


Subcellular Biochemistry 62

Stuart MacNeill *Editor*



The Eukaryotic Replisome: A Guide to Protein Structure and Function

 Springer

The Eukaryotic Replisome: A Guide to Protein Structure and Function

SUBCELLULAR BIOCHEMISTRY

SERIES EDITOR

J. ROBIN HARRIS, University of Mainz, Mainz, Germany

ASSISTANT EDITOR

P.J. QUINN, King's College London, London, U.K.

Recent Volumes in this Series

- Volume 33 **Bacterial Invasion into Eukaryotic Cells**
Tobias A. Oelschlaeger and Jorg Hacker
- Volume 34 **Fusion of Biological Membranes and Related Problems**
Edited by Herwig Hilderson and Stefan Fuller
- Volume 35 **Enzyme-Catalyzed Electron and Radical Transfer**
Andreas Holzenburg and Nigel S. Scrutton
- Volume 36 **Phospholipid Metabolism in Apoptosis**
Edited by Peter J. Quinn and Valerian E. Kagan
- Volume 37 **Membrane Dynamics and Domains**
Edited by P.J. Quinn
- Volume 38 **Alzheimer's Disease: Cellular and Molecular Aspects of Amyloid beta**
Edited by R. Harris and F. Fahrenholz
- Volume 39 **Biology of Inositols and Phosphoinositides**
Edited by A. Lahiri Majumder and B.B. Biswas
- Volume 40 **Reviews and Protocols in DT40 Research**
Edited by Jean-Marie Buerstedde and Shunichi Takeda
- Volume 41 **Chromatin and Disease**
Edited by Tapas K. Kundu and Dipak Dasgupta
- Volume 42 **Inflammation in the Pathogenesis of Chronic Diseases**
Edited by Randall E. Harris
- Volume 43 **Subcellular Proteomics**
Edited by Eric Bertrand and Michel Faupel
- Volume 44 **Peroxiredoxin Systems**
Edited by Leopold Flohd J. Robin Harris
- Volume 45 **Calcium Signalling and Disease**
Edited by Ernesto Carafoli and Marisa Brini
- Volume 46 **Creatine and Creatine Kinase in Health and Disease**
Edited by Gajja S. Salomons and Markus Wyss
- Volume 47 **Molecular Mechanisms of Parasite Invasion**
Edited by Barbara A. Burleigh and Dominique Soldati-Favre
- Volume 48 **The Coronin Family of Proteins**
Edited by Christoph S. Clemen, Ludwig Eichinger and Vasily Rybakina
- Volume 49 **Lipids in Health and Disease**
Edited by Peter J. Quinn and Xiaoyuan Wang
- Volume 50 **Genome Stability and Human Diseases**
Edited by Heinz Peter Nasheuer
- Volume 51 **Cholesterol Binding and Cholesterol Transport Proteins**
Edited by J. Robin Harris
- Volume 52 **A Handbook of Transcription Factors**
Edited by T.R. Hughes

For further volumes:

<http://www.springer.com/series/6515>

Stuart MacNeill
Editor

The Eukaryotic Replisome: A Guide to Protein Structure and Function

 Springer

Editor

Stuart MacNeill
School of Biology
University of St Andrews
St Andrews, Fife, UK

ISSN 0306-0225

ISBN 978-94-007-4571-1

ISBN 978-94-007-4572-8 (eBook)

DOI 10.1007/978-94-007-4572-8

Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012945769

© Springer Science+Business Media Dordrecht 2012

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

High-fidelity genome duplication is fundamental to life and health. There are clear links between chromosome replication defects and genome instability, genetic disease and cancer in humans, making a detailed understanding of the molecular mechanisms of genome duplication vital for future advances in diagnosis, drug design, and treatment. The core cellular DNA replication machinery comprises around 40–50 individual conserved proteins, many of which are components of a series of elaborate molecular machines that interact with one another in a spatially and temporally coordinated manner to perform distinct functions at the replication fork, such as replication origin recognition, DNA unwinding, DNA synthesis and ligation. Our understanding of how these processes occur is now entering an exciting new phase as protein structure determination by X-ray crystallography allows us to view the molecular make-up of the eukaryotic replication machinery in unprecedented detail. High-resolution three-dimensional structures are now available for most of the key players in the replication process, allowing enzyme active sites and nucleic acid- and protein-interaction surfaces to be viewed at atomic resolution. Where crystal structures remain elusive, established methods such as single-particle reconstruction using cryo-electron microscopy (cryo-EM) and emerging techniques such as small angle X-ray scattering (SAXS) are increasingly being harnessed to provide important information on the overall shape of individual protein complexes and the organization of subunits therein.

The aim of this book is to provide a detailed guide to the structure and function of the key conserved components of the eukaryotic replisome with particular emphasis on how recent breakthroughs in protein structure determination have led to important insights into protein function, protein-protein interactions, and enzyme mechanism.

Chapter 1 offers a brief overview of the replication process in eukaryotic cells, from pre-RC formation in G1 through to Okazaki fragment processing at the end of S-phase. The role of individual proteins and protein complexes in these processes is summarized and the availability (or otherwise) of protein structural information highlighted. Chapter 2 explores the extent to which the proteins that make up the conserved machinery of chromosome replication in mammalian cells and in well-studied eukaryotic model organisms such as budding and fission yeasts, *Xenopus*

and *Drosophila*, are conserved across all eukaryotic evolution and reaches some thought-provoking conclusions. The remainder of the book takes the reader on a guided tour through the replication machinery, with each chapter focusing on an individual protein or protein complex. This systematic approach allows the structure and function of each factor to be considered at a level of detail that would otherwise be impossible and makes this single volume a truly comprehensive guide to the overall structure and function of the replisome, one that can serve as introduction to the complexities of the replication machinery for advanced undergraduate and post-graduate students and as an essential guide and companion for experienced researchers already working in the field.

As a final note, I would like to thank the authors for their hard work in preparing their uniformly excellent chapters and for their patience with the production process, my colleagues in St Andrews and elsewhere for their advice and encouragement, publishing editor Thijs van Vlijmen at Springer SBM for his help and support, and the Scottish Universities Life Sciences Alliance (SULSA) for funding.

St Andrews

Stuart MacNeill

Contents

1	Composition and Dynamics of the Eukaryotic Replisome: A Brief Overview	1
	Stuart MacNeill	
2	Evolutionary Diversification of Eukaryotic DNA Replication Machinery	19
	Stephen J. Aves, Yuan Liu, and Thomas A. Richards	
3	The Origin Recognition Complex: A Biochemical and Structural View	37
	Huilin Li and Bruce Stillman	
4	Archaeal Orc1/Cdc6 Proteins	59
	Stephen D. Bell	
5	Cdt1 and Geminin in DNA Replication Initiation	71
	Christophe Caillat and Anastassis Perrakis	
6	MCM Structure and Mechanics: What We Have Learned from Archaeal MCM	89
	Ian M. Slaymaker and Xiaojiang S. Chen	
7	The Eukaryotic Mcm2-7 Replicative Helicase	113
	Sriram Vijayraghavan and Anthony Schwacha	
8	The GINS Complex: Structure and Function	135
	Katsuhiko Kamada	
9	The Pol α-Primase Complex	157
	Luca Pellegrini	
10	The Structure and Function of Replication Protein A in DNA Replication	171
	Aishwarya Prakash and Gloria E.O. Borgstahl	

11 Structural Biology of Replication Initiation Factor Mcm10..... 197
Wenyue Du, Melissa E. Stauffer, and Brandt F. Eichman

12 Structure and Function of Eukaryotic DNA Polymerase δ 217
Tahir H. Tahirov

13 DNA Polymerase ϵ 237
Matthew Hogg and Erik Johansson

14 The RFC Clamp Loader: Structure and Function 259
Nina Y. Yao and Mike O’Donnell

**15 PCNA Structure and Function: Insights from Structures
of PCNA Complexes and Post-translationally Modified PCNA..... 281**
Lynne M. Dieckman, Bret D. Freudenthal, and M. Todd Washington

**16 The Wonders of Flap Endonucleases: Structure,
Function, Mechanism and Regulation..... 301**
L. David Finger, John M. Atack, Susan Tsutakawa, Scott Classen,
John Tainer, Jane Grasby, and Binghui Shen

17 DNA Ligase I, the Replicative DNA Ligase 327
Timothy R.L. Howes and Alan E. Tomkinson

Index..... 343

Chapter 1

Composition and Dynamics of the Eukaryotic Replisome: A Brief Overview

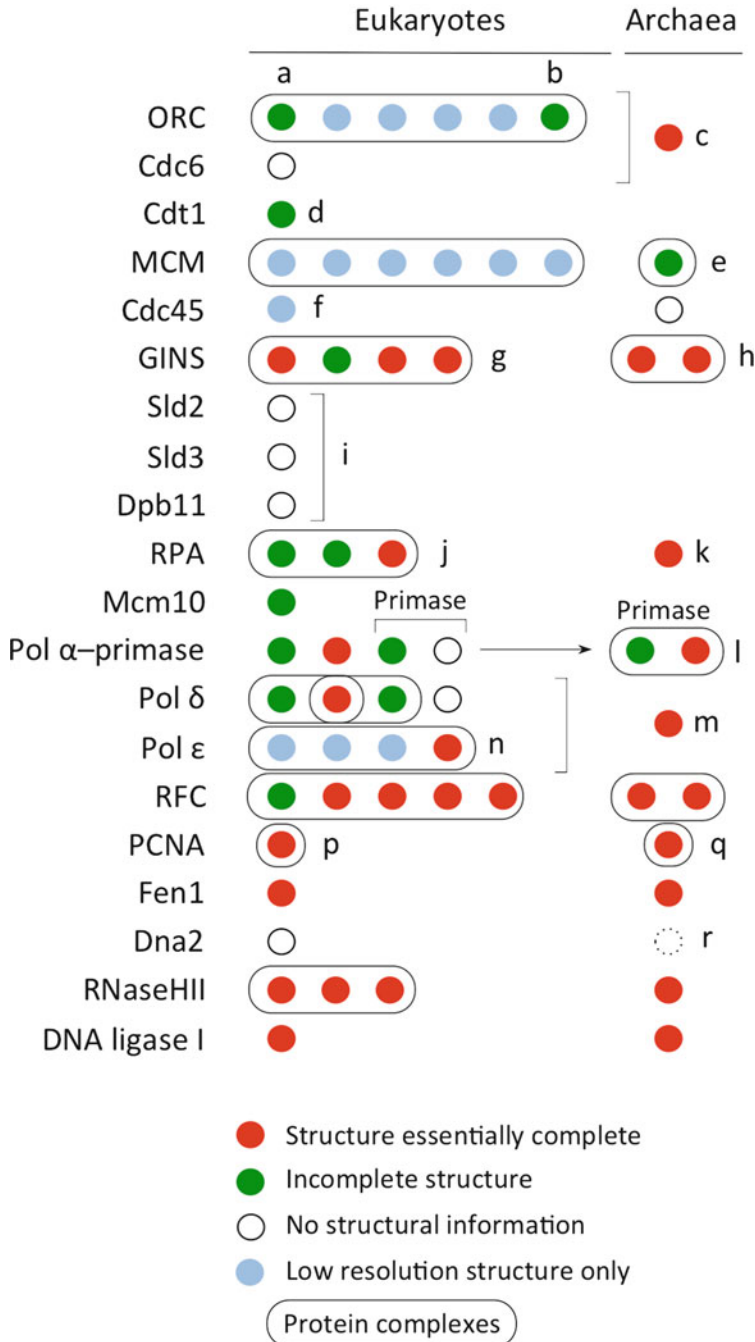
Stuart MacNeill

Abstract High-fidelity chromosomal DNA replication is vital for maintaining the integrity of the genetic material in all forms of cellular life. In eukaryotic cells, around 40–50 distinct conserved polypeptides are essential for chromosome replication, the majority of which are themselves component parts of a series of elaborate molecular machines that comprise the replication apparatus or replisome. How these complexes are assembled, what structures they adopt, how they perform their functions, and how those functions are regulated, are key questions for understanding how genome duplication occurs. Here I present a brief overview of current knowledge of the composition of the replisome and the dynamic molecular events that underlie chromosomal DNA replication in eukaryotic cells.

1.1 Introduction

Chromosomal DNA replication in all cells requires the complex interplay of variety of essential and non-essential protein factors in a temporally and spatially coordinated manner. In eukaryotes, chromosome replication as such (that is, the templated synthesis of new DNA on leading and lagging strands in a semi-discontinuous manner) occurs during S phase of the cell cycle, although some of the molecular events that lead up to the initiation of S phase (such as assembly of pre-replicative complexes, or pre-RCs, discussed below) take place in G1 phase (Bell and Dutta 2002) and it is likely that the final steps of the replication process take place in what is conventionally thought of as G2 (Lygeros et al. 2008). Checkpoints (Hartwell and Weinert 1989)

S. MacNeill (✉)
Biomedical Sciences Research Complex, School of Biology,
University of St Andrews, North Haugh, St Andrews, Fife, KY16 9ST, UK
e-mail: stuart.macneill@st-andrews.ac.uk



ensure that cells do not enter mitosis (M phase) until replication is complete (Labib and De Piccoli 2011) and, under normal circumstances, elaborate regulatory mechanisms ensure that each part of the genome is replicated once and only once during the cell cycle (Blow and Dutta 2005), thereby preventing unwanted amplification of individual genes or larger regions of the chromosomes.

The following sections outline what is known of the functions of key conserved components of the eukaryotic replication machinery (replisome), highlighting the current state of knowledge of the structure of these diverse factors (summarised schematically in Fig. 1.1). Detailed descriptions of those factors for which structural information is available can be found in Chaps. 3–17, while Chap. 2 takes a phylogenetic view of the extent to which the replication machinery is conserved across the major eukaryotic sub-groups.

1.2 Replication Origins and the Origin Recognition Complex

Eukaryotic chromosome replication is initiated at multiple replication origins on each chromosome. Origin structure and function (reviewed by Cvetič and Walter 2005; Lucas and Raghuraman 2003; MacAlpine and Bell 2005) lies largely outside the scope of this volume but has been studied in greatest detail in the budding yeast *Saccharomyces cerevisiae*. Briefly, the budding yeast genome contains around 300–500 origins, equivalent to one origin every 30–50 kb, but not every origin is activated (fired) in every cell cycle. Those origins that do fire, do so with characteristic timing – some origins fire reproducibly early in S phase, for example. Exactly what controls the timing of origin firing is unclear: chromatin accessibility clearly plays a role (reviewed by Mechali 2010) as does the availability of the replication initiation factors Cdc45, Sld2, Sld3, Sld7 and Dpb11, discussed below (Mantiero et al. 2011;



Fig. 1.1 Structures of eukaryotic and archaeal replisome components. Key: (a) structure of BAH domain of Orc1 only, (b) structure of middle domain of Orc6 only, (c) the archaeal Orc1/Cdc6 protein is regarded as hybrid of eukaryotic ORC and Cdc6 proteins – structures solved include various Cdc6/Orc1 proteins bound to DNA, (d) a partial structure of Cdt1 in complex with the replication inhibitor geminin has also been solved, (e) archaeal MCM is a homohexamer, (f) Cdc45 structure has been examined by SAXS only, (g) the B-domain of the Psf1 subunit is invisible in GINS structures, (h) the structure of a Gins15₂Gins23₂ tetrameric GINS has been solved – other archaeal GINS complexes are Gins15₄ homotetramers, (i) no known archaeal homologues, (j) various complexes featuring one, two or three subunits but no complete structure, (k) archaea RPA's are highly heterogeneous in composition but several near complete structures have been solved, (l) structure of PriS-PriL dimer solved without PriL C-terminal domain, (m) structures of several monomeric archaeal PolB enzymes are known, (n) structure of Dpb4 protein solved in complex with the chromatin remodelling factor Dls1 – otherwise only a cryo-EM structure for Pol ε complex is available, (p) structures of both modified and unmodified PCNA solved with and without bound PIP peptides, (q) structures of homotrimeric and heterotrimeric archaeal PCNA complexes available, and (r) potential homologues in some species only. See text and individual chapters for details and references

Tanaka and Araki 2011; Tanaka et al. 2011). Budding yeast origins are relatively short (<200 bp) and include a well conserved DNA sequence element (the ACS or ARS-consensus sequence); origins in other well-studied species, including the fission yeast *Schizosaccharomyces pombe*, are significantly more complex in nature and do not contain a conserved sequence at their core.

Origins are bound by the origin recognition complex (ORC), a widely conserved six-subunit protein complex (Duncker et al. 2009) (described in detail in Chap. 3, this volume). Precise details of how ORC recognises origin sequences are still unclear and the differences in origin structure apparent across eukaryotic evolution make it highly likely that the details will differ from species to species. With the notable exceptions of the N-terminal BAH (bromo-adjacent homology) domain of the budding yeast and mouse Orc1 proteins (Hou et al. 2005; Hsu et al. 2005; Kuo et al. 2012; Zhang et al. 2002) and the middle domain of the human Orc6 protein (the least conserved of the ORC subunits) (Liu et al. 2011), no crystal structures of ORC subunits have been reported, although recent cryo-EM studies (discussed in detail in Chap. 3) have provided significant insights into ORC structure at lower resolution (Fig. 1.1). The BAH domain of metazoan Orc1 (but not its yeast counterparts) recognises and binds specifically to histone H4 dimethylated at lysine 20 (K4K20me2), thereby linking the process of replication origin licensing to chromatin modification status (Kuo et al. 2012). Mutations the human Orc1 BAH domain have been implicated in Meier–Gorlin syndrome (MGS), a form of primordial dwarfism. Also implicated in MGS are mutations in Orc4 and Orc6, and in the Cdc6 and Cdt1 proteins described in Sect. 1.3 below (Bicknell et al. 2011a, b; de Munnik et al. 2012; Guernsey et al. 2011). The middle domain of Orc6 is similar in structure to part of the eukaryotic transcriptional factor TFIIB, allowing DNA binding by Orc6 to be modeled; mutation of residues implicated in DNA binding in this way results in reduced DNA replication in cultured cells (Liu et al. 2011). The observed similarity between Orc6 and TFIIB has led to the suggestion that Orc6 may have role in correctly positioning ORC at origins in the same manner that TFIIB functions to position the transcription PIC (pre-initiation complex) at promoters (Liu et al. 2011).

1.3 Formation of the Pre-RC at Origins

During the G1 phase of the budding yeast cell cycle, ORC is bound by the AAA+ family protein Cdc6, which then recruits two additional factors, Cdt1 and the MCM (minichromosome maintenance) helicase, to form the pre-RC (pre-replicative complex) on DNA (reviewed by Blow and Dutta 2005). Regulated pre-RC formation (also known as replication licensing) is crucial for maintaining once-per-cell-cycle replication control and in vertebrates the geminin protein (described in detail in Chap. 5) plays an important role in regulating this process. Structures have been solved for several archaeal Cdc6 proteins (see Fig. 1.1 and Chap. 4 – note that these archaeal proteins share features of eukaryotic Cdc6 and ORC, and are commonly

known as Cdc6/Orc1, and that ORC itself is absent from archaea) and for both Cdt1 and geminin (Chap. 5). In eukaryotes, the MCM helicase is a heterohexameric complex comprising six related subunits, Mcm2-Mcm7, whereas in most archaea, MCM is a homohexamer. Each MCM subunit is a member of the AAA+ protein superfamily (Duderstadt and Berger 2008; Hanson and Whiteheart 2005). Crystal structures of several archaeal MCM proteins have been solved, in whole or in part, but none of the eukaryotic subunits (Fig. 1.1). The structures of the archaeal MCM proteins are discussed in Chap. 6 and mechanistic studies of the eukaryotic enzyme in Chap. 7.

The mechanism of assembly of the pre-RC is currently the subject of considerable interest (Labib 2011). Two recent cryo-EM studies have provided significant insights into this process by showing that budding yeast MCM is loaded at the origin as a head-to-head double hexamer (Evrin et al. 2009; Remus et al. 2009). Prior to loading, budding yeast MCM binds Cdt1 to form a heptameric MCM•Cdt1 unit and it appears that the single ORC complex present at an individual origin loads two of these to produce the head-to-head double hexamer (Takara and Bell 2011). Interestingly, double hexamers are also seen at licensed replication origins in *Xenopus* egg extracts too, suggesting that a similar loading mechanism may also be used at metazoan origins (Gambus et al. 2011). ORC remains bound at origins throughout the yeast cell cycle but once MCM is activated in S phase, Cdc6 and Cdt1 are lost and individual MCM hexamers move off with the replication forks (Gambus et al. 2006; Yardimci et al. 2010). MCM movement requires a marked remodelling of the MCM complex: initial loading clearly takes place on double-stranded origin DNA but a recent series of elegant biochemical experiments appear to indicate that the active helicase translocates 3'–5' on the single-stranded leading strand template DNA (Fu et al. 2011), a process that requires that the DNA duplex at the origin is melted and one strand (the lagging strand template) is excluded from the central channel of the helicase.

Activation of MCM helicase activity to facilitate DNA unwinding is a complicated and highly regulated process that requires MCM to associate with two additional factors, Cdc45 and GINS, to form the CMG (Cdc45–MCM–GINS) complex (Costa et al. 2011; Ilves et al. 2010; Moyer et al. 2006). The precise roles of the Cdc45 and GINS components of the CMG are not yet known. The near-complete crystal structure of the tetrameric human GINS complex has been solved (Chang et al. 2007; Choi et al. 2007; Kamada et al. 2007) (see Fig. 1.1, Chap. 8) and cryo-EM studies have provided a low-resolution view of CMG complex structure (Costa et al. 2011). No crystal structures are available for Cdc45 but this protein has recently been reported to be related to the DHH family of phosphoesterases that includes the bacterial RecJ nuclease (Krastanova et al. 2012; Makarova et al. 2012; Sanchez-Pulido and Ponting 2011). The structure of RecJ can be modelled into the SAXS (small angle X-ray scattering) structure of human Cdc45 (Krastanova et al. 2012). Eukaryotic Cdc45 does not possess nuclease activity but at least some archaeal Cdc45 homologues, exemplified by the GAN (GINS-associated nuclease) protein, do (Li et al. 2011).

Prior to associating with MCM, Cdc45 is found in a complex with the Sld3 protein (Kamimura et al. 2001), whereas GINS is part of the pre-LC (or pre-loading complex) together with Sld2, Dpb11 and DNA polymerase ϵ (Pol ϵ) (Muramatsu et al. 2010). Sld2 and Sld3 bind to Dpb11 directly via the latter's BRCT domains, but only when phosphorylated by the S-CDK (the S phase cyclin-dependent kinase) (Tanaka et al. 2007; Zegerman and Diffley 2007). Somehow, this phosphorylation leads to disassembly of the Sld3-Cdc45 and pre-LC complexes and formation of the CMG. No structural information is presently available for Sld2, Sld3 or Dpb11, and no archaeal homologues of these proteins are apparent. Once the CMG is formed, DNA unwinding can occur and replication can begin in earnest. Recent results have shown that while it is not required for CMG assembly at origins, the conserved Mcm10 protein appears to be required for CMG translocation and for origin unwinding, as the trimeric single-stranded DNA binding factor RPA is not recruited to origins when Mcm10 is depleted (Watase et al. 2012). Partial X-ray structures are available for both RPA (Chap. 10) and for Mcm10 (Chap. 11).

1.4 The Replisome Progression Complex

The CMG complex forms the heart of the molecular assembly (called the replisome progression complex, RPC) at each replication fork (Gambus et al. 2006, 2009). Experiments in budding yeast have shown that each RPC contains a single MCM hexamer only, presumably as part of the CMG, together with a number of other proteins including the Tof1–Csm3 complex that is required for forks to pause at protein–DNA barriers, the histone chaperone FACT, the checkpoint mediator Mrc1, the type I topoisomerase Top1, the Mcm10 and Ctf4 proteins known to bind DNA polymerase α –primase (Pol α –primase) and Pol α –primase itself. Interestingly, purified RPCs do not contain either Pol ϵ (despite this factor's early involvement in CMG assembly as part of the pre-LC) nor DNA polymerase δ (Pol δ), suggesting that these polymerases (which are known to be present at moving replication forks – see Aparicio et al. 1999) are only loosely associated with the helicase machinery (Gambus et al. 2006, 2009). Consistent with this, experiments with *Xenopus* egg extracts have shown that it is possible to physically uncouple the helicase and polymerase activities by addition of the polymerase inhibitor aphidicolin (Pacek et al. 2006).

1.5 The Replicative Polymerases

Once the origin DNA is unwound, presumably by the action of the CMG complex in the RPC, templated DNA synthesis can begin. DNA polymerases cannot synthesise DNA *de novo*, but can only extend from a pre-existing 3'OH group. To generate the appropriate 3' end, a short (10–15 nucleotide) RNA primer is synthesised by a specialised RNA polymerase enzyme known as primase – unlike DNA polymerases,

RNA polymerases have no difficulty in initiating RNA synthesis *de novo* to form the 5' ends of nascent RNA transcripts. In eukaryotes, primase forms a stable tetrameric complex with DNA polymerase α , the Pol α -primase complex (see Chap. 9 for detailed discussion). Once the primase component of the Pol α -primase complex has synthesised the short RNA, Pol α recognises the newly formed 3'OH and the nascent strand is extended by a further 20–25 nucleotides of DNA. The process of RNA-DNA primer synthesis by Pol α -primase is required to initiate leading strand replication at the origin and also to initiate synthesis of each and every Okazaki fragment on the discontinuously synthesised lagging strand.

