729 particles) and CMG-Polε exo- (2,629 particles). 2D class averages show that N-Pol2 is flexible and the non-catalytic lobe of Polε anchors to the AAA+ tier of CMG. IMAGIC multi-reference alignment used.

This is based on a recently developed assay by Patrik Eickhoff and Dr. Thomas Miller in the lab, where a preformed replication fork is immobilised onto magnetic beads via a desthiobiotin-streptavidin interaction and used as bait to capture DNA binding

complexes. The fork used in these experiments contains a poly-T leading strand template and CG rich lagging strand template based on previously published sequences (Petojevic et al., 2015) (see **Table 2.2**). After binding DNA to the beads, I added CMG-ATPγS followed by Polε-exo- in a 3:1 ratio (**Figure 5.5a**). Nucleoprotein complexes were washed twice to remove weakly bound complexes before a biotin elution strategy was utilised, where biotin outcompetes the weaker streptavidin-desthiobiotin interaction and releases the DNA bound complexes.

As shown from the silver stained SDS-PAGE gel in **Figure 5.5b**, only a small subset of CMGε-exo- complexes binds to the DNA (compare lane 1 to lanes 2-3). This could be because the pre-assembled CMG bound to ATPγS forms a topologically closed complex as evidenced by 2D class averages in **Figure 5.2d** and earlier structural characterisation of CMG-ATPγS-DNA (see *section 4.2.2*). Therefore, instead of loading onto the DNA using an open gate akin to MCM-Cdt1, the CMG may have to thread through the leading strand template via its central channel. Furthermore, elution of DNA bound complexes was approximately 50% efficient. Despite this particle limitation, negative stain EM analysis of the elution revealed a homogenous set of two different populations: isolated Polε-exo- (bound to DNA) and CMG decorated with an additional lobe (**Figure 5.5c**). To interpret the extra density on the CMG, I subjected these particles to 2D classification in RELION-2 (Kimanius et al., 2016). This showed the same conformations of isolated Polε-exo- as described earlier in **Figure 5.4b** and a second set of CMGε-exo- classes, exhibiting a large array of views (**Figure 5.5d**).

The Polε-exo- density is tethered to the CMG on the AAA+ tier of the MCM, consistent with previous assignments on the DNA-free wild-type CMGε and CMGε-Δcat complexes (Sun et al., 2015, Zhou et al., 2017). Although one domain of Polε-exo- is firmly anchored to the helicase, a second more flexible density can also be visualised from the 2D classes (appears as a blurry entity in lower-right panel classes of **Figure 5.5d**). To interpret the identity of these lobes, I carried out the same CMG fork affinity assay described in **Figure 5.5a** but with either wild-type Polε or Polε-Δcat. After single particle image processing, I aligned two representative classes from each of the CMGε and CMGε-Δcat datasets and compared them with CMGε-exo- (**Figure 5.5e**). This comparison revealed that the well-resolved lobe in CMGε-exo-

belongs to the non-catalytic domains of the leading strand polymerase (best appreciated by comparing CMGε-Δcat and CMGε-exo- side views, **Figure 5.5e**). Although a minority of particles in the CMGε-exo- sample show a second N-Pol2 lobe that is completely invisible in the wild-type CMGε complex, it is likely that this region will be averaged out in a 3D reconstruction due to its flexibility. Indeed, a crosslinked CMGε complex previously published by our lab shows the full two lobes in the structure whereas a non-crosslinked sample, as is the case here, only resolved the non-catalytic domains (Sun et al., 2015).

In summary, CMGε-exo- (herein referred to as CMGε) forms a complex on DNA that is amenable to structural characterisation. The non-catalytic domains of Polε are firmly attached to the CMG whereas N-Pol2 is flexibly tethered and may be refractory to structural characterisation.

5.2.4 Sub-nanometre cryo-EM structure of CMGε on fork DNA

To gain mechanistic insights into the replisome during replication fork establishment and understand the roles of the essential Pol ε domains in this context, a CMGε-DNA structure was determined by cryo-EM. To this end, CMGε was reconstituted on a pre-formed model replication fork as described in *section 5.2.3* (**Figure 5.5**). The eluted sample was applied twice to lacey grids containing a layer of ultrathin carbon (Agar) to enhance particle concentration before cryo-grid preparation (by Dr. Julia Locke). To obtain sufficient particle quantities, two separate datasets were collected on a Titan Krios (FEI ThermoFisher) equipped with a K2 summit detector operated in counting mode (Gatan). Movies were aligned and summed using MotionCor 2 (Zheng et al., 2017), followed by semi-automatic picking of particles and manual erasing of carbon edges by EMAN2 (Tang et al., 2007). All particles from the first dataset were initially subjected to 3D classification without a preceding 2D classification step in RELION-2 (Kimanius et al., 2016). After rounds of cleaning by 3D classification and refinement, and combining these particles with a subset of '2D cleaned' particles from a second dataset, a CMGε structure at 5.0 Å global resolution was obtained in cryoSPARC (Punjani et al., 2017) (**Figure 5.6-5.8**). The local resolution of the structure exhibited significant variation, ranging from 4.5 Å in the core of the MCM AAA+ tier to 6-7 Å at the peripheral extra density (**Figure 5.8e**).

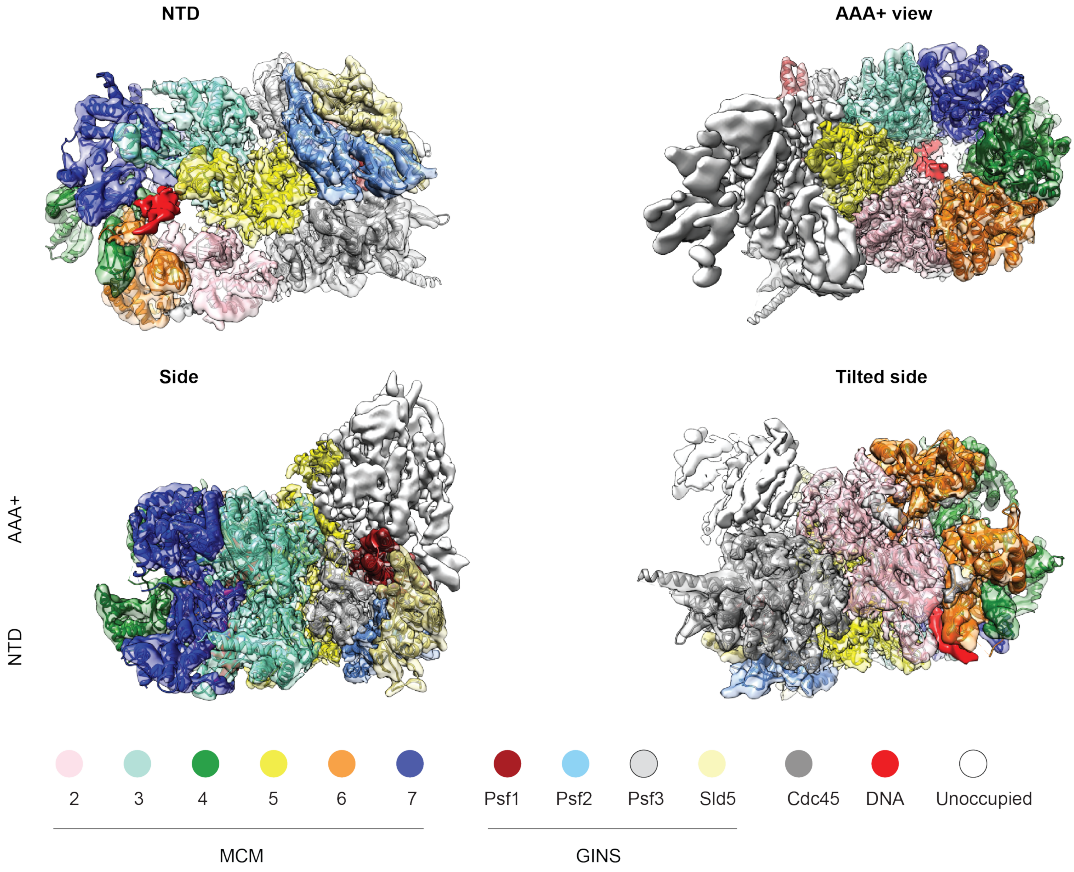
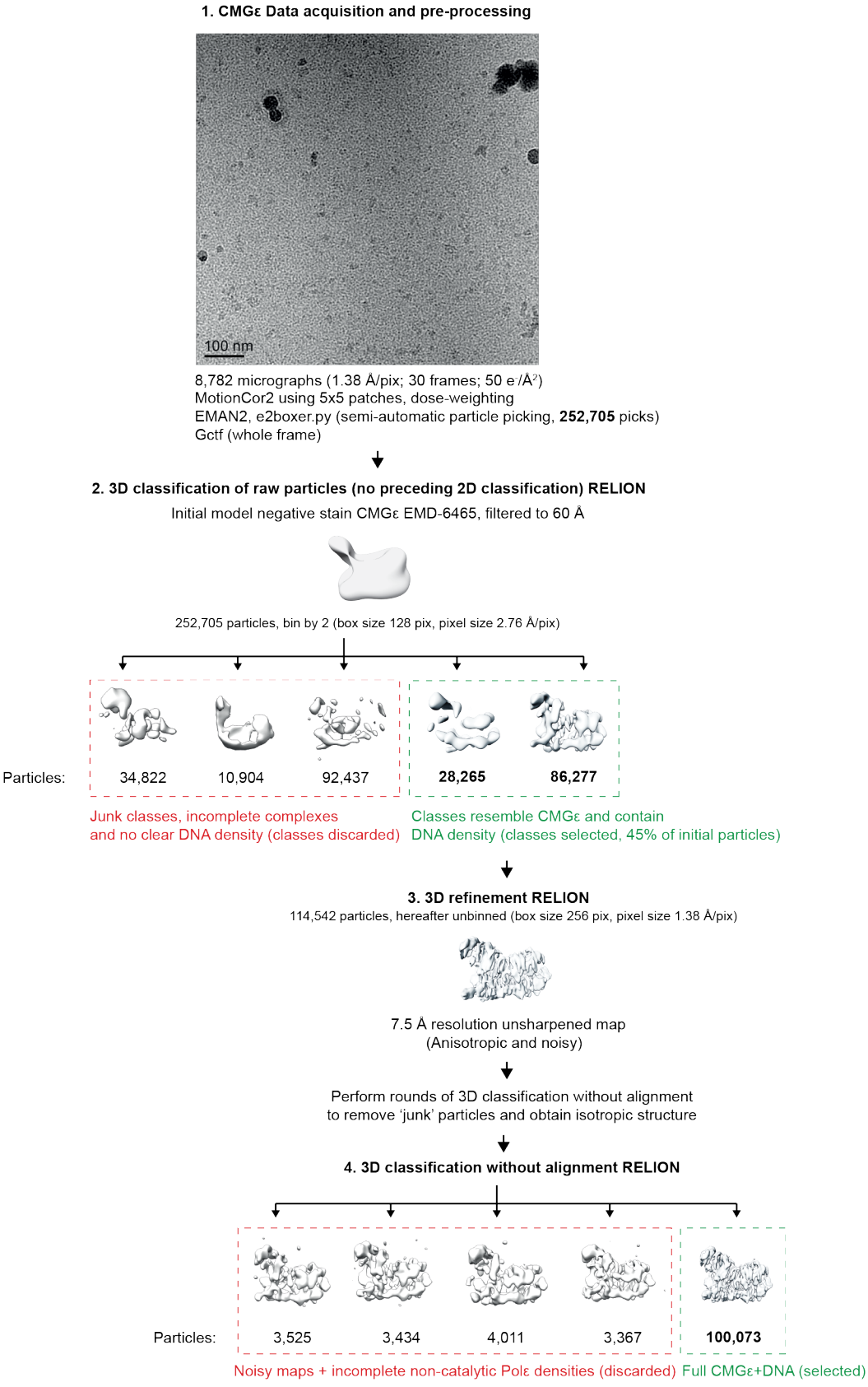


Figure 5.6: Cryo-EM structure of CMGε at 5.0 Å resolution
CMG real-space refined model was docked in to sharpened map; model based on PDB: 5u8s. White density indicates unoccupied regions, probably corresponding to the non-catalytic lobe of Polε.

However, this resolution range was sufficient for unambiguous docking of CMG sub-domains using atomic coordinates from a previously published CMG-DNA structure (PDB: 5u8s) (Georgescu et al., 2017) (**Figure 5.6**) and subsequent real-space refinement in PHENIX (Afonine et al., 2018) (by Dr. Andrew Purkiss). After assignment of all CMG subunits and DNA, a large unoccupied density decorating the AAA+ tier of the MCM was observed (**Figure 5.6**). This region was subsequently assigned to C-Pol2 and Dpb2 domains of the non-catalytic lobe of Polε as discussed further in *section 5.2.5* (Dpb3 and Dpb4 are flexible in this structure). As indicated by the negative stain 2D classes (**Figure 5.5d**) and predicted by not using crosslinking agents (Zhou et al., 2017, Sun et al., 2015), N-Pol2 is invisible in the CMGε structure. In the following sections, I will discuss the architecture of the non-catalytic lobe (*section 5.2.5*) before analysis of DNA and nucleotide binding by CMGε (*section 5.2.6)* and an in-depth overview of the CMG and Polε contacts (*section 5.2.7)*.


1. CMGε Data acquisition and pre-processing
100 nm
8,782 micrographs (1.38 Å/pix; 30 frames; 50 e⁻/Å²)
MotionCor2 using 5x5 patches, dose-weighting
EMAN2, e2boxer.py (semi-automatic particle picking, 252,705 picks)
Gctf (whole frame)
2. 3D classification of raw particles (no preceding 2D classification) RELION
Initial model negative stain CMGε EMD-6465, filtered to 60 Å
252,705 particles, bin by 2 (box size 128 pix, pixel size 2.76 Å/pix)
Particles:
34,822
10,904
92,437
28,265
86,277
Junk classes, incomplete complexes and no clear DNA density (classes discarded)
Classes resemble CMGε and contain DNA density (classes selected, 45% of initial particles)
3. 3D refinement RELION
114,542 particles, hereafter unbinned (box size 256 pix, pixel size 1.38 Å/pix)
7.5 Å resolution unsharpened map
(Anisotropic and noisy)
Perform rounds of 3D classification without alignment
to remove 'junk' particles and obtain isotropic structure
4. 3D classification without alignment RELION
Particles:
3,525
3,434
4,011
3,367
100,073
Noisy maps + incomplete non-catalytic Polε densities (discarded)
Full CMGε+DNA (selected)

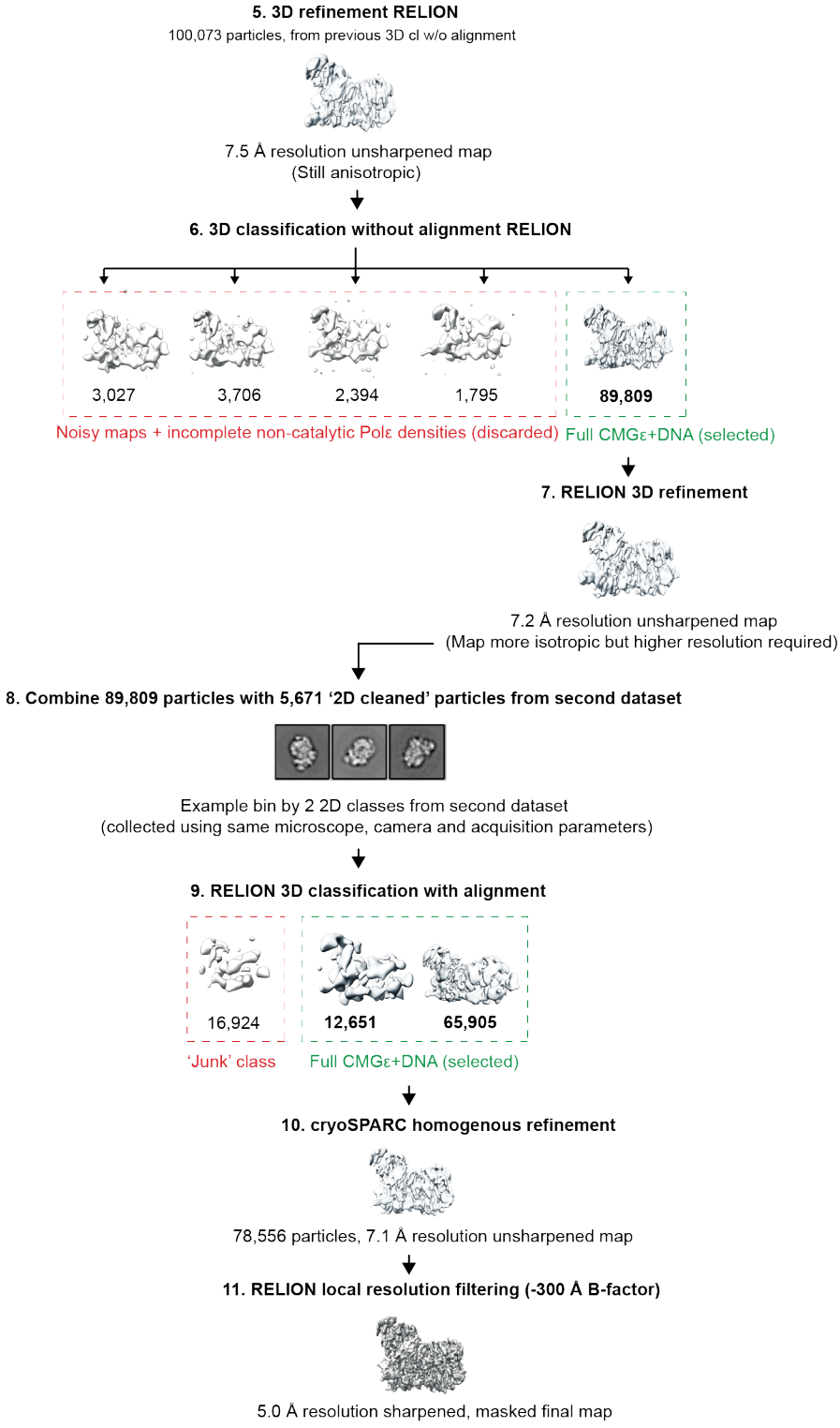


Figure 5.7: Single-particle CMG-DNA-Polε-ATPγS 3D reconstruction and classification overview
An initial dataset of CMGε was collected on a Titan Krios equipped with a K2 summit direct electron detector (counting mode). Particles were picked from non-dose weighted micrographs using EMAN2 e2boxer.py in a semi-automatic manner. Bin by 2 particles (2.76 Å/pixel size, 128 pixel box size) were extracted in RELION-2 from CTF corrected, dose-weighted micrographs (step 1). 3D classification was performed as a first particle cleaning step (step 2) and the resulting structures were inspected. Junk classes and classes with incomplete non-catalytic Polε density and no clear DNA inside the channel were discarded (red boxes). Only the two classes resembling the negative stain CMGε structure previously published (EMD-6465) and

containing DNA were selected (green box) for re-extraction to unbinned parameters (1.38 Å/pixel, 256 pixel box size). These particles were subjected to an initial 3D refinement round, which resulted in an anisotropic structure (step 3). Thus, rounds of 3D classification and refinement were performed to remove over-represented views and additional 'junk' particles in the dataset (steps 4-7). A second dataset was collected (3,095 movies) in an attempt to increase particle numbers contributing to the final structure. 5,671 particles from this second dataset was combined with the first dataset particles after rounds of 2D classification (step 8). Pooled particles were subjected to 3D classification with alignment (step 9) and the two classes resembling CMGε bound to DNA were combined for a final 3D refinement (step 10) with dynamic masking (cryoSPARC homogenous refinement). RELION local resolution filtering and sharpening (-300 Å B-factor) was applied yielding a structure at 5.0 Å resolution (gold-standard Fourier shell correlation, 0.143 criterion).

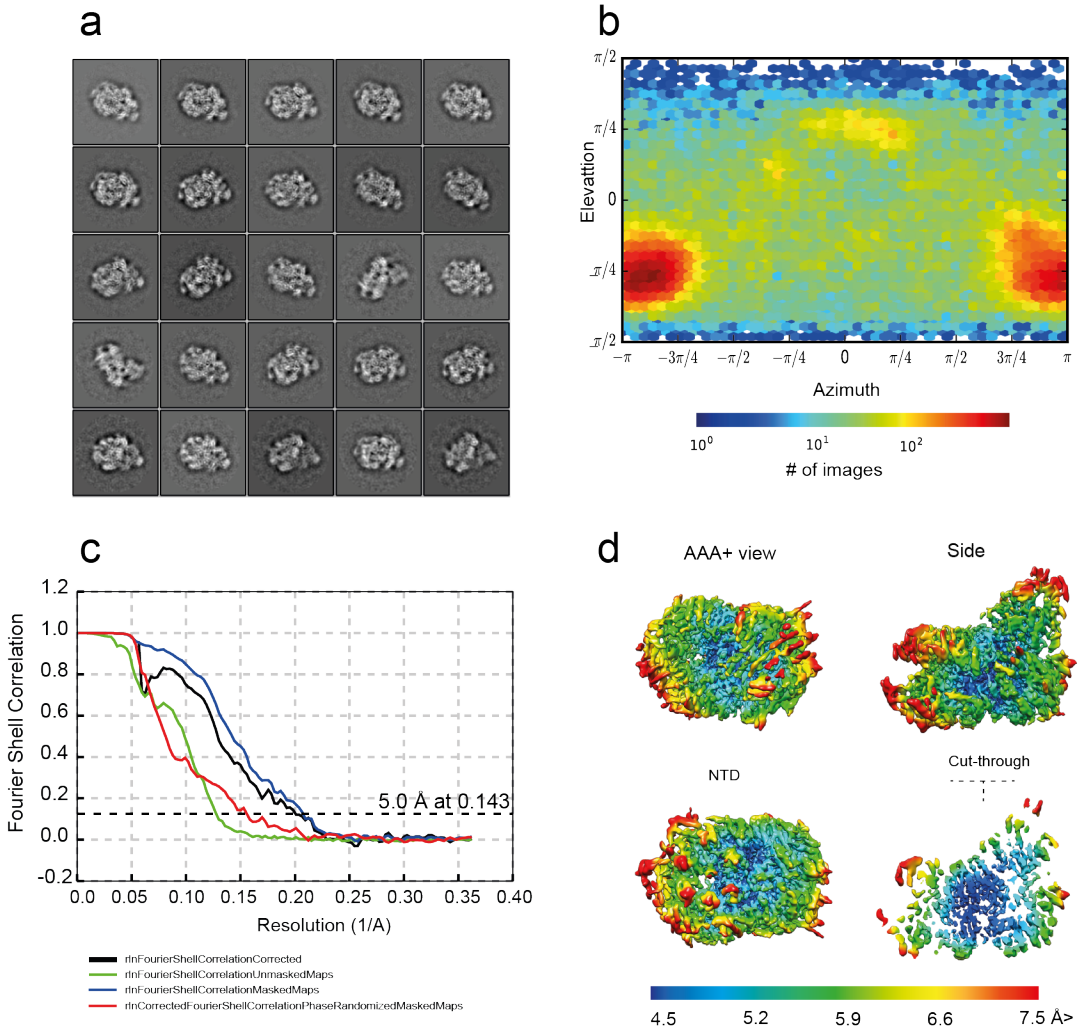


Figure 5.8: CMG-DNA-Polε-ATPγS cryo-EM, angular distribution and resolution estimation
(a) 2D classification in RELION-2 (Kimanius et al., 2016) of the 78,556 particles that contributed to the final refined structure. **(b)** Angular distribution. **(c)** Gold-standard Fourier Shell Correlation according to the 0.143 criterion. **(d)** Local resolution as estimated by RELION's LocRes function.

5.2.5 Structure of the essential, non-catalytic domain of Polε

To gain mechanistic insights into the non-catalytic but essential Polε domains (highlighted in **Figure 5.9a**), I docked in a recently determined atomic model of C-Pol2 bound to C-Dpb2 (obtained from a cryo-EM map of the isolated Polε-Δcat by Dr. Panchali Goswami) as a rigid body into the peripheral, additional density of the CMGε structure. This gave an unambiguous fit, showing that these non-catalytic domains do not undergo a conformational change upon CMG binding. **Figure 5.9b** shows three views of C-Pol2/C-Dpb2 bound to CMG, highlighting how Polε anchors the MCM at the AAA+ tier. Notably, a pore exists in between Polε and the NTD tier of MCM that could conceivably be bound by the flexible N-terminal Pol2 domain or other replisome components.

Segmentation of the C-Pol2/C-Dpb2 density allows a closer inspection of the inter subunit contacts within this lobe (**Figure 5.9c**). In line with a CMG anchoring role for the non-catalytic subunits of Polε (Zhou et al., 2017), the C-Pol2/C-Dpb2 density resembles an anchor shape, measuring at 100 x 90 Å (**Figure 5.9c**, left). C-Pol2 comprises a catalytically inactive DNA polymerase fold that was first predicted by phylogenetic analysis (Tahirov et al., 2009) and confirmed by the Polε-Δcat structure from our laboratory (Dr. Panchali Goswami, unpublished). The main C-Pol2 fold is followed by a ZF domain at the extreme C-terminus of Pol2, which stably binds C-Dpb2 (**Figure 5.9c**, right). In this arrangement, the ZF domain acts as bridge that connects the inactive polymerase fold of C-Pol2 to Dpb2. The importance of the zinc finger appendix is further underscored by mutagenesis of zinc-binding cysteines in the ZF1 element, which leads to abolished cell growth of yeast cells (Dua et al., 1999). Moreover, the architecture of C-Pol2/C-Dpb2 provides a rationale for why a deletion of the extreme C-terminal residues of Dpb2 is lethal and abolishes binding to the rest of Polε subunits (Isoz et al., 2012).