Once the 30–40 nucleotide RNA-DNA primer is completed, Pol α -primase is believed to play no further part in leading or lagging strand synthesis: Pol α -primase is not a processive enzyme nor does it possess the ability to proofread the DNA it synthesises, raising the possibility that the DNA segment of the primer might contain potentially mutagenic sequence errors. Instead, a polymerase switch occurs and the RNA-DNA primer is extended, apparently in a strand-specific manner, by DNA polymerase δ (Pol δ) and DNA polymerase ϵ (Pol ϵ) (see Chaps. 12 and 13, respectively, for details). Elegant genetic studies in yeast indicate that Pol δ is the lagging strand polymerase and Pol ϵ the leading strand polymerase (Nick McElhinny et al. 2008; Pursell et al. 2007, reviewed by Kunkel and Burgers 2008; Stillman 2008).

Like Pol α -primase, both Pol δ and Pol ϵ are multi-subunit enzymes, comprising in each case a family B polymerase catalytic subunit and a number of smaller subunits, one of which, the B-subunit, is also distantly related between all three replicative enzymes (reviewed by Johansson and MacNeill 2010). The catalytic subunits of Pol δ and Pol ϵ possess both polymerase and 3'–5' exonuclease (proofreading) activities and replicate DNA with high fidelity. Combined with earlier studies of distantly related bacteriophage and archaeal family B polymerase enzymes, recent structural studies of the Pol δ catalytic subunit Pol3 (discussed in detail in Chap. 12) have shed considerable light on Pol δ enzyme mechanism and in particular on how the enzyme discriminates between correctly and incorrectly incorporated bases and how an incorrectly incorporated base triggers movement of the nascent strand from the polymerase to the exonuclease active site (Swan et al. 2009). The structures of the B- and part of the C-subunit of Pol δ have also solved (Baranovskiy et al. 2008), as has the structure of the C-terminal domain of Pol11, the catalytic subunit of Pol α , bound to its B-subunit Pol12 (Klinge et al. 2009), and the iron-sulphur cluster domain of the large subunit of primase (Sauguet et al. 2010; Vaithiyalingam et al. 2010) (see Fig. 1.1).

Despite being implicated as playing a key role in leading strand synthesis, genetic studies in both yeasts have shown that the entire catalytic domain of Pol ϵ can be deleted without loss of cell viability, although chromosome replication is significantly slowed under these conditions and the cells display a variety of additional defects (Kesti et al. 1999). This behaviour is only seen when the catalytic domain is absent and not with catalytically inactive full-length Pol ϵ proteins, suggesting that deleting the catalytic domain is necessary to free-up sufficient space to allow access of Pol δ to the leading strand substrate.

1.6 Sliding Clamp and Clamp Loader Complexes

Processivity is a vital characteristic of Pol δ and Pol ϵ but is not an intrinsic property of these enzymes. Instead, processivity is acquired through interaction with a separate processivity factor, the conserved sliding clamp PCNA (proliferating cell nuclear antigen). PCNA is a ring-shaped trimer that is able to encircle and slide along double-stranded DNA (Ludwig and Walkinshaw 2006). Eukaryotic PCNA is a homotrimer whereas both homo- and heterotrimeric PCNAs are found in archaea. In bacteria, the sliding clamp is a dimer known as the β -sliding clamp. Both dimeric and trimeric PCNA display six-fold symmetry (Ludwig and Walkinshaw 2006). In addition to being essential for polymerase processivity, PCNA acts as a stable platform onto which a large number of DNA replication and repair factors are assembled (Tsurimoto 2006). In recent years, much progress has been made in dissecting the regulation of PCNA function by post-translational modification (PCNA is ubiquitylated, SUMOylated and phosphorylated) and structures of modified and modified PCNA complexes have been solved (see Chap. 14).

In order to be loaded onto DNA, the PCNA ring must be opened and closed around the duplex. In eukaryotes, PCNA ring opening and closing is accomplished by replication factor C, a pentameric clamp loader complex that comprises a large subunit Rfc1 and four small subunits Rfc2–Rfc5 (see Chap. 15 for details) (Majka and Burgers 2004). PCNA loading is an ATP-dependent process and each RFC subunit is a member of the AAA+ family of ATPases and ATP binding proteins. After ORC, Cdc6 and MCM, RFC is the fourth key component of the replication machinery to be a member of the AAA+ protein superfamily (Duderstadt and Berger 2008; Snider et al. 2008). The core structure of yeast RFC in a complex with PCNA has been solved (Bowman et al. 2004) as has an NMR structure for the N-terminal BRCT domain of the large Rfc1 subunit (Kobayashi et al. 2006).

1.7 Okazaki Fragment Processing

The last stage in the replication process requires sees Okazaki fragments on the lagging strand being processed, to remove the 5' RNA primer and the short stretch of potentially error-containing DNA synthesised by Pol α –primase prior to the polymerase switch, and finally joined. PCNA plays an important role here too, as at least three of the key enzymes implicated in these reactions (the nucleases Fen1 and RNaseHII, and DNA ligase I) bind directly to PCNA via a conserved short sequence motif known as a PIP (PCNA interacting protein) motif (Warbrick 1998). A number of structures of eukaryotic and archaeal PCNA bound to PIP box peptides have been reported, beginning with the structure of human PCNA bound to a PIP motif peptide derived from the mammalian cell cycle and DNA replication inhibitor p21^{Cip1} (Gulbis et al. 1996) (see Chap. 15). DNA polymerase δ also contacts PCNA via PIP motifs; in budding yeast these are found in all three subunits of the Pol δ complex (Acharya et al. 2011).

While several, perhaps all, of the key players are known (Fen1, RNaseHIII and the helicase-nuclease Dna2), the precise contributions made by each to Okazaki fragment processing remain somewhat unclear, a reflection of the complexity of the task at hand: the 5' end of each Okazaki fragment is a unique species and it is likely that different enzymes, or combination of enzymes, are involved in processing different classes of 5' end (see Henry et al. 2010; Pike et al. 2009, 2010; Stewart et al. 2008, 2009, for recent contributions to this field and detailed discussion). The structures of human Fen1 protein complexed to PCNA (Sakurai et al. 2005) and to a 5' DNA flap structure (Tsutakawa et al. 2011) have been solved, as have archaeal Fen1 and RNaseHIII structures (Chapados et al. 2001, 2004; Hosfield et al. 1998; Hwang et al. 1998; Lai et al. 2000) (see Chap. 16 for detailed discussion of Fen1 structure and function). The structure of trimeric human and mouse RNaseHIII enzymes have also been determined (Figiel et al. 2011; Shaban et al. 2010), allowing mutations implicated in the human auto-inflammatory disorder Aicardi–Goutières Syndrome (AGS) to be mapped.

The final step in the process of Okazaki fragment maturation sees DNA ligase I seal the nicks in the processed DNA, thereby producing a continuous nascent strand (see Chap. 17). In yeast, this ATP-dependent DNA ligase family member is essential for the completion of nuclear chromosomal DNA replication and plays an essential role in mitochondrial replication also (Donahue et al. 2001; Martin and MacNeill 2004; Willer et al. 1999). DNA ligase I has also been shown to be essential for mouse development (Bentley et al. 1996, 2002; Petrini et al. 1995) and a human patient with a DNA ligase I deficiency and various developmental and growth abnormalities has been identified, underlining the importance of this enzyme for maintaining genome integrity (Barnes et al. 1992; Webster et al. 1992). The structure of human DNA ligase I has been solved (Pascal et al. 2004). In addition, archaea also possess ATP-dependent DNA ligases and the structures of several of these (as well as the structure of several ATP-dependent DNA ligases from eukaryotic viruses and bacteriophage) have been determined (Kim et al. 2009; Nishida et al. 2006; Pascal et al. 2006), providing insights into the conserved ligase catalytic mechanism (see Chap. 17 for details). DNA ligase I also interacts with PCNA and with the clamp loader RFC, an interaction that is regulated by phosphorylation of the ligase (Vijayakumar et al. 2009).

1.8 Model Systems for the Studying Eukaryotic Replication

Much of what we know about the enzymes and mechanisms of eukaryotic chromosome replication has come from studies of a relatively small number of model systems and organisms, chosen for their tractability to genetic and/or biochemical analysis, or for their simplicity. As much of the replication machinery is conserved across species (at least within the range of well-studied eukaryotic organisms – see Chap. 2, this volume, for a wider discussion of this point) detailed understanding of protein function has often come from complementary and non-overlapping

approaches in diverse models. Indeed, it is not unreasonable to assert that full appreciation of protein function can only come from multi-disciplinary multi-organism approaches. The following sections briefly summarise the advantages and disadvantages of the most widely used model systems.

1.8.1 SV40

A number of the protein factors essential for chromosome replication in eukaryotes were first identified in studies that made use of the ability of mammalian cell extracts to successfully replicate plasmids carrying the SV40 viral replication origin *in vitro* (Waga and Stillman 1998). SV40 is a polyoma virus with a circular double-stranded DNA (dsDNA) genome. The virus encodes only a single protein that is required for replication of its genome (large T-antigen, a hexameric DNA helicase that also recognises the viral replication origin), all other replication factors being encoded by the host cell (Waga and Stillman 1998). Biochemical fractionation of host cell extracts led to identification of a number of protein factors that were later shown to be essential for chromosome replication also: these included DNA polymerase α -primase (Pol α -primase) and DNA polymerase δ (Pol δ), the single-stranded DNA binding factor RPA (replication protein A), the sliding clamp DNA polymerase processivity factor PCNA (proliferating cell nuclear antigen) and the sliding clamp loader replication factor C (RFC). Factors required for replication origin recognition and/or DNA unwinding (or for the regulation of these processes) were not identified in these studies, as these functions were provided by exogenously added T-antigen. Studies on T-antigen continue to provide valuable insights into DNA helicase function that are highly relevant to our understanding of the catalytic core of the cellular replicative helicase, the MCM complex (see Chaps. 6 and 7).

1.8.2 Yeast

Studies on the budding yeast *S. cerevisiae* and the distantly related fission yeast *Schiz. pombe* have proved vital for our understanding of the biology of chromosomal DNA replication in eukaryotes. Both organisms are genetically tractable and screens for conditional-lethal mutants (and in particular, temperature-sensitive mutants) have led to the identification of many essential replication factors in both organisms, including key regulatory factors that could not be identified in the SV40 system. In addition to being genetically tractable, both yeasts are easy to grow in the laboratory, have short generation times (2–3 h is typical), have sequenced genomes of ~12.5 Mb, and are amenable to a wide range of molecular and cell biological and biochemical applications. In addition, the cell cycle of budding yeast cells can be readily synchronised by addition and removal of the mating pheromone α -factor, allowing studies of the timing of replication events. Most of what is known about

replication origin function, the regulation of replication initiation and the molecular composition of the replication apparatus has come from studies with budding yeast.

1.8.3 *Xenopus*

The African clawed frog *Xenopus laevis* is arguably the most important biochemical model for eukaryotic chromosome replication currently in widespread use. Central to the utility of the *Xenopus* system is the ability of egg cytoplasmic extracts to faithfully replicate exogenously added sperm nuclei or purified DNA in a once-per-cell-cycle manner (Blow et al. 1987). Further modifications to this system, such as the teasing apart of key regulatory steps by separating the egg cytoplasmic extract into a high-speed supernatant fraction (HSS), which supports pre-RC formation (see below), and a nucleoplasmic extract (NPE) that stimulates replication initiation (Walter et al. 1998), have proved immensely valuable – for example, for dissecting key steps in the regulation of replication initiation, for identifying components of the replicative helicase (sometimes called the unwindosome) (Pacek et al. 2006) and for characterising the properties of the latter (Fu et al. 2011).

1.8.4 *Archaea*

The yeast and *Xenopus* systems offer the opportunity of relatively straightforward genetic and/or biochemical analysis of the processes of eukaryotic chromosome replication. However, although both are regarded as simple models for higher eukaryotic (i.e. mammalian) replication, in reality the composition of the replication machinery in these systems is probably as complex as that found in human cells. This complexity (typified by the number of multiprotein complexes on the list of factors known to be essential for chromosome replication) creates many problems, especially for biochemical and structural analysis. The archaea provide a partial solution to this problem. These organisms make up the third domain of life on Earth and form a sister group to the eukaryotes; the components of the archaeal DNA replication machinery resemble their eukaryotic counterparts but are frequently simpler in structure (Barry and Bell 2006) (Fig. 1.1). The MCM helicase, for example, a heterohexameric in eukaryotes, is homohexameric in archaea (see Chaps. 6 and 7, this volume). Archaea also have the added advantage that proteins from thermophilic or hyperthermophilic organisms can often be efficiently expressed and purified in recombinant form and are well suited to both biochemical and structural analysis. To date, for example, the only high-resolution structures of the Cdc6 and MCM proteins are those of archaeal organisms (see Chaps. 4 and 6, this volume). With tools for molecular genetic analysis now becoming available for a number of archaeal species (Leigh et al. 2011), there is no doubt that archaeal systems still have much to offer the eukaryotic replication community.

1.8.5 Other Model Systems

In addition to the yeasts, *Xenopus* and the archaea, significant recent insights into the eukaryotic replication machinery have come from the studies on the fruit fly *Drosophila melanogaster* (Costa et al. 2011; Ilves et al. 2010), from the nematode *Caenorhabditis elegans* (Sonneville et al. 2012) and from the kinetoplastid *Trypanosoma brucei* (Dang and Li 2011) amongst others. *Drosophila* has proved particularly important recently for studies on the CMG complex (Costa et al. 2011; Ilves et al. 2010; Moyer et al. 2006) and characterisation of the *T. brucei* CMG complex was also recently described (Dang and Li 2011), offering a rare glimpse of replication enzyme function in a less well-studied early-branching eukaryal sub-group.

1.9 Conclusions

Structure determination is changing the face of eukaryotic replication research but there is still some way to go before the change is complete. Figure 1.1 provides a visual summary of current knowledge of the three-dimensional structures of the eukaryotic and archaeal DNA replication factors, highlighting some significant gaps in the information at hand. These include the absence of high resolution structures for the origin recognition complex (ORC), the heterohexameric MCM helicase and DNA polymerase ϵ . Perhaps most strikingly, despite their importance as key substrates of the S phase cyclin-dependent kinase (CDK), nothing of yet known of the structures of the Sld2, Sld3 and Dpb11 proteins (although there is presumably scope for modelling the latter's BRCT domains). Given the rate at which high-resolution structures have been obtained in the last 2–3 years, however, it is highly likely that it will not be long before at least some of these gaps are filled.

Acknowledgements I am grateful to colleagues in St Andrews and elsewhere for comments on the manuscript. Replication research in my lab is funded by the Scottish Universities Life Sciences Alliance (SULSA).

References

- Acharya N, Klassen R, Johnson RE, Prakash L, Prakash S (2011) PCNA binding domains in all three subunits of yeast DNA polymerase δ modulate its function in DNA replication. *Proc Natl Acad Sci USA* 108:17927–17932
- Aparicio OM, Stout AM, Bell SP (1999) Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. *Proc Natl Acad Sci USA* 96:9130–9135
- Baranovskiy AG, Babayeva ND, Liston VG, Rogozin IB, Koonin EV, Pavlov YI, Vassilyev DG, Tahirov TH (2008) X-ray structure of the complex of regulatory subunits of human DNA polymerase δ . *Cell Cycle* 7:3026–3036
- Barnes DE, Tomkinson AE, Lehmann AR, Webster AD, Lindahl T (1992) Mutations in the DNA ligase I gene of an individual with immunodeficiencies and cellular hypersensitivity to DNA damaging agents. *Cell* 69:495–503

- Barry ER, Bell SD (2006) DNA replication in the archaea. *Microbiol Mol Biol Rev* 70:876–887
- Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem* 71:333–374
- Bentley D, Selfridge J, Millar JK, Samuel K, Hole N, Ansell JD, Melton DW (1996) DNA ligase I is required for fetal liver erythropoiesis but is not essential for mammalian cell viability. *Nat Genet* 13:489–491
- Bentley DJ, Harrison C, Ketchen AM, Redhead NJ, Samuel K, Waterfall M, Ansell JD, Melton DW (2002) DNA ligase I null mouse cells show normal DNA repair activity but altered DNA replication and reduced genome stability. *J Cell Sci* 115:1551–1561
- Bicknell LS, Bongers EM, Leitch A, Brown S, Schoots J, Harley ME, Aftimos S, Al-Aama JY, Bober M, Brown PA, van Bokhoven H, Dean J, Edrees AY, Feingold M, Fryer A, Hoefsloot LH, Kau N, Knoers NV, Mackenzie J, Opitz JM, Sarda P, Ross A, Temple IK, Toutain A, Wise CA, Wright M, Jackson AP (2011a) Mutations in the pre-replication complex cause Meier-Gorlin syndrome. *Nat Genet* 43:356–359
- Bicknell LS, Walker S, Klingseisen A, Stiff T, Leitch A, Kerzendorfer C, Martin CA, Yeyati P, Al Sanna N, Bober M, Johnson D, Wise C, Jackson AP, O'Driscoll M, Jeggo PA (2011b) Mutations in *ORC1*, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. *Nat Genet* 43:350–355
- Blow JJ, Dutta A (2005) Preventing re-replication of chromosomal DNA. *Nat Rev Mol Cell Biol* 6:476–486
- Blow JJ, Dilworth SM, Dingwall C, Mills AD, Laskey RA (1987) Chromosome replication in cell-free systems from *Xenopus* eggs. *Philos Trans R Soc Lond B Biol Sci* 317:483–494
- Bowman GD, O'Donnell M, Kuriyan J (2004) Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature* 429:724–730
- Chang YP, Wang G, Bermudez V, Hurwitz J, Chen XS (2007) Crystal structure of the GINS complex and functional insights into its role in DNA replication. *Proc Natl Acad Sci U S A* 104:12685–12690
- Chapados BR, Chai Q, Hosfield DJ, Qiu J, Shen B, Tainer JA (2001) Structural biochemistry of a type 2 RNase H: RNA primer recognition and removal during DNA replication. *J Mol Biol* 307:541–556
- Chapados BR, Hosfield DJ, Han S, Qiu J, Yelent B, Shen B, Tainer JA (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* 116:39–50
- Choi JM, Lim HS, Kim JJ, Song OK, Cho Y (2007) Crystal structure of the human GINS complex. *Genes Dev* 21:1316–1321
- Costa A, Ilves I, Tamberg N, Petojevic T, Nogales E, Botchan MR, Berger JM (2011) The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol* 18:471–477
- Cvetic C, Walter JC (2005) Eukaryotic origins of DNA replication: could you please be more specific? *Semin Cell Dev Biol* 16:343–353
- Dang HQ, Li Z (2011) The Cdc45-Mcm2-7-GINS protein complex in trypanosomes regulates DNA replication and interacts with two Orc1-like proteins in the origin recognition complex. *J Biol Chem* 286:32424–32435
- de Munnik SA, Bicknell LS, Aftimos S, Al-Aama JY, van Bever Y, Bober MB, Clayton-Smith J, Edrees AY, Feingold M, Fryer A, van Hagen JM, Hennekam RC, Jansweijer MC, Johnson D, Kant SG, Opitz JM, Ramadevi AR, Reardon W, Ross A, Sarda P, Schrander-Stumpel CT, Schoots J, Temple IK, Terhal PA, Toutain A, Wise CA, Wright M, Skidmore DL, Samuels ME, Hoefsloot LH, Knoers NV, Brunner HG, Jackson AP, Bongers EM (2012) Meier-Gorlin syndrome genotype-phenotype studies: 35 individuals with pre-replication complex gene mutations and 10 without molecular diagnosis. *Eur J Hum Genet* 20:598–606
- Donahue SL, Corner BE, Bordone L, Campbell C (2001) Mitochondrial DNA ligase function in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 29:1582–1589
- Duderstadt KE, Berger JM (2008) AAA+ATPases in the initiation of DNA replication. *Crit Rev Biochem Mol Biol* 43:163–187
- Duncker BP, Chesnokov IN, McConkey BJ (2009) The origin recognition complex protein family. *Genome Biol* 10:214

- 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 USA* 106:20240–20245
- Figiel M, Chon H, Cerritelli SM, Cybulska M, Crouch RJ, Nowotny M (2011) The structural and biochemical characterization of human RNase H2 complex reveals the molecular basis for substrate recognition and Aicardi-Goutières syndrome defects. *J Biol Chem* 286:10540–10550
- Fu YV, Yardimci H, Long DT, Ho TV, Guainazzi A, Bermudez VP, Hurwitz J, van Oijen A, Scharer OD, Walter JC (2011) Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell* 146:931–941
- Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, Labib K (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol* 8:358–366
- Gambus A, van Deursen F, Polychronopoulos D, Foltman M, Jones RC, Edmondson RD, Calzada A, Labib K (2009) A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase α within the eukaryotic replisome. *EMBO J* 28:2992–3004
- Gambus A, Khoudoli GA, Jones RC, Blow JJ (2011) MCM2-7 form double hexamers at licensed origins in *Xenopus* egg extract. *J Biol Chem* 286:11855–11864
- Guemsey DL, Matsuoka M, Jiang H, Evans S, Macgillivray C, Nightingale M, Perry S, Ferguson M, LeBlanc M, Paquette J, Patry L, Rideout AL, Thomas A, Orr A, McMaster CR, Michaud JL, Deal C, Langlois S, Superneau DW, Parkash S, Ludman M, Skidmore DL, Samuels ME (2011) Mutations in origin recognition complex gene *ORC4* cause Meier-Gorlin syndrome. *Nat Genet* 43:360–364
- Gulbis JM, Kelman Z, Hurwitz J, Odonnell M, Kuriyan J (1996) Structure of the C-terminal region of p21^{Waf1/Cip1} complexed with human PCNA. *Cell* 87:297–306
- Hanson PI, Whiteheart SW (2005) AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* 6:519–529
- Hartwell LH, Weinert TA (1989) Checkpoints: controls that ensure the order of cell-cycle events. *Science* 246:629–634
- Henry RA, Balakrishnan L, Ying-Lin ST, Campbell JL, Bambara RA (2010) Components of the secondary pathway stimulate the primary pathway of eukaryotic Okazaki fragment processing. *J Biol Chem* 285:28496–28505
- Hosfield DJ, Mol CD, Shen B, Tainer JA (1998) Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell* 95:135–146
- Hou Z, Bernstein DA, Fox CA, Keck JL (2005) Structural basis of the Sir1-origin recognition complex interaction in transcriptional silencing. *Proc Natl Acad Sci USA* 102:8489–8494
- Hsu HC, Stillman B, Xu RM (2005) Structural basis for origin recognition complex 1 protein-silence information regulator 1 protein interaction in epigenetic silencing. *Proc Natl Acad Sci USA* 102:8519–8524
- Hwang KY, Baek K, Kim HY, Cho Y (1998) The crystal structure of flap endonuclease-1 from *Methanococcus jannaschii*. *Nat Struct Biol* 5:707–713
- Ilves I, Petojevic T, Pesavento JJ, Botchan MR (2010) Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell* 37:247–258
- Johansson E, MacNeill SA (2010) The eukaryotic replicative DNA polymerases take shape. *Trends Biochem Sci* 35:339–347
- 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* 14:388–396
- Kamimura Y, Tak YS, Sugino A, Araki H (2001) Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in *Saccharomyces cerevisiae*. *EMBO J* 20:2097–2107
- Kesti T, Flick K, Keranen S, Syafoja JE, Wittenberg C (1999) DNA polymerase ϵ catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. *Mol Cell* 3:679–685
- Kim DJ, Kim O, Kim HW, Kim HS, Lee SJ, Suh SW (2009) ATP-dependent DNA ligase from *Archaeoglobus fulgidus* displays a tightly closed conformation. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 65:544–550