Analysis of the docked coordinates revealed an unoccupied density emanating from C-Dpb2 and extending towards the GINS subunits (**Figure 5.9c**, right). This region was previously unstructured and invisible in the isolated Polε-Δcat structure. Although the local resolution at this region is relatively low, docking in a yeast homology model of N-Dpb2 (based on an NMR structure of the human variant),

which comprises a globular 4-helix bundle (Nuutinen et al., 2008), yielded a good fit against the cryo-EM density. Therefore, CMG binding stabilises N-Dpb2, a critical region of Polε involved in helicase assembly and DNA replication initiation (discussed further in *section 5.2.7*) (Sengupta et al., 2013). Furthermore, the thin connecting density between the two Dpb2 domains rationalises why N-Dpb2 region is flexible in the absence of the helicase (Baranovskiy et al., 2017) (**Figure 5.9c**, right).

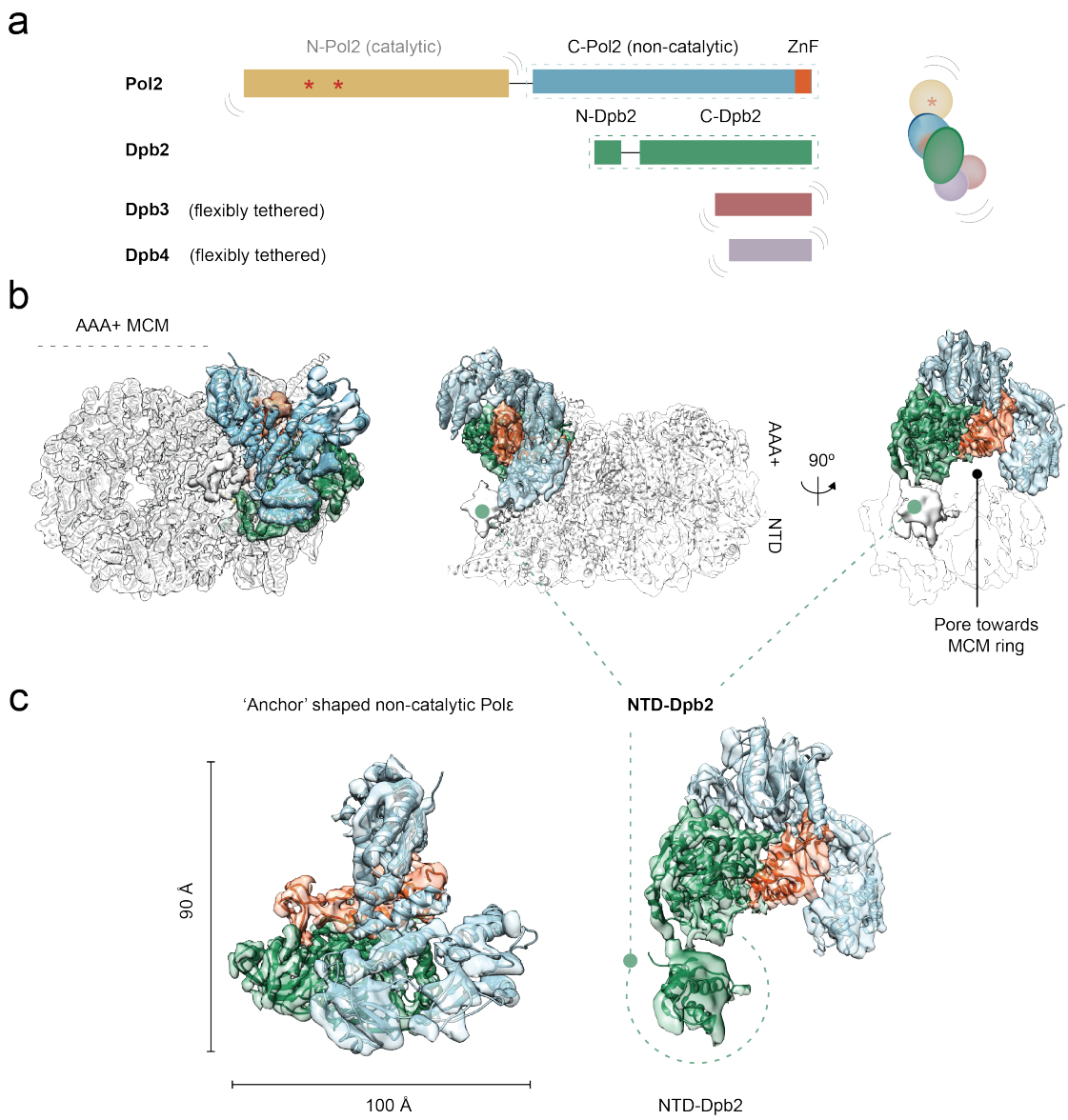


Figure 5.9: Structure of the non-catalytic lobe of Polε
(a) Domain architecture of full-length Polε. Flexible domains are greyed out whereas domains resolved in the CMGε structure are highlighted with a dashed box. Cartoon representation illustrates the rigid versus flexible elements. **(b)** Three views of CMGε where the CMG density and coordinates are greyed out. The real-space refined model of C-Pol2/C-Dpb2 (by Dr. Panchali Goswami and Dr. Andrew Purkiss) was docked in leaving behind an unoccupied density. **(c)** Segmented density corresponding to the Polε C-Pol2/C-Dpb2 (left) and C-Pol2/Dpb2 (right). The unoccupied density highlighted in (b) fits the coordinates of the N-Dpb2 NMR structure (PDB: 2v6z).

5.2.6 DNA binding and nucleotide occupancy of CMGε

Several structural studies have characterized DNA binding within the isolated CMG (*see section 4.2.6*) (Georgescu et al., 2017) but how CMGε binds to a replication fork is completely unknown. Close inspection of the CMGε map reveals the presence of both duplex and single-stranded DNA within the complex (**Figure 5.10**). Single-stranded DNA binds the highly conserved PS1 loop spiral staircase of Mcm6-2-5-3 at the AAA+ tier domains (**Figure 5.10a**) whereas duplex DNA is contacted by the NTD tier in close proximity to the OB-fold loops of Mcm7-4-6 (**Figure 5.10b**). Although the lagging strand template cannot be resolved, the fork nexus is held down by the Mcm4 OB fold, which may constitute the pivot at which the two strands split.

These DNA binding regions are similar to previous observations of a CMG-ATP-fork DNA structure (Georgescu et al., 2017), although in CMGε there is a slight rotation of the double stranded portion of the DNA as the leading strand template enters the channel. Importantly, DNA does not bind the C-Pol2/Dpb2 modules, in line with unpublished gel shift assays in the lab (by Dr. Panchali Goswami). The observation of the duplex DNA junction in the structure was initially surprising as a previous report indicated that non-hydrolysable ATP conditions prevents successful fork engagement (see *section 4.2.6,* (Georgescu et al., 2017)). Indeed, the only other time the fork nexus had been visualised was when isolated CMG was incubated with ATP and roadblocked fork DNA (Georgescu et al., 2017). Collectively, this argues that Polε promotes a conformational change in the CMG that leads to productive fork binding.

In support of Polε mediated CMG stabilisation, the local resolution at the interfaces which are contacted by C-Pol2 and Dpb2 (Mcm2-5-3) is much higher compared to the looser interfaces in the rest of the MCM ring (**Figure 5.11a**). In line with such a structural asymmetry in the MCM ring of CMGε, nucleotide density corresponding to ATPγS, was found at the Mcm2-5 and Mcm5-3 interfaces but not in the other half of the MCM ring (modelling and real-space refinement of atomic coordinates by Dr. Andrew Purkiss) (**Figure 5.11b**). In contrast, both the isolated CMG bound to single-stranded DNA-AMPPNP and CMG bound to forked DNA-ATP showed an additional site (Mcm6-2) to be nucleotide occupied (Georgescu et al., 2017).

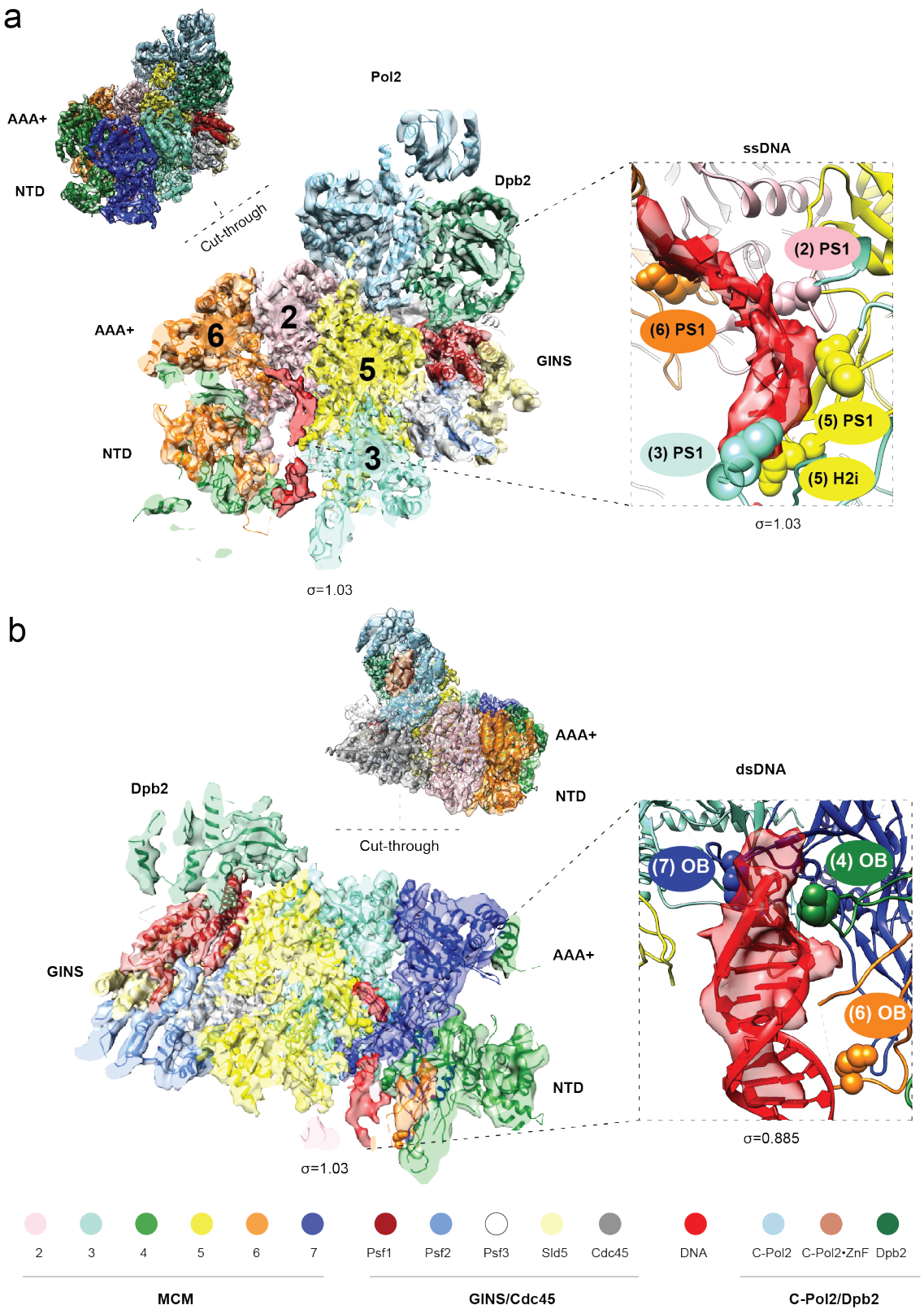


Figure 5.10: DNA binding in CMGε
(a) Focus on the single-stranded DNA resolved in the CMGε complex. A zoom in shows that single-stranded DNA is bound by the PS1 spiral of Mcm3 (K499)-6 (K665)-2 (K633)-5 (K506) and also by the H2i of Mcm5 (R460). **(b)** Focus on the fork nexus of DNA. A zoom in at slightly lower contour level than the middle panel shows that the duplex junction is in close proximity to the OB folds of Mcm7 (K364), 4 (K449) and 6 (K428).

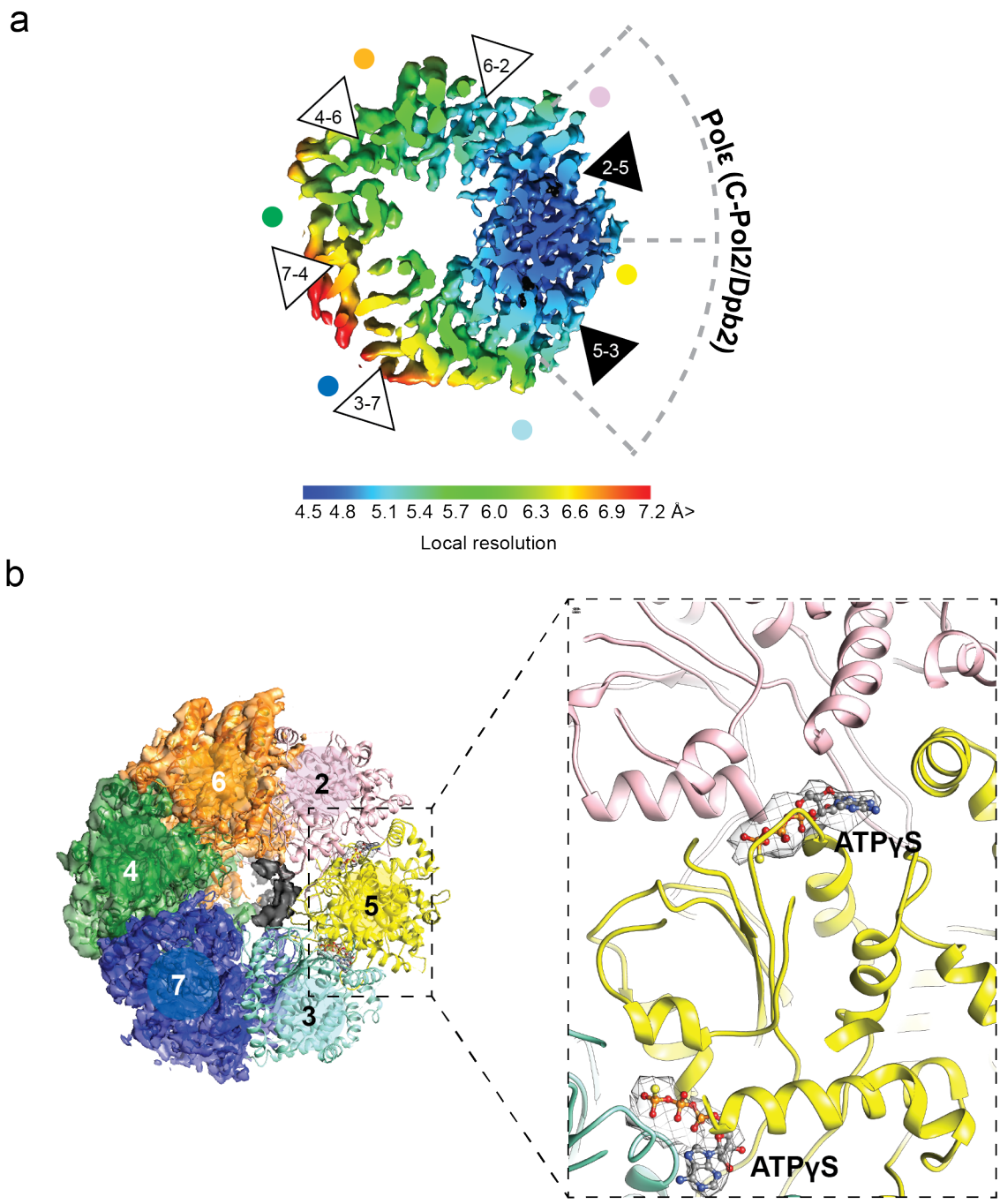


Figure 5.11: Polε mediated compaction of MCM ring and nucleotide occupancy in CMGε
(a) Local resolution estimation shows that the Mcm2-5-3 interfaces at the AAA+ tier are higher where C-Pol2/Dpb2 contacts the MCM ring compared to the opposite face of the ring. **(b)** Nucleotide occupancy in the AAA+ tier. The MCM subunits that are contacted by Polε are also where density that fits ATPγS is found (modelling of nucleotide by Dr. Andrew Purkiss).

A striking conformational change that may further promote productive fork engagement of the CMG by Polε is the WHD of Mcm5. Whereas in the CMG-DNA (*section 4.2.2*) and apo-CMG structures (Yuan et al., 2016) Mcm5 WHD partially occludes the channel, in the CMGε structure this region is captured by C-Pol2 and

therefore lifted away from where DNA would exit the AAA+ pore (the interactions of CMG with C-Pol2 and Dpb2 is discussed in further detail in *section 5.2.7*). Moreover, the train experiments performed in *section 4.2.10* using the *in vitro* reconstitution of origin firing suggest that Polε remains stably bound to CMG during translocation. Single particle image analysis from two different repeats showed Polε bound to a translocating CMG in 75-100% cases. However, whether Polε directly stimulates the ATPase activity of CMG remains to be demonstrated.

In summary, CMGε binds single-stranded DNA inside the channel around highly conserved ATPase pore loops and the duplex DNA junction is visualized for the first time using ATPγS conditions. One of the roles of C-Pol2/Dpb2 may be to compact the MCM ring and seize the Mcm5 WHD away from the central channel to enable productive fork engagement.

5.2.7 Analysis of the CMGε interaction network

To investigate the spatial relationship between CMG and C-Pol2/Dpb2, I mapped all of the interaction points within the complex. Collectively, C-Pol2/Dpb2 interact extensively with both the MCM and GINS subunits of the CMG (**Figure 5.12**). A notable contact is N-Dpb2 binding the B-domain of Psf1 (**Figure 5.12a**). This interaction, shown in more detail in **Figure 5.12b**, was previously identified in a biochemical study, which also characterised the importance of the Psf1-Dpb2 relationship in helicase assembly and replication initiation (Sengupta et al., 2013). In this context, it is interesting to note that both the Psf1 B-domain and N-Dpb2 are flexible as isolated entities unless they are incorporated into the CMG (Chang et al., 2007, Choi et al., 2007, Kamada et al., 2007, Baranovskiy et al., 2017).

A second set of contacts occur between C-Pol2/Dpb2 and the AAA+ tier of the MCM ring. As alluded to before, in CMGε, the Mcm5 WHD is moved away from its MCM channel restrictive position in apo-CMG and CMG-DNA (Yuan et al., 2016)(see *section 4.2.2*). This conformational change is promoted by the firm brace of Mcm5 WHD in between the C-Pol2 inactive polymerase domain and the C-Pol2 ZF appendix (**Figure 5.12c**).

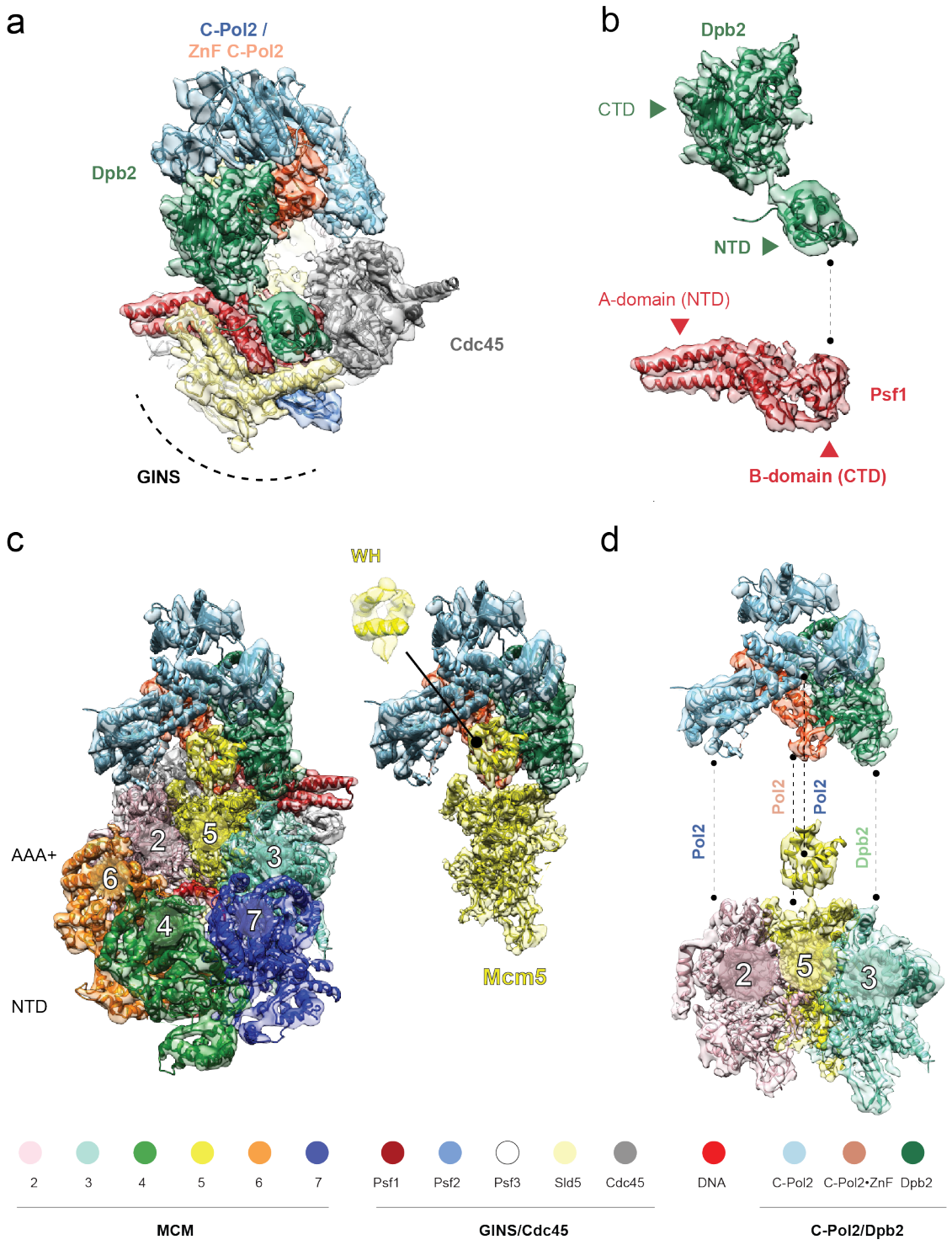


Figure 5.12: Analysis of the CMGε interaction network
(a) N-Dpb2 extends towards and contacts GINS. **(b)** N-Dpb2 contacts the Psf1 B-domain. **(c)** Mcm5 WHD stably binds in between C-Pol2 and Dpb2. **(d)** C-Pol2 and Dpb2 stabilise the Mcm2-5-3 AAA+ interfaces, forming four main contacts.

Together with the ZF appendix also contacting the Mcm2 AAA+ subdomain and the inactive polymerase C-Pol2 fold contacting the Mcm5 AAA+ module, this serves to firmly shut the Mcm2-5 gate at ATPase tier (**Figure 5.12d**). Further stabilization of the MCM ring is mediated by C-Dpb2 interacting with the Mcm3-5 AAA+ domains (**Figure 5.12d**). Therefore, akin to Cdc45 and GINS keeping a rigid brace of the NTD Mcm2-5-3 sites, Pol2 and Dpb2 maintain a closed configuration of the same MCM subunits but at the AAA+ tier. Furthermore, this constitutive binding mode of C-Pol2/Dpb2 to the CMG (at least in the ATPγS condition) might explain why only a closed conformer of the CMGε complex could be identified (**Figure 5.7**) whereas the isolated CMG-DNA dataset comprised a subset of particles with an open Mcm2-5 ATPase gate (*see section 4.2.5*). Additionally, by interacting with three AAA+ modules, C-Pol2/Dpb2 may be responsible for the observed nucleotide exchange step (ADP release to ATP binding) that occurs during CMG assembly (Douglas et al., 2018).

5.2.8 A model for the role of C-Pol2 and Dpb2 in CMG formation

GINS and Polε are recruited together onto the origin as part of the same pre-LC complex and in this context, GINS recruitment is dependent on Polε (Muramatsu et al., 2010). In the CMGε structure, C-Pol2 mainly interacts with the MCM ring whereas Dpb2 mainly contacts GINS (**Figure 5.12**). Such interactions could be key to bridging the GINS co-activator with the helicase motor during CMG assembly. Analysis of the CMGε interaction network raised the hypothesis that contacts between GINS and MCM mediated by C-Pol2/Dpb2 reflect the minimal set of contacts required for CMG assembly during origin activation. To test the hypothesis that C-Pol2/Dpb2 are sufficient for helicase assembly, CMG complexes were assembled by *in vitro* reconstitution using purified proteins in the presence of various Polε deletion mutants (experiments by Dr. Max Douglas, Polε purification by Dr. Panchali Goswami). After loading MCM onto immobilised DNA and adding in the full complement of origin firing factors, it was observed that full-length Pol2, C-Pol2 or Dpb2 subunits in isolation were insufficient to support CMG formation. In contrast, wild-type Polε supported full assembly as did C-Pol2/Dpb2 and Polε-Δcat, albeit less efficiently (**Figure 5.13a**). This result indicates that C-Pol2/Dpb2 can promote CMG assembly in the absence of N-Pol2 and the ancillary Dpb3/4 subunits. The CMG assembled by these various

deletion mutants were subsequently tested for DNA replication by supplementing the full complement of proteins required for leading and lagging strand synthesis (Yeeles et al., 2017). As shown in **Figure 5.13b**, only wild-type Polε, C-Pol2/Dpb2 and Polε-Δcat supported DNA replication in this system. In the case of C-Pol2/Dpb2 and Polε-Δcat, where the catalytic Polε subunit is absent, a *trans* acting Polδ takes over leading strand synthesis but the products are much shorter, in line with the notion that N-Pol2 is required to for efficient and accurate DNA replication (Yeeles et al., 2017). Collectively, these results show that C-Pol2/Dpb2 are essential and sufficient for CMG formation and that such a complex could support DNA replication, albeit less efficiently than the wild-type Polε control.

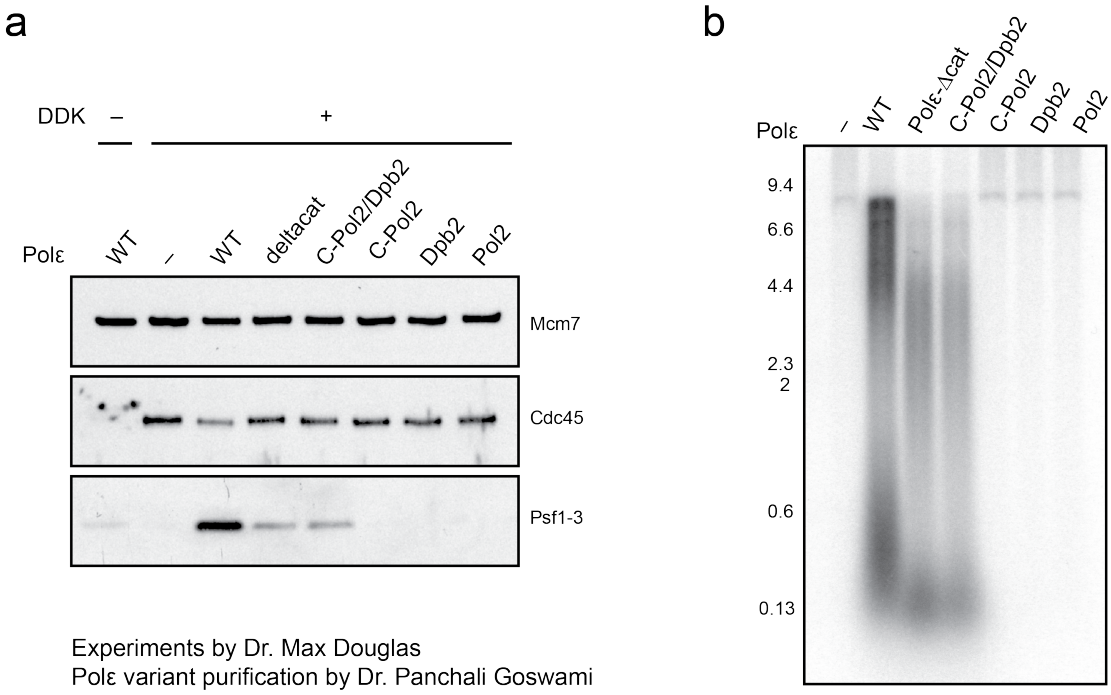


Figure 5.13: C-Pol2/Dpb2 are the minimal Polε subunits required for CMG assembly and replication
(a) CMG assembly reconstitution *in vitro* shows that wild-type Polε, Δcat and C-Pol2/Dpb2 constructs can all support recruitment of GINS although the latter two Polε variants are less efficient. Assay: first, MCM was loaded on biotinylated 2.8 kb linear DNA containing an ARS1 site (60 ng) immobilized on M-280 streptavidin beads. MCM was loaded by adding 37.5 nM ORC, 50 nM Cdc6, and 100 nM Mcm2-7/Cdt1 in buffer containing 25 mM Hepes pH 7.6, 250 mM K-glutamate, 10 mM Mg-acetate, 0.02% NP-40S, 8% glycerol, 400 µg/ml BSA, 5 mM ATP and 1 mM DTT in a 10 µl volume (10 minute incubation at 30ºC). DDK was subsequently added to a 50 nM concentration and reaction was incubated for 5 minutes. Firing factors were mixed (50 nM Dpb11, 50 nM GINS, 50 nM Cdc45, 30 nM Polε, 20 nM Clb5-Cdc28, 2.5 nM Mcm10, 30 nM Sld3/7 and 55 nM Sld2) and added for an incubation of 10 min at 30ºC. Beads were washed twice in buffer (25 mM HEPES 7.6, 5 mM Mg-acetate, 0.02% NP-40-S, 10% glycerol, and 250 mM K-glutamate) before Western blot analysis. **(b)** DNA replication *in vitro* shows that C-Pol2/Dpb2 and C-Pol2 are sufficient for DNA replication but wild-type Polε is required for complete leading and lagging strand synthesis. 4 nM 8.2 kb ARS1 containing template was used for all reactions. MCMs were loaded for 20 min at 30ºC (5 mM ATP, 45 nM Cdc6, 22.5 nM ORC, 100 nM Cdt1-Mcm2-7 in buffer containing 25 mM Hepes-KOH pH 7.6, 100 K-glutamate, 10 mM Mg-acetate, 100 µg/ml BSA, 1 mM DTT, 0.01% NP40S), DDK was added to 20 nM and reactions were incubated for 20 min. Nucleotide was added to 200 µM CTP, GTP and UTP as well as 80 µM dCTP, dATP and dTTP and 33 nM α32P-dCTP before addition of replication factors 25 nM Sld3/7, 20

nM S-CDK, 50 nM Sld2, 30 nM Dpb11, 210 nM GINS, 40 nM Cdc45, 20 nM Polε [either wild-type, drop out variant or mutant Polε], 5 nM Mcm10, 20 nM RFC, 10 nM PCNA, 10 nM Topo I, 20 nM Topo II, 20 nM Ctf4, 100 nM RPA, 20 nM Polα, 10 nM Mrc1 and 10 nM Polδ). Following a 40 minute incubation, reactions were quenched using EDTA and the products were analysed using a 0.6% denaturing alkaline agarose gel. See sections *2.4.5-6* for detailed protocols. Data in this figure is from Drs. Max Douglas and Pannchali Goswami.

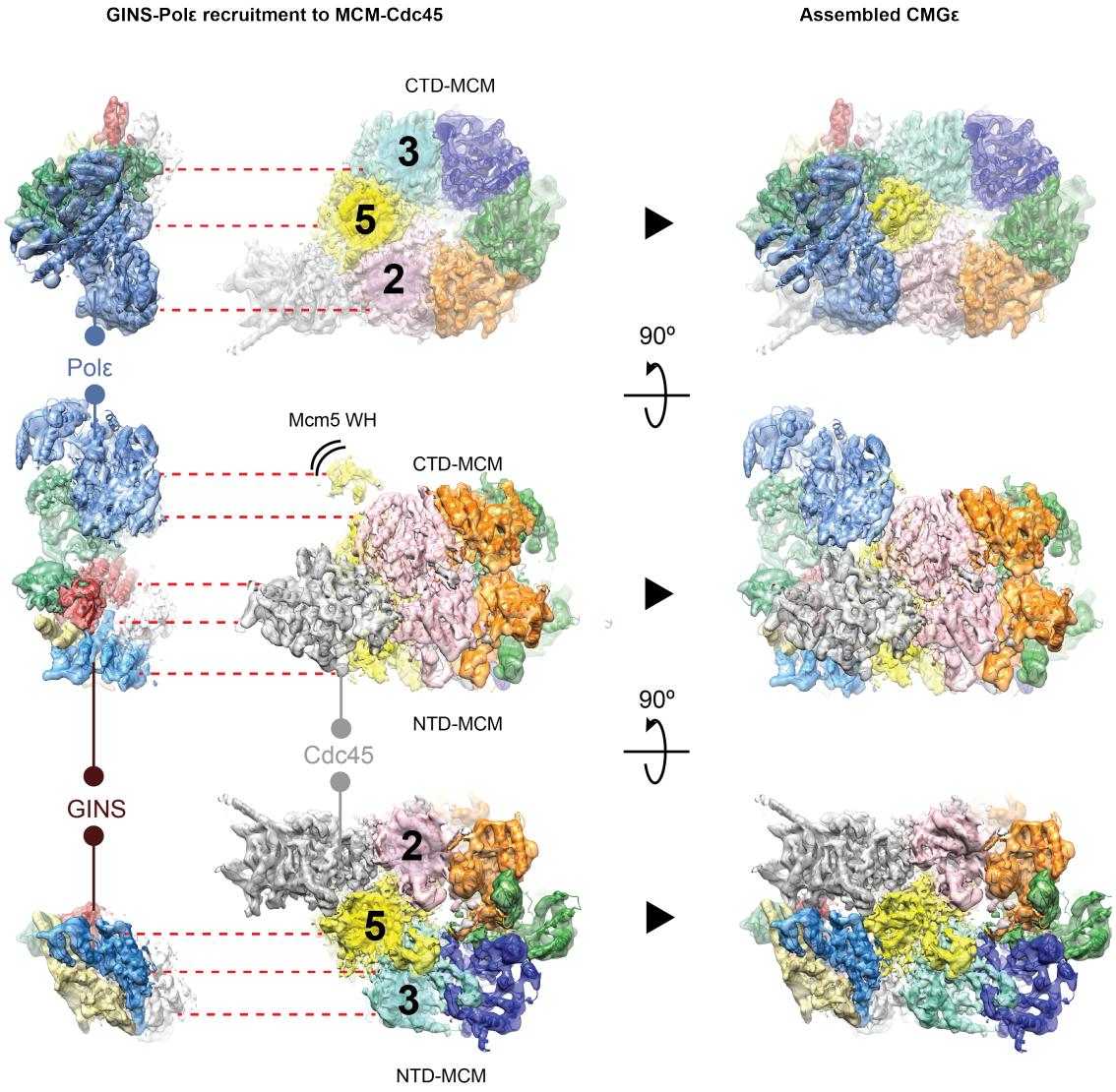


Figure 5.14: Proposed model for the concerted GINS and Polε incorporation onto MCM-Cdc45
Cdc45 and GINS binding stabilises the NTD Mcm2-5-3 interfaces whereas C-Pol2 and Dpb2 binding stabilises the AAA+ tier. The biochemical and structural interaction of N-Dpb2 with Psf1 B-domain is compatible with the simultaneous recruitment of GINS and Polε to assemble and eventually activate the replicative helicase.

Together, the cryo-EM structure and the biochemical findings by my colleagues allow a model to be proposed for C-Pol2/Dpb2 in CMG formation and replication fork establishment (**Figure 5.14**). After DDK dependent and Sld3/7 mediated Cdc45 recruitment, GINS and Polε are recruited together to MCM-Cdc45 as part of the

fragile pre-LC complex, that also comprises Sld2, Dpb11 and depends on CDK phosphorylation (Muramatsu et al., 2010). In this complex, GINS-Polε are tethered together by the N-Dpb2-Psf1 interaction (Sengupta et al., 2013). A simultaneous incorporation of GINS-Polε into the CMG would cause double hexamer destabilisation as shown in *sections 3.5.2* and *4.2.4* due to the Psf2-Mcm5 A-domain interaction. Bridging of C-Pol2/Dpb2 to the hexamer is mainly strengthened by the C-Pol2 interactions at the AAA+ tier, which keeps a compact conformation of the MCM motor to enable productive replication fork engagement. Together, C-Pol2/Dpb2 and Cdc45-GINS rigidly brace the Mcm2-5-3 sites at the AAA+ and NTD tiers, respectively, to reconstitute key MCM sites which are essential for helicase activity (Ilves et al., 2010) and topologically restrain the DNA from escaping the channel. However, several factors, including Sld3/7, Sld2 and Dpb11 are missing from this structural model. As such, a complete view of Polε-dependent GINS recruitment can only be obtained once CMG has been assembled step-wise using the full complement of origin firing factors to recapitulate the events that occur *in vivo*.

5.3 Conclusions

In this chapter, I adopted a combined biochemical and cryo-EM strategy to structurally characterise a CMG-Polε complex bound to fork DNA. The key findings from this study corroborate that Polε has a dual role in the replisome (Sengupta et al., 2013). In the first role, C-Pol2/Dpb2 bridges the GINS-MCM interaction, which could be critical for incorporation of GINS into MCM-Cdc45 (**Figure 5.12**) and to promote nucleotide exchange during CMG assembly (Douglas et al., 2018). In a second role, C-Pol2/Dpb2 stabilises the two essential Mcm2-5 and Mcm5-3 ATPase sites in the MCM motor, which may be responsible for the productive replication fork engagement observed (**Figure 5.10**). It remains to be established whether Polε directly stimulates CMG for translocation although the train experiments in *section 4.2.10* indicate that the translocating helicase is almost always Polε bound.

The isolation of only one compact CMGε conformer indicates a constitutive binding mode by C-Pol2/Dpb2. In contrast, the CMG-DNA structures presented in Chapter 4 and apo-CMG show two different configurations of the Mcm2-5 sites. How CMGε unwinds DNA awaits further studies where ATP is used. However, the CMGε

structure suggests that large Mcm2-5 re-configurations could disrupt C-Pol2/Dpb2 binding. Therefore, it is more likely that unwinding involves subtle conformational changes of the DNA-binding pore loops and rotation of the DNA fork (as observed when comparing isolated CMG-DNA to CMGε-DNA) instead of the previously proposed nodding pump-jack mechanism (discussed further in Chapter 6).

Chapter 6. Discussion

6.1 Introduction

Eukaryotic origin activation occurs in three distinct stages: origin licensing (Chapter 3), helicase assembly (Chapter 3-4) and replication fork establishment (Chapter 4-5). During the G1 phase of the cell cycle, the motor of the replicative helicase, MCM, is loaded onto origin DNA as an inactive head-to-head double hexamer (Remus et al., 2009, Evrin et al., 2009). During S-phase, origin firing factors, including the protein kinases DDK and CDK, and co-activators Cdc45, GINS and Polε convert each MCM complex into two assembled CMG holo-helicases (Yeeles et al., 2015). Each CMG is subsequently activated by Mcm10 to hydrolyse ATP for bi-directional fork unwinding (Yeeles et al., 2015, Douglas et al., 2018). At the replication fork, the CMG forms the core of the replisome (Gambus et al., 2006), which contains all of the proteins required for fast and accurate DNA replication, including Polα-primase, Polδ, Polε, Mrc1, Csm3/Tof1, PCNA, FACT and RPA (Yeeles et al., 2017, Kurat et al., 2017). Although CMG activity is fundamental to DNA replication initiation and progression, the molecular mechanism of helicase activation is still unclear.

6.2 Key conclusions

During my PhD studies, I studied several helicase intermediates by EM to investigate the molecular basis for CMG activation. To this end, I first focused on characterising the inactive licensed origin and then structurally characterised the conformational changes that occur during CMG assembly and replication fork establishment. I interchangeably analysed MCM complexes in different stages of replication initiation either using the *in vitro* reconstitution assay of origin firing (in collaboration with Dr. Max Douglas from Dr. John Diffley's laboratory) or by purifying and reconstituting active DNA bound CMG complexes via overexpression systems.

In Chapter 3, I determined the structure of the budding yeast double hexamer bound to DNA. Due to the tilted and offset nature of the MCM-MCM interaction, I showed that the double hexamer binds bent but intact double-stranded DNA (**Figure 3.12**). This goes against a previous hypothesis, which predicted origin melting to occur at

this stage (Li et al., 2015). I subsequently mapped all of the DNA-MCM interaction points, showing that each hexamer binds both the Watson and Crick strands equally (**Figure 3.12**). This could represent a mechanism to prevent premature directional translocation alongside the firm binding of MCM to ADP (discussed further in *Section 6.3*).

One of the earliest events in MCM activation is the DDK mediated recruitment of Sld3/7 to licensed origins (Heller et al., 2011, Tanaka et al., 2011b). This step enables incorporation of the helicase co-activator Cdc45 onto MCM via Sld3/7 (Deegan et al., 2016). To better understand the function of DDK in this process, I determined the cryo-EM structure of the DDK phosphorylated MCM double hexamer. This structure showed no conformational changes in the core MCM particle relative to the unmodified double hexamer but revealed a symmetrical density at lower contour level emanating from the flexible N-terminal tails of Mcm4, 6. As discussed in *section 6.4*, this could be used as a structural platform for a Sld3/7 dimer of dimers to recruit two Cdc45 molecules.

Using the reconstituted *in vitro* origin firing assay and negative stain EM, I was then able to provide the first structural characterisation of fired origins. This assay also allowed me to substantiate that assembly and activation of the CMG are two separate events. In this context, GINS recruitment to MCM-Cdc45 leads to double hexamer separation irrespective of Mcm10. Furthermore, assembled and inactive CMGs separate by hundreds of base pairs, likely by passive diffusion.

In Chapter 4, I went on to analyse the structure of *D. melanogaster* CMG bound single-stranded DNA, which enabled insights into how the MCM motor gets activated by Cdc45 and GINS. Furthermore, I identified single-stranded DNA inside the CMG channel, which provided evidence for a strand exclusion mechanism for unwinding. Understanding the structure of a catalytically active CMG allowed modelling of the protein re-configurations that cause origin melting upon switching from the inactive (double hexamer, Chapter 3) to active MCM (CMG).

The final sections of this study are focussed on replication fork establishment. In particular, I addressed a debate concerning the polarity of the translocating CMG at

a replication fork (in collaboration with Dr. Max Douglas). I showed using negative stain EM that CMG actively pushes against a row of MCM 'trains' which have been prevented from falling off the DNA by covalently linked roadblocks. This showed that the CMG is oriented NTD tier first at the fork junction and has several important implications for replication initiation and replisome organisation (discussed in *section 6.5.5*).

In Chapter 5, I presented the structure of CMGε bound to fork DNA. This structure provided insights into the spatial relationship between the non-catalytic domains of Polε and the CMG, revealing how the leading strand polymerase stabilises the ATPase modules of MCM. A speculative model was further proposed for GINS recruitment to MCM-Cdc45 mediated by Polε based on structure-guided *in vitro* CMG assembly assays (discussed further in *sections 6.4.3* and *6.5.6*).

In the following sections I will discuss these results in further detail in the context of previously published literature before proposing a provisional model for helicase activation.

6.3 Origin licensing

6.3.1 Structure of the licensed origin

In eukaryotes, the temporal separation of helicase loading, where the MCM is intrinsically inactive, and origin firing, where the CMG readily binds and hydrolyses ATP, is imperative to ensure that replication strictly occurs only once per cell cycle (Diffley, 2001). In order to understand how to activate the helicase, an understanding of the post-catalytic inactive MCM state is required. Although the minimal components required for MCM double hexamer formation have been known for several years (Remus et al., 2009, Evrin et al., 2009), the reasons a licensed origin is inactive have remained unclear. Results presented in Chapter 3, combined with two recent cryo-EM studies of the MCM double hexamer both with and without DNA (Li et al., 2015, Noguchi et al., 2017) are beginning to provide the mechanism for restricted Mcm2-7 activity during origin licensing.

A striking feature of the double hexamer architecture is the tilted conformation of the complex both at the MCM-MCM interface and between the NTD and AAA+ tiers in a single hexamer (Sun et al., 2014, Li et al., 2015, Noguchi et al., 2017, Abid Ali et al., 2017). In this arrangement, the double hexamer interface is formed by a network of Zf domains that assemble into an inter-digitated, staggered brace at the N-terminal tiers. Collectively, this causes a 14º offset of the two rings and a left-handed twist of the AAA+ modules relative to the NTD domains within each hexamer (**Figure 3.1**) (Li et al., 2015).

In contrast to the eukaryotic MCM double hexamer, the AAA+ homo-hexameric archaeal MCM (Gomez-Llorente et al., 2005, Costa et al., 2006, Fletcher et al., 2005) and viral L-Tag (Valle et al., 2000) double hexamers are arranged into near co-axial structures along the vertical axis and are loaded as active assemblies (Chong et al., 2000, Smelkova and Borowiec, 1997). It is likely that the unique features of each MCM subunit within the Mcm2-7 hetero-hexamer mediate higher levels of regulation to ensure distinction of initiation and activation events during DNA replication. In line with this, eukaryotic MCM proteins comprise extensions at both the N-terminal and C-terminal domains that may promote the 'off' conformation of the double hexamer (Li et al., 2015). For example, MCM4 and MCM6 contain a long linker stretch between the OB-fold and AAA+ domain that may induce the shift seen between the two tiers of a single hexamer. Similarly, the unstructured and disordered N-terminal loops of Mcm2, 4 and 6 may have a role in promoting the tilt at the double hexamer interface (Noguchi et al., 2017, Li et al., 2015). Indeed, our structure of the DDK phosphorylated MCM corroborates this notion, showing a flexibly tethered density that probably corresponds to the N-terminal tails of Mcm4 and Mcm6 adjacent to the tilted double hexamer interface (**Figure 3.7**).

Importantly, some of these eukaryotic specific MCM insertions are involved in inter-hexamer contacts to strengthen the inactive double hexamer configuration. This is exemplified by Mcm7, which contains a yeast specific long insertion of approximately 70 residues that folds into a helix-turn-helix emanating from the A-domain. The peripheral Mcm7 alpha helix interacts with the A-domain of Mcm5 from the opposing ring, which is the same Mcm5 site that Psf2 binds to during origin firing factor recruitment (**Figure 3.1c**, inset and **Figure 4.12**). As such, this may explain why the

double hexamer only dissociates upon GINS recruitment, the second co-activator essential for CMG assembly (discussed further below).

A key question is whether the tilted double hexamer and left-handed twist of the AAA+ domain relative to the NTD inactivates the ATPase activity of MCM. ATP hydrolysis experiments of single MCM hexamers versus double hexamers have shown a stark reduction in ATPase activity of the latter (10 pmol/min vs 1.5 pmol/min, respectively) (Sun et al., 2014). One hypothesis to explain this observation is that uncoupling of the ATP hydrolysis motifs (Walker A, Walker B and arginine finger) occurs in the double hexamer due to the aforementioned shift in the AAA+ tier. In line with this idea, catalytically active intermediates OCCM and CMG exhibit near co-axial tiers (Yuan et al., 2016, Yuan et al., 2017).

Alternatively, it could be that the double hexamer form is inherently locked in a compact state which prevents the continuum of conformations required for cycles of ATP binding and hydrolysis. Consistent with this notion, the AAA+ interfaces are tightly packed, where the largest gap between two subunits is 4 Å (Li et al., 2015). Although ATP hydrolysis is essential for double hexamer loading onto origin DNA (Coster et al., 2014), once formed the MCM remains stably bound to ADP (Noguchi et al., 2017, Li et al., 2015) and is refractory to binding or hydrolysing additional ATP in a nucleotide exchange experiment (Douglas et al., 2018). In agreement with these observations, our double hexamer structure is found to bind tightly to nucleotide (most likely ADP) at the Mcm2-5 and Mcm3-5 sites despite extensive high salt washes using nucleotide-free buffers (**Figure 3.1e**).

Aside from nucleotide occupying Mcm2-5 and Mcm5-3 in the double hexamer, the rest of the Mcm3-7, 7-4 and 4-6 sites are nucleotide free. Mutagenesis studies in *D. melanogaster* CMG have shown that Mcm2-5 and Mcm3-5 are indispensable for helicase activity whereas the rest of the sites can be mutated with minimal effect (Ilves et al., 2010). However, the Mcm3-7, 7-4 and 4-6 sites are required for double hexamer formation (Coster et al., 2014) and in line with this, the earlier OCCM intermediate binds non-hyrolysable ATP nucleotide at the Mcm3-7, 7-4, 4-6 and 6-2 sites (Yuan et al., 2017). This analysis therefore shows that different ATPase interfaces contribute to different stages of helicase activation (**Figure 6.1**).

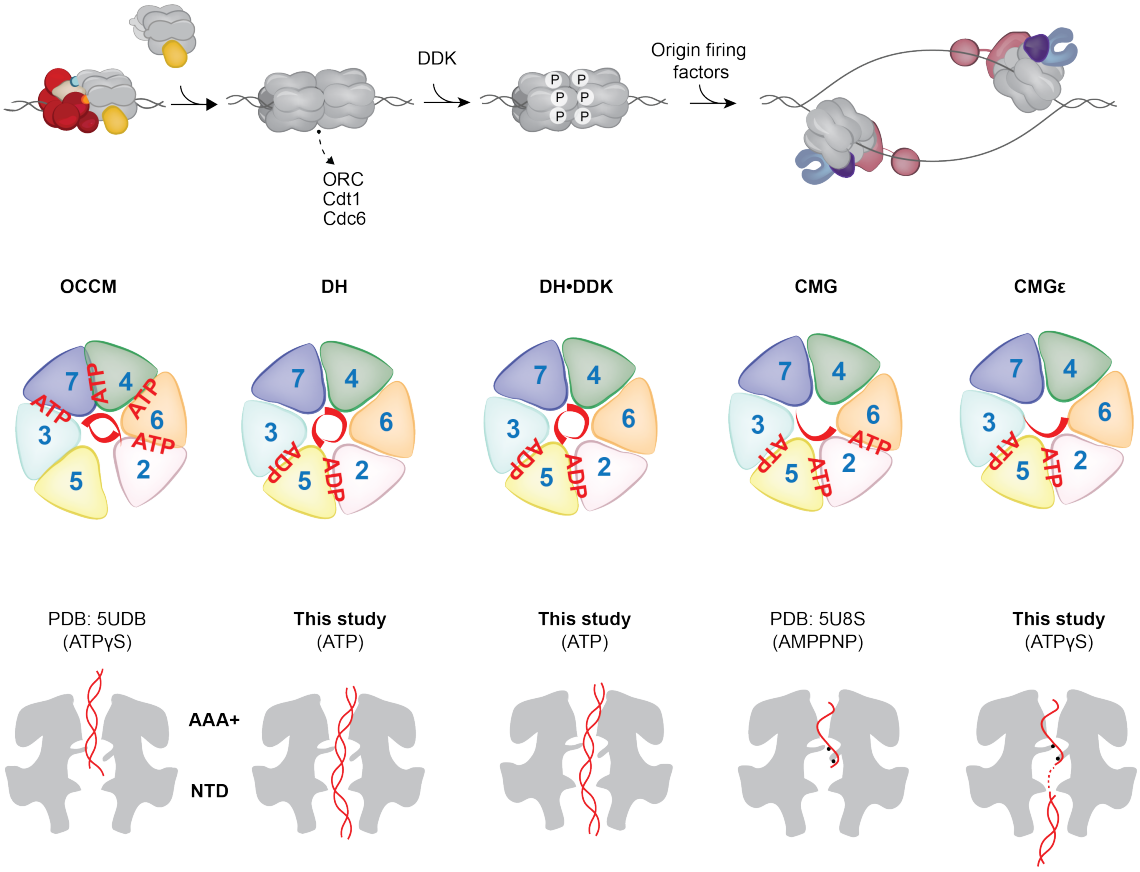


Figure 6.1: Comparison of ATPase state and DNA binding modes of different MCM intermediates
MCM nucleotide occupancy and DNA topology transitions during origin licensing and CMG assembly. Transitions necessitate nucleotide exchange, origin melting, Mcm2-5 gate opening, double hexamer separation and strand exclusion events. Each unique MCM subunit plays a part in this process, providing a rationale for the hetero-hexameric organisation of the eukaryotic replicative helicase. OCCM (Yuan et al., 2017) and CMG (Georgescu et al., 2017) structures were used to depict the first and fourth panel while the rest of the diagrams are based on this study.