- Klinge S, Nunez-Ramirez R, Llorca O, Pellegrini L (2009) 3D architecture of DNA Pol α reveals the functional core of multi-subunit replicative polymerases. *EMBO J* 28:1978–1987
- Kobayashi M, Figaroa F, Meeuwenoord N, Jansen LE, Siegal G (2006) Characterization of the DNA binding and structural properties of the BRCT region of human replication factor C p140 subunit. *J Biol Chem* 281:4308–4317
- Krastanova I, Sannino V, Amenitsch H, Gileadi O, Pisani FM, 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* 287:4121–4128
- Kunkel TA, Burgers PM (2008) Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol* 18:521–527
- Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, Chen JK, Patel DJ, Gozani O (2012) The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* 484:115–119
- Labib K (2011) Building a double hexamer of DNA helicase at eukaryotic replication origins. *EMBO J* 30:4853–4855
- Labib K, De Piccoli G (2011) Surviving chromosome replication: the many roles of the S-phase checkpoint pathway. *Philos Trans R Soc Lond B Biol Sci* 366:3554–3561
- Lai L, Yokota H, Hung LW, Kim R, Kim SH (2000) Crystal structure of archaeal RNase HIII: a homologue of human major RNase H. *Structure* 8:897–904
- Leigh JA, Albers SV, Atomi H, Allers T (2011) Model organisms for genetics in the domain Archaea: methanogens, halophiles, *Thermococcales* and *Sulfolobales*. *FEMS Microbiol Rev* 35:577–608
- Li Z, Pan M, Santangelo TJ, Chemnitz W, Yuan W, Edwards JL, Hurwitz J, Reeve JN, Kelman Z (2011) A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res* 39:6114–6123
- Liu S, Balasov M, Wang H, Wu L, Chesnokov IN, Liu Y (2011) Structural analysis of human Orc6 protein reveals a homology with transcription factor TFIIB. *Proc Natl Acad Sci USA* 108:7373–7378
- Lucas IA, Raghuraman MK (2003) The dynamics of chromosome replication in yeast. *Curr Top Dev Biol* 55:1–73
- Ludwig C, Walkinshaw M (2006) The structure of PCNA and its homologues. In: Lee H (ed) *Proliferating cell nuclear antigen (PCNA)*. Research Signpost, Kerala, pp 1–24
- Lygeros J, Koutroumpas K, Dimopoulos S, Legouras I, Kouretas P, Heichinger C, Nurse P, Lygerou Z (2008) Stochastic hybrid modeling of DNA replication across a complete genome. *Proc Natl Acad Sci USA* 105:12295–12300
- MacAlpine DM, Bell SP (2005) A genomic view of eukaryotic DNA replication. *Chromosome Res* 13:309–326
- Majka J, Burgers PM (2004) The PCNA-RFC families of DNA clamps and clamp loaders. *Prog Nucleic Acid Res Mol Biol* 78:227–260
- Makarova KS, Koonin EV, Kelman Z (2012) The archaeal CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol Direct* 7:7
- Mantiero D, Mackenzie A, Donaldson A, Zegerman P (2011) Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *EMBO J* 30:4805–4814
- Martin IV, MacNeill SA (2004) Functional analysis of subcellular localization and protein-protein interaction sequences in the essential DNA ligase I protein of fission yeast. *Nucleic Acids Res* 32:632–642
- Mechali M (2010) Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat Rev Mol Cell Biol* 11:728–738
- Moyer SE, Lewis PW, Botchan MR (2006) Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci USA* 103:10236–10241

- Muramatsu S, Hirai K, Tak YS, Kamimura Y, Araki H (2010) CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol ϵ , and GINS in budding yeast. *Genes Dev* 24: 602–612
- Nick McElhinny SA, Gordenin DA, Stith CM, Burgers PM, Kunkel TA (2008) Division of labor at the eukaryotic replication fork. *Mol Cell* 30:137–144
- Nishida H, Kiyonari S, Ishino Y, Morikawa K (2006) The closed structure of an archaeal DNA ligase from *Pyrococcus furiosus*. *J Mol Biol* 360:956–967
- Pacek M, Tutter AV, Kubota Y, Takisawa H, Walter JC (2006) Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol Cell* 21:581–587
- Pascal JM, O'Brien PJ, Tomkinson AE, Ellenberger T (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* 432:473–478
- Pascal JM, Tsodikov OV, Hura GL, Song W, Cotner EA, Classen S, Tomkinson AE, Tainer JA, Ellenberger T (2006) A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. *Mol Cell* 24:279–291
- Petrini JHJ, Xiao YH, Weaver DT (1995) DNA ligase I mediates essential functions in mammalian-cells. *Mol Cell Biol* 15:4303–4308
- Pike JE, Burgers PM, Campbell JL, Bambara RA (2009) Pif1 helicase lengthens some Okazaki fragment flaps necessitating Dna2 nuclease/helicase action in the two-nuclease processing pathway. *J Biol Chem* 284:25170–25180
- Pike JE, Henry RA, Burgers PM, Campbell JL, Bambara RA (2010) An alternative pathway for Okazaki fragment processing: resolution of fold-back flaps by Pif1 helicase. *J Biol Chem* 285:41712–41723
- Pursell ZF, Isoz I, Lundstrom EB, Johansson E, Kunkel TA (2007) Yeast DNA polymerase ϵ participates in leading-strand DNA replication. *Science* 317:127–130
- Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JF (2009) Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139:719–730
- Sakurai S, Kitano K, Yamaguchi H, Hamada K, Okada K, Fukuda K, Uchida M, Ohtsuka E, Morioka H, Hakoshima T (2005) Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J* 24:683–693
- Sanchez-Pulido L, Ponting CP (2011) Cdc45: the missing RecJ ortholog in eukaryotes? *Bioinformatics* 27:1885–1888
- Sauguet L, Klinge S, Perera RL, Maman JD, Pellegrini L (2010) Shared active site architecture between the large subunit of eukaryotic primase and DNA photolyase. *PLoS One* 5:e10083
- Shaban NM, Harvey S, Perrino FW, Hollis T (2010) The structure of the mammalian RNase H2 complex provides insight into RNA-DNA hybrid processing to prevent immune dysfunction. *J Biol Chem* 285:3617–3624
- Snider J, Thibault G, Houry WA (2008) The AAA+ superfamily of functionally diverse proteins. *Genome Biol* 9:216
- Sonneville R, Querenet M, Craig A, Gartner A, Blow JJ (2012) The dynamics of replication licensing in live *Caenorhabditis elegans* embryos. *J Cell Biol* 196:233–246
- Stewart JA, Miller AS, Campbell JL, Bambara RA (2008) Dynamic removal of replication protein A by Dna2 facilitates primer cleavage during Okazaki fragment processing in *Saccharomyces cerevisiae*. *J Biol Chem* 283:31356–31365
- Stewart JA, Campbell JL, Bambara RA (2009) Significance of the dissociation of Dna2 by flap endonuclease 1 to Okazaki fragment processing in *Saccharomyces cerevisiae*. *J Biol Chem* 284:8283–8291
- Stillman B (2008) DNA polymerases at the replication fork in eukaryotes. *Mol Cell* 30:259–260
- Swan MK, Johnson RE, Prakash L, Prakash S, Aggarwal AK (2009) Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase δ . *Nat Struct Mol Biol* 16:979–986
- Takara TJ, Bell SP (2011) Multiple Cdt1 molecules act at each origin to load replication-competent Mcm2-7 helicases. *EMBO J* 30:4885–4896
- Tanaka S, Araki H (2011) Multiple regulatory mechanisms to inhibit untimely initiation of DNA replication are important for stable genome maintenance. *PLoS Genet* 7:e1002136

- 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: 328–332
- Tanaka S, Nakato R, Katou Y, Shirahige K, Araki H (2011) Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Curr Biol* 21:2055–2063
- Tsurimoto T (2006) PCNA-interacting proteins. In: Lee H (ed) *Proliferating cell nuclear antigen (PCNA)*. Research Signpost, Kerala, pp 25–49
- Tsutakawa SE, Classen S, Chapados BR, Arvai AS, Finger LD, Guenther G, Tomlinson CG, Thompson P, Sarker AH, Shen B, Cooper PK, Grasby JA, Tainer JA (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell* 145:198–211
- Vaithiyalingam S, Warren EM, Eichman BF, Chazin WJ (2010) Insights into eukaryotic DNA priming from the structure and functional interactions of the 4Fe-4S cluster domain of human DNA primase. *Proc Natl Acad Sci USA* 107:13684–13689
- Vijayakumar S, Dziegielewska B, Levin DS, Song W, Yin J, Yang A, Matsumoto Y, Bermudez VP, Hurwitz J, Tomkinson AE (2009) Phosphorylation of human DNA ligase I regulates its interaction with replication factor C and its participation in DNA replication and DNA repair. *Mol Cell Biol* 29:2042–2052
- Waga S, Stillman B (1998) The DNA replication fork in eukaryotic cells. *Annu Rev Biochem* 67:721–751
- Walter J, Sun L, Newport J (1998) Regulated chromosomal DNA replication in the absence of a nucleus. *Mol Cell* 1:519–529
- Warbrick E (1998) PCNA binding through a conserved motif. *Bioessays* 20:195–199
- Watae G, Takisawa H, Kanemaki MT (2012) Mcm10 plays a role in functioning of the eukaryotic replicative DNA helicase, Cdc45-Mcm-GINS. *Curr Biol* 22:343–349
- Webster AD, Barnes DE, Arlett CF, Lehmann AR, Lindahl T (1992) Growth retardation and immunodeficiency in a patient with mutations in the DNA ligase I gene. *Lancet* 339:1508–1509
- Willer M, Rainey M, Pullen T, Stirling CJ (1999) The yeast *CDC9* gene encodes both a nuclear and a mitochondrial form of DNA ligase I. *Curr Biol* 9:1085–1094
- Yardimci H, Loveland AB, Habuchi S, van Oijen AM, Walter JC (2010) Uncoupling of sister replisomes during eukaryotic DNA replication. *Mol Cell* 40:834–840
- Zegerman P, Diffley JF (2007) Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* 445:281–285
- Zhang Z, Hayashi MK, Merkel O, Stillman B, Xu RM (2002) Structure and function of the BAH-containing domain of Orc1p in epigenetic silencing. *EMBO J* 21:4600–4611

Chapter 2

Evolutionary Diversification of Eukaryotic DNA Replication Machinery

Stephen J. Aves, Yuan Liu, and Thomas A. Richards

Abstract DNA replication research to date has focused on model organisms such as the vertebrate *Xenopus laevis* and the yeast species *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. However, animals and fungi both belong to the Opisthokonta, one of about six eukaryotic phylogenetic ‘supergroups’, and therefore represent only a fraction of eukaryotic diversity. To explore evolutionary diversification of the eukaryotic DNA replication machinery a bioinformatic approach was used to investigate the presence or absence of yeast/animal replisome components in other eukaryotic taxa. A comparative genomic survey was undertaken of 59 DNA replication proteins in a diverse range of 36 eukaryotes from all six supergroups. Twenty-three proteins including Mcm2–7, Cdc45, RPA1, primase, some DNA polymerase subunits, RFC1–5, PCNA and Fen1 are present in all species examined. A further 20 proteins are present in all six eukaryotic supergroups, although not necessarily in every species: with the exception of RNase H2B and the fork protection complex component Timeless/Tof1, all of these are members of anciently derived paralogous families such as ORC, MCM, GINS or RPA. Together these form a set of 43 proteins that must have been present in the last common eukaryotic ancestor (LCEA). This minimal LCEA replisome is significantly more complex than the related replisome in Archaea, indicating evolutionary events including

S.J. Aves (✉)

Biosciences, College of Life and Environmental Sciences, University of Exeter,
Geoffrey Pope Building, Stocker Road, Exeter, EX4 4QD, UK
e-mail: s.j.aves@exeter.ac.uk

Y. Liu • T.A. Richards

Biosciences, College of Life and Environmental Sciences, University of Exeter,
Geoffrey Pope Building, Stocker Road, Exeter, EX4 4QD, UK

Department of Zoology, Natural History Museum,
Cromwell Road, London, SW7 5BD, UK
e-mail: yl274@exeter.ac.uk; thomr@nhm.ac.uk

duplications of DNA replication genes in the LCEA lineage. This pattern of early diversification of the DNA replisome in the LCEA is consistent with similar patterns seen in the early evolution of other complex eukaryotic cellular features.

Keywords Comparative genomics • Last common eukaryotic ancestor • Opisthokonta • Phylogeny • Supergroup

2.1 Introduction

Most of our knowledge of eukaryotic DNA replication comes from studies on model organisms such as the fungus *S. cerevisiae* and the animal *X. laevis*. But fungi and animals belong to just one of the six major eukaryotic ‘supergroups’ (Adl et al. 2005; Simpson and Roger 2004), so variation and diversification in DNA replication systems remain largely unexplored in the diversity of eukaryotic life. This diversity covers numerous biological forms including important parasite groups, keystone species in environmental processes, and independent lineages that have evolved multicellularity, cellular differentiation and a range of reproductive systems. The recent rise in availability of genome sequence data from a range of eukaryotes allows bioinformatic investigation of the extent to which the yeast/animal replisome components are present, absent, or expanded by gene duplication in other eukaryotic groups. This comparative genomic approach is proving an important tool for understanding the evolution and diversification of numerous cellular systems (Dacks and Field 2007; Dacks et al. 2008; DeGrasse et al. 2009; Hodges et al. 2010; Ramesh et al. 2005; Richards and Cavalier-Smith 2005; Wickstead et al. 2010), providing insight into how they operate and also identifying differentially distributed gene targets for therapeutic agents. This chapter will apply similar approaches to the diversification of DNA replication machinery in extant eukaryotes and the last common eukaryotic ancestor (LCEA). As part of this work we will also compare the eukaryotic form to its homologous counterpart in Archaea, giving insight into the ancestral diversification of this core cellular system.

2.2 Eukaryotic Diversity

Eukaryotes have unique features such as a nucleus and other complex cell structures, but also share many cellular and molecular characteristics with one or both of the other two domains of life, the Archaea (formerly, archaebacteria) and the Bacteria (eubacteria). The evolutionary origin of eukaryotes is hotly debated with a number of contesting hypotheses (Embley and Martin 2006; Martin et al. 2001; Martin and Muller 1998), many of which posit that this ancient transition involved endosymbiotic event(s) between two or more prokaryotes, one of which was a member, close relative or ancestor of the Archaea (Martin 2005; Martin et al. 2001).

Indeed, some have claimed an archaeon was the progenitor of the nucleus and represented the first endosymbiotic event in the eukaryotic lineage (Lake and Rivera 1994). Regardless of the details of eukaryogenesis, the similarities of the eukaryote and Archaea DNA replisome and the non-homologous nature of the bacterial replisome are certainly consistent with shared ancestry between Archaea and at least a subsection of primary eukaryotic conglomerations. Whether this subsection derives from an ancestor within the Archaea, or whether Eukarya and Archaea share a common ancestor (the so-called ‘two primary domains’ or ‘three primary domains’ (2D or 3D) scenarios), is the subject of much debate (Gribaldo et al. 2010). What is certain, however, is that many complex cellular characters evolved after the initial conglomeration event(s) in the early eukaryotic lineage and before the diversification of the last common eukaryotic ancestor (LCEA) into extant and sampled taxa. These complex cellular characters include diverse elements of the cytoskeleton (Richards and Cavalier-Smith 2005; Wickstead and Gull 2011; Wickstead et al. 2010), nuclear pore complexes (DeGrasse et al. 2009), elements of the endomembrane system (Dacks and Field 2007; Dacks et al. 2008), centrioles (Hodges et al. 2010) and many genes encoding the machinery of meiosis (Ramesh et al. 2005).

Evolutionary and taxonomic explanations for the diversity of present-day eukaryotic forms are in a state of flux, with different datasets and rival hypotheses identifying a number of different phylogenetic trees and taxonomic hierarchies. These phylogenetic trees reveal between three and eight major eukaryotic clades, the exact number depending on the analysis performed and the dataset used (Bapteste et al. 2002; Burki et al. 2007, 2008; Hampl et al. 2009; Rodriguez-Ezpeleta et al. 2005, 2007). Animals and fungi, together with some unicellular organisms such as free-living choanoflagellates, parasitic Ichthyosporea, and amoeboid organisms known as nucleariids, belong to the **Opisthokonta**, which is currently recognised as one of the six major eukaryotic phylogenetic ‘supergroups’ (Adl et al. 2005; Simpson and Roger 2004). ‘Opisthokont’ means ‘posterior flagellum’ and refers to the characteristic single rear organ of motility possessed by some animal and fungal cells (think sperm, or the motile zoospores of chytrid fungi) and represents one of the most consistently recovered phylogenetic groupings (Burki et al. 2007, 2008). Flattened mitochondrial cristae are the other ancestral defining feature of this supergroup (Patterson 1999). These cytological characteristics and molecular phylogenies have been used to demonstrate that this group represents a holophyletic clade (Cavalier-Smith 2003; Lang et al. 2002), which helps to explain why yeasts are useful model organisms for biomedical studies. However, we note that both yeast species commonly used for experimental study have undergone relatively recent gene loss events, in some cases limiting their use as comparative models; we discuss examples of this below. For comparative genomics, the opisthokonts represent one of the best sampled groups, with over 100 fungal genomes reported and numerous animal genomes representing the wide diversity of metazoan forms. Increasing effort has been applied to genome sequencing of single cellular relatives of the fungi and animals, including the choanoflagellate *Monosiga brevicollis* (King et al. 2008), while a sequencing initiative to sample further opisthokont taxa

that branch in and around the fungi and the animal radiations is also underway (Ruiz-Trillo et al. 2007).

A range of molecular evidence suggests that the opisthokonts form a sister branch to the **Amoebozoa** supergroup (Baptiste et al. 2002; Burki et al. 2008; Richards and Cavalier-Smith 2005), which includes diverse forms of amoebic protozoa. In terms of genome projects this supergroup is less well represented, with genomes of the cellular slime mould *Dictyostelium discoideum* and the anaerobic dysentery pathogen *Entamoeba histolytica* completed, and that of *Acanthamoeba castellanii* underway.

The positions of the remaining groups, and indeed the number of major clades and how they branch relative to the root of the eukaryotes, remain unclear. However, recognised major groups include the **Plantae** supergroup (also known as Archaeplastida – referring to the ancient primary endosymbiosis of a cyanobacterium – (Adl et al. 2005; Gould et al. 2008)). This contains the familiar land plants (e.g. *Arabidopsis thaliana* and the moss *Physcomitrella patens* genomes) and green algae (e.g. *Chlamydomonas reinhardtii* and *Ostreococcus tauri* genomes), as well as the red algae (rhodophytes – e.g. *Cyanidioschyzon merolae* genome), and a small group of unicellular algae, the glaucophytes. Other algal groups can be found in the **Chromalveolata**, **Rhizaria** and **Excavata**, and are all the product of multiple secondary and/or tertiary endosymbiotic transfers of plastids (Archibald 2009).

The supergroup **Chromalveolata** has changed in terms of constituent groups on a number of occasions. It was originally proposed as a major grouping united by an ancient secondary endosymbiosis of a red alga (Cavalier-Smith 2000). This larger grouping (sometimes called Chromista (Cavalier-Smith 1987, 1998)) has undergone a number of revisions (Burki et al. 2007, 2008) and recent phylogenetic data suggest that there were two separate red algal endosymbioses (Baurain et al. 2010). As such, current versions of the Chromalveolata encompass the alveolates and the stramenopiles which include for example the photosynthetic diatoms (e.g. *Thalassiosira pseudonana* and *Phaeodactylum tricornerutum* genomes), brown algae (e.g. *Ectocarpus siliculosus* and the microalga *Aureococcus anophagefferens*), dinoflagellates, *Chromera* and their non-photosynthetic relatives such as the oomycete potato blight pathogen *Phytophthora*, ciliates (e.g. *Tetrahymena* and *Paramecium*), and parasitic apicomplexa. Many of the apicomplexa possess a remnant plastid organelle, the apicoplast, for example the causative agents of toxoplasmosis and malaria (e.g. *Toxoplasma gondii* and *Plasmodium falciparum* genomes).

Also traditionally included within the Chromalveolata are a group now sometimes referred to as ‘Hacrobia’ – the haptophytes and cryptomonads (cryptophytes). Haptophytes include the coccolithophores, such as *Emiliania huxleyi*, which are ecologically and geologically important phytoplankton, capable of forming huge blooms and whose calcareous platelets form a major constituent of chalk and limestone sedimentary rocks. The Hacrobia acquired their plastids from a red algal endosymbiosis, and current data suggest they constitute a monophyletic group (Okamoto et al. 2009; Patron et al. 2007) along with several heterotrophic protists e.g. the Katablepharids and Telonemids (Burki et al. 2008). At present Hacrobia are

poorly represented by genome sequences and are in a state of phylogenetic limbo as recent analyses suggest the possibility that they may belong to the Plantae supergroup rather than the Chromalveolata (Burki et al. 2008; Hampl et al. 2009; Patron et al. 2007); they are not included in this analysis.

The **Rhizaria** supergroup was defined from molecular data (Archibald et al. 2003; Bass et al. 2005) and unites a diversity of planktonic and benthic heterotrophs with phototrophs derived from another secondary endosymbiosis, in this case a green algal endosymbiosis (e.g. *Bigelowiella natans* for which the genome is currently being sequenced). Some phylogenetic studies indicate affinity between the Rhizaria and certain chromalveolate groups (Burki et al. 2007), but deep evolutionary relationships between the supergroups remain controversial and the Rhizaria will be treated as a separate supergroup in this discussion consistent with the current taxonomic framework (Adl et al. 2005).

The final supergroup, the **Excavata**, comprises mainly flagellates with a wide diversity of morphological forms, most notably the agents that cause sleeping sickness (e.g. *Trypanosoma brucei* genome), giardiasis (e.g. *Giardia intestinalis* genome), and trichomoniasis (e.g. *Trichomonas vaginalis* genome). The Excavata has been a contentious grouping because they share no single defining morphological character – rather they possess a suite of overlapping cellular characters (Simpson et al. 2006). Attempts to test the phylogenetic relationships of these groups have been greatly affected by artefacts such as long-branch attraction (Philippe 2000; Rodriguez-Ezpeleta et al. 2007). However, a recent phylogenomic analysis focused on correcting such artefacts supports the monophyly of the Excavata and confirms a subsection of the excavates including the Discoba (e.g. *Trypanosoma*, *Naegleria* and *Euglena*), metamonads and *Malawimonas* is monophyletic when only slowly-evolving sites are sampled for phylogenetic analysis (Hampl et al. 2009; Rodriguez-Ezpeleta et al. 2007). The status of this group remains controversial however: it includes several long-branch forming taxa, which group together in the Metamonada (e.g. *Giardia* and *Trichomonas*) (Cavalier-Smith 2003). This very group has been suggested to include the primary branch in the eukaryotic radiation (Morrison et al. 2007), implying the root of the eukaryotes may lie within a subsection of the excavates and this may therefore not be a holophyletic group when rooted.

Even from the brief outline presented here it can be seen that there is a huge diversity of eukaryotic life, and that each of the supergroups contains organisms of great ecological and medical importance. To what extent is the process of DNA replication conserved or diverged across these taxa? Notwithstanding some experimental data for plants (Bryant 2010) and trypanosomes (e.g. Dang and Li 2011), little replication research has been carried out on non-animal/fungi organisms, so this question is being addressed by bioinformatic studies using completed genome sequences. These comparisons also enable us to identify which features of the DNA replication system are conserved and ancestral to all sampled eukaryotic forms, and which features are derived. Such analysis is important for comparisons with prokaryotic replication systems, understanding how the replisome has diversified as cell complexity has evolved, and identifying therapeutic targets.

2.3 Conservation of Replisome Proteins

A comparative genomic survey of MCM proteins (see Chaps. 6 and 7 for detailed description) in a diverse range of 36 eukaryotes from all six supergroups is shown in Fig. 2.1. BLAST, PSI-BLAST (Altschul et al. 1997), and local Pfam (Bateman et al. 2004) searches using Hidden Markov Models were performed to identify MCM orthologues, with phylogenetic analysis to confirm the identities of the individual MCM paralogues (Liu et al. 2009). In cases of apparent absence, Expressed Sequence Tag (EST) and Genome Survey Sequences (GSS) data of closely-related species were also searched. This analysis enabled us to identify the distribution of DNA replication proteins across the extant eukaryotes. We do not use these data to identify duplication events within each DNA replication subfamily; as such all references to gene duplications refer to anciently derived paralogues present in the LCEA.

All six of the Mcm2–7 helicase subunits were found to be present in all 36 eukaryotes sampled, consistent with the essential roles of all six subunits in the replicative helicase. However, the same pattern was not observed for the Mcm10 replisome protein (see Chap. 11) which in animals/fungi is required for replication initiation and elongation (Gambus et al. 2006; Moore and Aves 2008; Pacek et al. 2006). Mcm10, which is not related to Mcm2–7 as has no identifiable sequence similarity, appears absent from at least some species in three supergroups, and from both Amoebozoa species sampled. While it cannot of course be ruled out that homologues were not detected due to low homology, or that individual genome sequences may not have 100% coverage, this implies that, although Mcm10 has widespread distribution across the eukaryotes, in some species its replication roles are either not required or are provided by other factors.

The Mcm2–7 paralogues Mcm8, Mcm9 and MCM-BP also show widespread but patchy distributions across the eukaryotes, implying gene loss events in more than one lineage (Fig. 2.1). These proteins have received relatively little experimental attention, possibly because they are absent in *S. cerevisiae*, but in vertebrates they have been reported to function in aspects of DNA replication (Gozuacik et al. 2003; Kinoshita et al. 2008; Lutzmann and Mechali 2008; Maiorano et al. 2005; Volkening and Hoffmann 2005). Particularly notable in Fig. 2.1 is the concordant pattern of presence/absence of Mcm8 and Mcm9: in all but one case the absence of one gene corresponds with the absence of the other. This suggests that Mcm8 and Mcm9 may have associated functions in the cell. Phylogenetic analysis groups Mcm8 and Mcm9 as sister paralogues indicating that they also share co-ancestry. The one exception to the co-ordinate loss pattern is *Drosophila melanogaster* in which Mcm8 is present but Mcm9 is absent. This may be the exception that proves the rule however, because closer inspection reveals that all *Drosophila* species have a highly divergent Mcm8 which has a meiotic role; *Drosophila* therefore may not be a good model for Mcm8 in other organisms (Blanton et al. 2005; Liu et al. 2009; Matsubayashi and Yamamoto 2003).