6.3.2 The eukaryotic MCM double hexamer binds duplex DNA

The MCM double hexamer binds 62 base-pairs of double stranded DNA, which span the entire length of the channel, closely matching a previous estimation (Remus et al., 2009). The observation of intact DNA throughout the MCM double hexamer goes against a prevailing hypothesis that the offset NTD tiers would constrict the channel and thereby melt the encircled DNA (Li et al., 2015). This is also in contrast to the observed melting of DNA in the SV40 L-Tag double hexamer (Gai et al., 2016). Additionally, we found no configuration of the double hexamer where DNA would thread through one ring and exit through a side channel of the second ring (Hizume et al., 2017). Instead, we and others have shown that the DNA bends or ‘zig-zags’ to fit within the kinked central channel of the double hexamer (**Figure 3.9**) (Noguchi et al., 2017). Intriguingly, only ZF domains from one side of the ring (Mcm5-2-6) contact

the duplex DNA at the N-terminal tier (**Figure 3.9d**). This places the DNA proximal to each Mcm2-5 gate, which is a site known to be required for loading (Samel et al., 2014) and may be necessary for extrusion of the lagging strand template upon activation. However, in the double hexamer this gate is tightly shut, which may serve to block premature strand separation.

Close inspection of the DNA bound double hexamer structure reveals 14 different highly conserved loops from all six MCM protomers of each ring binding equally to both the Watson and Crick strands, primarily through electrostatic interactions of lysine/arginine residues and the phosphate backbone of DNA (**Figure 3.9d**). The extent of these tight interactions suggests that the MCM stably grips the DNA, which would be required for a complex that licenses origins in G1 and gets activated much later in S-phase. Furthermore, the promiscuous strand engagement mode of the double hexamer may prevent premature strand selection bias and subsequent directional translocation by the MCM. The SV40 Large T antigen double hexamer has also been shown to bind both strands of double helix (Gai et al., 2016). This may therefore represent a common DNA binding modality across AAA+ double hexamers during replication initiation.

The numerous DNA-MCM contacts may also explain why a high salt wash is required to destabilise the electrostatic nucleoprotein interactions and enable free sliding of the double hexamers on naked DNA (Remus et al., 2009). In line with this, after double hexamer loading, we performed a high salt wash at 0.3 KCl to remove weakly bound loading factors and observed only a subset of MCM complexes still bound by DNA (**Figure 3.9b**). Although it is likely that *in vivo* the extent of sliding is limited by physiological salt concentrations (~0.1M NaCl) and by regularly spaced nucleosomes, passive mobility of the double hexamer may be an important functional facet. For example, many more double hexamers are loaded than are fired and the 'reserve' double hexamers could be pushed by the CMG during unperturbed replication fork progression as observed from our train experiments (**Figure 4.20**) (Donovan et al., 1997). These double hexamers could then be used to restore stalled forks and allow replication restart without having to re-load extra double hexamers during S-phase (Remus et al., 2009). It has also been shown that RNA polymerase can push double hexamers *in vivo* such that origin firing occurs distally to where the

MCM was first loaded (Gros et al., 2015). Thus, sliding can re-define the origin to overcome the challenge by a transcription-replication collision.

In summary, the structure of the MCM double hexamer provides a rationale for the observed helicase inactivity at the G1 phase. First, the tilted, but compact arrangement of double hexamer ensures stable binding to ADP and prevents additional ATP binding/hydrolysis events. Second, strong inter-hexamer interactions prevent separation of the two rings. Finally, the double hexamer stably binds duplex DNA and a locked Mcm2-5 gate blocks untimely strand extrusion.

6.4 CMG assembly

CMG assembly is a highly regulated process that requires eight origin firing factors in total (Yeeles et al., 2015). By taking advantage of the *in vitro* reconstituted origin firing assay, we assessed the structural changes that accompany CMG formation. DDK phosphorylation is critical for both Cdc45 and GINS recruitment to subsequently assemble a CMG complex (Yeeles et al., 2015). As such, we first investigated the structure of the DDK treated double hexamer to analyse the MCM conformational changes that may occur at this stage.

6.4.1 DDK phosphorylation stabilises the N-terminal tails of Mcm4,6 to create a recruitment platform for origin firing factors

Although DDK specifically acts on DNA bound double hexamers rather than soluble MCM (Francis et al., 2009, Rohou and Grigorieff, 2015, Sheu and Stillman, 2006), it is a promiscuous kinase that poly-phosphorylates the N-terminal tails of Mcm2, 4 and 6 (Randell et al., 2010). The structure of the licensed origin may explain why DDK specifically binds the double hexamer conformation of MCM (Francis et al., 2009). Whereas Cdc7 binds to Mcm4, the Dbf4 subunit has been shown to bind Mcm2 (Ramer et al., 2013). In the double hexamer, but not in the isolated MCM, these subunits are neighbouring *in trans* (**Figure 3.1c**).

To specifically investigate the MCM conformational changes after phosphorylation, we performed a high salt wash to remove any bound kinase after DDK treatment. Our findings reveal no reconfigurations in the core MCM particle compared to the

unmodified double hexamer, consistent with previous low-resolution studies (On et al., 2014, Sun et al., 2014). We further show no difference in nucleotide binding state or DNA conformation upon DDK phosphorylation (**Figure 3.11b**). This is further in agreement with nucleotide exchange and origin unwinding experiments showing that DDK is insufficient to cause rebinding of ATP or melting of DNA by the MCM (Douglas et al., 2018). Instead, a previously invisible density in the unmodified double hexamer becomes structured upon DDK phosphorylation at lower contour level, which probably corresponds to the long and flexible N-terminal tails of Mcm4 and 6. These densities arrange into a symmetrical, yet flexibility tethered platform that could be important for downstream recruitment of origin firing factors.

The Sld3/7 dimer of hetero-dimers is thought to bind the phosphorylated tails of Mcm4 and Mcm6 to recruit two Cdc45 co-activators to the double hexamer (Deegan et al., 2016, Itou et al., 2015).

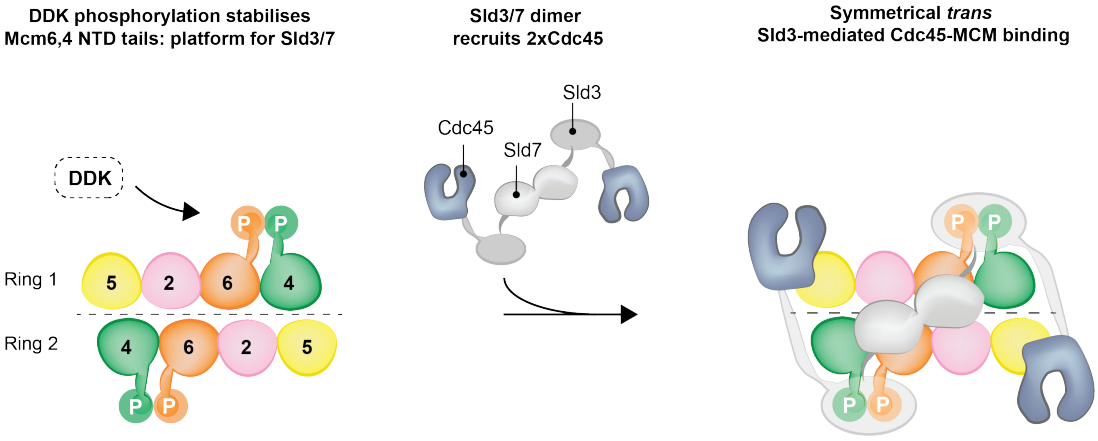


Figure 6.2: DDK phosphorylation of Mcm4,6 creates a docking platform for Sld3/7-Cdc45
Phospho-Mcm4,6 form a flexibly linked density that would be suitably positioned for the Sld3/7 dimer of hetero-dimers, which is also predicted to form a symmetrical quaternary structure. Sld3 bound to Mcm4,6 of one hexamer places Cdc45 in *trans* on Mcm5 from the opposing hexamer in a symmetrical fashion. Model based on DDK phosphorylated MCM double hexamer from this study, studies of Mcm4,6 phospho-peptide binding to Sld3 (Deegan et al., 2016) and crystal structure of Sld3/7 (Itou et al., 2015).

In this model, two Sld3 proteins (connected by an Sld7 dimer) concomitantly bind the two sets of phospho-Mcm4,6 platforms (Itou et al., 2014, Deegan et al., 2016) (**Figure 6.2**). Our DDK phosphorylated double hexamer shows that such a binding pattern is plausible, as the two extending densities are symmetrical (**Figure 3.7**). Each Sld3 bound to Mcm4,6 in turn would guide the recruitment of Cdc45 to the

directly opposing Mcm2-5 sites on the second ring of the double hexamer. A second alternative is that Sld3 places Cdc45 within the same ring that Sld3 binds to. However, by placing Cdc45 on the *in trans* Mcm2-5 site, Sld3 could promote the coordinated assembly of two CMG helicases from the single double hexamer, which would favour bi-directional unwinding after full origin firing. Together, this model suggests that the symmetry of the double hexamer may provide the basis for concerted helicase assembly.

6.4.2 GINS recruitment finalises holo-helicase assembly and causes double hexamer separation

Given that DDK phosphorylation does not dissociate the two MCMs, we next asked at what stage double hexamer separation takes place using drop out experiments of key origin firing factors from the *in vitro* DNA replication assay (DDK, Sld2, Sld3/7, Dpb11, Mcm10 were separately omitted). As shown in **Figure 3.13**, GINS recruitment leads to the dissociation of the double hexamers as it is simultaneously incorporated into CMG. This was a surprising observation in light of earlier *in vivo* data that suggested an Mcm10 mediated splitting of the two MCM rings for activation (Quan et al., 2015). Our findings are consistent with a recent systematic ChIP study showing that pre-IC assembly acts a 'molecular switch' that separates the helicases ahead of Mcm10 recruitment (Miyazawa-Onami et al., 2017). A reason for the observed MCM separation could be that GINS (Psf2 subunit) competes for the same Mcm5 A-domain that is protected by Mcm7 from the opposing ring in the double hexamer (**Figure 4.12**). Therefore, a conformational rearrangement has to occur to expose the Mcm5 A-domain and allow binding of GINS. The separated CMGs in the absence of Mcm10 were as much as 400 bp apart but it is likely that this represents passively diffused helicases. *In vivo* this separation may be limited by nucleosome positioning. Furthermore, the ability of Mcm10 to bind to the double hexamer prior to activation means that Mcm10 could be poised for immediate activation of separated CMG helicases (Douglas and Diffley, 2016).

It must be emphasised at this stage that the assay does not distinguish whether GINS alone or in combination with Polε cause the double hexamer disassembly. However, the finding that double hexamers remain intact until the second MCM co-

activator is incorporated may be important to ensure the aforementioned symmetrical and concerted CMG assembly.

6.4.3 GINS and Polε are recruited together to the helicase

GINS and Polε are recruited to the MCM as part of the same pre-LC complex, composed of Sld2, Dpbl11, GINS and Polε (Muramatsu et al., 2010). In the CMGε structure, whereas C-Pol2 mainly contacts the AAA+ tier, Dpb2 links the MCM and GINS subunits. In particular, N-Dpb2 extends towards and binds the B-domain of Psf1. Notably, several structural studies have shown that Psf1 and Dpb2 are flexible in isolation (Choi et al., 2007, Chang et al., 2007, Kamada et al., 2007, Tahirov et al., 2009). However, the CMGε structure provides visual evidence that together, B-domain of Psf1 and N-Dpb2 interact and become resolved. The fact that only C-Pol2/Dpb2 was visible in the CMGε structure even though the full length Polε protein was used raised the hypothesis that the Polε essential, non-catalytic domains are sufficient to form the CMG. In an unpublished *in vitro* reconstitution of CMG assembly by Dr. Panchali Goswami and Dr. Max Douglas, it was shown that C-Pol2 and Dpb2 in combination were indeed the minimal requirements to build a CMG complex (at least from the range of mutants tested [wt Polε, Dpb2, Polε-Δcat, Pol2, C-Pol2]) (**Figure 5.13**). Together with the cryo-EM structure of CMGε, these observations provide a rationale for why C-Pol2/Dpb2 of Polε are essential for replication (Dua et al., 1999, Araki et al., 1991a) despite losing their catalytic activity (Baranovskiy et al., 2017, Tahirov et al., 2009). However, this is presently a speculative model; to formally test the role of Polε in CMG assembly future structural characterisation of pre-LC recruitment to MCM-Sld3/7-Cdc45 is required.

6.5 CMG activation and replication fork establishment

6.5.1 Mcm10 converts assembled CMG complexes to active holo-helicases

If Mcm10 is not involved in either CMG assembly nor double hexamer separation, what is its precise activating role in DNA replication? A recent report suggests that Mcm10 stimulates ATP turnover by the CMG while untwisting a further 1.5 turns of the double helix compared to the 0.6-0.7 helical turns in the assembled, inactive CMG (Douglas et al., 2018). This extent of unwinding exerted by Mcm10 (~15 nm) is thought to be sufficient to completely exclude the lagging strand template from the

central CMG channel. In support of a lagging strand exclusion role, Mcm10 has been shown to bind to the N-terminus of Mcm2 to capture a conformational change at this subunit between the A-domain and OB-fold (Looke et al., 2017) and strongly binds single-stranded DNA (Robertson et al., 2008). Mcm2 together with Mcm5 form a natural opening in the CMG complex (Samel et al., 2014); Mcm10 interaction at this site may therefore promote the open Mcm2-5 gate for lagging strand template extrusion (Looke et al., 2017). The fact that double hexamer separation has already occurred prior to CMG activation by Mcm10 (*section 3.5*) means that the Mcm2-5 sites may be more available in contrast to their locked configuration at the pre-RC.

In line with these suggested roles, our train experiments showed unequivocally that translocation by CMG depends on Mcm10 (**Figure 4.21**). However, alignment of 2D classes between –Mcm10 and complete origin firing reactions suggest no conformational change in the helicase upon Mcm10 interaction (**Figure 3.12**). Furthermore, our experiments indicate that Mcm10 is sensitive to high salt washes (in contrast to a recent report of an overexpressed CMG-Mcm10 complex (Langston et al., 2017)) and a stable CMG-Mcm10 intermediate could not be identified under high or low salt wash conditions (**Figure 3.12**). Thus, the reported long-term association of Mcm10 with the replisome (Looke et al., 2017) is likely to be of a transient and dynamic nature.

Collectively, these experiments show that the *in vitro* reconstitution assay is a rich source to probe the role of various origin firing factors in helicase assembly and activation.

6.5.2 The role of Cdc45 and GINS in the active CMG helicase

Recent years have seen several subnanometre cryo-EM structures describing the active CMG helicase from budding yeast and *D. melanogaster* systems (Yuan et al., 2016, Georgescu et al., 2017, Abid Ali et al., 2016) either isolated or in association with Polε (Chapter 5). This has provided an understanding of the roles of the key Cdc45 and GINS auxiliary factors in activating the MCM, how nucleotide binding is synchronised with DNA fork engagement and the importance of Polε in stabilizing the replicative helicase. In the following sections, I will first discuss what we have

learned from studying the isolated CMG bound to single-stranded DNA before discussion of the CMGε structure.

The cryo-EM structure of the CMG shows that Cdc45 wedges in between the A-domains of Mcm2-5, thereby safely guarding the gate and preventing DNA from slipping out of the central channel (Petojevic et al., 2015). In the CMG, Cdc45 binding to the helicase is stabilized by interactions with Psf1 and Psf2 of GINS. Prior to GINS recruitment, Cdc45 is presumably only held in place by a Cdc45 CID interacting domain contacting the Mcm2-5 A-domains. This supports the observation that a Sld3/7 chaperone is initially required to keep Cdc45 stabilized on the double hexamer before GINS gets recruited (Simon et al., 2016).

Surprisingly, a large portion of Cdc45 is exposed to solvent in the CMG complex, which supports binding to other replisome components. For example, the protruding alpha helix 6 identified in this study has been shown to bind to the N-Pol2 by both structural and crosslinking/mass spectrometry studies (Zhou et al., 2017, Sun et al., 2015). Moreover, this Cdc45 alpha helix has also been reported to bind RPA. In this role, Cdc45 guides RPA at the replication fork to prevent nuclease attack of the nascent single-stranded DNA (Szambowska et al., 2017). It remains to be seen in a CMG-DNA structure whether Cdc45 can bind DNA as shown by a crosslinking study (Petojevic et al., 2015). Unlike its ancestral RecJ homologue which binds single-stranded DNA between open DDH and DDHA1 domains (Wakamatsu et al., 2010), Cdc45 adopts a constitutively closed conformation of these domains within the CMG (**Figure 4.10**).

Soon after the discovery of GINS 15 years ago, the role of this tetramer at the replication fork began to emerge, with evidence of GINS-mediated replisome interactions (Gambus et al., 2006), activation of helicase ATPase activity (Costa et al., 2011, Ilves et al., 2010) and coupling the helicase to the leading strand polymerase (Zhou et al., 2017, Sun et al., 2015, Sengupta et al., 2013). Both the isolated CMG and CMGε structures presented here build on these observations. In the CMG Psf2 and Psf3 interact with the Mcm3-5 A-domains whereas the previously disordered Psf1 in the crystal structures becomes structured and bind Cdc45 (**Figure 4.11**). This region of Psf1 is also critical for binding to the Dpb2 subunit of the non-

catalytic lobe of Polε and is thought to be involved in initiation when the leading strand polymerase and GINS are recruited together (Sengupta et al., 2013) (**Figure 5.12**) (further discussed below). Sld5 is the only GINS subunit not bound to Cdc45 or MCM. Instead, Sld5 stably binds the Ctf4 trimer (Samora et al., 2016, Simon et al., 2014) linking the CMG to Polα-primase (Simon et al., 2014), sister chromatid cohesion (Samora et al., 2016) and other factors involved in rDNA regulation and Okazaki fragment processing, such as Tof2 and Dna2 (Villa et al., 2016). Thus, Cdc45 and GINS both stabilize the MCM motor and couple unwinding to replication and wider cellular processes.

6.5.3 The active MCM ring binds single-stranded DNA

Whether the CMG translocates on single or double stranded DNA was poorly understood at the time I set out to address this question. In the CMG-DNA sample, the predominant compact ATPase conformer showed additional density lining the channel at the AAA+ tier of the replicative helicase. Only one strand of DNA extracted from B-form DNA matched the chirality and thickness for this density, providing evidence for the CMG translocating on single-stranded DNA. The DNA spirals around a PS1 staircase, emanating from the Mcm4-6-7 subunits and in close proximity to PS1 of Mcm3. The lysine/arginine residues at the tips of these loops are extremely well conserved from archaea to human, and their mutation reduce DNA binding and abrogate helicase activity in both archaeal and fly systems (McGeoch et al., 2005, Petojevic et al., 2015). This structure explains these phenotypes by providing visual evidence for their previously predicted role in DNA binding using biochemical/biophysical assays (McGeoch et al., 2005). Additional density inside the channel was also found in a second conformer (relaxed ATPase), identified by comparison to DNA-free ATP-CMG structures (Abid Ali et al., 2016). This density matched the thickness of single-stranded DNA although the low resolution precluded fitting of DNA coordinates. Notably, the DNA in this conformer was in close proximity to the NTD OB-fold hairpins of Mcm3-5-4-7. Mutation of arginine/lysine residues at these loops substantially reduced DNA binding and DNA unwinding as shown in archaea (McGeoch et al., 2005, Barry et al., 2009) and yeast (Froelich et al., 2014).

Comparison of the yeast CMG structure (Georgescu et al., 2017) with the *D. melanogaster* compact ATPase CMG structure, reveal a near identical architecture of the MCM and co-activators. Furthermore, single-stranded DNA is found inside the AAA+ tier of the MCM channel in both structures. However, in the yeast CMG, DNA binds at the Mcm2-5-3 side of the ring, interacting with PS1 and H2i loops. This difference in DNA position may reflect a species difference or a more plausible explanation is that two different DNA translocation states have been captured by these structures. In support of the latter notion, superimposition of the two structures shows that single-stranded DNA seems to be derived from the same leading strand template in a B-form conformation (Georgescu et al., 2017). If they are two different translocation intermediates, then this shows that AAA+ modules from all MCM subunits can bind DNA during different stages of translocation.

With only single-stranded DNA detected inside the channel a strand exclusion mechanism for unwinding is favoured. Determining the structure of CMGε bound to fork DNA further strengthened support for this model, as the duplex DNA junction is located against the NTD tier (in particular held in place by Mcm4 OB-fold) but never fully inside the channel (consistent with observations from CMG-DNA-ATP structure (Georgescu et al., 2017)). On the other hand, the lagging strand template has so far eluded all structural characterisation. However, tagging of the 5' end of a fork (*section 4.2.7*) coupled with crosslinking studies of CMG-DNA (Petojevic et al., 2015) indicate that this filament is free to move outside the helicase.

6.5.4 MCM reconfigurations from origin licensing to replication fork establishment inform origin melting and DNA translocation

The MCM motor undergoes several conformational and compositional changes as it transitions from the inactive double hexamer to the active helicase. First, the motor goes from stably binding ADP at Mcm2-5 and Mcm3-5 in the inactive MCM to binding ATP at an additional Mcm2-6 site in the active CMG (Georgescu et al., 2017, Douglas et al., 2018) (**Figure 6.1**). Indeed, it has recently been shown that ADP release from MCM only occurs upon origin firing factor recruitment (after DDK phosphorylation and before activation by Mcm10) (Douglas et al., 2018). Consistent with this result,

comparison of the inactive MCM to the MCM from the CMG shows that the AAA+ domains re-align with the NTD during this transition (*see section 4.2.9*). Such a relaxation of the AAA+ twist in the active helicase may promote ATP reuptake and subsequent hydrolysis upon Mcm10 recruitment (Sun et al., 2014, Georgescu et al., 2017).

Additionally, during the MCM to CMG transition, the MCM B-domains constrict and narrow the central channel while the peripheral A-domain rotate. Collectively, these protein reconfigurations cause a translation and rotation of the DNA that is consistent with the observation of melting at this stage (Douglas et al., 2018). In this model, the DNA binding loops of Mcm6 (PS1) and Mcm7 (OB) move in opposite directions while Mcm4 (OB) remains as a fixed pivot at the duplex junction. As a result of stretching one strand, the double helix is deformed and the derived mechanism is compatible with 0.6-0.7 turns of double helix untwisted (Douglas et al., 2018).

The propensity of CMG to stretch DNA was also shown using single molecule FRET, where upon CMG binding, single-stranded DNA was stabilized in an extended configuration using either ATP or ATPγS in the reaction buffer. Our FRET studies combined with nucleotide exchange experiments (Douglas et al., 2018) indicate that ATP binding is sufficient to drive reconfiguration of the encircled DNA. Such an ATP-binding dependent stretch-to-open mechanism for origin melting was also shown to occur by the bacterial initiator DnaA (Duderstadt et al., 2011). However, in bacteria DnaA is an initiator and after melting, the bacterial replicative helicase DnaB is loaded onto single-stranded DNA. Therefore, although a similar mechanism may melt the DNA across the prokaryotic and eukaryotic systems, distinct components are involved. In SV40, melting by L-Tag is also ATP-binding dependent (Dean and Hurwitz, 1991), although L-Tag is sufficient on its own to recognize origins, load onto DNA and partially melt an AT-rich region of the double helix before activating bi-directional replication fork establishment (Gai et al., 2016, Valle et al., 2006). The complexity in the eukaryotic system where MCM loading, CMG assembly and replication fork establishment are separated events under the control of cell cycle clues and origin firing factor recruitment may allow higher levels of regulation to maintain genome stability.

6.5.5 ATPase remodeling in the CMG

As a motor that performs mechanical work, the MCM ring comprises highly dynamic ATPase modules to allow cycles of ATP binding and hydrolysis to power replication fork progression. To obtain the cryo-EM structures of the CMG-DNA presented in Chapter 4, I used a non-hydrolysable ATP analogue to stabilise DNA binding within the helicase. As a result, the vast majority of particles (74%) classified into a compact ATPase state. A second, relaxed ATPase conformer was identified although its functional role in DNA translocation was difficult to interpret as nucleotide occupancy could not be discerned. The main conformational difference was the Mcm2-5 gate, where it was either closed (compact ATPase) or open (relaxed ATPase). Similar conformers were also identified in an apo CMG structure from budding yeast (Yuan et al., 2016). The identification of only two ATPase and DNA-binding states led us and others to postulate an inch-worming mechanism for DNA translocation (Abid Ali et al., 2016, Yuan et al., 2016). This model suggests that the CMG grips DNA at the AAA+ tier (representing the compact ATPase conformer) then switches to the second configuration (relaxed ATPase conformer) where DNA is released from the AAA+ and handed off to the NTD. In this way, the CMG relaxes and tightens the Mcm2-5 gate to 'inch-worm' or move with a nodding 'pump-jack' motion along the DNA.

However, since publication of these CMG structures (Yuan et al., 2016, Abid Ali et al., 2016) the new CMGε structure bound to forked DNA (Chapter 5) has shown that such an inch-worming mechanism is incompatible with Polε binding. The pump-jack model in particular posits fluctuation between open, spiral and closed, planar ATPase states concomitant with Mcm5 WHD domain movement. All the conformational changes that drive the relaxed ATPase state would disrupt the Mcm2-5-Pol2 interaction sites. Therefore, in hindsight, the relaxed ATPase state most likely represents a paused intermediate, as would occur during fork stalling (Petojevic et al., 2015). To address the mechanism of helicase unwinding in the future, a CMGε-ATP complex bound to DNA may be required coupled with cryo-EM image processing classification approaches to identify the different conformations in the translocation cycle.