MCM binding protein (MCM-BP) shares only limited homology with Mcm2–9 (Sakwe et al. 2007). However, MCM-BP interacts with MCM proteins and, at least

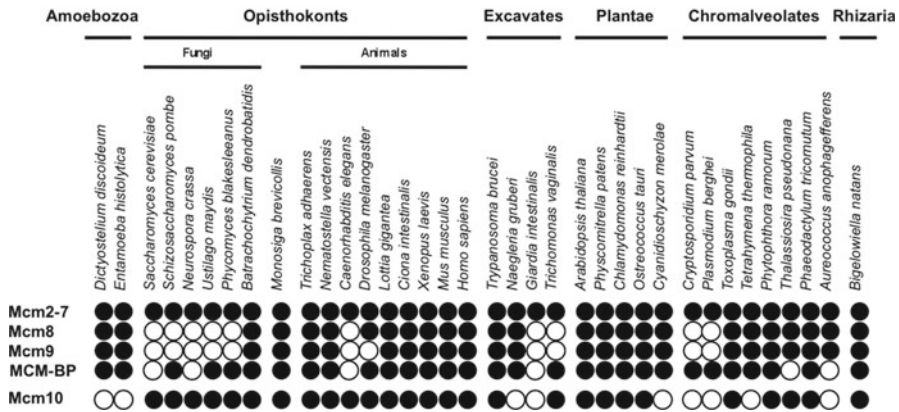


Fig. 2.1 Distribution of MCM proteins in eukaryotes. *Black circles* indicate detections and *white circles* indicate no homologue detected in a comparative genomic survey of 36 species (Figure adapted from Liu et al. 2009)

in animals and fission yeast, can form an alternative complex in which Mcm2 is replaced by MCM-BP (MCM^{MCM-BP}) (Ding and Forsburg 2011; Li et al. 2011; Nishiyama et al. 2011; Sakwe et al. 2007; Takahashi et al. 2008). *Xenopus* MCM-BP has been reported to participate in unloading of the Mcm2–7 complex from chromatin in late S-phase (Nishiyama et al. 2011). MCM-BP is widely distributed across eukaryote taxa but its patchy distribution is different from that of Mcm8/9 and also from that of Mcm10 (Fig. 2.1); this suggests that it does not function in association with these proteins and that its roles are dispensable, or are provided by other components in species such as *S. cerevisiae* and *Caenorhabditis elegans* that lack MCM-BP.

Comparative genomic surveys of 50 other replisome proteins, carried out across a diversity of eukaryotes as for the MCM proteins, are summarised in Fig. 2.2. It can be seen that some replication proteins, like Mcm2–7, are completely conserved in all species sampled – these include Cdc45, RPA1, primase, some DNA polymerase subunits, RFC1–5, PCNA and Fen1 – and are likely to be conserved because they perform a core function in the DNA replisome such as DNA unwinding, single-strand DNA binding, priming, DNA synthesis, clamp loading (where PCNA is the sliding clamp, see Chap. 15) or Okazaki fragment processing.

Other gene families, like Mcm8 and Mcm9, have a widespread distribution across all six supergroups but are absent from individual species, suggesting they have a shared and ancient ancestry but have been lost on multiple occasions. The third category of proteins is those, like Mcm10, which appear to be absent from one or more supergroups, although in almost all cases they have taxonomic distributions well beyond the opisthokonts. This demonstrates a high degree of conservation of the replisome system in eukaryotes. Figure 2.2 also indicates those eukaryotic replisome proteins which have homologues in Archaea.

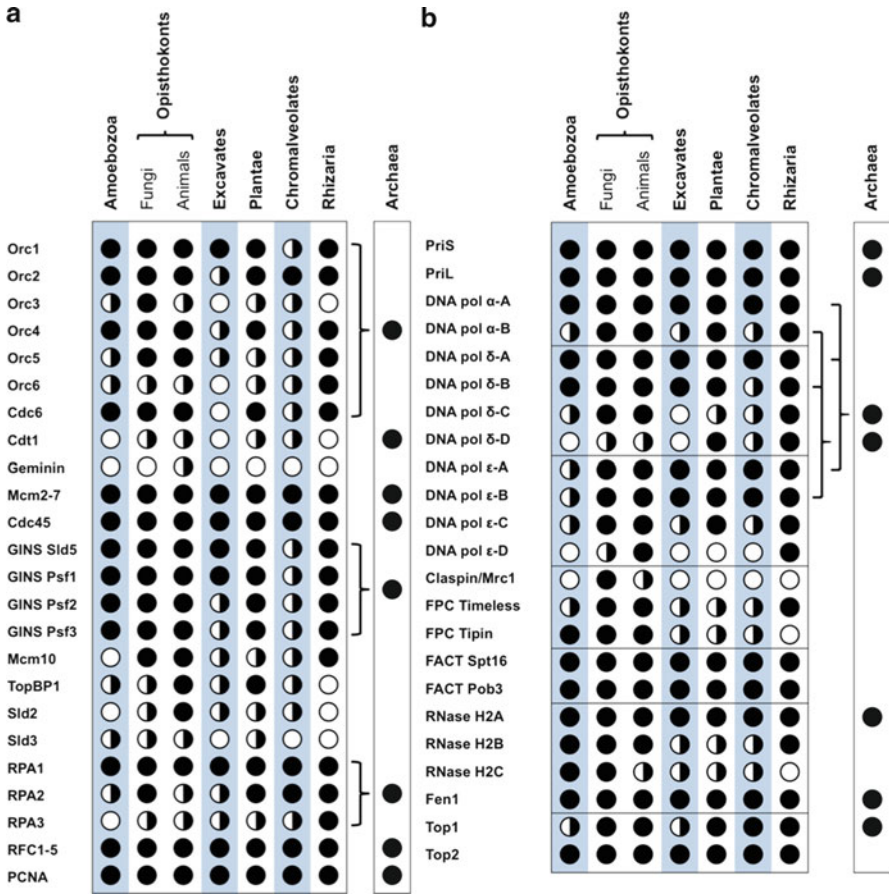


Fig. 2.2 Distribution of DNA replication proteins across eukaryotic supergroups. *Black dot* indicates proteins present in all species; *black/white dot* indicates proteins present in some species; *white dot* indicates undetected proteins. See Fig. 2.1 for genomes analysed. Replication proteins with established archaeal homologues are indicated (final column: *black dots*). Lines connecting the eukaryote rows to the Archaea rows indicate paralogue relationships. (a) Initiation, sliding clamp and clamp loader proteins. Distributions of Mcm8, Mcm9 and MCM-BP are in Fig. 2.1. (b) DNA synthesis and associated proteins. DNA polymerase subunits labelled ‘A’ and primase subunit PriS are catalytic; ‘DNA pol ε-C’ and ‘DNA pol ε-D’ designate Dpb3 and Dpb4 subunits respectively. *FACT* Facilitates Chromatin Transcription; *FPC* fork protection complex

2.4 Indispensable Replisome Proteins

Many replisome proteins are found in all eukaryotic species (indicated by a row of black dots in Fig. 2.2). These proteins therefore appear to be indispensable components of the replisome: certainly they have remained steadfast components of

the eukaryotic genome during the at least one billion years of evolution that has generated the huge diversity of eukaryotic forms (Berney and Pawlowski 2006; Parfrey et al. 2011). We predict these indispensable proteins provide key functions in the DNA replication process. Interestingly, almost all of them have homologues in archaeal genomes (Fig. 2.2).

The set of indispensable replication proteins includes the Mcm2–7 hexamer plus its accessory factor Cdc45, and the largest subunit of the RPA single-stranded DNA binding protein (see Chap. 10). These represent the key initiation function of DNA unwinding.

For the DNA synthesis functions, the sliding clamp PCNA plus all five subunits of the clamp loader RFC are completely conserved in eukaryotes (Fig. 2.2) (Chia et al. 2010), as are both primase subunits, the catalytic subunit of the initiating DNA polymerase α and the catalytic subunit of the processive DNA polymerase δ (see Chaps. 9 and 12). With the exception of the dysentery pathogen *Entamoeba histolytica*, the catalytic and B-subunit of the leading-strand processive DNA polymerase ϵ (Chap. 13) are also completely conserved in eukaryotes. Together these represent all the key activities for DNA synthesis on leading and lagging strands. For processing Okazaki fragments on the lagging strand, indispensable replication proteins ribonuclease H2A and flap endonuclease Fen1 (Chap. 16) are conserved (Fig. 2.2) and although not part of this study, it is likely that DNA ligase I (Chap. 17) can also be added to this list (Ellenberger and Tomkinson 2008). And for chromatin configuration, topoisomerase IIA (Top2) is conserved, as is the FACT (facilitates chromatin transcription) complex of Spt16 and Pob3/SSRP1 for histone interactions and nucleosome disassembly/reassembly (Formosa 2012).

Virtually all of the key indispensable replication proteins outlined above have homologues in Archaea but not in Bacteria (Barry and Bell 2006; Chia et al. 2010; Edgell and Doolittle 1997; Forterre and Gabelle 2009; Johansson and MacNeill 2010; MacNeill 2011; Marinsek et al. 2006; Robbins et al. 2005; Robinson and Bell 2007); only the FACT complex and possibly topoisomerase type IIA (Forterre and Gabelle 2009) appear to be eukaryotic innovations. Again this confirms that the DNA replisome was derived from a lineage within Archaea or a close relative, consistent with models of eukaryotic genesis that suggest an Archaea or Archaea-like entity contributed to the primary eukaryotic conglomeration. In many cases the eukaryotic core replication apparatus contains paralogues which in many Archaea are represented by a single ancestrally derived orthologue, for example the Mcm2–7 heterohexamer is present in all eukaryotes whereas many Archaea have a homohexameric replicative helicase; in those cases where Archaea possess multiple MCM proteins these are best explained by Archaea-specific gene duplications (Chia et al. 2010; Liu et al. 2009). Conserved eukaryotic paralogues such as Mcm2–7 arose by early gene duplication events of an archaeal-like MCM after this gene family was acquired by the eukaryotic progenitor cell prior to the LCEA (Liu et al. 2009). These wider observations suggest that a pattern of ancient gene duplication was important in the early evolution of the eukaryotic DNA replisome prior to the LCEA.

2.5 Replisome Proteins Present in All Eukaryotic Supergroups

In addition to the ‘indispensable’ eukaryotic replisome proteins, many other proteins are present in members of all six eukaryotic supergroups, although missing from particular species. These ‘anciently acquired but dispensable’ proteins must therefore represent gene products which were present in the LCEA but have been lost from different lineages; for example, Mcm8 and 9 have been lost on at least five occasions in evolutionary history (Liu et al. 2009). Each ‘anciently acquired but dispensable’ protein must either not be absolutely required for DNA replication, or its function can be substituted by other protein(s). In this context, it is notable that all but two of these 20 proteins are members of anciently derived paralogous gene families (ORC/Cdc6; Mcm2–9; GINS; RPA; DNA pol B; topoisomerase IB) (Figs. 2.1 and 2.2) with only RNase H2B and the fork protection complex (FPC) subunit Timeless (Tim1) having no evidence of ancient gene duplication and paralogues but being differentially lost. Note that these proteins are ‘dispensable’ only in an evolutionary sense: in any one species they may be performing an essential function (e.g. Orc6 is essential in *S. cerevisiae* (Li and Herskowitz 1993) but absent from the related ascomycete fungus *Neurospora crassa*). Examples of replication proteins in this ‘anciently acquired but dispensable’ category are ORC subunits Orc1, Orc2, Orc4 and Orc5; RPA subunit Rpa2; ribonuclease H2B; topoisomerase IB (Top1); the regulatory B-subunits of all three replicative DNA polymerases; and the Dpb3 subunit of DNA polymerase ϵ .

Interestingly, the individual ORC/Cdc6 and GINS subunits (see Chaps. 3 and 8) appear to be dispensable. While all species sampled in the Amoebozoa, Opisthokonta, Plantae and Rhizaria possess all four GINS subunits, individual subunits are absent in particular species of the Excavata and the Chromalveolata (Fig. 2.2). It is noteworthy that many Archaea possess only one GINS protein in their replisomes which has homology to two eukaryotic GINS subunits (Yoshimochi et al. 2008). In eukaryotes, the GINS and ORC/Cdc6 complexes are the only anciently derived paralogous gene families amongst the DNA replication proteins which do not contain at least one ‘indispensable’ member (Fig. 2.2).

Aside from the RNase H2B subunit, the Timeless (Tim1) protein of the FPC is the only protein with no evidence of anciently derived paralogues which is ‘anciently acquired but dispensable’. The FPC appears to be a eukaryotic innovation which is conserved across all supergroups but may be dispensable, in whole or in part, in particular species. The two components of the FPC, Timeless (Tof1 in *S. cerevisiae*; Swi1 in *S. pombe*) and Tipin (*ScCsm3*; *SpSwi3*) together function in yeasts and Metazoa to stabilise the paused replisome, activate the replication checkpoint and facilitate chromatin cohesion, thereby contributing to genome stability (Leman et al. 2010; McFarlane et al. 2010). It may be that in certain species both subunits are not required, or this function may be provided in a different manner, or may be less important due to the biology of the organism e.g. faster generation time or tolerance of higher mutation rates.

2.6 Replisome Proteins Not Present in All Supergroups

A minority of replisome proteins are only present in some supergroups. Some, like Mcm10, TopBP1/Dpb11, ORC subunits Orc3 and Orc6, RPA subunit 3, RNase H2C and subunits of DNA polymerases δ and ϵ , have widespread distribution and may possibly have been present in the LCEA but have not been detected in one or two supergroups to date (Figs. 2.1 and 2.2). A few proteins have a more limited distribution and may represent regulatory variations between taxa despite conserved DNA replication mechanisms (Errico and Costanzo 2010; Kearsey and Cotterill 2003). For example the FPC-interacting checkpoint mediator protein Claspin/Mrc1 is limited to opisthokonts, and geminin is an animal-specific inhibitor of the MCM loading factor Cdt1 (Fig. 2.2). It is possible that alternative factors act as regulators of Cdt1 in different eukaryotic taxa, such as the GEM protein in plants (Caro et al. 2007; Caro and Gutierrez 2007).

An alternative explanation for a limited distribution of a regulatory replication protein is that it may be poorly conserved at the sequence level and therefore difficult to detect across supergroups using bioinformatic methods. Sld3 is a case in point: this replication initiation protein was initially thought to be restricted to fungi, but experimental clues and advanced bioinformatic analysis revealed homology with the vertebrate Treslin/Ticrr protein (Kumagai et al. 2010; Sansam et al. 2010) and identified Sld3 homologues in the Plantae and Amoebozoa supergroups (Sanchez-Pulido et al. 2010). Sld3 function as well as structure is conserved between yeast and vertebrates: in yeast, phosphorylation of Sld3 and Sld2/Drc1 by cyclin-dependent kinase (CDK) leads to the formation of a ternary complex with the BRCT-domain protein Dpb11, which is required for CMG complex formation and initiation of DNA replication (Tanaka et al. 2007; Zegerman and Diffley 2007). Similarly, CDK-dependent phosphorylation of Treslin/Ticrr is required for binding to BRCT-domains of TopBP1, the vertebrate Dpb11, and initiation of DNA replication in both *Xenopus* and humans (Boos et al. 2011; Kumagai et al. 2010, 2011). Sld3 phosphorylation sites and the binding region of Dpb11 are conserved in metazoans: phosphorylated Treslin/Ticrr binds to BRCT repeats 1 and 2 of TopBP1, which are homologous to the Sld3-binding BRCT repeats 1 and 2 in Dpb11 (Boos et al. 2011).

The Dpb11 protein has homologues in at least five eukaryotic supergroups (also known as Mei1 in *Arabidopsis*; Mus101 in *Drosophila*; Rad4/Cut5 in *S. pombe*; TopBP1 in humans)(Garcia et al. 2005) which suggests that Sld3 and Sld2 may also be widely conserved. However, the situation for Sld2 is not straightforward in that its apparent animal homologue, the RecQL4 helicase, only shares homology in the N-terminal domain and, although it is required for initiation of DNA replication (Im et al. 2009; Matsuno et al. 2006; Sangrithi et al. 2005; Xu et al. 2009), it is not clear if CDK phosphorylation is conserved (Boos et al. 2011). Other TopBP1-binding proteins may also play a role in initiation of vertebrate DNA replication (Balestrini et al. 2010; Chowdhury et al. 2010). The extent of RecQL4 functional similarity with yeast Sld2 therefore remains to be determined (Masai 2011).

2.7 A Complex Ancestral Replisome

An important evolutionary point about replisome proteins represented in all six supergroups, regardless of their dispensability or otherwise, is that these must all have been present in the LCEA. This assumes that horizontal gene transfer is not a factor (Keeling and Palmer 2008; Richards et al. 2011) which is consistent with the complexity hypothesis which suggests gene transfer is rare in DNA replication-encoding gene families (Cotton and McInerney 2010; Jain et al. 1999). It is thus possible to deduce a core replisome present in the LCEA from the sum of the ‘indispensable’ and ‘anciently acquired but dispensable’ replication proteins (Fig. 2.3). It is immediately clear that this is much more complex than the ‘core’ archaeal replisome, i.e. involving additional novel gene families and duplicated members of the archaeal form. This indicates that many events occurred early in the evolution of the eukaryotic cell to produce the replisome of the LCEA, most notably a series of gene duplications to give rise to anciently derived paralogues of single proteins (MCM, GINS, RPA, B-family DNA polymerase, etc.) present in replisomes of extant Archaea. This observation is consistent with many other cellular systems e.g. nuclear pore complexes, membrane trafficking systems, molecular motors, protein complexes that control meiosis, where a large proportion

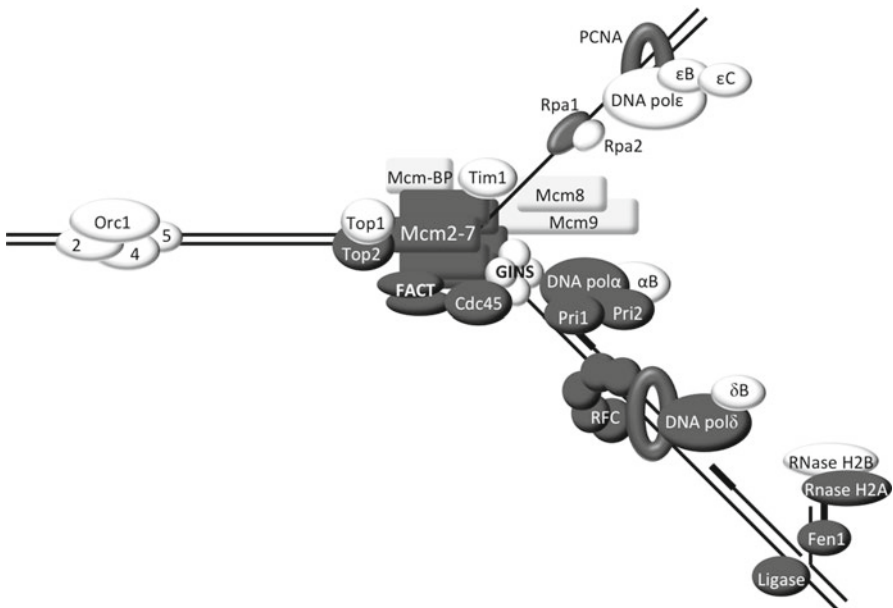


Fig. 2.3 Schematic diagram of the possible replisome in the LCEA with ‘indispensable’ proteins in *black* and others (‘anciently acquired but dispensable’) in *white*. DNA ligase I was not part of this study but is included in this diagram as it is likely to be conserved in eukaryotes (Ellenberger and Tomkinson 2008)

of the features are derived in the LCEA (Dacks and Field 2007; Dacks et al. 2008; DeGrass et al. 2009; Hodges et al. 2010; Ramesh et al. 2005; Richards and Cavalier-Smith 2005; Wickstead et al. 2010).

2.8 Conclusions

A high level of conservation across all six eukaryotic phylogenetic supergroups indicates that the last common eukaryotic ancestor (LCEA) possessed a complex DNA replication machinery comprising at least 43 proteins. Twenty-three of these ancestral replication proteins appear to be indispensable, in that they are present in the genome of all species sampled; the remaining 20 have been lost in some taxa implying that their function is not essential or can be provided by other factors. The replisome of the LCEA was significantly more complex than replisomes of related Archaea, possessing novel eukaryotic components and multiple paralogues. This indicates evolutionary events including gene duplications in the lineage leading to the LCEA, paralleling the acquisition of other complex cellular features in early eukaryotic evolution.

DNA replication research to date has been heavily concentrated on model opisthokonts. Studies should now be carried out on representatives of other phylogenetic supergroups to both test bioinformatic predictions and to seek other DNA replication components within the diversity of eukaryotic life.

References

- Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor MF (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* 52:399–451
- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Archibald JM, Longet D, Pawlowski J, Keeling PJ (2003) A novel polyubiquitin structure in Cercozoa and Foraminifera: evidence for a new eukaryotic supergroup. *Mol Biol Evol* Jan; 20(1):62–6. PubMed PMID: 12519907.
- Archibald JM (2009) The puzzle of plastid evolution. *Curr Biol* 19:R81–R88
- Balestrini A, Cosentino C, Errico A, Garner E, Costanzo V (2010) GEMC1 is a TopBP1-interacting protein required for chromosomal DNA replication. *Nat Cell Biol* 12:484–156
- Baptiste E, Brinkmann H, Lee JA, Moore DV, Sensen CW, Gordon P, Duruffe L, Gaasterland T, Lopez P, Muller M, Philippe H (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proc Natl Acad Sci USA* 99:1414–1419
- Barry ER, Bell SD (2006) DNA replication in the archaea. *Microbiol Mol Biol Rev* 70:876–887

- Bass D, Moreira D, López-García P, Polet S, Chao EE, von der Heyden S, Pawlowski J, Cavalier-Smith T (2005) Polyubiquitin insertions and the phylogeny of Cercozoa and Rhizaria. *Protist*. Aug; 156(2):149-61. PubMed PMID: 16171183.
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, Studholme DJ, Yeats C, Eddy SR (2004) The Pfam protein families database. *Nucleic Acids Res* 32:D138–D141
- Baurain D, Brinkmann H, Petersen J, Rodriguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H (2010) Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol* 27:1698–1709
- Berney C, Pawlowski J (2006) A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc Biol Sci* 273:1867–1872
- Blanton HL, Radford SJ, McMahan S, Kearney HM, Ibrahim JG, Sekelsky J (2005) REC, *Drosophila* MCM8, drives formation of meiotic crossovers. *PLoS Genet* 1:343–354
- Boos D, Sanchez-Pulido L, Rappas M, Pearl LH, Oliver AW, Ponting CP, Diffley JFX (2011) Regulation of DNA replication through Sld3-Dpb11 interaction is conserved from yeast to humans. *Curr Biol* 21:1152–1157
- Bryant JA (2010) Replication of nuclear DNA. In: Lüttge U, Beyschlag W, Büdel B, Francis D (eds) *Progress in botany*, vol 71. Springer, Berlin/Heidelberg, pp 25–60
- Burki F, Shalchian-Tabrizi K, Minge M, Skjaveland A, Nikolaev SI, Jakobsen KS, Pawlowski J (2007) Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One* 2:e790
- Burki F, Shalchian-Tabrizi K, Pawlowski J (2008) Phylogenomics reveals a new ‘megagroup’ including most photosynthetic eukaryotes. *Biol Lett* 4:366–369
- Caro E, Gutierrez C (2007) A green GEM: intriguing analogies with animal geminin. *Trends Cell Biol* 17:580–585
- Caro E, Castellano MM, Gutierrez C (2007) A chromatin link that couples cell division to root epidermis patterning in *Arabidopsis*. *Nature* 447:213–216
- Cavalier-Smith T (1987) The kingdom Chromista: origin and systematics. In: Round FE, Chapman DJ (eds) *Progress in phycological research*, vol 4. BioPress, Bristol, pp 309–348
- Cavalier-Smith T (1998) A revised six-kingdom system of life. *Biol Rev Camb Philos Soc* 73:203–266
- Cavalier-Smith T (2000) Membrane heredity and early chloroplast evolution. *Trends Plant Sci* 5: 174–182
- Cavalier-Smith T (2003) The excavate protozoan phyla Metamonada Grasse emend. (Anaeromonadea, Parabasalia, Carpediemonas, Eopharyngia) and Loukozoa emend. (Jakobea, Malawimonas): their evolutionary affinities and new higher taxa. *Int J Syst Evol Microbiol* 53:1741–1758
- Chia N, Cann I, Olsen GJ (2010) Evolution of DNA replication protein complexes in eukaryotes and archaea. *PLoS One* 5:e10866
- Chowdhury A, Liu G, Kemp M, Chen X, Katrangi N, Myers S, Ghosh M, Yao J, Gao Y, Bubulya P, Leffak M (2010) The DNA unwinding element binding protein DUE-B interacts with Cdc45 in preinitiation complex formation. *Mol Cell Biol* 30:1495–1507
- Cotton JA, McInerney JO (2010) Eukaryotic genes of archaeobacterial origin are more important than the more numerous eubacterial genes, irrespective of function. *Proc Natl Acad Sci USA* 107:17252–17255
- Dacks JB, Field MC (2007) Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *J Cell Sci* 120:2977–2985
- Dacks JB, Poon PP, Field MC (2008) Phylogeny of endocytic components yields insight into the process of nonendosymbiotic organelle evolution. *Proc Natl Acad Sci USA* 105:588–593
- Dang HQ, Li Z (2011) The Cdc45-Mcm2-7-GINS protein complex in trypanosomes regulates DNA replication and interacts with two Orc1-like proteins in the origin recognition complex. *J Biol Chem* 286:32424–32435
- DeGrasse JA, DuBois KN, Devos D, Siegel TN, Sali A, Field MC, Rout MP, Chait BT (2009) Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor. *Mol Cell Proteomics* 8:2119–2130
- Ding L, Forsburg SL (2011) *Schizosaccharomyces pombe* minichromosome maintenance-binding protein (MCM-BP) antagonizes MCM helicase. *J Biol Chem* 286:32918–32930