6.5.6 Polε stimulates productive fork engagement by CMG

By comparing the yeast CMGε structure presented in Chapter 5 to yeast CMG-DNA (Georgescu et al., 2017), a conformational change in the AAA+ tier becomes apparent. In particular, the Mcm5-3 and Mcm2-5 sites bound by ATPγS tighten while all other interfaces are empty upon Polε binding. Importantly, in the presence of Polε, both the single-stranded DNA and fork nexus are clearly engaged by the CMG. In contrast, using similar conditions for the isolated CMG (fork DNA and non-hydrolysable ATP analogues) failed to resolve the duplex DNA junction (*see section 4.2.8)* (Georgescu et al., 2017, Abid Ali et al., 2016). This suggests that Polε might promote productive fork engagement by the CMG. In support of this notion and further showing that Polε remains CMG bound during unwinding, our train experiments show that the translocating helicase is almost always tethered to Polε (75-100% cases from different repeats) (**Figure 4.20**). Several other studies suggest that Polε directly enhances CMG unwinding activity (whereas Polδ does not) (Kang et al., 2012, Yeeles et al., 2017) and vice versa, CMG stimulates Polε synthesis rates (Bermudez et al., 2011, Langston et al., 2014). Although not directly proven by the CMGε structure, the tight association of C-Pol2/Dpb2 with the replicative helicase is compatible with such a synergistic stimulation model.

In line with their catalytic inactivity, C-Pol2 and Dpb2 do not bind DNA as shown by the CMGε structure (**Figure 5.10**). Inspection of the isolated C-Pol2/Dpb2 and comparison to Polα (Perera et al., 2013) provides a rationale for this observation: the C-terminal ZF of Pol2 occludes the polymerase fold to block the DNA binding groove and the inner surface of the polymerase fold is less positively charged. Instead of DNA binding, C-Pol2 and Dpb2 perform a helicase anchoring role where they stabilise the AAA+ tier analogous to the binding of Cdc45 and GINS at the NTD tier. Indeed, it seems that while C-Pol2/Dpb2 have lost their polymerase/exonuclease activity they may have evolved these new structural roles akin to Cdc45 which has lost its ancestral exonuclease activity but stabilises the MCM ring.

6.5.7 Flexibility of N-Pol2 is important for polymerase function

As further shown by the CMGε structure, the N-Pol2 module of Polε is invisible (despite extensive classification attempts to visualise this region). This is in line with

crosslinking studies showing that the isolated Polε contains two lobes that are only thinly connected (Sun et al., 2015) and negative stain EM studies that demonstrate the movement of the two N-Pol2 and C-Pol2 modules with respect to each other (*see section 5.2.2*) (Zhou et al., 2017, Georgescu et al., 2017). The flexibility of N-Pol2 within the integrated replisome many confer a distinct advantage for both unperturbed and stressed replication fork progression. For example, by temporarily dissociating from the 3' terminus, Polε via its N-Pol2 domain may leave the space required for PCNA loading yet still remain stably bound to the replisome via its strong C-Pol2/Dpb2-CMG interaction. In turn, many PCNA dependent processes can occur, including improved processivity (Chilkova et al., 2007), mismatch repair (Pluciennik et al., 2010) and assembly of *de novo* nucleosomes by Caf1 (Moggs et al., 2000).

Additionally, the N-Pol2 movement may enable polymerase switching with Polδ, a concept that has recently been characterised both biochemically (Yeeles et al., 2017) and structurally using the full length CMGε structures (Zhou et al., 2017). In this model, Polδ-PCNA can initiate leading strand synthesis as CMGε begins slowly unwinding from the origin and moves away from the Polα primed 3' end. Once Polδ catches up with the CMGε, a polymerase switch is thought to occur to speed up fork progression, where N-Pol2 takes over leading strand extension while C-Pol2/Dpb2 simultaneously stimulate CMG fork unwinding (alongside PCNA, Mrc1 and Csm3/Tof1) (Yeeles et al., 2017). Thus, plasticity of Polε is imperative to its proper function within the replisome.

6.5.8 Polarity of the CMG helicase at the replication fork

The orientation of the active MCM replicative helicase has been a contentious matter in the field for several years (Costa et al., 2014, Georgescu et al., 2017, Rothenberg et al., 2007, McGeoch et al., 2005, Froelich et al., 2014, Trakselis, 2016). In Chapter 4 (*section 4.2.10*), Dr. Max Douglas and I showed that the CMG translocates with its NTD tier ahead of the AAA+ motor. This is consistent with observations from the recent CMG structure determined in the presence of ATP and roadblocked fork DNA (Georgescu et al., 2017) and has several implications for both unwinding and replisome organization (**Figure 6.3**).

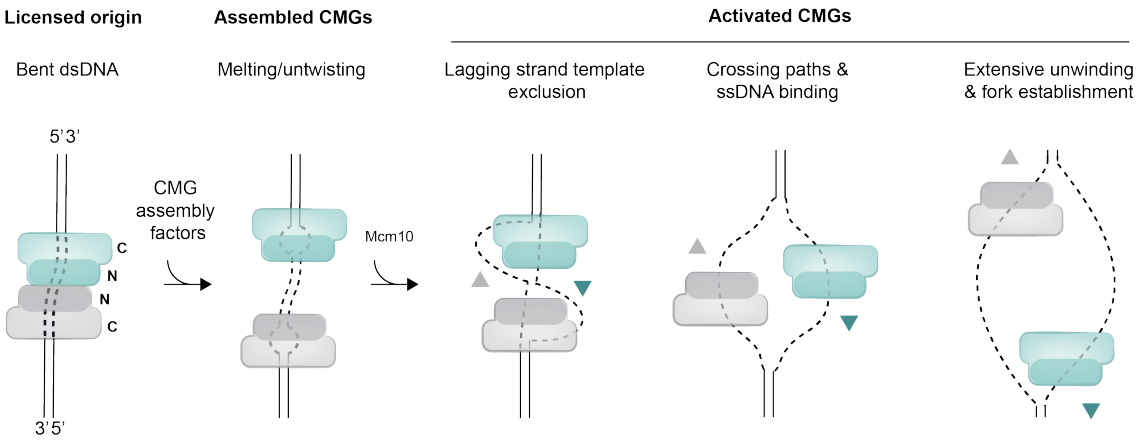


Figure 6.3: Polarity of CMG helicase suggests a model for origin activation
Double hexamers are stacked in an NTD-NTD configuration (as shown in Chapter 3). CMG assembly leads to origin melting (as proposed in Chapter 4 and (Douglas et al., 2018)). Mcm10 recruitment thereafter causes full strand extrusion and CMG ATP hydrolysis activation (Douglas et al., 2018). With an NTD tier directionality, the two helicases have to cross paths, ensuring all origin DNA is unwound. Active CMGs bi-directionally unwind DNA during genome duplication. Model adapted from (Georgescu et al., 2017).

First, an NTD ahead of AAA+ orientation means that the activated CMG helicases have to cross paths as the MCMs are initially tethered at the NTD interfaces (left, **Figure 6.3**). This requirement dictates that each CMG translocates on one strand and excludes the lagging strand template, consistent with biochemical experiments (Fu et al., 2011) and cryo-EM studies (*sections 4.2.6 and 5.2.6*) (Georgescu et al., 2017). By crossing paths, the CMGs ensure that all origin DNA is unwound and subsequently replicated, which is key to maintaining genome stability. Importantly, this polarity also means that both CMGs have to be activated concertedly, because one CMG binding to duplex DNA would block the path of a second CMG translocating on single-stranded DNA. In a similar fashion, the bacterial DnaB helicases have been found to pass each other over the origin to initiate DNA replication (Fang et al., 1999).

The bacterial Rho and DnaB helicases track in a 5'-3' polarity whereas the eukaryotic based CMG and E1 translocate in a 3'-5' directionality. However, the four helicases bind nucleic acid with the same polarity through their motors irrespective of their RecA-type and AAA+ ATPase divergent folds (Itsathitphaisarn et al., 2012, Enemark and Joshua-Tor, 2006, Thomsen and Berger, 2009, Georgescu et al., 2017). Why the eukaryotic-based CMG and E1 translocate with their NTD ahead of the AAA+ while bacterial helicases unwind with their motor domains first on the opposite strand of the fork remain unanswered questions.

At the replisome (**Figure 6.4**), this revised CMG orientation means that Polα-primase (bound via Ctf4) is ahead of the fork. This assignment is evidenced by negative stain analysis of co-purifying CMG-Ctf4 complexes *in section 5.2.1* and other structural studies (Sun et al., 2015, Simon et al., 2014) showing that Ctf4 binds the NTD tier of the helicase. By its close proximity to the fork nexus, the primosome is ideally situated to capture the excluded lagging strand template and create the initial primers immediately.

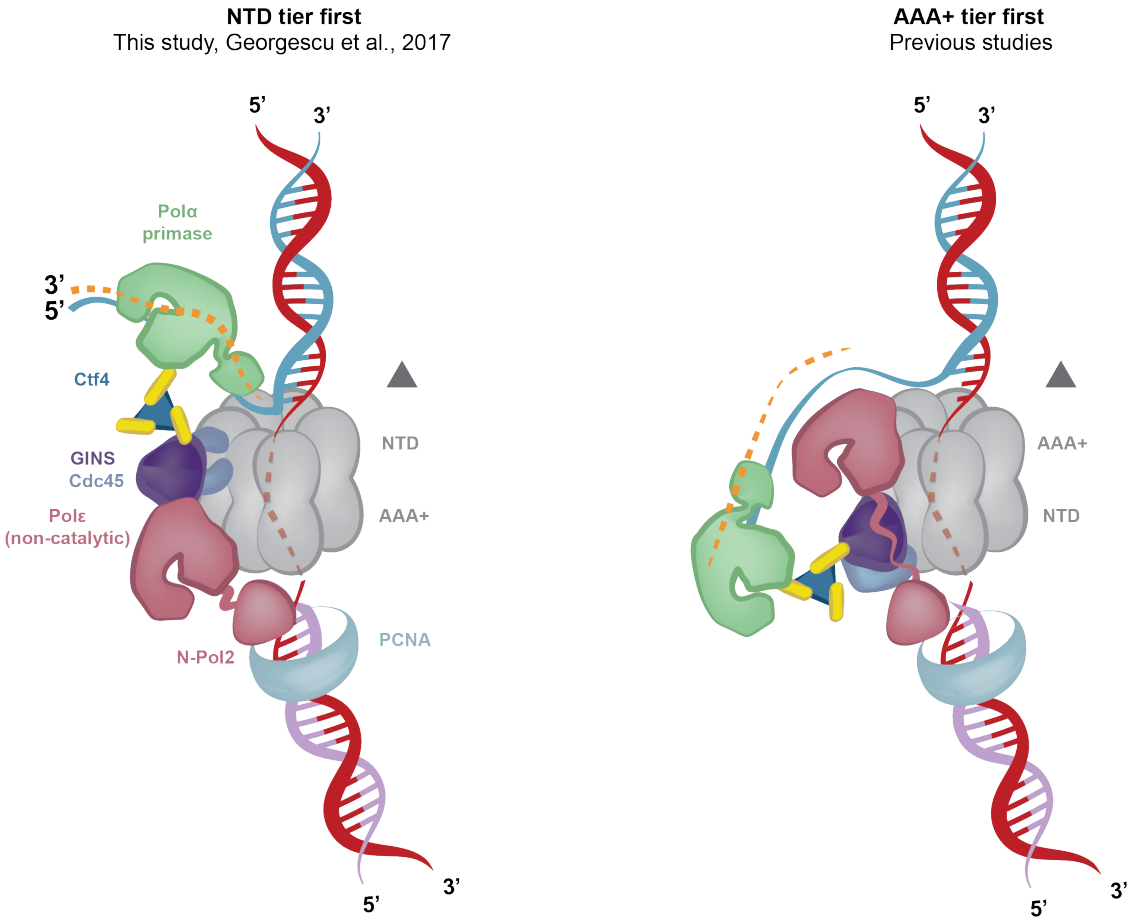


Figure 6.4: Revised polarity of the CMG suggests an efficient replisome architecture
Polα-primase interacts with CMG via Ct4 on the NTD tier whereas Polε binds directly to the AAA+ tier of the CMG (Chapter 5). NTD tier ahead of the AAA+ tier means that the lagging strand template can be directly captured by the primosome for immediate priming while the Polε catalytic domain, flexibly linked to the C-Pol2/Dpb2 CMG anchor, captures the leading strand template exiting from the ATPase pore.

In contrast, Polε would be behind the fork as shown by its binding to the AAA+ tier in the CMGε structure (Chapter 5), train experiments (see *section 4.2.10*) and two previous replisome studies (Zhou et al., 2017, Sun et al., 2015). As such, Polε is suitably poised to use the single-stranded DNA that exits the CMG to extend the leading strand. However, this is only a simplified view of replisome architecture

based on available structural studies (Zhou et al., 2017, Sun et al., 2015). Many more components are involved for processive translocation and replication, including RFC, PCNA, Polδ, FACT, Mcm10, Mrc1 and Csm3/Tof1 (Yeeles et al., 2017, Kurat et al., 2017, Looke et al., 2017). Future structural studies are required to map their interaction sites relative to the replicative helicase.

6.6 A provisional model for helicase assembly and activation

Together, the data presented in this study combined with our current understanding of CMG assembly and origin firing by other groups (Noguchi et al., 2017, Yeeles et al., 2015, Georgescu et al., 2017, Yuan et al., 2016, Deegan et al., 2016, Douglas et al., 2018), enables me to propose a model for helicase activation and replication fork establishment (**Figure 6.5**). First, MCM is loaded onto origins along the genome as an inactive double hexamer that encircles double stranded DNA (origin licensing, **Figure 6.5**). The stability of the licensed origin is maintained through extensive ZF interactions at the N-terminal double hexamer interface as well as the inter-hexamer Mcm5-7 interaction. Upon DDK phosphorylation, the previously disordered Mcm4 and Mcm6 N-terminal tails become resolved as flexibly tethered densities and form a symmetrical platform that could be used by the Sld3/7 chaperone to concertedly recruit two copies of Cdc45 to the double hexamer (Deegan et al., 2016, Itou et al., 2015) (CMG assembly, **Figure 6.5**). The double hexamer withstands origin firing factor recruitment and remains intact as a dimer until GINS binds Mcm5, which releases the Mcm5-7 latch and leads to the separation of two fully assembled CMGs. Polε has a critical role in CMG assembly, where it may cause nucleotide exchange (Douglas et al., 2018) as well as connecting GINS to MCM within the holo-helicase (Sengupta et al., 2013). Concomitant with separation of the two helicases, CMG melts the DNA via a stretching mechanism that involves the rotation of the duplex DNA and counteracting movements of Mcm7 and Mcm6 ATPase pore loops. The CMG is then activated by Mcm10 to promote ATP hydrolysis and extrusion of the lagging strand template via a poorly understood mechanism (Douglas et al., 2018). During replication fork establishment, Polε fastens the Mcm2-5-3 AAA+ sites and captures the Mcm5 WHD to promote productive binding to the fork nexus. As the helicase unwinds, Polε remains strongly bound to the translocating CMG, which unwinds with its NTD tier first and its AAA+ motor pushing from behind.

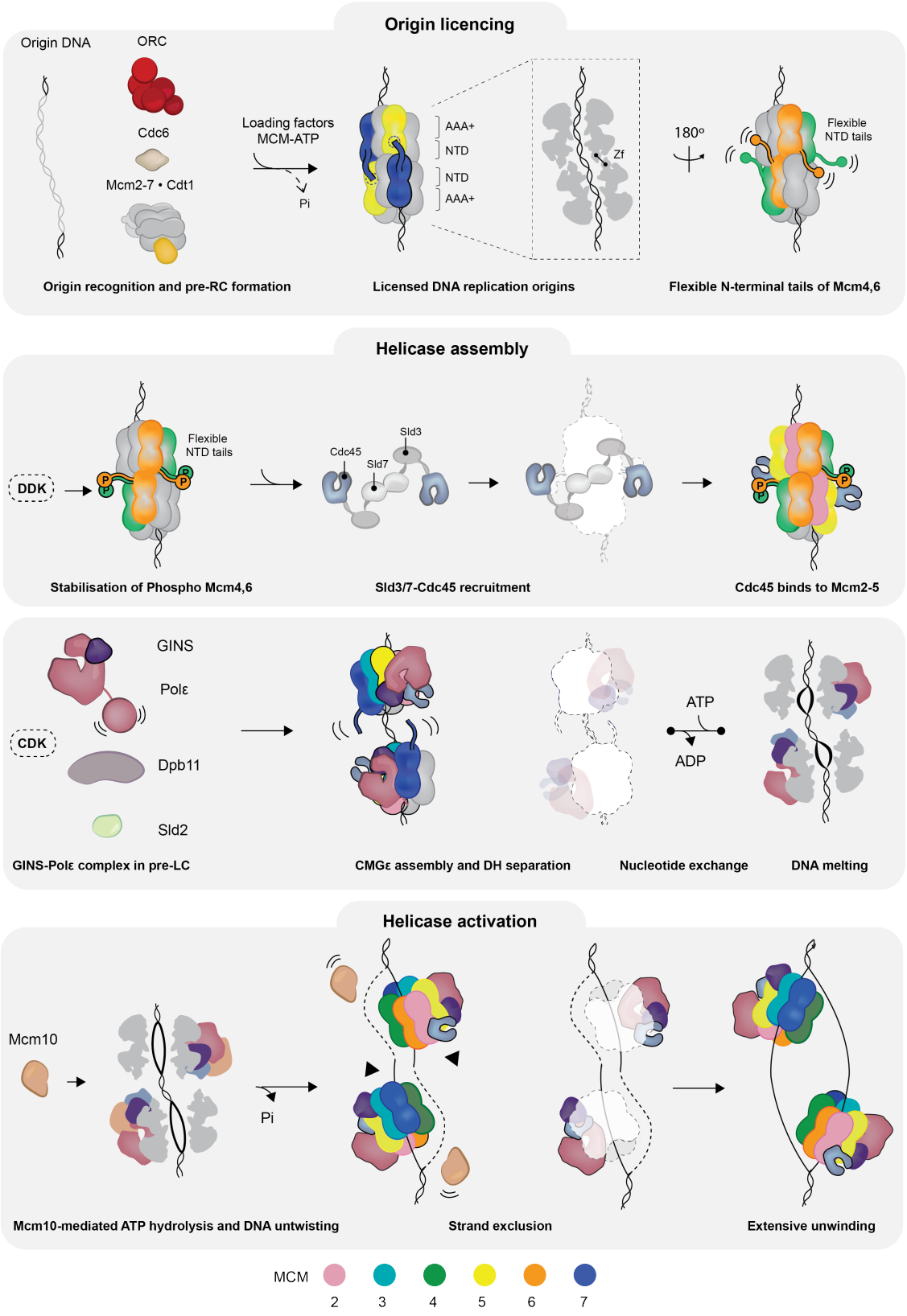


Figure 6.5: Provisional model for helicase activation
See section 6.6 for description. Model based on results presented in this thesis and (Deegan et al., 2016, Douglas et al., 2018, Itou et al., 2015, Muramatsu et al., 2010).

6.7 Future directions

Exciting progress has been recently made in the field of DNA replication, in particular due to technological advances in cryo-EM (Kuhlbrandt, 2014) and the cumulative efforts over decades of research to reconstitute eukaryotic DNA replication *in vitro* (Kurat et al., 2017, Yeeles et al., 2015, Yeeles et al., 2017, Georgescu et al., 2015). In particular, cryo-EM structures of ORC (Li et al., 2018) , OCCM (Yuan et al., 2017), MCM double hexamer (Li et al., 2015, Noguchi et al., 2017, Abid Ali et al., 2017) and the CMG helicase (Abid Ali et al., 2016, Yuan et al., 2016, Georgescu et al., 2017) have provided mechanistic insight into pre-RC formation and CMG activation.

Future studies in the field will likely be devoted towards (i) capturing the various MCM intermediates leading up to CMG formation (ii) origin melting and strand extrusion (iii) coordination of ATP hydrolysis with unwinding and (iv) replisome progression. While the negative stain EM experiments of the reconstituted origin firing reaction demonstrated the premise of using this assay to structurally characterise helicase activation, high resolution cryo-EM studies are required to better understand the molecular basis for DDK and CDK regulated events. Key intermediates to study would be a MCM-Sld3/7-Cdc45 complex to understand whether CMG formation occurs in the predicted symmetrical fashion presented in *section 6.4.1*. Additionally, it will be important to address the Polε/GINS association as part of the pre-LC to gain a better understanding of how the second co-activator is incorporated into the helicase.

Furthermore, while I suggested a stretch-to-open model for origin melting by comparing the double hexamer to the assembled CMG, a structure that captures the path of the melted strand (e.g. a complete origin firing reaction sans Mcm10) is required to provide evidence for such a mechanism. Similarly, it will be important to identify the mechanism by which Mcm10 activates the helicase, in particular the strand extrusion and promotion of ATP hydrolysis by the CMG described in a recent biochemical study (Douglas et al., 2018). What conformational change does Mcm10 exert on the helicase that enables full melting of the DNA and ATP hydrolysis? Moreover, an important area of research is to understand how the active CMG synchronises ATP hydrolysis with translocation on single-stranded DNA. It may be

important to address helicase unwinding mode in the context of the replisome to gain insights into how the active CMG coordinates unwinding with organisation of the replication fork.

Studying such large, flexible and occasionally transient complexes will be a challenge. However, the very recent development of time-resolved electron microscopy may enable events that occur within milliseconds to be captured (Frank, 2017). Furthermore, a new development in cryo-EM image processing software allows multi-body refinement of macromolecular machines and thereby structural characterisation of dynamic proteins (Nakane et al., 2018). Lastly, while all of the experiments carried out in this thesis were performed on naked DNA, *in vivo* replisome progression occurs in the context of chromatin (Miller and Costa, 2017). It will be of essence in the future to understand how nucleosomes reshuffle ahead of the advancing CMG helicase and get reassembled on replicated DNA.

Reference List

ABID ALI, F. & COSTA, A. 2016. The MCM Helicase Motor of the Eukaryotic Replisome. *J Mol Biol,* 428**,** 1822-32.
ABID ALI, F., DOUGLAS, M. E., LOCKE, J., PYE, V. E., NANS, A., DIFFLEY, J. F. X. & COSTA, A. 2017. Cryo-EM structure of a licensed DNA replication origin. *Nat Commun,* 8**,** 2241.
ABID ALI, F., RENAULT, L., GANNON, J., GAHLON, H. L., KOTECHA, A., ZHOU, J. C., RUEDA, D. & COSTA, A. 2016. Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate. *Nat Commun,* 7**,** 10708.
ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr,* 66**,** 213-21.
AFONINE, P. V., POON, B. K., READ, R. J., SOBOLEV, O. V., TERWILLIGER, T. C., URZHUMTSEV, A. & ADAMS, P. D. 2018. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol,* 74**,** 531-544.
AKSENOVA, A., VOLKOV, K., MACELUCH, J., PURSELL, Z. F., ROGOZIN, I. B., KUNKEL, T. A., PAVLOV, Y. I. & JOHANSSON, E. 2010. Mismatch repair-independent increase in spontaneous mutagenesis in yeast lacking non-essential subunits of DNA polymerase epsilon. *PLoS Genet,* 6**,** e1001209.
APARICIO, O. M., WEINSTEIN, D. M. & 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.
ARAKI, H., HAMATAKE, R. K., JOHNSTON, L. H. & SUGINO, A. 1991a. DPB2, the gene encoding DNA polymerase II subunit B, is required for chromosome replication in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A,* 88**,** 4601-5.
ARAKI, H., HAMATAKE, R. K., MORRISON, A., JOHNSON, A. L., JOHNSTON, L. H. & SUGINO, A. 1991b. Cloning DPB3, the gene encoding the third subunit of DNA polymerase II of Saccharomyces cerevisiae. *Nucleic Acids Res,* 19**,** 4867-72.
ARAKI, H., LEEM, S. H., PHONGDARA, A. & SUGINO, A. 1995. Dpb11, which interacts with DNA polymerase II(epsilon) in Saccharomyces cerevisiae, has a dual role in S-phase progression and at a cell cycle checkpoint. *Proc Natl Acad Sci U S A,* 92**,** 11791-5.
ARAVIND, L., ANANTHARAMAN, V., BALAJI, S., BABU, M. M. & IYER, L. M. 2005. The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev,* 29**,** 231-62.
ARAVIND, L. & KOONIN, E. V. 1998. Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res,* 26**,** 3746-52.
ARIAS, E. E. & WALTER, J. C. 2007. Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes Dev,* 21**,** 497-518.
ASTURIAS, F. J., CHEUNG, I. K., SABOURI, N., CHILKOVA, O., WEPPLO, D. & JOHANSSON, E. 2006. Structure of Saccharomyces cerevisiae DNA polymerase epsilon by cryo-electron microscopy. *Nat Struct Mol Biol,* 13**,** 35-43.
AVES, S. J., TONGUE, N., FOSTER, A. J. & HART, E. A. 1998. The essential schizosaccharomyces pombe cdc23 DNA replication gene shares structural and functional homology with the Saccharomyces cerevisiae DNA43 (MCM10) gene. *Curr Genet,* 34**,** 164-71.
BAI, X. C., RAJENDRA, E., YANG, G., SHI, Y. & SCHERES, S. H. 2015. Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife,* 4.