- Edgell DR, Doolittle WF (1997) Archaea and the origin(s) of DNA replication proteins. *Cell* 89:995–998
- Ellenberger T, Tomkinson AE (2008) Eukaryotic DNA ligases: structural and functional insights. *Annu Rev Biochem* 77:313–338
- Embley TM, Martin W (2006) Eukaryotic evolution, changes and challenges. *Nature* 440:623–630
- Errico A, Costanzo V (2010) Differences in the DNA replication of unicellular eukaryotes and metazoans: known unknowns. *EMBO Rep* 11:270–278
- Formosa T (2012) The role of FACT in making and breaking nucleosomes. *Biochim Biophys Acta* 1819:247–255
- Forterre P, Gabelle D (2009) Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Res* 37:679–692
- Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, Labib K (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol* 8:358–366
- Garcia V, Furuya K, Carr AM (2005) Identification and functional analysis of TopBP1 and its homologs. *DNA Repair* 4:1227–1239
- Gould SB, Waller RR, McFadden GI (2008) Plastid evolution. *Annu Rev Plant Biol* 59:491–517
- Gozuacik D, Chami M, Lagorce D, Faivre J, Murakami Y, Poch O, Biermann E, Knippers R, Brechot C, Paterlini-Brechot P (2003) Identification and functional characterization of a new member of the human Mcm protein family: hMcm8. *Nucleic Acids Res* 31:570–579
- Gribaldo S, Poole AM, Daubin V, Forterre P, Brochier-Armanet C (2010) The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat Rev Microbiol* 8:743–752
- Hampel V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups”. *Proc Natl Acad Sci USA* 106:3859–3864
- Hodges ME, Scheumann N, Wickstead B, Langdale JA, Gull K (2010) Reconstructing the evolutionary history of the centriole from protein components. *J Cell Sci* 123:1407–1413
- Im J-S, Ki S-H, Farina A, Jung 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 USA* 106:15628–15632
- Jain R, Rivera MC, Lake JA (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci USA* 96:3801–3806
- Johansson E, MacNeill SA (2010) The eukaryotic replicative DNA polymerases take shape. *Trends Biochem Sci* 35:339–347
- Kearsey SE, Cotterill S (2003) Enigmatic variations: divergent modes of regulating eukaryotic DNA replication. *Mol Cell* 12:1067–1075
- Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* 9:605–618
- King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzow R, Dirks W, Good M, Goodstein D, Lemons D, Li WQ, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, Rokhsar D, Sequencing JGI (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451:783–788
- Kinoshita Y, Johnson EM, Gordon RE, Negri-Bell H, Evans MT, Coolbaugh J, Rosario-Peralta Y, Samet J, Slusser E, Birkenbach MP, Daniel DC (2008) Colocalization of MCM8 and MCM7 with proteins involved in distinct aspects of DNA replication. *Microsc Res Tech* 71:288–297
- Kumagai A, Shevchenko A, Shevchenko A, Dunphy WG (2010) Treslin collaborates with TopBP1 in triggering the initiation of DNA replication. *Cell* 140:349–359
- Kumagai A, Shevchenko A, Shevchenko A, Dunphy WG (2011) Direct regulation of Treslin by cyclin-dependent kinase is essential for the onset of DNA replication. *J Cell Biol* 193:995–1007
- Lake JA, Rivera MC (1994) Was the nucleus the first endosymbiont? *Proc Natl Acad Sci USA* 91:2880–2881

- Lang BF, O'Kelly C, Nerad T, Gray MW, Burger G (2002) The closest unicellular relatives of animals. *Curr Biol* 12:1773–1778
- Leman AR, Noguchi C, Lee CY, Noguchi E (2010) Human Timeless and Tipin stabilize replication forks and facilitate sister-chromatid cohesion. *J Cell Sci* 123:660–670
- Li JJ, Herskowitz I (1993) Isolation of Orc6, a component of the yeast origin recognition complex by a one-hybrid system. *Science* 262:1870–1874
- Li J-J, Schnick J, Hayles J, MacNeill SA (2011) Purification and functional inactivation of the fission yeast MCM^{MCM-BP} complex. *FEBS Lett* 585:3850–3855
- Liu Y, Richards TA, Aves SJ (2009) Ancient diversification of eukaryotic MCM DNA replication proteins. *BMC Evol Biol* 9:60
- Lutzmann M, Mechali M (2008) MCM9 binds Cdt1 and is required for the assembly of prereplication complexes. *Mol Cell* 31:190–200
- MacNeill SA (2011) Protein-protein interactions in the archaeal core replisome. *Biochem Soc Trans* 39:163–168
- Maiorano D, Cuvier O, Danis E, Mechali M (2005) MCM8 is an MCM2-7-related protein that functions as a DNA helicase during replication elongation and not initiation. *Cell* 120:315–328
- Marinsek N, Barry ER, Makarova KS, Dionne I, Koonin EV, Bell SD (2006) GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep* 7:539–545
- Martin W (2005) Archaeobacteria (Archaea) and the origin of the eukaryotic nucleus. *Curr Opin Microbiol* 8:630–637
- Martin W, Muller M (1998) The hydrogen hypothesis for the first eukaryote. *Nature* 392:37–41
- Martin W, Hoffmeister M, Rotte C, Henze K (2001) An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biol Chem* 382:1521–1539
- Masai H (2011) RecQL4: a helicase linking formation and maintenance of a replication fork. *J Biochem* 149:629–631
- Matsubayashi H, Yamamoto MT (2003) REC, a new member of the MCM-related protein family, is required for meiotic recombination in *Drosophila*. *Genes Genet Syst* 78:363–371
- Matsuno K, Kumano M, Kubota Y, Hashimoto Y, Takisawa H (2006) The N-terminal noncatalytic region of *Xenopus* RecQ4 is required for chromatin binding of DNA polymerase alpha in the initiation of DNA replication. *Mol Cell Biol* 26:4843–4852
- McFarlane RJ, Mian S, Dalgaard JZ (2010) The many facets of the Tim-Tipin protein families' roles in chromosome biology. *Cell Cycle* 9:700–705
- Moore K, Aves SJ (2008) Mcm10 and DNA replication in fission yeast. *SEB Exp Biol Ser* 59:45–69
- Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, Davids BJ, Dawson SC, Elmendorf HG, Hehl AB, Holder ME, Huse SM, Kim UU, Lasek-Nesselquist E, Manning G, Nigam A, Nixon JEJ, Palm D, Passamaneck NE, Prabhu A, Reich CI, Reiner DS, Samuelson J, Svard SG, Sogin ML (2007) Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* 317:1921–1926
- Nishiyama A, Frappier L, Mechali M (2011) MCM-BP regulates unloading of the MCM2-7 helicase in late S phase. *Genes Dev* 25:165–175
- Okamoto N, Chantangsi C, Horak A, Leander BS, Keeling PJ (2009) Molecular phylogeny and description of the novel katablepharid *Roombia truncata* gen. et sp. nov., and establishment of the Hacrobia taxon nov. *PLoS One* 4:e7080
- Pacek M, Tutter AV, Kubota Y, Takisawa H, Walter JC (2006) Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol Cell* 21:581–587
- Parfrey LW, Lahr DJG, Knoll AH, Katz LA (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci USA* 108:13624–13629
- Patron NJ, Inagaki Y, Keeling PJ (2007) Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr Biol* 17:887–891
- Patterson DJ (1999) The diversity of eukaryotes. *Am Nat* 154:S96–S124
- Philippe H (2000) Opinion: long branch attraction and protist phylogeny. *Protist* 151:307–316

- Ramesh MA, Malik SB, Logsdon JM (2005) A phylogenomic inventory of meiotic genes: evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr Biol* 15:185–191
- Richards TA, Cavalier-Smith T (2005) Myosin domain evolution and the primary divergence of eukaryotes. *Nature* 436:1113–1118
- Richards TA, Leonard G, Soanes DM, Talbot NJ (2011) Gene transfer into the fungi. *Fungal Biol Rev* 25:98–110
- Robbins JB, McKinney MC, Guzman CE, Sriratana B, Fitz-Gibbon S, Ha T, Cann IKO (2005) The euryarchaeota, nature's medium for engineering of single-stranded DNA-binding proteins. *J Biol Chem* 280:15325–15339
- Robinson NP, Bell SD (2007) Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc Natl Acad Sci USA* 104:5806–5811
- Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Loffelhardt W, Bohnert HJ, Philippe H, Lang BF (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol* 15:1325–1330
- Rodriguez-Ezpeleta N, Brinkmann H, Burger G, Roger AJ, Gray MW, Philippe H, Lang BF (2007) Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Curr Biol* 17:1420–1425
- Ruiz-Trillo I, Burger G, Holland PWH, King N, Lang BF, Roger AJ, Gray MW (2007) The origins of multicellularity: a multi-taxon genome initiative. *Trends Genet* 23:113–118
- Sakwe AM, Nguyen T, Athanasopoulos V, Shire K, Frappier L (2007) Identification and characterization of a novel component of the human minichromosome maintenance complex. *Mol Cell Biol* 27:3044–3055
- Sanchez-Pulido L, Diffley JFX, Ponting CP (2010) Homology explains the functional similarities of Treslin/Ticrr and Sld3. *Curr Biol* 20:R509–R510
- Sangrithi MN, Bernal JA, Madine M, Philpott A, Lee J, Dunphy WG, Venkitaraman AR (2005) Initiation of DNA replication requires the RECQL4 protein mutated in Rothmund-Thomson syndrome. *Cell* 121:887–898
- Sansam CL, Cruz NM, Danielian PS, Amsterdam A, Lau ML, Hopkins N, Lees JA (2010) A vertebrate gene, *ticrr*, is an essential checkpoint and replication regulator. *Genes Dev* 24:183–194
- Simpson AGB, Roger AJ (2004) The real 'kingdoms' of eukaryotes. *Curr Biol* 14:R693–R696
- Simpson AG, Inagaki Y, Roger AJ (2006) Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. *Mol Biol Evol* 23:615–625
- Takahashi N, Lammens T, Boudolf V, Maes S, Yoshizumi T, De Jaeger G, Witters E, Inze D, De Veylder L (2008) The DNA replication checkpoint aids survival of plants deficient in the novel replisome factor ETG1. *EMBO J* 27:1840–1851
- 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* 445:328–332
- Volkening M, Hoffmann I (2005) Involvement of human MCM8 in prereplication complex assembly by recruiting hcd6 to chromatin. *Mol Cell Biol* 25:1560–1568
- Wickstead B, Gull K (2011) The evolution of the cytoskeleton. *J Cell Biol* 194:513–525
- Wickstead B, Gull K, Richards TA (2010) Patterns of kinesin evolution reveal a complex ancestral eukaryote with a multifunctional cytoskeleton. *BMC Evol Biol* 10:110
- Xu Y, Lei Z, Huang H, Dui W, Liang X, Ma J, Jiao R (2009) dRecQ4 is required for DNA synthesis and essential for cell proliferation in *Drosophila*. *PLoS One* 4:e6107
- 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* 283:1601–1609
- Zegerman P, Diffley JFX (2007) Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* 445:281–285

Chapter 3

The Origin Recognition Complex: A Biochemical and Structural View

Huilin Li and Bruce Stillman

Abstract The origin recognition complex (ORC) was first discovered in the baker's yeast in 1992. Identification of ORC opened up a path for subsequent molecular level investigations on how eukaryotic cells initiate and control genome duplication in each cell cycle. Twenty years after the first biochemical isolation, ORC is now taking on a three-dimensional shape, although a very blurry shape at the moment, thanks to the recent electron microscopy and image reconstruction efforts. In this chapter, we outline the current biochemical knowledge about ORC from several eukaryotic systems, with emphasis on the most recent structural and biochemical studies. Despite many species-specific properties, an emerging consensus is that ORC is an ATP-dependent machine that recruits other key proteins to form pre-replicative complexes (pre-RCs) at many origins of DNA replication, enabling the subsequent initiation of DNA replication in S phase.

Keywords DNA replication • Replicative helicase loader • Eukaryotes • Replication initiators

H. Li (✉)
Department of Biochemistry and Cell Biology,
Stony Brook University, Stony Brook, NY 11794, USA

Biology Department, Brookhaven National
Laboratory, Upton, NY 11973, USA
e-mail: hli@bnl.gov

B. Stillman
Cold Spring Harbor Laboratory, Cold Spring Harbor,
NY 11724, USA
e-mail: bstillman@cshl.edu

3.1 Introduction

At the most fundamental level, the concept that at least one origin of DNA replication and multiple protein factors are required to initiate the physical process of genome duplication is conserved across the three domains of life (Kawakami and Katayama 2010; Mendez and Stillman 2003). In eukaryotes, an origin of DNA replication is a stretch of DNA sequence where the Origin Recognition Complex (Bell and Stillman 1992) (ORC) binds and subsequently recruits other factors to establish a pre-replicative complex (pre-RC) (Bell and Dutta 2002; Bielinsky and Gerbi 2001; DePamphilis 2003; Diffley and Labib 2002; Kawakami and Katayama 2010; Mendez and Stillman 2003; Scholefield et al. 2011; Cocker et al. 1996). In budding yeast, the genetically defined replication origins are usually near and often overlap with the replication start sites (Marahrens and Stillman 1992; Rao et al. 1994; Bielinsky and Gerbi 1998; Brewer and Fangman 1987; Theis and Newlon 1994); in mammalian species start sites have not been genetically characterized in sufficient detail but have been reported as new ORC binding sites (Abdurashidova et al. 2003; Bielinsky and Gerbi 2001). In *Drosophila*, in which origins of DNA replication in a metazoan species have been studied in most detail, ORC binding sites are located near actual sites of initiation of DNA replication and also correspond to start sites of transcription (Austin et al. 1999; Kim et al. 2011; Royzman et al. 1999; Xie and Orr-Weaver 2008; Beall et al. 2002; Bielinsky et al. 2001; Chesnokov et al. 1999; Gossen et al. 1995; MacAlpine et al. 2010; Sher et al. 2012). Simple organisms such as viruses and bacteria use a single origin of replication (Kawakami and Katayama 2010). In eukaryotes, due to their vastly expanded genome size and the hierarchical structure of the chromosomes, there are hundreds to tens of thousands of replication origins, depending on the organism (Gilbert 1998, 2010; Ryba et al. 2010; Cvetic and Walter 2005; DePamphilis et al. 2006). The existence and utilization of the great number of origins is likely to ensure that these large genomes can be duplicated within a reasonable time frame and certainly within a single cell division cycle.

Aside from the fact that origins are located near to ORC binding sites and promote initiation of DNA replication, there is little consensus among eukaryotic species as to what constitutes an origin of replication. This is further complicated because ORC lacks sufficient DNA binding specificity to predict the location of origins of DNA replication, even in the budding yeasts where there is some sequence specificity to ORC binding (Chang et al. 2011). In this regard, the yeast *S. cerevisiae* is perhaps an exception rather than the norm. In *S. cerevisiae*, the replication origins are well-defined and ScORC binds to the genetically defined origins, but even here the consensus sequence is rather variable (Chang et al. 2011; Marahrens and Stillman 1992; Theis and Newlon 2001; Bell and Stillman 1992). The *S. cerevisiae* origins of replication are autonomously replicating sequences (ARS), 100–150 bp long, and constitute three of four elements termed A, B1, and B2, with an auxiliary element in some origins called B3 (Marahrens and Stillman 1992). Element A contains the AT-rich 11 bp ARS consensus sequence (ACS) that is the most conserved and A and

B1 contribute to ORC binding specificity (Bell and Stillman 1992; Deshpande and Newlon 1992; Van Houten and Newlon 1990). The B2 element contains the double strand DNA unwinding element (DUE) where DNA replication starts and this element is required for loading the pre-RC and DNA helicase component Mcm2–7 [mini-chromosome maintenance subunits 2–7] (Zou and Stillman 2000; Wilmes and Bell 2002). B3 is an accessory sequence 22 bp long and at the *ARS1* origin binds the transcription factor Abf1 (Marahrens and Stillman 1992). There are several hundred origins in *S. cerevisiae* and they all share the same ACS sequence and the general three-element architecture. In other eukaryotic organisms such as *Schizosaccharomyces pombe*, *Drosophila melanogaster*, *Xenopus laevis* and *Homo sapiens*, the origin sequence pattern is not so well defined, except for the fact that they are generally contain AT rich sequences. It is clear now that, in these organisms, certain features outside ORC may be more important than ORC in defining replication origins. These additional determinants may include the local chromatin structure such as nucleosome positioning (Aggarwal and Calvi 2004; Calvi et al. 2007; Chang et al. 2011; Eaton et al. 2010; Lipford and Bell 2001; MacAlpine et al. 2010; Zou et al. 2006), chromatin modifications (Eaton et al. 2011; Liu et al. 2012; Weber et al. 2008; Hartl et al. 2007), transcription regulation (Karnani et al. 2010; MacAlpine et al. 2004), extra protein or RNA partners (Norseen et al. 2008; Thomae et al. 2008; Bartke et al. 2010; Shen et al. 2010), and potentially physical properties, such as the rigidity or the malleability of the DNA fragment (Cao et al. 2008; Huang and Kowalski 1993; Natale et al. 1993).

In contrast to the great divergence in the number and sequence of the eukaryotic replication origins, ORC, the ATP-dependent molecular machine that binds to those origins and helps to execute DNA replication, is well conserved throughout evolution, at least at the amino acid sequence level (Gavin et al. 1995; Tugal et al. 1998; Speck et al. 2005; Clarey et al. 2006). There are, however, considerable differences between species with regard to the stability and composition of ORC subunits during the cell division cycle, a topic that is addressed below with a discussion of selected species. Furthermore, ORC has functions and activities well beyond DNA replication (Sasaki and Gilbert 2007). Here we limit our discussion to the principal role of ORC in replication initiation, although a summary of ORC activities and structures is summarized at the end of this review.

ORC is composed of six protein subunits Orc1–6, named initially in the yeast according to their molecular masses, with Orc1 being the largest subunit (120 kDa) and Orc6 the smallest (50 kDa) (Bell et al. 1995; Bell and Stillman 1992). ORC subunits in other eukaryotes are named according to their function and amino acid sequence conservation with their yeast counterparts (Fig. 3.1). ORC is an ATPase and its binding to origin DNA is usually ATP-dependent (Klemm et al. 1997; Lee and Bell 2000; Speck et al. 2005; Takenaka et al. 2004; Bell and Stillman 1992). Despite the overall conservation, ORC has sufficiently evolved to warrant us to discuss ORC from different species individually. In the following sections, we will briefly survey four ORCs from *S. cerevisiae*, *S. pombe*, *D. melanogaster*, and *H. sapiens*, outlining some biochemical and structural studies that have been reported.

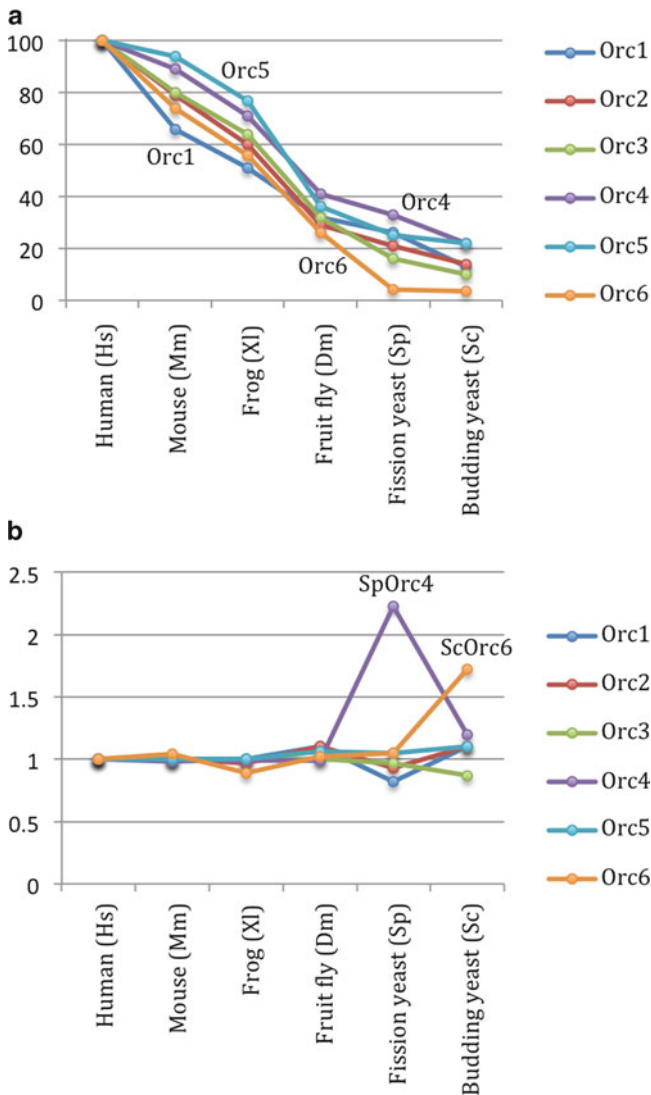


Fig. 3.1 Conservation of the six ORC subunits among six selected eukaryotic species. **(a)** Sequence identity of Orc1 through Orc6 as compared to the corresponding human subunits. **(b)** The relative amino acid sequence length of Orc1 to Orc6 with that of the human subunit set at 1

3.2 The *S. cerevisiae* ORC

The most distinctive feature of ScORC is that it forms a hetero-hexamers consisting of the Orc1, Orc2, Orc3, Orc4, Orc5 and Orc6 subunits, forming a stable structure throughout the cell division cycle that constitutively binds all of the origins of replication (Gibson et al. 2006). ScORC binds in a DNA sequence-specific manner,

although compared to sequence-specific transcription factors, ScORC has low binding affinity for its highly variable recognition site, which consists primarily of the A and B1 elements of the double strand DNA origins (Rao and Stillman 1995; Rowley et al. 1995). ScORC exhibits a high affinity for single strand DNA (ssDNA) ($K_d \approx 10^{-8}$) in a sequence non-specific manner and without the requirement for ATP, as long as the ssDNA is longer than 80 bases (Lee et al. 2000; Clarey et al. 2006; Lee and Bell 1997; Speck et al. 2005).

Among the six subunits, the first five subunits, Orc1 through Orc5, all contain a predicted AAA+ domain, and are essential for DNA binding, although only Orc1, Orc2, Orc4, and Orc5 appear to have direct contact to the origin DNA (Clarey et al. 2006; Lee and Bell 1997; Speck et al. 2005). This suggests that Orc3 may function to glue the DNA-contacting subunits together as a stable complex, but not bind DNA directly. Orc6 is the only subunit that does not contain a predicted AAA+ domain, and does not bind DNA. However, Orc6 is an essential subunit for DNA replication because Orc6 recruits multiple Cdt1 molecules during repeated loading of the replicative helicase core, the Mcm2–7 hexamer (Asano et al. 2007; Chen and Bell 2011; Chen et al. 2007; Takara and Bell 2011).

The largest subunit Orc1 is unique among ORC subunits, because it has a N-terminal 235-residue bromo-adjacent homology (BAH) domain that interacts with the C-terminal region of silencing regulator Sir1 (Zhang et al. 2002; Bose et al. 2004; Fox et al. 1997; Gardner et al. 1999; Hou et al. 2005; Ozaydin and Rine 2010; Triolo and Sternglanz 1996). The BAH domain is not essential, but its presence in Orc1 can influence origin binding specificity, as it also does in human ORC (Muller et al. 2010; Noguchi et al. 2006). This latter specificity may be related to the known nucleosome binding properties of the BAH domain present in the Orc1-related protein called Sir3 (Armache et al. 2011; Hickman and Rusche 2010). Sir3 is a regulator of the silent mating type genes in yeast and maintains certain mating type gene loci transcriptionally silent. The BAH domain of Orc1 binds to the Sir1 protein, which is also required for efficient silencing of the silent mating type loci. In the absence of Sir1 or the Orc1 BAH domain, different epigenetic states of mating type gene expression are established (Bell et al. 1993, 1995; Pillus and Rine 1989, 2004; Zhang et al. 2002). The crystal structure of the Sir3 AAA+ domain has recently been determined and although it does not have an ATPase activity (unlike the Orc1 AAA+ domain), the Sir3 AAA+ domain has evolved to bind to its partner Sir4 (silent information regulator protein) and to chromatin containing non-methylated H3K79 residues (Ehrentraut et al. 2011). Thus, after duplication of the Orc1 gene, the Sir3 allele evolved to acquire diverse biochemical functions for an AAA+ protein, even though it retained the same overall structural features of the AAA+ domain.

Both Orc1 and Orc5 can bind ATP (Klemm et al. 1997; Klemm and Bell 2001; Takehara et al. 2008), but the ATPase activity of ScORC primarily resides in the Orc1 subunit, and relies on the presence of an arginine finger in Orc4 (Bowers et al. 2004; Randell et al. 2006). Orc1 has a predicted classic AAA+ATPase domain, with functional Walker A and Walker B motifs and the ATPase activity controlled by the insertion of an arginine residue present in the Orc4 subunit into the active site of the Orc1 ATP binding site (Bowers et al. 2004). Orc1 ATPase activity is required for loading of multiple subunits of the Mcm2–7 hexamer and is blocked by origin-specific

double-stranded DNA (Klemm et al. 1997; Randell et al. 2006). Thus the ATPase activity of ORC is required for DNA replication. Many ORC subunits, including Orc1 have a predicted winged-helix (WH) domain that may contribute to DNA binding (Clarey et al. 2006; Speck et al. 2005), just like the DnaA protein uses both its AAA+ and its Helix-turn-helix domains for DNA sequence specific DNA binding (Kawakami and Katayama 2010).

ScORC is subject to cyclin-dependent kinase (CDK) activity regulation and is a substrate for the cell cycle regulatory kinase (Nguyen et al. 2001; Weinreich et al. 2001; Wilmes et al. 2004). ScOrc2 and ScOrc6 are phosphorylated by CDK only during S and G2 phases of the cell cycle. Phosphorylation on Orc2 enhances Orc5 to bind to ATP (Makise et al. 2009). Phosphorylation of Orc6 by CDK prevents it from interacting with Cdt1, thus regulating the Mcm2–7 helicase loading (Chen and Bell 2011). CDK phosphorylation sites in ORC can be altered without much phenotypic consequence, unless additional sites in Cdc6 and Mcm2–7 are simultaneously altered, leading to over-replication of DNA from certain origins of DNA replication in the genome (Nguyen et al. 2001).

The Orc1 subunit, in addition to having a primary amino acid sequence related to Sir3, is also highly related to Cdc6. In fact many archaea species only have a single Orc1/Cdc6 protein that binds to the origin in a DNA sequence-dependent manner (see Chap. 4, this volume) and has amino acid sequence similarity to both Orc1 and Cdc6 (Capaldi and Berger 2004; Duncker et al. 2009; Gaudier et al. 2007; Tada et al. 2008). In *S. cerevisiae*, the Cdc6 protein is an AAA+ ATPase that is required for Mcm2–7 loading onto the pre-RC, but then Cdc6 is degraded by ubiquitin-mediated proteolysis at the G1 to S phase transition, dependent on the activation of the S-phase cyclin-CDKs (Clb5-Cdc28 and Clb6-Cdc28) (Cocker et al. 1996; Liang et al. 1995; Santocanale and Diffley 1996; Duncker et al. 1999; Perkins et al. 2001; Drury et al. 1997, 2000; Piatti et al. 1996). The destruction of Cdc6, while leaving ORC intact, is part of the mechanism that ensures that the pre-RC cannot be reassembled in S and G2 phases, thereby limiting the initiation of DNA replication to once per cell division.

Recent electron microscopic (EM) studies of the intact ScORC revealed a bipartite structure about 120 Å wide and 160 Å long (Fig. 3.2) (Speck et al. 2005). Assuming that DNA binds along the length of the protein complex, the length of the ScORC structure is sufficient to interact with the observed 48 bp DNA fragment in the presence of ATP, as observed in the DNase I footprint assay (Bell and Stillman 1992; Speck and Stillman 2007). Systematic subunit mapping with a maltose binding protein fused at the N-terminus or the C-terminus of the individual ORC subunits showed that Orc1, 4, and 5 are located to one domain of the bi-globular structure, whereas Orc2 and Orc3 are located at the other end (Chen et al. 2008). Therefore, the most likely arrangement of the first five subunits is Orc1-Orc4-Orc5-Orc3-Orc2, with Orc1 and Orc2 at the two extremes of the pseudo-ringed structure. Such an architecture is consistent with *in vivo* and *in vitro* subunit interaction assays revealing direct contact between Orc2 and Orc3, and between Orc4 and Orc5, between Orc1 and Orc4, and between Orc4–5 and Orc2–3 (Chen et al. 2008; Matsuda et al. 2007). The physical position of Orc6 was mapped by comparing the EM structures of the

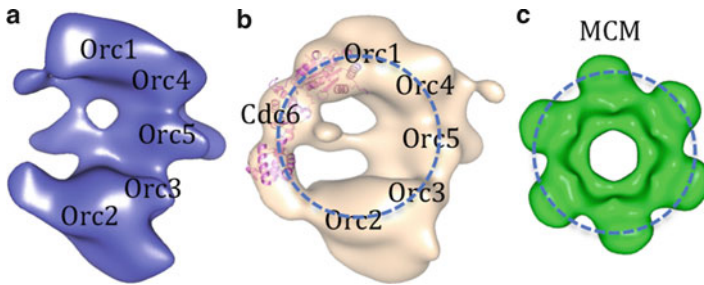


Fig. 3.2 EM structures of (a) ScORC, (b) ORC-Cdc6 and (c) archaeal Mcm. (a, b) ScORC structure. The approximate locations of the ORC subunits (a) are marked based on the observed MBP location on ORC complex with MBP fused to N- or C-terminus of subunits, one subunit and one terminus at a time (Chen et al. 2008). Cdc6 binding to the left side of ORC, forming a ring-like feature as illustrated by a *dashed blue circle* (b). The crystal structure of an archaeal Cdc6 ortholog (PDB ID 1FNN) is docked into the EM density assigned to ScCdc6, and shown by *pink ribbons*. (c) The low-pass filtered crystal structure of the hexameric N-terminal domain of an archaeal Mcm (PDB ID 1LTL) (Figure modified from Speck et al. (2005))

intact ORC and Orc1–5 sub-complex missing Orc6 (Chen et al. 2008). In the 3D difference density map, two density peaks were identified: a larger one at the lower Orc2–3 lobe and a smaller one near the upper Orc1–4–5 lobe. This observation suggests that Orc6 binds mainly with Orc2–3 at the lower lobe but reaches up to Orc1–4–5 lobe. A weak interaction between Orc6 and Orc5 was indeed found in the yeast two-hybrid analysis (Matsuda et al. 2007).

S. cerevisiae Cdc6 is another important replication initiation factor beyond ORC. It is predicted to contain an AAA+ domain, including the DNA-binding initiator-specific motif (ISM), and the DNA-binding winged helix domain, features well-conserved among known replication initiators (Dueber et al. 2007, 2011; Gaudier et al. 2007). In the absence of Cdc6, ORC is merely an origin binder and cannot load the Mcm2–7 complex. In order to establish a pre-RC at an origin, ORC has to be transformed from a passive origin binder or origin marker to the active replicative helicase loader. It appears that Cdc6 throws a molecular switch in ORC that enables such a transformation (Lee and Bell 2000; Speck et al. 2005). Cdc6 binding to ORC *in vitro* introduces an extended pre-RC-like footprint on several replication origins. Formation of the pre-RC signature DNA footprint is dependent on specific origin sequence, and the intact ATPase activity of ORC as well as that of Cdc6 (Speck et al. 2005; Speck and Stillman 2007). The ATP hydrolysis activity of ORC is inhibited upon binding to the specific origin DNAs, and the ATP hydrolysis activity of Cdc6 is suppressed upon binding to ORC that is bound on specific origin DNAs. Presumably, the ATP molecules in the ORC-Cdc6-DNA assembly are preserved for use in the subsequent recruitment and loading of Mcm2–7 helicase that are likely to be a series of energy-requiring molecular events (Evrin et al. 2009; Remus et al. 2009; Lee and Bell 2000; Randell et al. 2006). EM studies showed that Cdc6 binds to the side of the bipartite ORC, in close contact with Orc1, forming an apparently

ring-like feature in the complex (Fig. 3.2). The ring-like feature may function as the landing pad for the ring-shaped Mcm2–7 helicase core complex. Since Mcm2–7 is first loaded on the dsDNA rather than ssDNA (Evrin et al. 2009; Remus et al. 2009), it is unlikely that ScORC will melt the dsDNA. Therefore, the function of initial dsDNA melting will then have to be executed by the replicative helicase.

Following the determination of the structure of ORC and the ORC-Cdc6 structure using transmission electron microscopy, a recent study has shown a higher resolution structure of ORC-Cdc6 bound to origin DNA [*ARS1*] (Sun et al. 2012). This structure shows that upon binding Cdc6 in an ATP-dependent manner, the BAH domain of Orc1 moves considerably to the back of the ring-like complex and new density, most likely the Orc6 subunit, protrudes to the front of the complex. Orc6 has been shown to bind to two copies of the Cdt1 protein and thus the major conformational change in Orc1 may facilitate the binding of two Ctd1-Mcm2–7 heptamers to the origin DNA. Although the DNA was not visible in the cryo-EM structure of ORC-Cdc6-DNA complex, modeling of the archaeal Orc1/Cdc6 crystal structure into the individual ORC-Cdc6 subunits allows the prediction that the DNA forms a highly bent conformation within the ORC-Cdc6 complex, explaining the extended protection from nuclease in DNA footprinting experiments. The bending of the DNA is also consistent with other EM studies of ORC bound to DNA (Chastain et al. 2004).

3.3 The *S. pombe* ORC

It was thought that SpORC, like ScORC, remains bound to chromatin throughout the cell cycle (DePamphilis 2005; Kong and DePamphilis 2002). But a recent study showed that SpORC behaves much like the metazoan ORC rather than ScORC and binds replication origin periodically during the cell cycle, with the binding peaking at M to G1 transition stage (Wu and Nurse 2009). Like in the budding yeast, SpORC, SpCdc18 and the SpCdt1 are required for Mcm loading and pre-RC assembly upon exit from mitosis (Kearsey et al. 2000; Kong and DePamphilis 2002; Moon et al. 1999; Nishitani and Nurse 1997; Ogawa et al. 1999; Takahashi et al. 2003).

SpOrc1, 2, and 5 subunits are highly conserved with their counterparts from *S. cerevisiae* (Fig. 3.1) (Moon et al. 1999). SpOrc4 is unique among ORC proteins in that it has an N-terminal extension containing nine AT-hook motifs that are not found in budding yeast or metazoan Orc4 homologs (Chuang and Kelly 1999). The SpOrc4 AT-hooks specifically bind the minor groove of AT-rich DNA tracts, and are necessary and sufficient for the DNA binding activity of SpOrc4 (Chuang et al. 2002; Gaczynska et al. 2004; Lee et al. 2001). Indeed, it appears that the SpOrc4 AT-hooks are solely responsible for the DNA binding activity of the entire SpORC complex, as deletion of the Orc4 AT-hooks not only abolishes the DNA binding of Orc4, but that of the SpORC as well (Gaczynska et al. 2004). However, the AT-hook mediated initial binding of SpORC to origin DNA is salt-sensitive, and this interaction is gradually converted to a salt-stable binding state in which the topology of origin

DNA is changed into a negatively-supercoiled or under-wound state (Houchens et al. 2008). In agreement with this suggested DNA topology change, an atomic force microscopy measurement of SpORC bound to the *ars1* containing DNA fragments revealed shortening of the DNA length by 140 bp, a length sufficient for wrapping around SpORC by two turns (Gaczynska et al. 2004). SpCdc18, the homolog of ScCdc6, interacts with Cdt1 and together they further enhance the binding stability of SpORC on origin DNA, as if the SpCdc18-Cdt1 binary complex is an additional origin determinant in *S. pombe* (Kelly et al. 1993; Houchens et al. 2008). No structural characterization of the purified SpORC has been reported so far. Because of the unique AT-hooks in SpOrc4, the mechanism of origin recognition of SpORC may be different from that of other eukaryotic systems.

3.4 The *D. melanogaster* ORC

DmORC can be isolated from a *Drosophila* embryo nuclear extract as a stable complex (Gossen et al. 1995). But the presence of DmORC is cell cycle-dependent, and is regulated by the degradation of Orc1 via the ubiquitin proteasome pathway at the late M phase (Araki et al. 2003). DmOrc1 is resynthesized during late G1-phase. DmORC is also an ATPase, and like ScOrc1, DmOrc1 is essential for ATP hydrolysis and for ATP-dependent DNA binding (Chesnokov et al. 2001). In contrast to ScOrc6, which is not required for DNA binding, the DmOrc6 is required for the DNA binding of DmORC and is an integral part of the DmORC complex (Chesnokov et al. 2001). The DmOrc6 alone has DNA binding activity, likely due to the predicted TFIIB-like DNA binding domain in the smallest subunit (Liu et al. 2011). Mutations to the predicted DNA binding region abolish its DNA binding activity (Liu et al. 2011). DmOrc6 contains a C-terminal domain that is important for cytokinesis and binds to the septin protein that mediates closure of the cytokinesis furrow at the end of cell division; this feature seems to be conserved among metazoans (Chesnokov et al. 2003; Huijbregts et al. 2009; Prasanth et al. 2002).

Under EM, DmORC is an elongated structure with dimension of 170 Å by 115 Å, similar to ScORC (Clarey et al. 2006) (Fig. 3.3). The notable feature of DmORC is a spiral crescent that encompasses a 25 Å channel. Because there has been no biochemical or structural reports on the subunit arrangement, it is not clear if DmORC shares the same architecture with ScORC. But the overall dimension and basic shape of the EM reconstruction of the two protein complexes appear to be comparable, although not exactly the same. Similar features include the open-ring and the middle location of Orc5 that was mapped in ScORC by MBP-fusion approach, and in DmORC by a specific antibody (Fig. 3.3b).

DmORC binds DNA with little sequence specificity. DmORC localizes to open chromatin regions that are depleted of nucleosomes (MacAlpine et al. 2004, 2010). Interestingly, DmORC binds the negatively supercoiled DNA 30-fold better than a linear or relaxed DNA and therefore DmORC may target the topology rather than the sequence of the origin DNA (Remus et al. 2004). Because of the lack of sequence

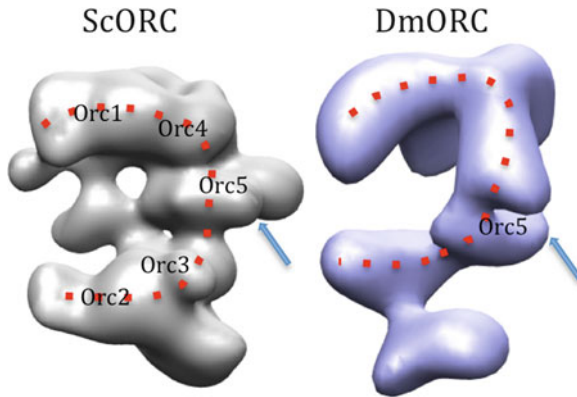


Fig. 3.3 Comparison of EM structures of ScORC (EM Data Bank ID: EMD-1156) and DmORC (EMD-4820) reveals the similar basic overall architecture, including the opened ring feature outlined by a *dashed curve*, and the middle location of Orc5 in both ORC complexes, but the difference in the structural details is also obvious

specificity, it has been unclear how DmORC binds origin DNA: which subunit does or does not directly contact DNA or what is the size of the DNA footprint of DmORC. An AFM study showed shortening of a linear DNA by ~130 bp upon DmORC binding. This length is similar to the 140 bp DNA shortening by the SpORC. Therefore, DmORC may also wrap around DNA, much like the SpORC (Clarey et al. 2006).

DmORC also is subject to CDK-mediated phosphorylation (Baldinger and Gossen 2009). DmOrc1 and DmOrc2 each contain several phosphorylation sites. Hyper-phosphorylation at these sites does not affect the integrity of DmORC, but prevents DmORC from binding to DNA (Remus et al. 2005). This is different from the yeast ORC where phosphorylation does not appear to affect their DNA binding activity. At the single molecule EM level, the hyper-phosphorylated DmORC is structurally indistinguishable from the dephosphorylated version (Fig. 3.3) (Clarey et al. 2006). Therefore, it can be concluded that phosphorylation does not introduce substantial conformational changes in DmORC. To reconcile the phosphorylation-induced interference with the DNA binding, one could imagine that extensive phosphorylation will significantly alter the physicochemical property of the surface of DmORC, even in the absence of large structural changes; this may interfere with DmORC interactions with DNA.

The location of DmORC to the entire genome has been reported using a chromatin-immunoprecipitation assay and mapping the associated DNA fragments using microarrays or deep sequencing (Calvi et al. 2007; MacAlpine et al. 2004, 2010; Spradling 1999). This association occurs as cells exit mitosis (Baldinger and Gossen 2009). DmORC is associated with regions of the genome enriched with the histone H3 variant H3.3 that is associated with transcribed regions and indeed, DmORC is associated with transcription initiation sites (MacAlpine et al. 2010). The complex is also associated with origins of DNA replication that are amplified within a single

cell division cycle during embryonic development, notably for the production of egg shell proteins in the follicle cells (Austin et al. 1999; Kim et al. 2011; Kim and Orr-Weaver 2011; Xie and Orr-Weaver 2008). Here, DmORC interactions with chromatin occur in regions enriched with the transcription factor E2F (Bosco et al. 2001; Royzman et al. 1999). The E2F transcription factor is associated with a larger complex called the Myb–MuvB (MMB)/dREAM complex that contains the Myb protein that is required for DNA replication and which is also associated with developmentally regulated gene expression (Georgette et al. 2007; Beall et al. 2002). Thus ORC may associate with specific chromatin structures that differ in different cell types, although this aspect of ORC binding has not been well investigated.

Mutations in DmORC subunits cause defects in DNA replication as expected, but cells are also observed to arrest in mitosis, although this has been attributed to DNA damage as a result of incomplete DNA replication (Chesnokov et al. 2001; Loupart et al. 2000; Pflumm and Botchan 2001). DmORC subunits, however, localize to centromeric heterochromatin and also bind the HP1 protein that is associated with heterochromatin (Badugu et al. 2005; Huang et al. 1998; Pak et al. 1997; Shareef et al. 2001, 2003).

3.5 The *H. sapiens* ORC

Human cells contain as many, if not more, origins of DNA replication than there are genes in the genome, although their usage is not well understood (Falaschi et al. 2007). Only a little over a dozen origins have been studied with any depth and even these have not been well characterized. One of the first specific origins identified was the lamin B2 origin (Abdurashidova et al. 2000). Recombinant human Orc4 alone was shown to bind the lamin B2 origin of DNA *in vitro* and in an ATP-independent manner (Stefanovic et al. 2003), perhaps reminiscent of the SpOrc4 in the origin DNA binding capacity as an individual ORC subunit, although the human protein contains no discernible AT-hooks. However it is unlikely that this biochemical interaction is functionally significant for DNA replication. Human ORC binds to the latent replication origin of Epstein–Barr virus in B cells where it is required for the maintenance of the EBV plasmid (Chaudhuri et al. 2001; Dhar et al. 2001b; Julien et al. 2004). ORC also has been reported to bind in HeLa S3 cells to intergenic AT-rich regions (Ladenburger et al. 2002) and to the *DBF4* promoter locus that is an efficient replication origin (Romero and Lee 2008). This origin of replication contains two initiation zones and two ORC binding sites that are approximately 400 bp apart. The two-zone origin appears to promote a novel mode of asymmetric bidirectional replication at the *DBF4* origin. The recombinant and purified human ORC has intrinsic DNA binding activity with preference for AT-rich sequences (Vashee et al. 2003). The purified human ORC seems to be capable of promoting the initiation of DNA replication from any DNA sequence *in vitro*, with no preference for human origin sequences, when DNA is added to ORC-depleted *Xenopus* egg extracts (Vashee et al. 2003). This specificity of initiation of DNA replication reflects the lack of DNA sequence specificity for DNA injected into activated *Xenopus* eggs

(Harland and Laskey 1980) and may say more about the nature of the egg system than origin specificity in human cells.

The six subunits of human ORC were first identified by sequence similarity to their yeast counterparts (Dhar et al. 2001a; Dhar and Dutta 2000; Gavin et al. 1995; Siddiqui and Stillman 2007; Tugal et al. 1998; Vashee et al. 2001) (see also Fig. 3.1). The assembly of ORC and the stability of the complex are both ATP-dependent. In the absence of ATP, the full complex does not assemble, and in the absence of ATP the assembled complex is so fragile that it cannot survive glycerol gradient fractionation (Ranjan and Gossen 2006). ATP can be replaced by ATP γ S for the purpose of assembly or maintenance of structural integrity, suggesting ATP is a structural cofactor for ORC assembly. The assembly of human ORC is a stepwise process *in vitro*: first, Orc2 and Orc3 form a binary complex, then the binary complex recruits Orc5. The newly formed ternary complex subsequently recruits Orc4, forming a quaternary complex. Incorporation of Orc4 into the growing complexes is ATP-dependent (Ranjan and Gossen 2006; Siddiqui and Stillman 2007). The Orc2–5 quaternary complex in turn recruits Orc1. ATP binding by Orc1 is not essential for this step, but ATP binding of Orc4 is essential for assembly of both Orc1–5 and Orc2–5 (Siddiqui and Stillman 2007). It is possible that Orc4 is physically located between Orc1 and Orc5, and ATP binding converts Orc4 into an assembly competent configuration with which Orc1 can interact from one side, and Orc5 interacts from the other. Mutations in the ATP binding sites of Orc4 and Orc5 impair complex assembly. Thus human ORC is unique in that ATP is not only required for its function in replication initiation, but also in the assembly and stability of the complex. Although the ATP binding motifs in Orc1, Orc4, and Orc5 subunits are important for replication activity (Giordano-Coltart et al. 2005), the ATPase activity of human ORC is largely contained in Orc1.

The HsORC is a very dynamic complex *in vivo* and indeed the period of the cell division cycle during which ORC exists as a complete complex may be temporally restricted to G1 phase (Kreitz et al. 2001; Mendez et al. 2002; Tatsumi et al. 2003). The HsOrc1 subunit is degraded at the G1 to S phase transition in a Skp2 ubiquitin ligase-dependent manner, only to be re-appear as cells enter mitosis (Mendez et al. 2002; Tatsumi et al. 2003). The Orc2, Orc3, Orc4, Orc5 and Orc6 subunits are displaced from chromatin as cells progress through S phase, but a heterodimer of Orc2 and Orc3 remains bound to the centromere and functions during mitosis (Craig et al. 2003; Prasanth et al. 2004; Siddiqui and Stillman 2007). Thus HsORC is a very dynamic complex with respect to the cell division cycle. It has also been reported to be associated with centrosomes where it controls the cyclin E-CDK2-dependent duplication of centrioles (Hemerly et al. 2009). Orc6 and Orc1 each contain their own NLS, and are targeted to nucleus independently of Orc2–5. Such a mechanism allows for the formation of different sub-complexes for the different functions (Ghosh et al. 2011). Additional proteins other than ORC subunits, such as the WD-repeat protein ORCA (LRWD1) may enhance human ORC binding to origins and facilitate pre-RC assembly (Bartke et al. 2010; Chakraborty et al. 2011; Shen et al. 2010; Vermeulen et al. 2010). The LRWD1 protein is interesting as it is associated with repressive marks on histone H3, such as H3K9 and H3K27 methylation,

suggesting that the ORC may recognize specialized chromatin structures via histone modifications.

Overexpressing all six subunits yields only Orc1–5 sub-complex, with Orc6 only loosely associated with the complex. Orc6 joins Orc1–5 only at the G1/M phase, and dissociates from Orc1–5 in the S phase. Orc6 contains a middle domain with a structure similar to the helical domain of the TFIIB transcription factor (Liu et al. 2011). Interestingly, this middle domain of Orc6 binds to dsDNA. Orc6 has also been shown to interact directly with Orc3 (Siddiqui and Stillman 2007); this is different to the situation in *S. cerevisiae* where Orc6 binds Orc2 (Sun et al. 2012). Other than that, the subunit interaction pattern of the human ORC appears to be similar to that of the ScORC, suggesting a conserved ORC architecture across evolution.

A prominent feature of human ORC is the transient association of both Orc1 and Orc6 with the Orc2–5 quaternary core complex (Dhar et al. 2001a; Siddiqui and Stillman 2007; Vashee et al. 2001, 2003). In addition to sites that correspond to origins of DNA replication, ORC also binds to heterochromatin and interacts with the heterochromatin protein HP1 (Chakraborty et al. 2011; Duncker et al. 2009; Lidonnici et al. 2004; Prasanth et al. 2010; Wallace and Orr-Weaver 2005). Human Orc1–5, together with HP1, localizes to heterochromatin, and may be involved in organizing higher order chromatin structure (Prasanth et al. 2010). Interestingly, the association of the Orc2, Orc3, Orc4 and Orc5 subunits at sites of heterochromatin is very dynamic, with a half-life of association of approximately 4–5 s *in vivo*. This dynamic association of ORC subunits is similar to the dynamic association of HP1 to heterochromatin. In stark contrast, the Orc1 subunit, once bound, is stably bound to heterochromatin, suggesting that this subunit has a distinct interaction with either DNA or HP1. Both the Orc1 and Orc3 subunits bind directly to HP1, perhaps explaining the different chromatin binding kinetics, but Orc1 must interact with other components of the chromatin other than HP1.