BARANOVSKIY, A. G., GU, J., BABAYEVA, N. D., KURINOV, I., PAVLOV, Y. I. & TAHIROV, T. H. 2017. Crystal structure of the human Pol B-subunit in complex with the C-terminal domain of the catalytic subunit. *J Biol Chem,* 292**,** 15717-15730.
BARRY, E. R. & BELL, S. D. 2006. DNA replication in the archaea. *Microbiol Mol Biol Rev,* 70**,** 876-87.
BARRY, E. R., LOVETT, J. E., COSTA, A., LEA, S. M. & BELL, S. D. 2009. Intersubunit allosteric communication mediated by a conserved loop in the MCM helicase. *Proc Natl Acad Sci U S A,* 106**,** 1051-6.
BARRY, E. R., MCGEOCH, A. T., KELMAN, Z. & BELL, S. D. 2007. Archaeal MCM has separable processivity, substrate choice and helicase domains. *Nucleic Acids Res,* 35**,** 988-98.
BEATTIE, T. R. & BELL, S. D. 2011. Molecular machines in archaeal DNA replication. *Curr Opin Chem Biol,* 15**,** 614-9.
BELL, S. D. & BOTCHAN, M. R. 2013. The minichromosome maintenance replicative helicase. *Cold Spring Harb Perspect Biol,* 5**,** a012807.
BELL, S. P., KOBAYASHI, R. & STILLMAN, B. 1993. Yeast origin recognition complex functions in transcription silencing and DNA replication. *Science,* 262**,** 1844-9.
BELL, S. P. & LABIB, K. 2016. Chromosome Duplication in Saccharomyces cerevisiae. *Genetics,* 203**,** 1027-67.
BELL, S. P. & STILLMAN, B. 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature,* 357**,** 128-34.
BERMUDEZ, V. P., FARINA, A., RAGHAVAN, V., TAPPIN, I. & HURWITZ, J. 2011. Studies on human DNA polymerase epsilon and GINS complex and their role in DNA replication. *J Biol Chem,* 286**,** 28963-77.
BLEICHERT, F., BOTCHAN, M. R. & BERGER, J. M. 2015. Crystal structure of the eukaryotic origin recognition complex. *Nature,* 519**,** 321-6.
BLEICHERT, F., BOTCHAN, M. R. & BERGER, J. M. 2017. Mechanisms for initiating cellular DNA replication. *Science,* 355.
BLOW, J. J. & LASKEY, R. A. 1986. Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of Xenopus eggs. *Cell,* 47**,** 577-87.
BLOW, J. J. & LASKEY, R. A. 1988. A role for the nuclear envelope in controlling DNA replication within the cell cycle. *Nature,* 332**,** 546-8.
BOCHMAN, M. L., BELL, S. P. & SCHWACHA, A. 2008. Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability. *Mol Cell Biol,* 28**,** 5865-73.
BOCHMAN, M. L. & 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. *J Biol Chem,* 282**,** 33795-804.
BOCHMAN, M. L. & SCHWACHA, A. 2008. The Mcm2-7 complex has in vitro helicase activity. *Mol Cell,* 31**,** 287-93.
BOCHMAN, M. L. & SCHWACHA, A. 2009. The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiol Mol Biol Rev,* 73**,** 652-83.
BOOS, D., SANCHEZ-PULIDO, L., RAPPAS, M., PEARL, L. H., OLIVER, A. W., PONTING, C. P. & DIFFLEY, J. F. 2011. Regulation of DNA replication through Sld3-Dpb11 interaction is conserved from yeast to humans. *Curr Biol,* 21**,** 1152-7.
BOOS, D., YEKEZARE, M. & DIFFLEY, J. F. 2013. Identification of a heteromeric complex that promotes DNA replication origin firing in human cells. *Science,* 340**,** 981-4.
BOUSSET, K. & DIFFLEY, J. F. 1998. The Cdc7 protein kinase is required for origin firing during S phase. *Genes Dev,* 12**,** 480-90.
BREWSTER, A. S., WANG, G., YU, X., GREENLEAF, W. B., CARAZO, J. M., TJAJADI, M., KLEIN, M. G. & CHEN, X. S. 2008. Crystal structure of a near-full-length

archaeal MCM: functional insights for an AAA+ hexameric helicase. *Proc Natl Acad Sci U S A,* 105**,** 20191-6.
BRUCK, I. & KAPLAN, D. L. 2013. Cdc45 protein-single-stranded DNA interaction is important for stalling the helicase during replication stress. *J Biol Chem,* 288**,** 7550-63.
BUENO, A. & RUSSELL, P. 1992. Dual functions of CDC6: a yeast protein required for DNA replication also inhibits nuclear division. *EMBO J,* 11**,** 2167-76.
BURGERS, P. M. J. & KUNKEL, T. A. 2017. Eukaryotic DNA Replication Fork. *Annu Rev Biochem,* 86**,** 417-438.
CARNEY, S. M., GOMATHINAYAGAM, S., LEUBA, S. H. & TRAKSELIS, M. A. 2017. Bacterial DnaB helicase interacts with the excluded strand to regulate unwinding. *J Biol Chem,* 292**,** 19001-19012.
CHANG, F., RIERA, A., EVRIN, C., SUN, J., LI, H., SPECK, C. & WEINREICH, M. 2015. Cdc6 ATPase activity disengages Cdc6 from the pre-replicative complex to promote DNA replication. *Elife,* 4.
CHANG, Y. P., WANG, G., BERMUDEZ, V., HURWITZ, J. & CHEN, X. S. 2007. Crystal structure of the GINS complex and functional insights into its role in DNA replication. *Proc Natl Acad Sci U S A,* 104**,** 12685-90.
CHEN, S., DE VRIES, M. A. & BELL, S. P. 2007. Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading. *Genes Dev,* 21**,** 2897-907.
CHENG, K., XU, H., CHEN, X., WANG, L., TIAN, B., ZHAO, Y. & HUA, Y. 2016. Structural basis for DNA 5 -end resection by RecJ. *Elife,* 5**,** e14294.
CHESNOKOV, I. N., CHESNOKOVA, O. N. & BOTCHAN, M. 2003. A cytokinetic function of Drosophila ORC6 protein resides in a domain distinct from its replication activity. *Proc Natl Acad Sci U S A,* 100**,** 9150-5.
CHILKOVA, O., STENLUND, P., ISOZ, I., STITH, C. M., GRABOWSKI, P., LUNDSTROM, E. B., BURGERS, P. M. & JOHANSSON, E. 2007. The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res,* 35**,** 6588-97.
CHOI, J. M., LIM, H. S., KIM, J. J., SONG, O. K. & CHO, Y. 2007. Crystal structure of the human GINS complex. *Genes Dev,* 21**,** 1316-21.
CHONG, J. P., HAYASHI, M. K., SIMON, M. N., XU, R. M. & STILLMAN, B. 2000. A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. *Proc Natl Acad Sci U S A,* 97**,** 1530-5.
CHRISTENSEN, T. W. & TYE, B. K. 2003. Drosophila MCM10 interacts with members of the prereplication complex and is required for proper chromosome condensation. *Mol Biol Cell,* 14**,** 2206-15.
CLAUSEN, A. R., LUJAN, S. A., BURKHOLDER, A. B., OREBAUGH, C. D., WILLIAMS, J. S., CLAUSEN, M. F., MALC, E. P., MIECZKOWSKI, P. A., FARGO, D. C., SMITH, D. J. & KUNKEL, T. A. 2015. Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. *Nat Struct Mol Biol,* 22**,** 185-91.
COCKER, J. H., PIATTI, S., SANTOCANALE, C., NASMYTH, K. & DIFFLEY, J. F. 1996. An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. *Nature,* 379**,** 180-2.
COLEMAN, T. R., CARPENTER, P. B. & 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.
COSTA, A., HOOD, I. V. & BERGER, J. M. 2013. Mechanisms for initiating cellular DNA replication. *Annu Rev Biochem,* 82**,** 25-54.
COSTA, A., ILVES, I., TAMBERG, N., PETOJEVIC, T., NOGALES, E., BOTCHAN, M. R. & BERGER, J. M. 2011. The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol,* 18**,** 471-7.

COSTA, A. & ONESTI, S. 2009. Structural biology of MCM helicases. *Crit Rev Biochem Mol Biol,* 44**,** 326-42.
COSTA, A., PAPE, T., VAN HEEL, M., BRICK, P., PATWARDHAN, A. & ONESTI, S. 2006. Structural basis of the Methanothermobacter thermautotrophicus MCM helicase activity. *Nucleic Acids Res,* 34**,** 5829-38.
COSTA, A., RENAULT, L., SWUEC, P., PETOJEVIC, T., PESAVENTO, J. J., ILVES, I., MACLELLAN-GIBSON, K., FLECK, R. A., BOTCHAN, M. R. & BERGER, J. M. 2014. DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. *Elife,* 3**,** e03273.
COSTA, A., VAN DUINEN, G., MEDAGLI, B., CHONG, J., SAKAKIBARA, N., KELMAN, Z., NAIR, S. K., PATWARDHAN, A. & ONESTI, S. 2008. Cryo-electron microscopy reveals a novel DNA-binding site on the MCM helicase. *EMBO J,* 27**,** 2250-8.
COSTER, G. & DIFFLEY, J. F. X. 2017. Bidirectional eukaryotic DNA replication is established by quasi-symmetrical helicase loading. *Science,* 357**,** 314-318.
COSTER, G., FRIGOLA, J., BEURON, F., MORRIS, E. P. & DIFFLEY, J. F. 2014. Origin licensing requires ATP binding and hydrolysis by the MCM replicative helicase. *Mol Cell,* 55**,** 666-77.
DALTON, S. & HOPWOOD, B. 1997. Characterization of Cdc47p-minichromosome maintenance complexes in Saccharomyces cerevisiae: identification of Cdc45p as a subunit. *Mol Cell Biol,* 17**,** 5867-75.
DAS-BRADOO, S., RICKE, R. M. & BIELINSKY, A. K. 2006. Interaction between PCNA and diubiquitinated Mcm10 is essential for cell growth in budding yeast. *Mol Cell Biol,* 26**,** 4806-17.
DAVEY, M. J., INDIANI, C. & O'DONNELL, M. 2003. Reconstitution of the Mcm2-7p heterohexamer, subunit arrangement, and ATP site architecture. *J Biol Chem,* 278**,** 4491-9.
DEAN, F. B. & HURWITZ, J. 1991. Simian virus 40 large T antigen untwists DNA at the origin of DNA replication. *J Biol Chem,* 266**,** 5062-71.
DEEGAN, T. D. & DIFFLEY, J. F. 2016. MCM: one ring to rule them all. *Curr Opin Struct Biol,* 37**,** 145-51.
DEEGAN, T. D., YEELES, J. T. & DIFFLEY, J. F. 2016. Phosphopeptide binding by Sld3 links Dbf4-dependent kinase to MCM replicative helicase activation. *EMBO J,* 35**,** 961-73.
DEVAULT, A., GUEYDON, E. & SCHWOB, E. 2008. Interplay between S-cyclin-dependent kinase and Dbf4-dependent kinase in controlling DNA replication through phosphorylation of yeast Mcm4 N-terminal domain. *Mol Biol Cell,* 19**,** 2267-77.
DEVAULT, A., VALLEN, E. A., YUAN, T., GREEN, S., BENSIMON, A. & SCHWOB, E. 2002. Identification of Tah11/Sid2 as the ortholog of the replication licensing factor Cdt1 in Saccharomyces cerevisiae. *Curr Biol,* 12**,** 689-94.
DEWAR, J. M., BUDZOWSKA, M. & WALTER, J. C. 2015. The mechanism of DNA replication termination in vertebrates. *Nature,* 525**,** 345-50.
DI PERNA, R., ARIA, V., DE FALCO, M., SANNINO, V., OKOROKOV, A. L., PISANI, F. M. & DE FELICE, M. 2013. The physical interaction of Mcm10 with Cdc45 modulates their DNA-binding properties. *Biochem J,* 454**,** 333-43.
DIFFLEY, J. F. 2001. DNA replication: building the perfect switch. *Curr Biol,* 11**,** R367-70.
DIFFLEY, J. F. & COCKER, J. H. 1992. Protein-DNA interactions at a yeast replication origin. *Nature,* 357**,** 169-72.
DIFFLEY, J. F., COCKER, J. H., DOWELL, S. J. & ROWLEY, A. 1994. Two steps in the assembly of complexes at yeast replication origins in vivo. *Cell,* 78**,** 303-16.

DIFFLEY, J. F. & STILLMAN, B. 1988. Purification of a yeast protein that binds to origins of DNA replication and a transcriptional silencer. *Proc Natl Acad Sci U S A,* 85**,** 2120-4.
DONALDSON, A. D., FANGMAN, W. L. & BREWER, B. J. 1998. Cdc7 is required throughout the yeast S phase to activate replication origins. *Genes Dev,* 12**,** 491-501.
DONOVAN, S., HARWOOD, J., DRURY, L. S. & DIFFLEY, J. F. 1997. Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proc Natl Acad Sci U S A,* 94**,** 5611-6.
DOUGLAS, M. E., ALI, F. A., COSTA, A. & DIFFLEY, J. F. X. 2018. The mechanism of eukaryotic CMG helicase activation. *Nature,* 555**,** 265-268.
DOUGLAS, M. E. & DIFFLEY, J. F. 2016. Recruitment of Mcm10 to Sites of Replication Initiation Requires Direct Binding to the Minichromosome Maintenance (MCM) Complex. *J Biol Chem,* 291**,** 5879-88.
DOWELL, S. J., ROMANOWSKI, P. & DIFFLEY, J. F. 1994. Interaction of Dbf4, the Cdc7 protein kinase regulatory subunit, with yeast replication origins in vivo. *Science,* 265**,** 1243-6.
DU, W., JOSEPHRAJAN, A., ADHIKARY, S., BOWLES, T., BIELINSKY, A. K. & EICHMAN, B. F. 2013. Mcm10 self-association is mediated by an N-terminal coiled-coil domain. *PLoS One,* 8**,** e70518.
DUA, R., LEVY, D. L. & CAMPBELL, J. L. 1998. Role of the putative zinc finger domain of Saccharomyces cerevisiae DNA polymerase epsilon in DNA replication and the S/M checkpoint pathway. *J Biol Chem,* 273**,** 30046-55.
DUA, R., LEVY, D. L. & CAMPBELL, J. L. 1999. Analysis of the essential functions of the C-terminal protein/protein interaction domain of Saccharomyces cerevisiae pol epsilon and its unexpected ability to support growth in the absence of the DNA polymerase domain. *J Biol Chem,* 274**,** 22283-8.
DUDERSTADT, K. E., CHUANG, K. & BERGER, J. M. 2011. DNA stretching by bacterial initiators promotes replication origin opening. *Nature,* 478**,** 209-13.
DUMAS, L. B., LUSSKY, J. P., MCFARLAND, E. J. & SHAMPAY, J. 1982. New temperature-sensitive mutants of Saccharomyces cerevisiae affecting DNA replication. *Mol Gen Genet,* 187**,** 42-6.
DUTTA, A. & BELL, S. P. 1997. Initiation of DNA replication in eukaryotic cells. *Annu Rev Cell Dev Biol,* 13**,** 293-332.
EATON, M. L., GALANI, K., KANG, S., BELL, S. P. & MACALPINE, D. M. 2010. Conserved nucleosome positioning defines replication origins. *Genes Dev,* 24**,** 748-53.
EISENBERG, S., KORZA, G., CARSON, J., LIACHKO, I. & TYE, B. K. 2009. Novel DNA binding properties of the Mcm10 protein from Saccharomyces cerevisiae. *J Biol Chem,* 284**,** 25412-20.
EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr,* 66**,** 486-501.
ENEMARK, E. J. & JOSHUA-TOR, L. 2006. Mechanism of DNA translocation in a replicative hexameric helicase. *Nature,* 442**,** 270-5.
ENEMARK, E. J. & JOSHUA-TOR, L. 2008. On helicases and other motor proteins. *Curr Opin Struct Biol,* 18**,** 243-57.
ERZBERGER, J. P. & BERGER, J. M. 2006. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct,* 35**,** 93-114.
EVRIN, C., CLARKE, P., ZECH, J., LURZ, R., SUN, J., UHLE, S., LI, H., STILLMAN, B. & SPECK, C. 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proc Natl Acad Sci U S A,* 106**,** 20240-5.
EVRIN, C., FERNANDEZ-CID, A., ZECH, J., HERRERA, M. C., RIERA, A., CLARKE, P., BRILL, S., LURZ, R. & SPECK, C. 2013. In the absence of ATPase activity, pre-RC

formation is blocked prior to MCM2-7 hexamer dimerization. *Nucleic Acids Res,* 41**,** 3162-72.
FAIRMAN-WILLIAMS, M. E., GUENTHER, U. P. & JANKOWSKY, E. 2010. SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol,* 20**,** 313-24.
FANG, L., DAVEY, M. J. & O'DONNELL, M. 1999. Replisome assembly at oriC, the replication origin of E. coli, reveals an explanation for initiation sites outside an origin. *Mol Cell,* 4**,** 541-53.
FENG, W. & D'URSO, G. 2001. Schizosaccharomyces pombe cells lacking the amino-terminal catalytic domains of DNA polymerase epsilon are viable but require the DNA damage checkpoint control. *Mol Cell Biol,* 21**,** 4495-504.
FERNANDEZ-CID, A., RIERA, A., TOGNETTI, S., HERRERA, M. C., SAMEL, S., EVRIN, C., WINKLER, C., GARDENAL, E., UHLE, S. & SPECK, C. 2013. An ORC/Cdc6/MCM2-7 complex is formed in a multistep reaction to serve as a platform for MCM double-hexamer assembly. *Mol Cell,* 50**,** 577-88.
FIEN, K., CHO, Y. S., LEE, J. K., RAYCHAUDHURI, S., TAPPIN, I. & HURWITZ, J. 2004. Primer utilization by DNA polymerase alpha-primase is influenced by its interaction with Mcm10p. *J Biol Chem,* 279**,** 16144-53.
FLETCHER, R. J., BISHOP, B. E., LEON, R. P., SCLAFANI, R. A., OGATA, C. M. & CHEN, X. S. 2003. The structure and function of MCM from archaeal M. Thermoautotrophicum. *Nat Struct Biol,* 10**,** 160-7.
FLETCHER, R. J., SHEN, J., GOMEZ-LLORENTE, Y., MARTIN, C. S., CARAZO, J. M. & CHEN, X. S. 2005. Double hexamer disruption and biochemical activities of Methanobacterium thermoautotrophicum MCM. *J Biol Chem,* 280**,** 42405-10.
FORSBURG, S. L. 2004. Eukaryotic MCM proteins: beyond replication initiation. *Microbiol Mol Biol Rev,* 68**,** 109-31.
FORSTER, F., HAN, B. G. & BECK, M. 2010. Visual proteomics. *Methods Enzymol,* 483**,** 215-43.
FOSS, M., MCNALLY, F. J., LAURENSON, P. & RINE, J. 1993. Origin recognition complex (ORC) in transcriptional silencing and DNA replication in S. cerevisiae. *Science,* 262**,** 1838-44.
FRANCIS, L. I., RANDELL, J. C., TAKARA, T. J., UCHIMA, L. & BELL, S. P. 2009. Incorporation into the prereplicative complex activates the Mcm2-7 helicase for Cdc7-Dbf4 phosphorylation. *Genes Dev,* 23**,** 643-54.
FRANK, J. 2017. Time-resolved cryo-electron microscopy: Recent progress. *J Struct Biol,* 200**,** 303-306.
FRICK, D. N. & RICHARDSON, C. C. 2001. DNA primases. *Annu Rev Biochem,* 70**,** 39-80.
FRIGOLA, J., HE, J., KINKELIN, K., PYE, V. E., RENAULT, L., DOUGLAS, M. E., REMUS, D., CHEREPANOV, P., COSTA, A. & DIFFLEY, J. F. X. 2017. Cdt1 stabilizes an open MCM ring for helicase loading. *Nat Commun,* 8**,** 15720.
FRIGOLA, J., REMUS, D., MEHANNA, A. & DIFFLEY, J. F. 2013. ATPase-dependent quality control of DNA replication origin licensing. *Nature,* 495**,** 339-43.
FROELICH, C. A., KANG, S., EPLING, L. B., BELL, S. P. & ENEMARK, E. J. 2014. A conserved MCM single-stranded DNA binding element is essential for replication initiation. *Elife,* 3**,** e01993.
FU, Y. V., YARDIMCI, H., LONG, D. T., HO, T. V., GUAINAZZI, A., BERMUDEZ, V. P., HURWITZ, J., VAN OIJEN, A., SCHARER, O. D. & WALTER, J. C. 2011. Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell,* 146**,** 931-41.
GAI, D., CHANG, Y. P. & CHEN, X. S. 2010. Origin DNA melting and unwinding in DNA replication. *Curr Opin Struct Biol,* 20**,** 756-62.
GAI, D., WANG, D., LI, S. X. & CHEN, X. S. 2016. The structure of SV40 large T hexameric helicase in complex with AT-rich origin DNA. *Elife,* 5.

GAI, D., ZHAO, R., LI, D., FINKIELSTEIN, C. V. & 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. & LABIB, K. 2006. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol,* 8**,** 358-66.
GAMBUS, A., VAN DEURSEN, F., POLYCHRONOPOULOS, D., FOLTMAN, M., JONES, R. C., EDMONDSON, R. D., CALZADA, A. & LABIB, K. 2009. A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome. *EMBO J,* 28**,** 2992-3004.
GANAI, R. A., OSTERMAN, P. & JOHANSSON, E. 2015. Yeast DNA polymerase catalytic core and holoenzyme have comparable catalytic rates. *J Biol Chem,* 290**,** 3825-35.
GAUDIER, M., SCHUWIRTH, B. S., WESTCOTT, S. L. & WIGLEY, D. B. 2007. Structural basis of DNA replication origin recognition by an ORC protein. *Science,* 317**,** 1213-6.
GAVIN, K. A., HIDAKA, M. & STILLMAN, B. 1995. Conserved initiator proteins in eukaryotes. *Science,* 270**,** 1667-71.
GEORGESCU, R., YUAN, Z., BAI, L., DE LUNA ALMEIDA SANTOS, R., SUN, J., ZHANG, D., YURIEVA, O., LI, H. & O'DONNELL, M. E. 2017. Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. *Proc Natl Acad Sci U S A,* 114**,** E697-E706.
GEORGESCU, R. E., LANGSTON, L., YAO, N. Y., YURIEVA, O., ZHANG, D., FINKELSTEIN, J., AGARWAL, T. & O'DONNELL, M. E. 2014. Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat Struct Mol Biol,* 21**,** 664-70.
GEORGESCU, R. E., SCHAUER, G. D., YAO, N. Y., LANGSTON, L. D., YURIEVA, O., ZHANG, D., FINKELSTEIN, J. & O'DONNELL, M. E. 2015. Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. *Elife,* 4**,** e04988.
GHOSH, M., LIU, G., RANDALL, G., BEVINGTON, J. & LEFFAK, M. 2004. Transcription factor binding and induced transcription alter chromosomal c-myc replicator activity. *Mol Cell Biol,* 24**,** 10193-207.
GILLESPIE, P. J., LI, A. & BLOW, J. J. 2001. Reconstitution of licensed replication origins on Xenopus sperm nuclei using purified proteins. *BMC Biochem,* 2**,** 15.
GOMEZ-LLORENTE, Y., FLETCHER, R. J., CHEN, X. S., CARAZO, J. M. & SAN MARTIN, C. 2005. Polymorphism and double hexamer structure in the archaeal minichromosome maintenance (MCM) helicase from Methanobacterium thermoautotrophicum. *J Biol Chem,* 280**,** 40909-15.
GRAHAM, B. W., SCHAUER, G. D., LEUBA, S. H. & 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 Res,* 39**,** 6585-95.
GRAHAM, B. W., TAO, Y., DODGE, K. L., THAXTON, C. T., OLASO, D., YOUNG, N. L., MARSHALL, A. G. & TRAKSELIS, M. A. 2016. DNA Interactions Probed by Hydrogen-Deuterium Exchange (HDX) Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Confirm External Binding Sites on the Minichromosomal Maintenance (MCM) Helicase. *J Biol Chem,* 291**,** 12467-80.
GROS, J., KUMAR, C., LYNCH, G., YADAV, T., WHITEHOUSE, I. & REMUS, D. 2015. Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. *Mol Cell,* 60**,** 797-807.
HA, T., ENDERLE, T., OGLETREE, D. F., CHEMLA, D. S., SELVIN, P. R. & WEISS, S. 1996. Probing the interaction between two single molecules: fluorescence resonance

energy transfer between a single donor and a single acceptor. *Proc Natl Acad Sci U S A,* 93**,** 6264-8.
HARAMI, G. M., GYIMESI, M. & KOVACS, M. 2013. From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins. *Trends Biochem Sci,* 38**,** 364-71.
HARDY, C. F., DRYGA, O., SEEMATTER, S., PAHL, P. M. & SCLAFANI, R. A. 1997. mcm5/cdc46-bob1 bypasses the requirement for the S phase activator Cdc7p. *Proc Natl Acad Sci U S A,* 94**,** 3151-5.
HARLAND, R. M. & LASKEY, R. A. 1980. Regulated replication of DNA microinjected into eggs of Xenopus laevis. *Cell,* 21**,** 761-71.
HARTWELL, L. H. 1976. Sequential function of gene products relative to DNA synthesis in the yeast cell cycle. *J Mol Biol,* 104**,** 803-17.
HE, H., LI, Y., DONG, Q., CHANG, A. Y., GAO, F., CHI, Z., SU, M., ZHANG, F., BAN, H., MARTIENSSEN, R., CHEN, Y. H. & LI, F. 2017. Coordinated regulation of heterochromatin inheritance by Dpb3-Dpb4 complex. *Proc Natl Acad Sci U S A,* 114**,** 12524-12529.
HELLER, R. C., KANG, S., LAM, W. M., CHEN, S., CHAN, C. S. & BELL, S. P. 2011. Eukaryotic origin-dependent DNA replication in vitro reveals sequential action of DDK and S-CDK kinases. *Cell,* 146**,** 80-91.
HERBIG, U., MARLAR, C. A. & FANNING, E. 1999. The Cdc6 nucleotide-binding site regulates its activity in DNA replication in human cells. *Mol Biol Cell,* 10**,** 2631-45.
HEYMANN, J. B. 2018. Guidelines for using Bsoft for high resolution reconstruction and validation of biomolecular structures from electron micrographs. *Protein Sci,* 27**,** 159-171.
HIZUME, K., KOMINAMI, H., KOBAYASHI, K., YAMADA, H. & ARAKI, H. 2017. Flexible DNA Path in the MCM Double Hexamer Loaded on DNA. *Biochemistry,* 56**,** 2435-2445.
HOFMANN, J. F. & BEACH, D. 1994. cdt1 is an essential target of the Cdc10/Sct1 transcription factor: requirement for DNA replication and inhibition of mitosis. *EMBO J,* 13**,** 425-34.
HOGG, M. & JOHANSSON, E. 2012. DNA polymerase epsilon. *Subcell Biochem,* 62**,** 237-57.
HOGG, M., OSTERMAN, P., BYLUND, G. O., GANAI, R. A., LUNDSTROM, E. B., SAUER-ERIKSSON, A. E. & JOHANSSON, E. 2014. Structural basis for processive DNA synthesis by yeast DNA polymerase varepsilon. *Nat Struct Mol Biol,* 21**,** 49-55.
HOHN, M., TANG, G., GOODYEAR, G., BALDWIN, P. R., HUANG, Z., PENCZEK, P. A., YANG, C., GLAESER, R. M., ADAMS, P. D. & LUDTKE, S. J. 2007. SPARX, a new environment for Cryo-EM image processing. *J Struct Biol,* 157**,** 47-55.
ILVES, I., PETOJEVIC, T., PESAVENTO, J. J. & BOTCHAN, M. R. 2010. Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell,* 37**,** 247-58.
IM, J. S., KI, S. H., FARINA, A., JUNG, D. S., HURWITZ, J. & LEE, J. K. 2009. Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins. *Proc Natl Acad Sci U S A,* 106**,** 15628-32.
ISHIMI, Y. 1997. A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex. *J Biol Chem,* 272**,** 24508-13.
ISOZ, I., PERSSON, U., VOLKOV, K. & JOHANSSON, E. 2012. The C-terminus of Dpb2 is required for interaction with Pol2 and for cell viability. *Nucleic Acids Res,* 40**,** 11545-53.
ITOU, H., MURAMATSU, S., SHIRAKIHARA, Y. & ARAKI, H. 2014. Crystal structure of the homology domain of the eukaryotic DNA replication proteins Sld3/Treslin. *Structure,* 22**,** 1341-1347.