3.6 Future Perspectives

Crystal structures of archaeal ORC proteins in complex with DNA have provided tantalizing clues about how the eukaryotic ORC may interact with DNA (Berquist and DasSarma 2003; Dueber et al. 2007; Duncker et al. 2009; Gaudier et al. 2007; Grainge et al. 2003; Wigley 2009) (see Chap. 4, this volume). But clearly, the crystal structure of ORC is a critical missing piece of information that, when available, will greatly advance the field. Many eukaryotic ORCs have been expressed in various heterologous systems and then purified. It is hoped that the next several years may see a series of crystal structures of the eukaryotic replication initiators, likely in the form of individual ORC subunits or as the stable sub-complexes. However, given the large size, the dynamic nature, and interaction with extended stretch of DNA sequences, the structure of the entire ORC, and its further assembly with other replication factors, such as the replicative helicases, could be exceedingly difficult

for such crystallographic approaches. We anticipate that single particle cryo-EM will continue to play an important role in elucidating the molecular mechanism of eukaryotic replication initiation. Indeed, the anticipated crystal structures of the individual subunits or sub-complexes will facilitate the interpretation of cryo-EM studies of the various replication initiation complexes, including ORC, Cdc6, Cdt1 bound to Mcm2–7 and the Mcm2–7 double hexamer. And without a doubt, EM and crystallographic observations will raise deeper mechanistic questions that will prompt more specific biochemical experiments.

Acknowledgements Research in the authors' labs is supported by grants from the National Institutes of Health, particularly CA13106 and GM45436 and GM74985.

References

- Abdurashidova G, Deganuto M, Klima R, Riva S, Biamonti G, Giacca M, Falaschi A (2000) Start sites of bidirectional DNA synthesis at the human lamin B2 origin. *Science* 287:2023–2026
- Abdurashidova G, Danailov MB, Ochem A, Triolo G, Djeliova V, Radulescu S, Vindigni A, Riva S, Falaschi A (2003) Localization of proteins bound to a replication origin of human DNA along the cell cycle. *EMBO J* 22:4294–4303
- Aggarwal BD, Calvi BR (2004) Chromatin regulates origin activity in *Drosophila* follicle cells. *Nature* 430:372–376
- Araki M, Wharton RP, Tang Z, Yu H, Asano M (2003) Degradation of origin recognition complex large subunit by the anaphase-promoting complex in *Drosophila*. *EMBO J* 22:6115–6126
- Armache KJ, Garlick JD, Canzio D, Narlikar GJ, Kingston RE (2011) Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* 334:977–982
- Asano T, Makise C, Takehara M, Mizushima T (2007) Interaction between ORC and Cdt1p of *Saccharomyces cerevisiae*. *FEMS Yeast Res* 7:1256–1262
- Austin RJ, Orr-Weaver TL, Bell SP (1999) *Drosophila* ORC specifically binds to ACE3, an origin of DNA replication control element. *Genes Dev* 13:2639–2649
- Badugu R, Yoo Y, Singh PB, Kellum R (2005) Mutations in the heterochromatin protein 1 (HP1) hinge domain affect HP1 protein interactions and chromosomal distribution. *Chromosoma* 113:370–384
- Baldinger T, Gossen M (2009) Binding of *Drosophila* ORC proteins to anaphase chromosomes requires cessation of mitotic cyclin-dependent kinase activity. *Mol Cell Biol* 29:140–149
- Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T (2010) Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* 143:470–484
- Beall EL, Manak JR, Zhou S, Bell M, Lipsick JS, Botchan MR (2002) Role for a *Drosophila* Myb-containing protein complex in site-specific DNA replication. *Nature* 420:833–837
- Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem* 71:333–374
- Bell SP, Stillman B (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* 357:128–134
- Bell SP, Kobayashi R, Stillman B (1993) Yeast origin recognition complex functions in transcription silencing and DNA replication. *Science* 262:1844–1849
- Bell SP, Mitchell J, Leber J, Kobayashi R, Stillman B (1995) The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. *Cell* 83:563–568
- Berquist BR, DasSarma S (2003) An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, *Halobacterium* sp. strain NRC-1. *J Bacteriol* 185: 5959–5966

- Bielinsky AK, Gerbi SA (1998) Discrete start sites for DNA synthesis in the yeast ARS1 origin. *Science* 279:95–98
- Bielinsky AK, Gerbi SA (2001) Where it all starts: eukaryotic origins of DNA replication. *J Cell Sci* 114:643–651
- Bielinsky AK, Blitzblau H, Beall EL, Ezrokhi M, Smith HS, Botchan MR, Gerbi SA (2001) Origin recognition complex binding to a metazoan replication origin. *Curr Biol* 11:1427–1431
- Bosco G, Du W, Orr-Weaver TL (2001) DNA replication control through interaction of E2F-RB and the origin recognition complex. *Nat Cell Biol* 3:289–295
- Bose ME, McConnell KH, Gardner-Aukema KA, Muller U, Weinreich M, Keck JL, Fox CA (2004) The origin recognition complex and Sir4 protein recruit Sir1p to yeast silent chromatin through independent interactions requiring a common Sir1p domain. *Mol Cell Biol* 24:774–786
- Bowers JL, Randell JC, Chen S, Bell SP (2004) ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. *Mol Cell* 16:967–978
- Brewer BJ, Fangman WL (1987) The localization of replication origins on ARS plasmids in *S. cerevisiae*. *Cell* 51:463–471
- Calvi BR, Byrnes BA, Kolpakas AJ (2007) Conservation of epigenetic regulation, ORC binding and developmental timing of DNA replication origins in the genus *Drosophila*. *Genetics* 177:1291–1301
- Cao XQ, Zeng J, Yan H (2008) Structural properties of replication origins in yeast DNA sequences. *Phys Biol* 5:036012
- Capaldi SA, Berger JM (2004) Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon *Methanothermobacter thermoautotrophicus*. *Nucleic Acids Res* 32:4821–4832
- Chakraborty A, Shen Z, Prasanth SG (2011) “ORCanization” on heterochromatin: linking DNA replication initiation to chromatin organization. *Epigenetics* 6:665–670
- Chang F, May CD, Hoggard T, Miller J, Fox CA, Weinreich M (2011) High-resolution analysis of four efficient yeast replication origins reveals new insights into the ORC and putative MCM binding elements. *Nucleic Acids Res* 39:6523–6535
- Chastain PD 2nd, Bowers JL, Lee DG, Bell SP, Griffith JD (2004) Mapping subunit location on the *Saccharomyces cerevisiae* origin recognition complex free and bound to DNA using a novel nanoscale biopointer. *J Biol Chem* 279:36354–36362
- Chaudhuri B, Xu H, Todorov I, Dutta A, Yates JL (2001) Human DNA replication initiation factors, ORC and MCM, associate with oriP of Epstein-Barr virus. *Proc Natl Acad Sci USA* 98:10085–10089
- Chen S, Bell SP (2011) CDK prevents Mcm2-7 helicase loading by inhibiting Cdt1 interaction with Orc6. *Genes Dev* 25:363–372
- Chen S, de Vries MA, Bell SP (2007) Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading. *Genes Dev* 21:2897–2907
- Chen Z, Speck C, Wendel P, Tang C, Stillman B, Li H (2008) The architecture of the DNA replication origin recognition complex in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 105:10326–10331
- Chesnokov I, Gossen M, Remus D, Botchan M (1999) Assembly of functionally active *Drosophila* origin recognition complex from recombinant proteins. *Genes Dev* 13:1289–1296
- Chesnokov I, Remus D, Botchan M (2001) Functional analysis of mutant and wild-type *Drosophila* origin recognition complex. *Proc Natl Acad Sci USA* 98:11997–12002
- Chesnokov IN, Chesnokova ON, Botchan M (2003) A cytokinetic function of *Drosophila* ORC6 protein resides in a domain distinct from its replication activity. *Proc Natl Acad Sci USA* 100:9150–9155
- Chuang RY, Kelly TJ (1999) The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc Natl Acad Sci USA* 96:2656–2661
- Chuang RY, Chretien L, Dai J, Kelly TJ (2002) Purification and characterization of the *Schizosaccharomyces pombe* origin recognition complex: interaction with origin DNA and Cdc18 protein. *J Biol Chem* 277:16920–16927

- Clarey MG, Erzberger JP, Grob P, Leschziner AE, Berger JM, Nogales E, Botchan M (2006) Nucleotide-dependent conformational changes in the DnaA-like core of the origin recognition complex. *Nat Struct Mol Biol* 13:684–690
- Cocker JH, Piatti S, Santocanale C, Nasmyth K, Diffley JF (1996) An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. *Nature* 379:180–182
- Craig JM, Earle E, Canham P, Wong LH, Anderson M, Choo KH (2003) Analysis of mammalian proteins involved in chromatin modification reveals new metaphase centromeric proteins and distinct chromosomal distribution patterns. *Hum Mol Genet* 12:3109–3121
- Cvetič C, Walter JC (2005) Eukaryotic origins of DNA replication: could you please be more specific? *Semin Cell Dev Biol* 16:343–353
- DePamphilis ML (2003) The ‘ORC cycle’: a novel pathway for regulating eukaryotic DNA replication. *Gene* 310:1–15
- DePamphilis ML (2005) Cell cycle dependent regulation of the origin recognition complex. *Cell Cycle* 4:70–79
- DePamphilis ML, Blow JJ, Ghosh S, Saha T, Noguchi K, Vassilev A (2006) Regulating the licensing of DNA replication origins in metazoa. *Curr Opin Cell Biol* 18:231–239
- Deshpande AM, Newlon CS (1992) The ARS consensus sequence is required for chromosomal origin function in *Saccharomyces cerevisiae*. *Mol Cell Biol* 12:4305–4313
- Dhar SK, Dutta A (2000) Identification and characterization of the human ORC6 homolog. *J Biol Chem* 275:34983–34988
- Dhar SK, Delmolino L, Dutta A (2001a) Architecture of the human origin recognition complex. *J Biol Chem* 276:29067–29071
- Dhar SK, Yoshida K, Machida Y, Khaira P, Chaudhuri B, Wohlschlegel JA, Leffak M, Yates J, Dutta A (2001b) Replication from *oriP* of Epstein-Barr virus requires human ORC and is inhibited by geminin. *Cell* 106:287–296
- Diffley JF, Labib K (2002) The chromosome replication cycle. *J Cell Sci* 115:869–872
- Drury LS, Perkins G, Diffley JF (1997) The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. *EMBO J* 16:5966–5976
- Drury LS, Perkins G, Diffley JF (2000) The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. *Curr Biol* 10:231–240
- Dueber EL, Corn JE, Bell SD, Berger JM (2007) Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science* 317:1210–1213
- Dueber EC, Costa A, Corn JE, Bell SD, Berger JM (2011) Molecular determinants of origin discrimination by Orc1 initiators in archaea. *Nucleic Acids Res* 39:3621–3631
- Duncker BP, Pasero P, Braguglia D, Heun P, Weinreich M, Gasser SM (1999) Cyclin B-cdk1 kinase stimulates ORC- and Cdc6-independent steps of semiconservative plasmid replication in yeast nuclear extracts. *Mol Cell Biol* 19:1226–1241
- Duncker BP, Chesnokov IN, McConkey BJ (2009) The origin recognition complex protein family. *Genome Biol* 10:214
- Eaton ML, Galani K, Kang S, Bell SP, MacAlpine DM (2010) Conserved nucleosome positioning defines replication origins. *Genes Dev* 24:748–753
- Eaton ML, Prinz JA, MacAlpine HK, Tretyakov G, Kharchenko PV, MacAlpine DM (2011) Chromatin signatures of the *Drosophila* replication program. *Genome Res* 21:164–174
- Ehrentauf S, Hassler M, Oppikofer M, Kueng S, Weber JM, Mueller JW, Gasser SM, Ladurner AG, Ehrenhofer-Murray AE (2011) Structural basis for the role of the Sir3 AAA+ domain in silencing: interaction with Sir4 and unmethylated histone H3K79. *Genes Dev* 25:1835–1846
- 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 USA* 106:20240–20245
- Falaschi A, Abdurashidova G, Sandoval O, Radulescu S, Biamonti G, Riva S (2007) Molecular and structural transactions at human DNA replication origins. *Cell Cycle* 6:1705–1712
- Fox CA, Ehrenhofer-Murray AE, Loo S, Rine J (1997) The origin recognition complex, SIR1, and the S phase requirement for silencing. *Science* 276:1547–1551

- Gaczynska M, Osmulski PA, Jiang Y, Lee JK, Bermudez V, Hurwitz J (2004) Atomic force microscopic analysis of the binding of the *Schizosaccharomyces pombe* origin recognition complex and the spOrc4 protein with origin DNA. *Proc Natl Acad Sci USA* 101:17952–17957
- Gardner KA, Rine J, Fox CA (1999) A region of the Sir1 protein dedicated to recognition of a silencer and required for interaction with the Orc1 protein in *Saccharomyces cerevisiae*. *Genetics* 151:31–44
- Gaudier M, Schuwirth BS, Westcott SL, Wigley DB (2007) Structural basis of DNA replication origin recognition by an ORC protein. *Science* 317:1213–1216
- Gavin KA, Hidaka M, Stillman B (1995) Conserved initiator proteins in eukaryotes. *Science* 270:1667–1671
- Georgette D, Ahn S, MacAlpine DM, Cheung E, Lewis PW, Beall EL, Bell SP, Speed T, Manak JR, Botchan MR (2007) Genomic profiling and expression studies reveal both positive and negative activities for the *Drosophila* Myb MuvB/dREAM complex in proliferating cells. *Genes Dev* 21:2880–2896
- Ghosh S, Vassilev AP, Zhang J, Zhao Y, DePamphilis ML (2011) Assembly of the human origin recognition complex occurs through independent nuclear localization of its components. *J Biol Chem* 286:23831–23841
- Gibson DG, Bell SP, Aparicio OM (2006) Cell cycle execution point analysis of ORC function and characterization of the checkpoint response to ORC inactivation in *Saccharomyces cerevisiae*. *Genes Cells* 11:557–573
- Gilbert DM (1998) Replication origins in yeast versus metazoa: separation of the haves and the have nots. *Curr Opin Genet Dev* 8:194–199
- Gilbert DM (2010) Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat Rev Genet* 11:673–684
- Giordano-Coltart J, Ying CY, Gautier J, Hurwitz J (2005) Studies of the properties of human origin recognition complex and its Walker A motif mutants. *Proc Natl Acad Sci USA* 102:69–74
- Gossen M, Pak DT, Hansen SK, Acharya JK, Botchan MR (1995) A *Drosophila* homolog of the yeast origin recognition complex. *Science* 270:1674–1677
- Grainge I, Scaife S, Wigley DB (2003) Biochemical analysis of components of the pre-replication complex of *Archaeoglobus fulgidus*. *Nucleic Acids Res* 31:4888–4898
- Harland RM, Laskey RA (1980) Regulated replication of DNA microinjected into eggs of *Xenopus laevis*. *Cell* 21:761–771
- Hartl T, Boswell C, Orr-Weaver TL, Bosco G (2007) Developmentally regulated histone modifications in *Drosophila* follicle cells: initiation of gene amplification is associated with histone H3 and H4 hyperacetylation and H1 phosphorylation. *Chromosoma* 116:197–214
- Hemerly AS, Prasanth SG, Siddiqui K, Stillman B (2009) Orc1 controls centriole and centrosome copy number in human cells. *Science* 323:789–793
- Hickman MA, Rusche LN (2010) Transcriptional silencing functions of the yeast protein Orc1/Sir3 subfunctionalized after gene duplication. *Proc Natl Acad Sci USA* 107:19384–19389
- Hou Z, Bernstein DA, Fox CA, Keck JL (2005) Structural basis of the Sir1-origin recognition complex interaction in transcriptional silencing. *Proc Natl Acad Sci USA* 102:8489–8494
- Houchens CR, Lu W, Chuang RY, Frattini MG, Fuller A, Simanek P, Kelly TJ (2008) Multiple mechanisms contribute to *Schizosaccharomyces pombe* origin recognition complex-DNA interactions. *J Biol Chem* 283:30216–30224
- Huang RY, Kowalski D (1993) A DNA unwinding element and an ARS consensus comprise a replication origin within a yeast chromosome. *EMBO J* 12:4521–4531
- Huang DW, Fanti L, Pak DT, Botchan MR, Pimpinelli S, Kellum R (1998) Distinct cytoplasmic and nuclear fractions of *Drosophila* heterochromatin protein 1: their phosphorylation levels and associations with origin recognition complex proteins. *J Cell Biol* 142:307–318
- Huijbrechts RP, Svitin A, Stinnett MW, Renfrow MB, Chesnokov I (2009) *Drosophila* Orc6 facilitates GTPase activity and filament formation of the septin complex. *Mol Biol Cell* 20:270–281
- Julien MD, Polonskaya Z, Hearing J (2004) Protein and sequence requirements for the recruitment of the human origin recognition complex to the latent cycle origin of DNA replication of Epstein-Barr virus *oriP*. *Virology* 326:317–328

- Karnani N, Taylor CM, Malhotra A, Dutta A (2010) Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol Biol Cell* 21:393–404
- Kawakami H, Katayama T (2010) DnaA, ORC, and Cdc6: similarity beyond the domains of life and diversity. *Biochem Cell Biol* 88:49–62
- Kearsey SE, Montgomery S, Labib K, Lindner K (2000) Chromatin binding of the fission yeast replication factor Mcm4 occurs during anaphase and requires ORC and cdc18. *EMBO J* 19:1681–1690
- Kelly TJ, Martin GS, Forsburg SL, Stephen RJ, Russo A, Nurse P (1993) The fission yeast *cdc18⁺* gene product couples S phase to START and mitosis. *Cell* 74:371–382
- Kim JC, Orr-Weaver TL (2011) Analysis of a *Drosophila* amplicon in follicle cells highlights the diversity of metazoan replication origins. *Proc Natl Acad Sci USA* 108:16681–16686
- Kim JC, Nordman J, Xie F, Kashevsky H, Eng T, Li S, MacAlpine DM, Orr-Weaver TL (2011) Integrative analysis of gene amplification in *Drosophila* follicle cells: parameters of origin activation and repression. *Genes Dev* 25:1384–1398
- Klemm RD, Bell SP (2001) ATP bound to the origin recognition complex is important for preRC formation. *Proc Natl Acad Sci USA* 98:8361–8367
- Klemm RD, Austin RJ, Bell SP (1997) Coordinate binding of ATP and origin DNA regulates the ATPase activity of the origin recognition complex. *Cell* 88:493–502
- Kong D, DePamphilis ML (2002) Site-specific ORC binding, pre-replication complex assembly and DNA synthesis at *Schizosaccharomyces pombe* replication origins. *EMBO J* 21:5567–5576
- Kreitz S, Ritz M, Baack M, Knippers R (2001) The human origin recognition complex protein 1 dissociates from chromatin during S phase in HeLa cells. *J Biol Chem* 276:6337–6342
- Ladenburger EM, Keller C, Knippers R (2002) Identification of a binding region for human origin recognition complex proteins 1 and 2 that coincides with an origin of DNA replication. *Mol Cell Biol* 22:1036–1048
- Lee DG, Bell SP (1997) Architecture of the yeast origin recognition complex bound to origins of DNA replication. *Mol Cell Biol* 17:7159–7168
- Lee DG, Bell SP (2000) ATPase switches controlling DNA replication initiation. *Curr Opin Cell Biol* 12:280–285
- Lee DG, Makhov AM, Klemm RD, Griffith JD, Bell SP (2000) Regulation of origin recognition complex conformation and ATPase activity: differential effects of single-stranded and double-stranded DNA binding. *EMBO J* 19:4774–4782
- Lee JK, Moon KY, Jiang Y, Hurwitz J (2001) The *Schizosaccharomyces pombe* origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. *Proc Natl Acad Sci USA* 98:13589–13594
- Liang C, Weinreich M, Stillman B (1995) ORC and Cdc6p interact and determine the frequency of initiation of DNA replication in the genome. *Cell* 81:667–676
- Lidonnici MR, Rossi R, Paixao S, Mendoza-Maldonado R, Paolinelli R, Arcangeli C, Giacca M, Biamonti G, Montecucco A (2004) Subnuclear distribution of the largest subunit of the human origin recognition complex during the cell cycle. *J Cell Sci* 117:5221–5231
- Lipford JR, Bell SP (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 IN, Liu Y (2011) Structural analysis of human Orc6 protein reveals a homology with transcription factor TFIIB. *Proc Natl Acad Sci USA* 108:7373–7378
- Liu J, McConnell K, Dixon M, Calvi BR (2012) Analysis of model replication origins in *Drosophila* reveals new aspects of the chromatin landscape and its relationship to origin activity and the prereplicative complex. *Mol Biol Cell* 23:200–212
- Loupart ML, Krause SA, Heck MS (2000) Aberrant replication timing induces defective chromosome condensation in *Drosophila* ORC2 mutants. *Curr Biol* 10:1547–1556
- MacAlpine DM, Rodriguez HK, Bell SP (2004) Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev* 18:3094–3105

- MacAlpine HK, Gordan R, Powell SK, Hartemink AJ, MacAlpine DM (2010) *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res* 20:201–211
- Makise M, Takehara M, Kuniyasu A, Matsui N, Nakayama H, Mizushima T (2009) Linkage between phosphorylation of the origin recognition complex and its ATP binding activity in *Saccharomyces cerevisiae*. *J Biol Chem* 284:3396–3407
- Marahrens Y, Stillman B (1992) A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* 255:817–823
- Matsuda K, Makise M, Sueyasu Y, Takehara M, Asano T, Mizushima T (2007) Yeast two-hybrid analysis of the origin recognition complex of *Saccharomyces cerevisiae*: interaction between subunits and identification of binding proteins. *FEMS Yeast Res* 7:1263–1269
- Mendez J, Stillman B (2003) Perpetuating the double helix: molecular machines at eukaryotic DNA replication origins. *Bioessays* 25:1158–1167
- Mendez J, Zou-Yang XH, Kim SY, Hidaka M, Tansey WP, Stillman B (2002) Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. *Mol Cell* 9:481–491
- Moon KY, Kong D, Lee JK, Raychaudhuri S, Hurwitz J (1999) Identification and reconstitution of the origin recognition complex from *Schizosaccharomyces pombe*. *Proc Natl Acad Sci USA* 96:12367–12372
- Muller P, Park S, Shor E, Huebert DJ, Warren CL, Ansari AZ, Weinreich M, Eaton ML, MacAlpine DM, Fox CA (2010) The conserved bromo-adjacent homology domain of yeast Orc1 functions in the selection of DNA replication origins within chromatin. *Genes Dev* 24:1418–1433
- Natale DA, Umek RM, Kowalski D (1993) Ease of DNA unwinding is a conserved property of yeast replication origins. *Nucleic Acids Res* 21:555–560
- Nguyen VQ, Co C, Li JJ (2001) Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* 411:1068–1073
- Nishitani H, Nurse P (1997) The cdc18 protein initiates DNA replication in fission yeast. *Prog Cell Cycle Res* 3:135–142
- Noguchi K, Vassilev A, Ghosh S, Yates JL, DePamphilis ML (2006) The BAH domain facilitates the ability of human Orc1 protein to activate replication origins *in vivo*. *EMBO J* 25:5372–5382
- Norseen J, Thomae A, Sridharan V, Aiyar A, Schepers A, Lieberman PM (2008) RNA-dependent recruitment of the origin recognition complex. *EMBO J* 27:3024–3035
- Ogawa Y, Takahashi T, Masukata H (1999) Association of fission yeast Orp1 and Mcm6 proteins with chromosomal replication origins. *Mol Cell Biol* 19:7228–7236
- Ozaydin B, Rine J (2010) Expanded roles of the origin recognition complex in the architecture and function of silenced chromatin in *Saccharomyces cerevisiae*. *Mol Cell Biol* 30:626–639
- Pak DT, Pflumm M, Chesnokov I, Huang DW, Kellum R, Marr J, Romanowski P, Botchan MR (1997) Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. *Cell* 91:311–323
- Perkins G, Drury LS, Diffley JF (2001) Separate SCF(CDC4) recognition elements target Cdc6 for proteolysis in S phase and mitosis. *EMBO J* 20:4836–4845
- Pflumm MF, Botchan MR (2001) Orc mutants arrest in metaphase with abnormally condensed chromosomes. *Development* 128:1697–1707
- Piatti S, Bohm T, Cocker JH, Diffley JF, Nasmyth K (1996) Activation of S-phase-promoting CDKs in late G1 defines a “point of no return” after which Cdc6 synthesis cannot promote DNA replication in yeast. *Genes Dev* 10:1516–1531
- Pillus L, Rine J (1989) Epigenetic inheritance of transcriptional states in *S. cerevisiae*. *Cell* 59:637–647
- Pillus L, Rine J (2004) SIR1 and the origin of epigenetic states in *Saccharomyces cerevisiae*. *Cold Spring Harb Symp Quant Biol* 69:259–265
- Prasanth SG, Prasanth KV, Stillman B (2002) Orc6 involved in DNA replication, chromosome segregation, and cytokinesis. *Science* 297:1026–1031