ITOU, H., SHIRAKIHARA, Y. & ARAKI, H. 2015. The quaternary structure of the eukaryotic DNA replication proteins Sld7 and Sld3. *Acta Crystallogr D Biol Crystallogr,* 71**,** 1649-56.
ITSATHITPHAISARN, O., WING, R. A., ELIASON, W. K., WANG, J. & STEITZ, T. A. 2012. The hexameric helicase DnaB adopts a nonplanar conformation during translocation. *Cell,* 151**,** 267-77.
IYER, L. M., LEIPE, D. D., KOONIN, E. V. & ARAVIND, L. 2004. Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol,* 146**,** 11-31.
JACKSON, A. L., PAHL, P. M., HARRISON, K., ROSAMOND, J. & SCLAFANI, R. A. 1993. Cell cycle regulation of the yeast Cdc7 protein kinase by association with the Dbf4 protein. *Mol Cell Biol,* 13**,** 2899-908.
JENKINSON, E. R. & CHONG, J. P. 2006. Minichromosome maintenance helicase activity is controlled by N- and C-terminal motifs and requires the ATPase domain helix-2 insert. *Proc Natl Acad Sci U S A,* 103**,** 7613-8.
JENKINSON, E. R., COSTA, A., LEECH, A. P., PATWARDHAN, A., ONESTI, S. & CHONG, J. P. 2009. Mutations in subdomain B of the minichromosome maintenance (MCM) helicase affect DNA binding and modulate conformational transitions. *J Biol Chem,* 284**,** 5654-61.
JOHANSSON, E. & DIXON, N. 2013. Replicative DNA polymerases. *Cold Spring Harb Perspect Biol,* 5.
JOHNSON, R. E., KLASSEN, R., PRAKASH, L. & PRAKASH, S. 2015. A Major Role of DNA Polymerase delta in Replication of Both the Leading and Lagging DNA Strands. *Mol Cell,* 59**,** 163-175.
KAMADA, K. 2012. The GINS complex: structure and function. *Subcell Biochem,* 62**,** 135-56.
KAMADA, K., KUBOTA, Y., ARATA, T., SHINDO, Y. & HANAOKA, F. 2007. Structure of the human GINS complex and its assembly and functional interface in replication initiation. *Nat Struct Mol Biol,* 14**,** 388-96.
KAMIMURA, Y., TAK, Y. S., SUGINO, A. & ARAKI, H. 2001. Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in Saccharomyces cerevisiae. *EMBO J,* 20**,** 2097-107.
KANEMAKI, M. & LABIB, K. 2006. Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. *EMBO J,* 25**,** 1753-63.
KANEMAKI, M., SANCHEZ-DIAZ, A., GAMBUS, A. & LABIB, K. 2003. Functional proteomic identification of DNA replication proteins by induced proteolysis in vivo. *Nature,* 423**,** 720-4.
KANG, S., WARNER, M. D. & BELL, S. P. 2014. Multiple functions for Mcm2-7 ATPase motifs during replication initiation. *Mol Cell,* 55**,** 655-65.
KANG, Y. H., GALAL, W. C., FARINA, A., TAPPIN, I. & 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. *Proc Natl Acad Sci U S A,* 109**,** 6042-7.
KANKE, M., KODAMA, Y., TAKAHASHI, T. S., NAKAGAWA, T. & MASUKATA, H. 2012. Mcm10 plays an essential role in origin DNA unwinding after loading of the CMG components. *EMBO J,* 31**,** 2182-94.
KAPLAN, D. L., DAVEY, M. J. & 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. *J Biol Chem,* 278**,** 49171-82.
KASIVISWANATHAN, R., SHIN, J. H., MELAMUD, E. & KELMAN, Z. 2004. Biochemical characterization of the Methanothermobacter thermautotrophicus minichromosome maintenance (MCM) helicase N-terminal domains. *J Biol Chem,* 279**,** 28358-66.

KELMAN, Z. & HURWITZ, J. 2003. Structural lessons in DNA replication from the third domain of life. *Nat Struct Biol,* 10**,** 148-50.
KELMAN, Z., LEE, J. K. & HURWITZ, J. 1999. The single minichromosome maintenance protein of Methanobacterium thermoautotrophicum DeltaH contains DNA helicase activity. *Proc Natl Acad Sci U S A,* 96**,** 14783-8.
KESTI, T., FLICK, K., KERANEN, S., SYVAOJA, J. E. & WITTENBERG, C. 1999. DNA polymerase epsilon catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. *Mol Cell,* 3**,** 679-85.
KIMANIUS, D., FORSBERG, B. O., SCHERES, S. H. & LINDAHL, E. 2016. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife,* 5.
KITADA, K., JOHNSTON, L. H., SUGINO, T. & SUGINO, A. 1992. Temperature-sensitive cdc7 mutations of Saccharomyces cerevisiae are suppressed by the DBF4 gene, which is required for the G1/S cell cycle transition. *Genetics,* 131**,** 21-9.
KLEMM, R. D., AUSTIN, R. J. & BELL, S. P. 1997. Coordinate binding of ATP and origin DNA regulates the ATPase activity of the origin recognition complex. *Cell,* 88**,** 493-502.
KOMATA, M., BANDO, M., ARAKI, H. & SHIRAHIGE, K. 2009. The direct binding of Mrc1, a checkpoint mediator, to Mcm6, a replication helicase, is essential for the replication checkpoint against methyl methanesulfonate-induced stress. *Mol Cell Biol,* 29**,** 5008-19.
KOONIN, E. V., WOLF, Y. I. & ARAVIND, L. 2000. Protein fold recognition using sequence profiles and its application in structural genomics. *Adv Protein Chem,* 54**,** 245-75.
KRASTANOVA, I., SANNINO, V., AMENITSCH, H., GILEADI, O., PISANI, F. M. & ONESTI, S. 2012. Structural and functional insights into the DNA replication factor Cdc45 reveal an evolutionary relationship to the DHH family of phosphoesterases. *J Biol Chem,* 287**,** 4121-8.
KUBOTA, Y., TAKASE, Y., KOMORI, Y., HASHIMOTO, Y., ARATA, T., KAMIMURA, Y., ARAKI, H. & TAKISAWA, H. 2003. A novel ring-like complex of Xenopus proteins essential for the initiation of DNA replication. *Genes Dev,* 17**,** 1141-52.
KUCHTA, R. D. & STENGEL, G. 2010. Mechanism and evolution of DNA primases. *Biochim Biophys Acta,* 1804**,** 1180-9.
KUCUKELBIR, A., SIGWORTH, F. J. & TAGARE, H. D. 2014. Quantifying the local resolution of cryo-EM density maps. *Nat Methods,* 11**,** 63-5.
KUHLBRANDT, W. 2014. Biochemistry. The resolution revolution. *Science,* 343**,** 1443-4.
KUMAGAI, A., SHEVCHENKO, A., SHEVCHENKO, A. & DUNPHY, W. G. 2010. Treslin collaborates with TopBP1 in triggering the initiation of DNA replication. *Cell,* 140**,** 349-59.
KUNKEL, T. A. & BURGERS, P. M. 2008. Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol,* 18**,** 521-7.
KUO, A. J., SONG, J., CHEUNG, P., ISHIBE-MURAKAMI, S., YAMAZOE, S., CHEN, J. K., PATEL, D. J. & GOZANI, O. 2012. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature,* 484**,** 115-9.
KURAT, C. F., YEELES, J. T. P., PATEL, H., EARLY, A. & DIFFLEY, J. F. X. 2017. Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates. *Mol Cell,* 65**,** 117-130.
LABIB, K. 2010. How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes Dev,* 24**,** 1208-19.
LABIB, K., DIFFLEY, J. F. & KEARSEY, S. E. 1999. G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus. *Nat Cell Biol,* 1**,** 415-22.

LABIB, K. & GAMBUS, A. 2007. A key role for the GINS complex at DNA replication forks. *Trends Cell Biol,* 17**,** 271-8.
LABIB, K., TERCERO, J. A. & DIFFLEY, J. F. 2000. Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science,* 288**,** 1643-7.
LAM, S. K., MA, X., SING, T. L., SHILTON, B. H., BRANDL, C. J. & DAVEY, M. J. 2013. The PS1 hairpin of Mcm3 is essential for viability and for DNA unwinding in vitro. *PLoS One,* 8**,** e82177.
LANGSTON, L. & O'DONNELL, M. 2017. Action of CMG with strand-specific DNA blocks supports an internal unwinding mode for the eukaryotic replicative helicase. *Elife,* 6.
LANGSTON, L. D., MAYLE, R., SCHAUER, G. D., YURIEVA, O., ZHANG, D., YAO, N. Y., GEORGESCU, R. E. & O'DONNELL, M. E. 2017. Mcm10 promotes rapid isomerization of CMG-DNA for replisome bypass of lagging strand DNA blocks. *Elife,* 6.
LANGSTON, L. D., ZHANG, D., YURIEVA, O., GEORGESCU, R. E., FINKELSTEIN, J., YAO, N. Y., INDIANI, C. & O'DONNELL, M. E. 2014. CMG helicase and DNA polymerase epsilon form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc Natl Acad Sci U S A,* 111**,** 15390-5.
LEIPE, D. D., ARAVIND, L. & KOONIN, E. V. 1999. Did DNA replication evolve twice independently? *Nucleic Acids Res,* 27**,** 3389-401.
LI, D., ZHAO, R., LILYESTROM, W., GAI, D., ZHANG, R., DECAPRIO, J. A., FANNING, E., JOCHIMIAK, A., SZAKONYI, G. & CHEN, X. S. 2003. Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature,* 423**,** 512-8.
LI, H. & STILLMAN, B. 2012. The origin recognition complex: a biochemical and structural view. *Subcell Biochem,* 62**,** 37-58.
LI, N., LAM, W. H., ZHAI, Y., CHENG, J., CHENG, E., ZHAO, Y., GAO, N. & TYE, B. K. 2018. Structure of the origin recognition complex bound to DNA replication origin. *Nature,* 559**,** 217-222.
LI, N., ZHAI, Y., ZHANG, Y., LI, W., YANG, M., LEI, J., TYE, B. K. & GAO, N. 2015. Structure of the eukaryotic MCM complex at 3.8 A. *Nature,* 524**,** 186-91.
LI, X., MOONEY, P., ZHENG, S., BOOTH, C. R., BRAUNFELD, M. B., GUBBENS, S., AGARD, D. A. & CHENG, Y. 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods,* 10**,** 584-90.
LI, Z., PAN, M., SANTANGELO, T. J., CHEMNITZ, W., YUAN, W., EDWARDS, J. L., HURWITZ, J., REEVE, J. N. & KELMAN, Z. 2011. A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res,* 39**,** 6114-23.
LIANG, C., WEINREICH, M. & STILLMAN, B. 1995. ORC and Cdc6p interact and determine the frequency of initiation of DNA replication in the genome. *Cell,* 81**,** 667-76.
LIPFORD, J. R. & BELL, S. P. 2001. Nucleosomes positioned by ORC facilitate the initiation of DNA replication. *Mol Cell,* 7**,** 21-30.
LIU, S., BALASOV, M., WANG, H., WU, L., CHESNOKOV, I. N. & LIU, Y. 2011. Structural analysis of human Orc6 protein reveals a homology with transcription factor TFIIB. *Proc Natl Acad Sci U S A,* 108**,** 7373-8.
LIU, W., PUCCI, B., ROSSI, M., PISANI, F. M. & LADENSTEIN, R. 2008. Structural analysis of the Sulfolobus solfataricus MCM protein N-terminal domain. *Nucleic Acids Res,* 36**,** 3235-43.
LIU, Y., RICHARDS, T. A. & AVES, S. J. 2009. Ancient diversification of eukaryotic MCM DNA replication proteins. *BMC Evol Biol,* 9**,** 60.
LOOKE, M., MALONEY, M. F. & BELL, S. P. 2017. Mcm10 regulates DNA replication elongation by stimulating the CMG replicative helicase. *Genes Dev,* 31**,** 291-305.

LYUBIMOV, A. Y., COSTA, A., BLEICHERT, F., BOTCHAN, M. R. & BERGER, J. M. 2012. ATP-dependent conformational dynamics underlie the functional asymmetry of the replicative helicase from a minimalist eukaryote. *Proc Natl Acad Sci U S A,* 109**,** 11999-2004.
LYUBIMOV, A. Y., STRYCHARSKA, M. & BERGER, J. M. 2011. The nuts and bolts of ring-translocase structure and mechanism. *Curr Opin Struct Biol,* 21**,** 240-8.
MACALPINE, H. K., GORDAN, R., POWELL, S. K., HARTEMINK, A. J. & MACALPINE, D. M. 2010. Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res,* 20**,** 201-11.
MAINE, G. T., SINHA, P. & TYE, B. K. 1984. Mutants of S. cerevisiae defective in the maintenance of minichromosomes. *Genetics,* 106**,** 365-85.
MAIORANO, D., MOREAU, J. & MECHALI, M. 2000. XCDT1 is required for the assembly of pre-replicative complexes in Xenopus laevis. *Nature,* 404**,** 622-5.
MAKAROVA, K. S., KOONIN, E. V. & KELMAN, Z. 2012. The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol Direct,* 7**,** 7.
MAKAROVA, K. S., WOLF, Y. I., MEKHEDOV, S. L., MIRKIN, B. G. & KOONIN, E. V. 2005. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res,* 33**,** 4626-38.
MARAHRENS, Y. & STILLMAN, B. 1992. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science,* 255**,** 817-23.
MARTIN, A., BAKER, T. A. & SAUER, R. T. 2005. Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature,* 437**,** 1115-20.
MASAI, H., TANIYAMA, C., OGINO, K., MATSUI, E., KAKUSHO, N., MATSUMOTO, S., KIM, J. M., ISHII, A., TANAKA, T., KOBAYASHI, T., TAMAI, K., OHTANI, K. & ARAI, K. 2006. Phosphorylation of MCM4 by Cdc7 kinase facilitates its interaction with Cdc45 on the chromatin. *J Biol Chem,* 281**,** 39249-61.
MASUMOTO, H., MURAMATSU, S., KAMIMURA, Y. & ARAKI, H. 2002. S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. *Nature,* 415**,** 651-5.
MASUMOTO, H., SUGINO, A. & ARAKI, H. 2000. Dpb11 controls the association between DNA polymerases alpha and epsilon and the autonomously replicating sequence region of budding yeast. *Mol Cell Biol,* 20**,** 2809-17.
MCGEOCH, A. T., TRAKSELIS, M. A., LASKEY, R. A. & BELL, S. D. 2005. Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism. *Nat Struct Mol Biol,* 12**,** 756-62.
MECHALI, M., YOSHIDA, K., COULOMBE, P. & PASERO, P. 2013. Genetic and epigenetic determinants of DNA replication origins, position and activation. *Curr Opin Genet Dev,* 23**,** 124-31.
MERCHANT, A. M., KAWASAKI, Y., CHEN, Y., LEI, M. & TYE, B. K. 1997. A lesion in the DNA replication initiation factor Mcm10 induces pausing of elongation forks through chromosomal replication origins in Saccharomyces cerevisiae. *Mol Cell Biol,* 17**,** 3261-71.
MICKLEM, G., ROWLEY, A., HARWOOD, J., NASMYTH, K. & DIFFLEY, J. F. 1993. Yeast origin recognition complex is involved in DNA replication and transcriptional silencing. *Nature,* 366**,** 87-9.
MILLER, J. M., ARACHEA, B. T., EPLING, L. B. & ENEMARK, E. J. 2014. Analysis of the crystal structure of an active MCM hexamer. *Elife,* 3**,** e03433.
MILLER, J. M. & ENEMARK, E. J. 2015. Archaeal MCM Proteins as an Analog for the Eukaryotic Mcm2-7 Helicase to Reveal Essential Features of Structure and Function. *Archaea,* 2015**,** 305497.
MILLER, T. C. & COSTA, A. 2017. The architecture and function of the chromatin replication machinery. *Curr Opin Struct Biol,* 47**,** 9-16.

MIYAZAWA-ONAMI, M., ARAKI, H. & TANAKA, S. 2017. Pre-initiation complex assembly functions as a molecular switch that splits the Mcm2-7 double hexamer. *EMBO Rep,* 18**,** 1752-1761.
MIZUSHIMA, T., TAKAHASHI, N. & STILLMAN, B. 2000. Cdc6p modulates the structure and DNA binding activity of the origin recognition complex in vitro. *Genes Dev,* 14**,** 1631-41.
MOGGS, J. G., GRANDI, P., QUIVY, J. P., JONSSON, Z. O., HUBSCHER, U., BECKER, P. B. & ALMOUZNI, G. 2000. A CAF-1-PCNA-mediated chromatin assembly pathway triggered by sensing DNA damage. *Mol Cell Biol,* 20**,** 1206-18.
MOIR, D., STEWART, S. E., OSMOND, B. C. & BOTSTEIN, D. 1982. Cold-sensitive cell-division-cycle mutants of yeast: isolation, properties, and pseudoreversion studies. *Genetics,* 100**,** 547-63.
MOREAU, M. J., MCGEOCH, A. T., LOWE, A. R., ITZHAKI, L. S. & BELL, S. D. 2007. ATPase site architecture and helicase mechanism of an archaeal MCM. *Mol Cell,* 28**,** 304-14.
MOYER, S. E., LEWIS, P. W. & BOTCHAN, M. R. 2006. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci U S A,* 103**,** 10236-10241.
MULLER, P., PARK, S., SHOR, E., HUEBERT, D. J., WARREN, C. L., ANSARI, A. Z., WEINREICH, M., EATON, M. L., MACALPINE, D. M. & FOX, C. A. 2010. The conserved bromo-adjacent homology domain of yeast Orc1 functions in the selection of DNA replication origins within chromatin. *Genes Dev,* 24**,** 1418-33.
MURAMATSU, S., HIRAI, K., TAK, Y. S., KAMIMURA, Y. & ARAKI, H. 2010. CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol (epsilon}, and GINS in budding yeast. *Genes Dev,* 24**,** 602-12.
MURZIN, A. G. 1993. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J,* 12**,** 861-7.
NAKANE, T., KIMANIUS, D., LINDAHL, E. & SCHERES, S. H. 2018. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife,* 7.
NASMYTH, K. & NURSE, P. 1981. Cell division cycle mutants altered in DNA replication and mitosis in the fission yeast Schizosaccharomyces pombe. *Mol Gen Genet,* 182**,** 119-24.
NEUWALD, A. F., ARAVIND, L., SPOUGE, J. L. & KOONIN, E. V. 1999. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res,* 9**,** 27-43.
NISHITANI, H., LYGEROU, Z., NISHIMOTO, T. & NURSE, P. 2000. The Cdt1 protein is required to license DNA for replication in fission yeast. *Nature,* 404**,** 625-8.
NOGUCHI, Y., YUAN, Z., BAI, L., SCHNEIDER, S., ZHAO, G., STILLMAN, B., SPECK, C. & LI, H. 2017. Cryo-EM structure of Mcm2-7 double hexamer on DNA suggests a lagging-strand DNA extrusion model. *Proc Natl Acad Sci U S A,* 114**,** E9529-E9538.
NUUTINEN, T., TOSSAVAINEN, H., FREDRIKSSON, K., PIRILA, P., PERMI, P., POSPIECH, H. & SYVAOJA, J. E. 2008. The solution structure of the amino-terminal domain of human DNA polymerase epsilon subunit B is homologous to C-domains of AAA+ proteins. *Nucleic Acids Res,* 36**,** 5102-10.
O'DONNELL, M., LANGSTON, L. & STILLMAN, B. 2013. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb Perspect Biol,* 5.
OHYA, T., MAKI, S., KAWASAKI, Y. & SUGINO, A. 2000. Structure and function of the fourth subunit (Dpb4p) of DNA polymerase epsilon in Saccharomyces cerevisiae. *Nucleic Acids Res,* 28**,** 3846-52.
ON, K. F., BEURON, F., FRITH, D., SNIJDERS, A. P., MORRIS, E. P. & DIFFLEY, J. F. 2014. Prereplicative complexes assembled in vitro support origin-dependent and independent DNA replication. *EMBO J,* 33**,** 605-20.

OYAMA, T., ISHINO, S., FUJINO, S., OGINO, H., SHIRAI, T., MAYANAGI, K., SAITO, M., NAGASAWA, N., ISHINO, Y. & MORIKAWA, K. 2011. Architectures of archaeal GINS complexes, essential DNA replication initiation factors. *BMC Biol,* 9**,** 28.
OYAMA, T., ISHINO, S., SHIRAI, T., YAMAGAMI, T., NAGATA, M., OGINO, H., KUSUNOKI, M. & ISHINO, Y. 2016. Atomic structure of an archaeal GAN suggests its dual roles as an exonuclease in DNA repair and a CMG component in DNA replication. *Nucleic Acids Res,* 44**,** 9505-9517.
PACEK, M., TUTTER, A. V., KUBOTA, Y., TAKISAWA, H. & WALTER, J. C. 2006. Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol Cell,* 21**,** 581-7.
PAIXAO, S., COLALUCA, I. N., CUBELLS, M., PEVERALI, F. A., DESTRO, A., GIADROSSI, S., GIACCA, M., FALASCHI, A., RIVA, S. & BIAMONTI, G. 2004. Modular structure of the human lamin B2 replicator. *Mol Cell Biol,* 24**,** 2958-67.
PARKER, M. W., BOTCHAN, M. R. & BERGER, J. M. 2017. Mechanisms and regulation of DNA replication initiation in eukaryotes. *Crit Rev Biochem Mol Biol,* 52**,** 107-144.
PATEL, S. S. & PICHA, K. M. 2000. Structure and function of hexameric helicases. *Annu Rev Biochem,* 69**,** 651-97.
PELLEGRINI, L. 2012. The Pol alpha-primase complex. *Subcell Biochem,* 62**,** 157-69.
PELLEGRINI, L. 2017. Structural insights into Cdc45 function: was there a nuclease at the heart of the ancestral replisome? *Biophys Chem,* 225**,** 10-14.
PERERA, R. L., TORELLA, R., KLINGE, S., KILKENNY, M. L., MAMAN, J. D. & PELLEGRINI, L. 2013. Mechanism for priming DNA synthesis by yeast DNA polymerase alpha. *Elife,* 2**,** e00482.
PERKINS, G. & DIFFLEY, J. F. 1998. Nucleotide-dependent prereplicative complex assembly by Cdc6p, a homolog of eukaryotic and prokaryotic clamp-loaders. *Mol Cell,* 2**,** 23-32.
PETOJEVIC, T., PESAVENTO, J. J., COSTA, A., LIANG, J., WANG, Z., BERGER, J. M. & BOTCHAN, M. R. 2015. Cdc45 (cell division cycle protein 45) guards the gate of the Eukaryote Replisome helicase stabilizing leading strand engagement. *Proc Natl Acad Sci U S A,* 112**,** E249-58.
PLUCIENNIK, A., DZANTIEV, L., IYER, R. R., CONSTANTIN, N., KADYROV, F. A. & MODRICH, P. 2010. PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair. *Proc Natl Acad Sci U S A,* 107**,** 16066-71.
POPLAWSKI, A., GRABOWSKI, B., LONG, S. E. & KELMAN, Z. 2001. The zinc finger domain of the archaeal minichromosome maintenance protein is required for helicase activity. *J Biol Chem,* 276**,** 49371-7.
POZO, P. N. & COOK, J. G. 2016. Regulation and Function of Cdt1; A Key Factor in Cell Proliferation and Genome Stability. *Genes (Basel),* 8.
PROKHOROVA, T. A. & BLOW, J. J. 2000. Sequential MCM/P1 subcomplex assembly is required to form a heterohexamer with replication licensing activity. *J Biol Chem,* 275**,** 2491-8.
PUNJANI, A., RUBINSTEIN, J. L., FLEET, D. J. & BRUBAKER, M. A. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods,* 14**,** 290-296.
PURSELL, Z. F., ISOZ, I., LUNDSTROM, E. B., JOHANSSON, E. & KUNKEL, T. A. 2007. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science,* 317**,** 127-30.
QUAN, Y., XIA, Y., LIU, L., CUI, J., LI, Z., CAO, Q., CHEN, X. S., CAMPBELL, J. L. & LOU, H. 2015. Cell-Cycle-Regulated Interaction between Mcm10 and Double Hexameric Mcm2-7 Is Required for Helicase Splitting and Activation during S Phase. *Cell Rep,* 13**,** 2576-2586.

RAMER, M. D., SUMAN, E. S., RICHTER, H., STANGER, K., SPRANGER, M., BIEBERSTEIN, N. & DUNCKER, B. P. 2013. Dbf4 and Cdc7 proteins promote DNA replication through interactions with distinct Mcm2-7 protein subunits. *J Biol Chem,* 288**,** 14926-35.
RANDELL, J., GOYAL, M., SAUNDERS, J. & REED, P. 2011. Effect of a context of concrete and abstract words on hallucinatory content in individuals scoring high in schizotypy. *J Behav Ther Exp Psychiatry,* 42**,** 149-53.
RANDELL, J. C., BOWERS, J. L., RODRIGUEZ, H. K. & BELL, S. P. 2006. Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol Cell,* 21**,** 29-39.
RANDELL, J. C., FAN, A., CHAN, C., FRANCIS, L. I., HELLER, R. C., GALANI, K. & BELL, S. P. 2010. Mec1 is one of multiple kinases that prime the Mcm2-7 helicase for phosphorylation by Cdc7. *Mol Cell,* 40**,** 353-63.
RAO, P. N. & JOHNSON, R. T. 1970. Mammalian cell fusion: studies on the regulation of DNA synthesis and mitosis. *Nature,* 225**,** 159-64.
REMUS, D., BEALL, E. L. & BOTCHAN, M. R. 2004. DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. *EMBO J,* 23**,** 897-907.
REMUS, D., BEURON, F., TOLUN, G., GRIFFITH, J. D., MORRIS, E. P. & DIFFLEY, J. F. 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell,* 139**,** 719-30.
REMUS, D. & DIFFLEY, J. F. 2009. Eukaryotic DNA replication control: lock and load, then fire. *Curr Opin Cell Biol,* 21**,** 771-7.
RICKE, R. M. & BIELINSKY, A. K. 2004. Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. *Mol Cell,* 16**,** 173-85.
ROBERTSON, P. D., WARREN, E. M., ZHANG, H., FRIEDMAN, D. B., LARY, J. W., COLE, J. L., TUTTER, A. V., WALTER, J. C., FANNING, E. & EICHMAN, B. F. 2008. Domain architecture and biochemical characterization of vertebrate Mcm10. *J Biol Chem,* 283**,** 3338-48.
ROBINSON, N. P. & BELL, S. D. 2005. Origins of DNA replication in the three domains of life. *FEBS J,* 272**,** 3757-66.
ROBINSON, N. P., DIONNE, I., LUNDGREN, M., MARSH, V. L., BERNANDER, R. & BELL, S. D. 2004. Identification of two origins of replication in the single chromosome of the archaeon Sulfolobus solfataricus. *Cell,* 116**,** 25-38.
ROHOU, A. & GRIGORIEFF, N. 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol,* 192**,** 216-21.
ROTHENBERG, E., TRAKSELIS, M. A., BELL, S. D. & HA, T. 2007. MCM forked substrate specificity involves dynamic interaction with the 5'-tail. *J Biol Chem,* 282**,** 34229-34.
ROWLEY, A., COCKER, J. H., HARWOOD, J. & DIFFLEY, J. F. 1995. Initiation complex assembly at budding yeast replication origins begins with the recognition of a bipartite sequence by limiting amounts of the initiator, ORC. *EMBO J,* 14**,** 2631-41.
SAIKRISHNAN, K., GRIFFITHS, S. P., COOK, N., COURT, R. & WIGLEY, D. B. 2008. DNA binding to RecD: role of the 1B domain in SF1B helicase activity. *EMBO J,* 27**,** 2222-9.
SAKAKIBARA, N., KASIVISWANATHAN, R., MELAMUD, E., HAN, M., SCHWARZ, F. P. & KELMAN, Z. 2008. Coupling of DNA binding and helicase activity is mediated by a conserved loop in the MCM protein. *Nucleic Acids Res,* 36**,** 1309-20.
SAMEL, S. A., FERNANDEZ-CID, A., SUN, J., RIERA, A., TOGNETTI, S., HERRERA, M. C., LI, H. & SPECK, C. 2014. A unique DNA entry gate serves for regulated loading of the eukaryotic replicative helicase MCM2-7 onto DNA. *Genes Dev,* 28**,** 1653-66.
SAMORA, C. P., SAKSOUK, J., GOSWAMI, P., WADE, B. O., SINGLETON, M. R., BATES, P. A., LENGRONNE, A., COSTA, A. & UHLMANN, F. 2016. Ctf4 Links DNA

Replication with Sister Chromatid Cohesion Establishment by Recruiting the Chl1 Helicase to the Replisome. *Mol Cell,* 63**,** 371-84.
SANCHEZ-PULIDO, L. & PONTING, C. P. 2011. Cdc45: the missing RecJ ortholog in eukaryotes? *Bioinformatics,* 27**,** 1885-8.
SCHERES, S. H. 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol,* 180**,** 519-30.
SCHERES, S. H. 2015. Semi-automated selection of cryo-EM particles in RELION-1.3. *J Struct Biol,* 189**,** 114-22.
SCHWACHA, A. & BELL, S. P. 2001. Interactions between two catalytically distinct MCM subgroups are essential for coordinated ATP hydrolysis and DNA replication. *Mol Cell,* 8**,** 1093-104.
SCLAFANI, R. A. 2000. Cdc7p-Dbf4p becomes famous in the cell cycle. *J Cell Sci,* 113 (Pt 12)**,** 2111-7.
SEMPLE, J. W., DA-SILVA, L. F., JERVIS, E. J., AH-KEE, J., AL-ATTAR, H., KUMMER, L., HEIKKILA, J. J., PASERO, P. & DUNCKER, B. P. 2006. An essential role for Orc6 in DNA replication through maintenance of pre-replicative complexes. *EMBO J,* 25**,** 5150-8.
SENGUPTA, S., VAN DEURSEN, F., DE PICCOLI, G. & LABIB, K. 2013. Dpb2 integrates the leading-strand DNA polymerase into the eukaryotic replisome. *Curr Biol,* 23**,** 543-52.
SHARMA, A., KAUR, M., KAR, A., RANADE, S. M. & SAXENA, S. 2010. Ultraviolet radiation stress triggers the down-regulation of essential replication factor Mcm10. *J Biol Chem,* 285**,** 8352-62.
SHEU, Y. J. & STILLMAN, B. 2006. Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. *Mol Cell,* 24**,** 101-13.
SHEU, Y. J. & STILLMAN, B. 2010. The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature,* 463**,** 113-7.
SHIMA, N., ALCARAZ, A., LIACHKO, I., BUSKE, T. R., ANDREWS, C. A., MUNROE, R. J., HARTFORD, S. A., TYE, B. K. & SCHIMENTI, J. C. 2007. A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. *Nat Genet,* 39**,** 93-8.
SIDDIQUI, K., ON, K. F. & DIFFLEY, J. F. 2013. Regulating DNA replication in eukarya. *Cold Spring Harb Perspect Biol,* 5.
SIMON, A. C., SANNINO, V., COSTANZO, V. & PELLEGRINI, L. 2016. Structure of human Cdc45 and implications for CMG helicase function. *Nat Commun,* 7**,** 11638.
SIMON, A. C., ZHOU, J. C., PERERA, R. L., VAN DEURSEN, F., EVRIN, C., IVANOVA, M. E., KILKENNY, M. L., RENAULT, L., KJAER, S., MATAK-VINKOVIC, D., LABIB, K., COSTA, A. & PELLEGRINI, L. 2014. A Ctf4 trimer couples the CMG helicase to DNA polymerase alpha in the eukaryotic replisome. *Nature,* 510**,** 293-297.
SINGLETON, M. R., DILLINGHAM, M. S., GAUDIER, M., KOWALCZYKOWSKI, S. C. & WIGLEY, D. B. 2004. Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature,* 432**,** 187-93.
SINGLETON, M. R., DILLINGHAM, M. S. & WIGLEY, D. B. 2007. Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem,* 76**,** 23-50.
SINGLETON, M. R., SAWAYA, M. R., ELLENBERGER, T. & WIGLEY, D. B. 2000. Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell,* 101**,** 589-600.
SKORDALAKES, E. & BERGER, J. M. 2003. Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell,* 114**,** 135-46.
SMELKOVA, N. V. & BOROWIEC, J. A. 1997. Dimerization of simian virus 40 T-antigen hexamers activates T-antigen DNA helicase activity. *J Virol,* 71**,** 8766-73.

SPECK, C., CHEN, Z., LI, H. & STILLMAN, B. 2005. ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nat Struct Mol Biol,* 12**,** 965-71.
SPECK, C. & STILLMAN, B. 2007. Cdc6 ATPase activity regulates ORC x Cdc6 stability and the selection of specific DNA sequences as origins of DNA replication. *J Biol Chem,* 282**,** 11705-14.
STINCHCOMB, D. T., STRUHL, K. & DAVIS, R. W. 1979. Isolation and characterisation of a yeast chromosomal replicator. *Nature,* 282**,** 39-43.
SUN, J., EVRIN, C., SAMEL, S. A., FERNANDEZ-CID, A., RIERA, A., KAWAKAMI, H., STILLMAN, B., SPECK, C. & LI, H. 2013. Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA. *Nat Struct Mol Biol,* 20**,** 944-51.
SUN, J., FERNANDEZ-CID, A., RIERA, A., TOGNETTI, S., YUAN, Z., STILLMAN, B., SPECK, C. & LI, H. 2014. Structural and mechanistic insights into Mcm2-7 double-hexamer assembly and function. *Genes Dev,* 28**,** 2291-303.
SUN, J., KAWAKAMI, H., ZECH, J., SPECK, C., STILLMAN, B. & LI, H. 2012. Cdc6-induced conformational changes in ORC bound to origin DNA revealed by cryo-electron microscopy. *Structure,* 20**,** 534-44.
SUN, J., SHI, Y., GEORGESCU, R. E., YUAN, Z., CHAIT, B. T., LI, H. & O'DONNELL, M. E. 2015. The architecture of a eukaryotic replisome. *Nat Struct Mol Biol,* 22**,** 976-82.
SUTERA, V. A., JR. & LOVETT, S. T. 2006. The role of replication initiation control in promoting survival of replication fork damage. *Mol Microbiol,* 60**,** 229-39.
SUYARI, O., KAWAI, M., IDA, H., YOSHIDA, H., SAKAGUCHI, K. & YAMAGUCHI, M. 2012. Differential requirement for the N-terminal catalytic domain of the DNA polymerase epsilon p255 subunit in the mitotic cell cycle and the endocycle. *Gene,* 495**,** 104-14.
SZAMBOWSKA, A., TESSMER, I., KURSULA, P., USSKILAT, C., PRUS, P., POSPIECH, H. & GROSSE, F. 2014. DNA binding properties of human Cdc45 suggest a function as molecular wedge for DNA unwinding. *Nucleic Acids Res,* 42**,** 2308-19.
SZAMBOWSKA, A., TESSMER, I., PRUS, P., SCHLOTT, B., POSPIECH, H. & GROSSE, F. 2017. Cdc45-induced loading of human RPA onto single-stranded DNA. *Nucleic Acids Res,* 45**,** 3217-3230.
TAHIROV, T. H., MAKAROVA, K. S., ROGOZIN, I. B., PAVLOV, Y. I. & KOONIN, E. V. 2009. Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. *Biol Direct,* 4**,** 11.
TAK, Y. S., TANAKA, Y., ENDO, S., KAMIMURA, Y. & ARAKI, H. 2006. A CDK-catalysed regulatory phosphorylation for formation of the DNA replication complex Sld2-Dpb11. *EMBO J,* 25**,** 1987-96.
TAKAHASHI, T. S., WIGLEY, D. B. & WALTER, J. C. 2005. Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. *Trends Biochem Sci,* 30**,** 437-44.
TAKAYAMA, Y., KAMIMURA, Y., OKAWA, M., MURAMATSU, S., SUGINO, A. & ARAKI, H. 2003. GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev,* 17**,** 1153-65.
TANAKA, S. & ARAKI, H. 2013. Helicase activation and establishment of replication forks at chromosomal origins of replication. *Cold Spring Harb Perspect Biol,* 5**,** a010371.
TANAKA, S. & DIFFLEY, J. F. 2002. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nat Cell Biol,* 4**,** 198-207.
TANAKA, S., KOMEDA, Y., UMEMORI, T., KUBOTA, Y., TAKISAWA, H. & ARAKI, H. 2013. Efficient initiation of DNA replication in eukaryotes requires Dpb11/TopBP1-GINS interaction. *Mol Cell Biol,* 33**,** 2614-22.

TANAKA, S., NAKATO, R., KATOU, Y., SHIRAHIGE, K. & ARAKI, H. 2011a. Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Curr Biol,* 21**,** 2055-63.
TANAKA, S., UMEMORI, T., HIRAI, K., MURAMATSU, S., KAMIMURA, Y. & ARAKI, H. 2007. CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature,* 445**,** 328-32.
TANAKA, T., KNAPP, D. & NASMYTH, K. 1997. Loading of an Mcm protein onto DNA replication origins is regulated by Cdc6p and CDKs. *Cell,* 90**,** 649-60.
TANAKA, T., UMEMORI, T., ENDO, S., MURAMATSU, S., KANEMAKI, M., KAMIMURA, Y., OBUSE, C. & ARAKI, H. 2011b. Sld7, an Sld3-associated protein required for efficient chromosomal DNA replication in budding yeast. *EMBO J,* 30**,** 2019-30.
TANG, G., PENG, L., BALDWIN, P. R., MANN, D. S., JIANG, W., REES, I. & LUDTKE, S. J. 2007. EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol,* 157**,** 38-46.
TAYLOR, M. R. G., SPIREK, M., CHAURASIYA, K. R., WARD, J. D., CARZANIGA, R., YU, X., EGELMAN, E. H., COLLINSON, L. M., RUEDA, D., KREJCI, L. & BOULTON, S. J. 2015. Rad51 Paralogs Remodel Pre-synaptic Rad51 Filaments to Stimulate Homologous Recombination. *Cell,* 162**,** 271-286.
THOMSEN, N. D. & BERGER, J. M. 2009. Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell,* 139**,** 523-34.
THU, Y. M. & BIELINSKY, A. K. 2013. Enigmatic roles of Mcm10 in DNA replication. *Trends Biochem Sci,* 38**,** 184-94.
TICAU, S., FRIEDMAN, L. J., CHAMPASA, K., CORREA, I. R., JR., GELLES, J. & BELL, S. P. 2017. Mechanism and timing of Mcm2-7 ring closure during DNA replication origin licensing. *Nat Struct Mol Biol,* 24**,** 309-315.
TICAU, S., FRIEDMAN, L. J., IVICA, N. A., GELLES, J. & BELL, S. P. 2015. Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell,* 161**,** 513-525.
TOCILJ, A., ON, K. F., YUAN, Z., SUN, J., ELKAYAM, E., LI, H., STILLMAN, B. & JOSHUA-TOR, L. 2017. Structure of the active form of human origin recognition complex and its ATPase motor module. *Elife,* 6.
TRAKSELIS, M. A. 2016. Structural Mechanisms of Hexameric Helicase Loading, Assembly, and Unwinding. *F1000Res,* 5.
TSUBOTA, T., TAJIMA, R., ODE, K., KUBOTA, H., FUKUHARA, N., KAWABATA, T., MAKI, S. & MAKI, H. 2006. Double-stranded DNA binding, an unusual property of DNA polymerase epsilon, promotes epigenetic silencing in Saccharomyces cerevisiae. *J Biol Chem,* 281**,** 32898-908.
VALLE, M., CHEN, X. S., DONATE, L. E., FANNING, E. & CARAZO, J. M. 2006. Structural basis for the cooperative assembly of large T antigen on the origin of replication. *J Mol Biol,* 357**,** 1295-305.
VALLE, M., GRUSS, C., HALMER, L., CARAZO, J. M. & DONATE, L. E. 2000. Large T-antigen double hexamers imaged at the simian virus 40 origin of replication. *Mol Cell Biol,* 20**,** 34-41.
VAN DEURSEN, F., SENGUPTA, S., DE PICCOLI, G., SANCHEZ-DIAZ, A. & LABIB, K. 2012. Mcm10 associates with the loaded DNA helicase at replication origins and defines a novel step in its activation. *EMBO J,* 31**,** 2195-206.
VAN HEEL, M., HARAUZ, G., ORLOVA, E. V., SCHMIDT, R. & SCHATZ, M. 1996. A new generation of the IMAGIC image processing system. *J Struct Biol,* 116**,** 17-24.
VAN HOUTEN, J. V. & NEWLON, C. S. 1990. Mutational analysis of the consensus sequence of a replication origin from yeast chromosome III. *Mol Cell Biol,* 10**,** 3917-25.

VENDITTI, P., COSTANZO, G., NEGRI, R. & CAMILLONI, G. 1994. ABFI contributes to the chromatin organization of Saccharomyces cerevisiae ARS1 B-domain. *Biochim Biophys Acta,* 1219**,** 677-89.
VILLA, F., SIMON, A. C., ORTIZ BAZAN, M. A., KILKENNY, M. L., WIRTHENSOHN, D., WIGHTMAN, M., MATAK-VINKOVIC, D., PELLEGRINI, L. & LABIB, K. 2016. Ctf4 Is a Hub in the Eukaryotic Replisome that Links Multiple CIP-Box Proteins to the CMG Helicase. *Mol Cell,* 63**,** 385-96.
WAKAMATSU, T., KITAMURA, Y., KOTERA, Y., NAKAGAWA, N., KURAMITSU, S. & MASUI, R. 2010. Structure of RecJ exonuclease defines its specificity for single-stranded DNA. *J Biol Chem,* 285**,** 9762-9.
WALTER, J. C. 2000. Evidence for sequential action of cdc7 and cdk2 protein kinases during initiation of DNA replication in Xenopus egg extracts. *J Biol Chem,* 275**,** 39773-8.
WAN, W. & BRIGGS, J. A. 2016. Cryo-Electron Tomography and Subtomogram Averaging. *Methods Enzymol,* 579**,** 329-67.
WARREN, E. M., HUANG, H., FANNING, E., CHAZIN, W. J. & EICHMAN, B. F. 2009. Physical interactions between Mcm10, DNA, and DNA polymerase alpha. *J Biol Chem,* 284**,** 24662-72.
WARREN, E. M., VAITHIYALINGAM, S., HAWORTH, J., GREER, B., BIELINSKY, A. K., CHAZIN, W. J. & EICHMAN, B. F. 2008. Structural basis for DNA binding by replication initiator Mcm10. *Structure,* 16**,** 1892-901.
WATASE, G., TAKISAWA, H. & KANEMAKI, M. T. 2012. Mcm10 plays a role in functioning of the eukaryotic replicative DNA helicase, Cdc45-Mcm-GINS. *Curr Biol,* 22**,** 343-9.
WEI, Z., LIU, C., WU, X., XU, N., ZHOU, B., LIANG, C. & ZHU, G. 2010. Characterization and structure determination of the Cdt1 binding domain of human minichromosome maintenance (Mcm) 6. *J Biol Chem,* 285**,** 12469-73.
WEINREICH, M. & STILLMAN, B. 1999. Cdc7p-Dbf4p kinase binds to chromatin during S phase and is regulated by both the APC and the RAD53 checkpoint pathway. *EMBO J,* 18**,** 5334-46.
WHITTAKER, A. J., ROYZMAN, I. & ORR-WEAVER, T. L. 2000. Drosophila double parked: a conserved, essential replication protein that colocalizes with the origin recognition complex and links DNA replication with mitosis and the down-regulation of S phase transcripts. *Genes Dev,* 14**,** 1765-76.
WICKHAM, T. J. & NEMEROW, G. R. 1993. Optimization of growth methods and recombinant protein production in BTI-Tn-5B1-4 insect cells using the baculovirus expression system. *Biotechnol Prog,* 9**,** 25-30.
WIGLEY, D. B. 2009. ORC proteins: marking the start. *Curr Opin Struct Biol,* 19**,** 72-8.
WILLIAMS, R. S., SHOHET, R. V. & STILLMAN, B. 1997. A human protein related to yeast Cdc6p. *Proc Natl Acad Sci U S A,* 94**,** 142-7.
WILMES, G. M. & BELL, S. P. 2002. The B2 element of the Saccharomyces cerevisiae ARS1 origin of replication requires specific sequences to facilitate pre-RC formation. *Proc Natl Acad Sci U S A,* 99**,** 101-6.
WOHLSCHLEGEL, J. A., DHAR, S. K., PROKHOROVA, T. A., DUTTA, A. & WALTER, J. C. 2002. Xenopus Mcm10 binds to origins of DNA replication after Mcm2-7 and stimulates origin binding of Cdc45. *Mol Cell,* 9**,** 233-40.
WOHLSCHLEGEL, J. A., DWYER, B. T., DHAR, S. K., CVETIC, C., WALTER, J. C. & DUTTA, A. 2000. Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. *Science,* 290**,** 2309-12.
WYRICK, J. J., APARICIO, J. G., CHEN, T., BARNETT, J. D., JENNINGS, E. G., YOUNG, R. A., BELL, S. P. & APARICIO, O. M. 2001. Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. *Science,* 294**,** 2357-60.

XU, Y., GRISTWOOD, T., HODGSON, B., TRINIDAD, J. C., ALBERS, S. V. & BELL, S. D. 2016. Archaeal orthologs of Cdc45 and GINS form a stable complex that stimulates the helicase activity of MCM. *Proc Natl Acad Sci U S A,* 113**,** 13390-13395.
YAO, N. Y. & O'DONNELL, M. E. 2016. Evolution of replication machines. *Crit Rev Biochem Mol Biol,* 51**,** 135-49.
YARDIMCI, H., LOVELAND, A. B., HABUCHI, S., VAN OIJEN, A. M. & WALTER, J. C. 2010. Uncoupling of sister replisomes during eukaryotic DNA replication. *Mol Cell,* 40**,** 834-40.
YARDIMCI, H., WANG, X., LOVELAND, A. B., TAPPIN, I., RUDNER, D. Z., HURWITZ, J., VAN OIJEN, A. M. & WALTER, J. C. 2012. Bypass of a protein barrier by a replicative DNA helicase. *Nature,* 492**,** 205-9.
YEELES, J. T., DEEGAN, T. D., JANSKA, A., EARLY, A. & DIFFLEY, J. F. 2015. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature,* 519**,** 431-5.
YEELES, J. T. P., JANSKA, A., EARLY, A. & DIFFLEY, J. F. X. 2017. How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication. *Mol Cell,* 65**,** 105-116.
YOSHIMOCHI, T., FUJIKANE, R., KAWANAMI, M., MATSUNAGA, F. & ISHINO, Y. 2008. The GINS complex from Pyrococcus furiosus stimulates the MCM helicase activity. *J Biol Chem,* 283**,** 1601-9.
YU, C., GAN, H., SERRA-CARDONA, A., ZHANG, L., GAN, S., SHARMA, S., JOHANSSON, E., CHABES, A., XU, R. M. & ZHANG, Z. 2018. A mechanism for preventing asymmetric histone segregation onto replicating DNA strands. *Science*.
YUAN, Z., BAI, L., SUN, J., GEORGESCU, R., LIU, J., O'DONNELL, M. E. & LI, H. 2016. Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. *Nat Struct Mol Biol,* 23**,** 217-24.
YUAN, Z., RIERA, A., BAI, L., SUN, J., NANDI, S., SPANOS, C., CHEN, Z. A., BARBON, M., RAPPSILBER, J., STILLMAN, B., SPECK, C. & LI, H. 2017. Structural basis of Mcm2-7 replicative helicase loading by ORC-Cdc6 and Cdt1. *Nat Struct Mol Biol,* 24**,** 316-324.
ZEGERMAN, P. & DIFFLEY, J. F. 2007. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature,* 445**,** 281-5.
ZHAI, Y., CHENG, E., WU, H., LI, N., YUNG, P. Y., GAO, N. & TYE, B. K. 2017a. Open-ringed structure of the Cdt1-Mcm2-7 complex as a precursor of the MCM double hexamer. *Nat Struct Mol Biol,* 24**,** 300-308.
ZHAI, Y., LI, N., JIANG, H., HUANG, X., GAO, N. & TYE, B. K. 2017b. Unique Roles of the Non-identical MCM Subunits in DNA Replication Licensing. *Mol Cell,* 67**,** 168-179.
ZHANG, K. 2016. Gctf: Real-time CTF determination and correction. *J Struct Biol,* 193**,** 1-12.
ZHENG, S. Q., PALOVCAK, E., ARMACHE, J. P., VERBA, K. A., CHENG, Y. & AGARD, D. A. 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods,* 14**,** 331-332.
ZHOU, J. C., JANSKA, A., GOSWAMI, P., RENAULT, L., ABID ALI, F., KOTECHA, A., DIFFLEY, J. F. X. & COSTA, A. 2017. CMG-Pol epsilon dynamics suggests a mechanism for the establishment of leading-strand synthesis in the eukaryotic replisome. *Proc Natl Acad Sci U S A,* 114**,** 4141-4146.
ZHU, W., UKOMADU, C., JHA, S., SENGA, T., DHAR, S. K., WOHLSCHLEGEL, J. A., NUTT, L. K., KORNBLUTH, S. & DUTTA, A. 2007. Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication. *Genes Dev,* 21**,** 2288-99.
ZOU, L. & STILLMAN, B. 1998. Formation of a preinitiation complex by S-phase cyclin CDK-dependent loading of Cdc45p onto chromatin. *Science,* 280**,** 593-6.