- Prasanth SG, Prasanth KV, Siddiqui K, Spector DL, Stillman B (2004) Human Orc2 localizes to centrosomes, centromeres and heterochromatin during chromosome inheritance. *EMBO J* 23:2651–2663
- Prasanth SG, Shen Z, Prasanth KV, Stillman B (2010) Human origin recognition complex is essential for HP1 binding to chromatin and heterochromatin organization. *Proc Natl Acad Sci USA* 107:15093–15098
- Randell JC, Bowers JL, Rodriguez HK, Bell SP (2006) Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol Cell* 21:29–39
- Ranjan A, Gossen M (2006) A structural role for ATP in the formation and stability of the human origin recognition complex. *Proc Natl Acad Sci USA* 103:4864–4869
- Rao H, Stillman B (1995) The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. *Proc Natl Acad Sci USA* 92:2224–2228
- Rao H, Marahrens Y, Stillman B (1994) Functional conservation of multiple elements in yeast chromosomal replicators. *Mol Cell Biol* 14:7643–7651
- Remus D, Beall EL, Botchan MR (2004) DNA topology, not DNA sequence, is a critical determinant for *Drosophila* ORC-DNA binding. *EMBO J* 23:897–907
- Remus D, Blanchette M, Rio DC, Botchan MR (2005) CDK phosphorylation inhibits the DNA-binding and ATP-hydrolysis activities of the *Drosophila* origin recognition complex. *J Biol Chem* 280:39740–39751
- Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JF (2009) Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139:719–730
- Romero J, Lee H (2008) Asymmetric bidirectional replication at the human DBF4 origin. *Nat Struct Mol Biol* 15:722–729
- Rowley A, Cocker JH, Harwood J, Diffley JF (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* 14:2631–2641
- Royzman I, Austin RJ, Bosco G, Bell SP, Orr-Weaver TL (1999) ORC localization in *Drosophila* follicle cells and the effects of mutations in dE2F and dDP. *Genes Dev* 13:827–840
- Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, Schulz TC, Robins AJ, Dalton S, Gilbert DM (2010) Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res* 20:761–770
- Santocanale C, Diffley JF (1996) ORC- and Cdc6-dependent complexes at active and inactive chromosomal replication origins in *Saccharomyces cerevisiae*. *EMBO J* 15:6671–6679
- Sasaki T, Gilbert DM (2007) The many faces of the origin recognition complex. *Curr Opin Cell Biol* 19:337–343
- Scholefield G, Veening JW, Murray H (2011) DnaA and ORC: more than DNA replication initiators. *Trends Cell Biol* 21:188–194
- Shareef MM, King C, Damaj M, Badagu R, Huang DW, Kellum R (2001) *Drosophila* heterochromatin protein 1 (HP1)/origin recognition complex (ORC) protein is associated with HP1 and ORC and functions in heterochromatin-induced silencing. *Mol Biol Cell* 12:1671–1685
- Shareef MM, Badagu R, Kellum R (2003) HP1/ORC complex and heterochromatin assembly. *Genetica* 117:127–134
- Shen Z, Sathyan KM, Geng Y, Zheng R, Chakraborty A, Freeman B, Wang F, Prasanth KV, Prasanth SG (2010) A WD-repeat protein stabilizes ORC binding to chromatin. *Mol Cell* 40:99–111
- Sher N, Bell GW, Li S, Nordman J, Eng T, Eaton ML, Macalpine DM, Orr-Weaver TL (2012) Developmental control of gene copy number by repression of replication initiation and fork progression. *Genome Res* 22:64–75
- Siddiqui K, Stillman B (2007) ATP-dependent assembly of the human origin recognition complex. *J Biol Chem* 282:32370–32383
- 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* 282:11705–11714

- Speck C, Chen Z, Li H, Stillman B (2005) ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nat Struct Mol Biol* 12:965–971
- Spradling AC (1999) ORC binding, gene amplification, and the nature of metazoan replication origins. *Genes Dev* 13:2619–2623
- Stefanovic D, Stanojic S, Vindigni A, Ochem A, Falaschi A (2003) *In vitro* protein-DNA interactions at the human lamin B2 replication origin. *J Biol Chem* 278:42737–42743
- 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* 20:534–544
- Tada S, Kundu LR, Enomoto T (2008) Insight into initiator-DNA interactions: a lesson from the archaeal ORC. *Bioessays* 30:208–211
- Takahashi T, Ohara E, Nishitani H, Masukata H (2003) Multiple ORC-binding sites are required for efficient MCM loading and origin firing in fission yeast. *EMBO J* 22:964–974
- Takara TJ, Bell SP (2011) Multiple Cdt1 molecules act at each origin to load replication-competent Mcm2-7 helicases. *EMBO J* 30:4885–4896
- Takehara M, Makise M, Takenaka H, Asano T, Mizushima T (2008) Analysis of mutant origin recognition complex with reduced ATPase activity *in vivo* and *in vitro*. *Biochem J* 413:535–543
- Takenaka H, Makise M, Kuwae W, Takahashi N, Tsuchiya T, Mizushima T (2004) ADP-binding to origin recognition complex of *Saccharomyces cerevisiae*. *J Mol Biol* 340:29–37
- Tatsumi Y, Ohta S, Kimura H, Tsurimoto T, Obuse C (2003) The ORC1 cycle in human cells: I. cell cycle-regulated oscillation of human ORC1. *J Biol Chem* 278:41528–41534
- Theis JF, Newlon CS (1994) Domain B of ARS307 contains two functional elements and contributes to chromosomal replication origin function. *Mol Cell Biol* 14:7652–7659
- Theis JF, Newlon CS (2001) Two compound replication origins in *Saccharomyces cerevisiae* contain redundant origin recognition complex binding sites. *Mol Cell Biol* 21:2790–2801
- Thomae AW, Pich D, Brocher J, Spindler MP, Berens C, Hock R, Hammerschmidt W, Schepers A (2008) Interaction between HMGA1a and the origin recognition complex creates site-specific replication origins. *Proc Natl Acad Sci USA* 105:1692–1697
- Triolo T, Sternglanz R (1996) Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing. *Nature* 381:251–253
- Tugal T, Zou-Yang XH, Gavin K, Pappin D, Canas B, Kobayashi R, Hunt T, Stillman B (1998) The Orc4p and Orc5p subunits of the *Xenopus* and human origin recognition complex are related to Orc1p and Cdc6p. *J Biol Chem* 273:32421–32429
- Van Houten JV, Newlon CS (1990) Mutational analysis of the consensus sequence of a replication origin from yeast chromosome III. *Mol Cell Biol* 10:3917–3925
- Vashee S, Simancek P, Challberg MD, Kelly TJ (2001) Assembly of the human origin recognition complex. *J Biol Chem* 276:26666–26673
- Vashee S, Cvetcic C, Lu W, Simancek P, Kelly TJ, Walter JC (2003) Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev* 17:1894–1908
- Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, Lee KK, Olsen JV, Hyman AA, Stunnenberg HG, Mann M (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* 142:967–980
- Wallace JA, Orr-Weaver TL (2005) Replication of heterochromatin: insights into mechanisms of epigenetic inheritance. *Chromosoma* 114:389–402
- Weber JM, Irlbacher H, Ehrenhofer-Murray AE (2008) Control of replication initiation by the Sum1/Rfm1/Hst1 histone deacetylase. *BMC Mol Biol* 9:100
- Weinreich M, Liang C, Chen HH, Stillman B (2001) Binding of cyclin-dependent kinases to ORC and Cdc6p regulates the chromosome replication cycle. *Proc Natl Acad Sci USA* 98:11211–11217
- Wigley DB (2009) ORC proteins: marking the start. *Curr Opin Struct Biol* 19:72–78
- Wilmes GM, Bell SP (2002) The B2 element of the *Saccharomyces cerevisiae* ARS1 origin of replication requires specific sequences to facilitate pre-RC formation. *Proc Natl Acad Sci USA* 99:101–106

- Wilmes GM, Archambault V, Austin RJ, Jacobson MD, Bell SP, Cross FR (2004) Interaction of the S-phase cyclin Clb5 with an “RXL” docking sequence in the initiator protein Orc6 provides an origin-localized replication control switch. *Genes Dev* 18:981–991
- Wu PY, Nurse P (2009) Establishing the program of origin firing during S phase in fission yeast. *Cell* 136:852–864
- Xie F, Orr-Weaver TL (2008) Isolation of a *Drosophila* amplification origin developmentally activated by transcription. *Proc Natl Acad Sci USA* 105:9651–9656
- Zhang Z, Hayashi MK, Merkel O, Stillman B, Xu RM (2002) Structure and function of the BAH-containing domain of Orc1p in epigenetic silencing. *EMBO J* 21:4600–4611
- Zou L, Stillman B (2000) Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbf4p kinase. *Mol Cell Biol* 20:3086–3096
- Zou Y, Yu Q, Bi X (2006) Asymmetric positioning of nucleosomes and directional establishment of transcriptionally silent chromatin by *Saccharomyces cerevisiae* silencers. *Mol Cell Biol* 26:7806–7819

Chapter 4

Archaeal Orc1/Cdc6 Proteins

Stephen D. Bell

Abstract The initiation of DNA replication in most archaeal genomes is mediated by proteins related to eukaryotic Orc1 and Cdc6. Archaeal replication origins have been mapped and their interactions with Orc1/Cdc6 proteins have been characterized at the biochemical level. Structural and biophysical studies have revealed the basic rules of sequence recognition by archaeal initiators.

Keywords DNA replication • Helicase loader • Archaea • Evolution • Initiators

4.1 Introduction

First put forward in 1963, the replicon hypothesis posits that defined sequences within genomes serve as *cis*-acting “replicator” or origin sequences and *trans*-acting “initiator” factors act upon these sites to mediate replication initiation (Jacob et al. 1963). In bacteria, the broadly-conserved DnaA protein fulfills the role of initiator. The eukaryotic counterpart of DnaA is the six-subunit origin recognition complex (ORC) composed of Orc1–Orc6. As detailed in Chap. 3 of this book, ORC interacts with origins and leads to the recruitment of the MCM replicative helicase, in a reaction that is dependent upon two additional factors, Cdc6 and Cdt1 (Bell and Dutta 2002). Interestingly, Orc1 and Cdc6 show a degree of sequence conservation, suggesting that they may have evolved from a common ancestor. When the first archaeal genome sequences became available, it was instantly apparent that archaea have a DNA replication machinery that is closely-related to that of eukarya and

S.D. Bell (✉)

Sir William Dunn School of Pathology, University of Oxford,
South Parks Road, Oxford OX1 3RE, UK
e-mail: stephen.bell@path.ox.ac.uk

clearly distinct from the analogous apparatus in bacteria. More specifically, genes encoding homolog(s) of the MCM helicase subunits, the two-subunit core DNA primase, the sliding clamp PCNA, the clamp loader RFC, the flap endonuclease Fen1 and the ATP-dependent DNA ligase I were found to be highly conserved between archaeal species (Edgell and Doolittle 1997). Intriguingly, however, the first archaeal genome to be sequenced, that of *Methanocaldococcus jannaschii*, did not reveal any clear candidates for initiator proteins. However, subsequent genomes of other archaeal species revealed one or more genes encoding proteins that were homologous to both Orc1 and Cdc6, possibly representative of the ancestral gene from which the distinct eukaryotic proteins evolved. In the following, I shall refer to the archaeal proteins generically as Orc1/Cdc6. Unfortunately, there has been no consensus policy adopted for the naming of these genes in archaeal genomes. This has resulted in a confusing and non-unified nomenclature with some projects calling orthologous proteins either Orc1 or Cdc6. To add to the confusion, some workers have named multiple Orc1/Cdc6 paralogs Orc1, Orc2, Orc3, etc., implying a non-existent relationship with the eukaryotic-specific ORC components Orc2, Orc3, etc.

Thus, with the exception of *M. jannaschii* and its relatives in the Methanococcales, archaea possess one or more Orc1/Cdc6 paralogs. To date the protein or proteins responsible for defining replication origins within the Methanococcales remain unknown.

4.2 Origins of DNA Replication in the Archaea

Four principal phyla of archaea have been identified thus far, the Crenarchaeota, Euryarchaeota, Thaumarchaeota and Korarchaeota (Brochier-Armanet et al. 2008; Elkins et al. 2008). While genes for Orc1/Cdc6 proteins are found in all four phyla, biochemical studies of these proteins have been restricted to the Crenarchaeota and Euryarchaeota and structural studies have been confined to crenarchaeal proteins. Nevertheless, as detailed below, the degree of sequence conservation both of the initiators and their DNA binding sites suggests that some general conclusions may be drawn despite the limited phylogenetic range of proteins sampled to date.

All archaea studied so far possess simple circular chromosomes that contain polycistronic transcription units and are thus reminiscent of the chromosome organization of most bacteria. This apparent parallel was strengthened with the first characterization of the replicon architecture of the chromosome of the euryarchaea from the genus *Pyrococcus*. Bioinformatic studies, in conjunction with *in vivo* DNA labeling studies, revealed that this organism, like bacteria, had a single origin of replication in its chromosome (Myllykallio et al. 2000). Furthermore, this origin was tightly linked to the gene for the single Orc1/Cdc6 homolog in *Pyrococcus*, again reminiscent of the linkage of *dnaA* genes with origins in many bacteria. However, as alluded to above, many archaea possess multiple Orc1/Cdc6 paralogs. For example, members of the genus *Sulfolobus* encode three such genes, now called *orc1-1*, *orc1-2* and *orc1-3*, in their single chromosome (She et al. 2001).

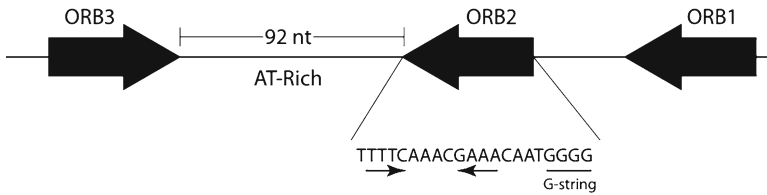


Fig. 4.1 Cartoon of the architecture of *S. solfataricus oriC1*. The three ORB element binding sites for Orc1–1 are indicated and the sequence of ORB2 is shown. The conserved dyad symmetric element (TTTC...GAAA) is indicated by *arrows* as is the polarity defining G-string element. Many archaeal origins share this arrangement where a central AT-rich region is flanked one or more ORB elements on each side

The *Sulfolobus orc1–1* gene is the clear ortholog of the single *Pyrococcus* Orc1/Cdc6. Neutral-neutral 2D agarose gel analyses of *S. solfataricus* revealed two replication origins: one, *oriC1*, adjacent to *orc1–1* as in *Pyrococcus*, and the second, *oriC2*, adjacent to *orc1–3*. No evidence for an origin within 15 kb of the *orc1–2* gene could be obtained (Robinson et al. 2004). A subsequent whole genome marker frequency analysis confirmed the existence of *oriC1* and *oriC2* and revealed a third origin in *Sulfolobus*, about 80 kb from the *orc1–2* gene (Lundgren et al. 2004). The position of *oriC3* was further mapped at high resolution by 2D gel analyses (Robinson et al. 2007). Remarkably, *oriC3* lies beside a gene encoding a divergent homolog of the eukaryal DNA replication initiation factor Cdt1. Studies using synchronized *Sulfolobus* cells have revealed that all three origins fire in every cell in every cell cycle during exponential growth. Furthermore, *oriC1* and *oriC3* fire highly synchronously while *oriC2* fires over a slightly broader temporal window (Duggin et al. 2008). How the coordinate control of origin firing is achieved is currently unknown. The existence and use of multiple replication origins per chromosome is not restricted to *Sulfolobus* species; two replication origins have been identified in another crenarchaeon, *Aeropyrum pernix*, and the main chromosome of the euryarchaeon *Haloferax volcanii* is replicated from at least two replication origins (Grainige et al. 2006; Norais et al. 2007; Robinson and Bell 2007). However, the stoichiometry of firing and timing of use of the origins in these species has yet to be evaluated.

Analysis of the sequence composition of the various replication origins reveals conservation of certain motifs between archaeal species (Fig. 4.1). The single origin of *Pyrococcus*, two origins of *Haloferax* and *oriC1* from *Sulfolobus* and *Aeropyrum* all contain conserved Origin Recognition Box (ORB) elements. These possess a dyad symmetric sequence flanked on one side by a run of three or more G bases (G-string). The G-string element therefore ascribes a polarity to the ORB element. Several origins have a common architecture where a central A-T rich region is flanked by ORB elements of inverted polarity in the two arms (Fig. 4.1) (Robinson et al. 2004).

Interestingly, *Sulfolobus oriC2* has a related mini-ORB element that lacks the G-string. Mini-ORB elements are also found at the single origin of replication in

the euryarchaeon *Methanothermobacter thermautotrophicum* (Capaldi and Berger 2004; Majernik and Chong 2008; Robinson et al. 2004). The ORB and mini-ORB elements are specific recognition sequences for *Sulfolobus* Orc1–1 and its orthologs from other species. The conservation of the binding site is sufficient to allow *Sulfolobus* Orc1–1 to bind specifically to the *Pyrococcus* origin *in vitro*, despite the phylum level divide between these organisms (Robinson et al. 2004). In addition to the Orc1–1-binding mini-ORB sites in *Sulfolobus oriC2*, this origin also possesses “C3” binding sites for Orc1–3. Orc1–3 has 35% sequence identity to Orc1–1 and its binding site contains a TTTC element that corresponds to one arm of the mini-ORB dyad. As described below, *oriC2* contains adjacent mini-ORB and C3 sites that bind Orc1–1 and Orc1–3 with a degree of positive cooperativity.

4.3 Orc1/Cdc6 Structure

Sequence analysis of the Orc1/Cdc6 proteins reveals that they possess a N-terminal AAA+ ATPase domain and a C-terminal winged-helix (wH) domain. This organisation is reminiscent of the bacterial DnaA protein that also contains a AAA+ fold followed by a DNA-binding domain, although in the case of DnaA this latter domain is a helix-turn-helix. AAA+ domains can be classified into seven distinct clades specified by characteristic embellishments on the core AAA+ fold (Erzberger and Berger 2006; Iyer et al. 2004). Importantly, DnaA, Orc1/Cdc6 and the eukaryotic Orc1 and Cdc6 all fall into the “Initiator” clade of AAA+ proteins – defined by the presence of an additional α -helix, termed the Initiator Specific Motif (ISM), that precedes the second α -helix of the core AAA+ fold. In DnaA it has been proposed that this additional α -helix serves as a steric wedge that helps drive the DnaA protein into a filamentous structure upon oligomerization (Erzberger et al. 2006).

The structures of Orc1/Cdc6 proteins from a number of archaeal species have been solved by X-ray crystallography. The first structure to be determined was that of Orc1/Cdc6 from the crenarchaeon *Pyrobaculum aerophilum* (Liu et al. 2000). This structure revealed a monomeric protein that had ADP bound in its active site. The tight ADP binding of this protein is found in many other archaeal Orc1/Cdc6s; indeed, a number of studies with recombinant Orc1/Cdc6s have found that it is necessary to employ a guanidinium hydrochloride-mediated denaturation/renaturation protocol to effect efficient exchange of ADP for ATP (Singleton et al. 2004).

This may be reflective of a switch-like regulation of the activity of the protein. Presumably, the protein when synthesized will bind to ATP, which it will then hydrolyse to ADP. If the cell regulates the timing of the synthesis of Orc1/Cdc6 during the cell cycle, a situation could be envisaged where a short window in time would be generated in which the ATP-bound form would be present. If one assumes that the ATP-bound form of the protein is the active form for initiation of replication, then a permissive period of the cell cycle would be dictated by the timing of synthesis and kinetics of ATP hydrolysis by the protein. It is also conceivable that

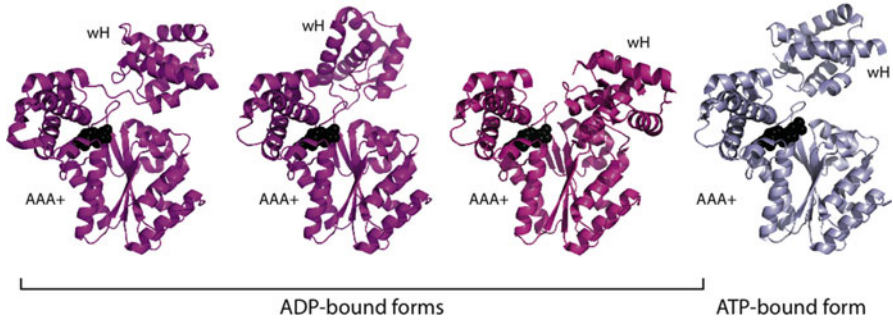


Fig. 4.2 Conformational variation of *A. pernix* Orc1–2 protein in different nucleotide bound states. ADP-bound forms of the protein are shown in magenta with the ADP in black. The ADPNP-bound form of the protein is shown in blue. The proteins were aligned on their AAA+ domains to highlight the distinct relative placement of the wH domains. Figure prepared using PDB files 1WSS and 1WST

specific nucleotide exchange factors may impinge upon this process in the cellular context, however, there is currently no evidence for the existence of such factors.

What are the consequences of ATP binding for the protein? Wigley and colleagues were able to determine the structures of an *A. pernix* Orc1–2 bound in the apo form, bound to ADP and bound to a non-hydrolysable analog of ATP, ADPNP (5'-adenylyl- β , γ -imidodiphosphate), by subjecting the protein to a denaturation/renaturation regimen before crystallization (Singleton et al. 2004). Interestingly, and in contrast to the situation with the *P. aerophilum* protein, the ADP-bound form of the *A. pernix* protein showed a range of distinct conformations (Fig. 4.2). While little change was observed within the AAA+ domain of the protein, the relative positioning of the wH domain varied, suggesting a degree of conformational flexibility in the ADP-bound form of the protein. In contrast, the ADPNP-bound form of the protein appeared to be much more conformationally constrained, with a locked position not seen in any of the ADP-bound forms of the protein being adopted.

4.4 Structures of Orc1/Cdc6s Bound to DNA

A major step forward in our understanding of the function of these proteins came in 2007 with the publication of two papers describing the structures of ADP-bound forms of Orc1/Cdc6s in complex with DNA (Dueber et al. 2007; Gaudier et al. 2007). One, from Wigley and colleagues, described the structure of *A. pernix* Orc1–1 bound to an ORB element derived from *A. pernix oriC1* (Fig. 4.3a). The second paper, from Berger and colleagues, described the complex of a heterodimer of *S. solfataricus* Orc1–1 and Orc1–3 bound to adjacent mini-ORB and C3 elements from that organism's *oriC2* (Fig. 4.3b). A key finding of both papers was the observation that the

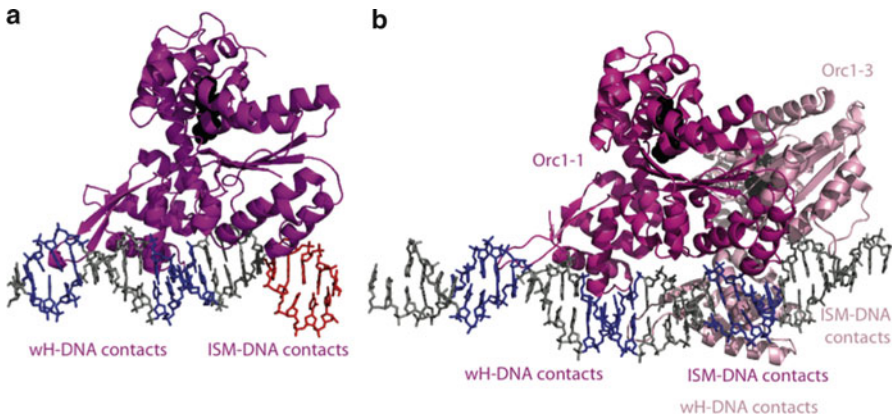


Fig. 4.3 Structures of Orc1/Cdc6 proteins bound to origin DNA. **(a)** *A. pernix* Orc1-1 bound to ORB4 from that organism's *oriC1* (PDB File 2V1U). The dyad symmetric residues in ORB4 (see Fig. 4.1) are shown in *blue* and the G-string in *red*. **(b)** The heterodimer of Orc1-1 (*magenta*) and Orc1-3 (*pink*) from *S. solfataricus* bound to adjacent mini-ORM and C3 sites from *oriC2*. The conserved dyad element of the mini-ORB and a related TTTC of the C3 site are shown in *blue*. Positions of contact between the proteins in DNA are shown below the diagram and colour coded as above

wH domain is not the sole DNA-binding interface in the proteins. All three proteins made additional contacts with DNA mediated by the ISM, the initiator clade signature alpha helix in the AAA+ domain. In the case of the *A. pernix* Orc1-1, this contact was made with the G-string element that is found on one side of the ORB element. Thus the *A. pernix* protein makes two sets of contacts with the DNA. This first is mediated by the wH domain. The recognition helix inserts deeply into the major groove, widening it by over 2 Å and the wing of the wH makes contact with the minor groove, also resulting in significant widening of over 5 Å. Intriguingly, only four base pairs within the recognition site are directly contacted by the protein, although there are a number of additional contacts made with the phosphodiester backbone. These unanticipated additional contacts between ISM and DNA are mediated by a short loop immediately following the ISM alpha helix that inserts into the minor groove of the G-string. This makes a single sequence-specific contact with one of the G-string's guanine residues and has the consequence of widening the minor groove. Thus, the net effect of Orc1-1 binding to the ORB element leads to considerable under-winding of the DNA in the complex and also to the introduction of a bend in the DNA of about 35°. Footprinting studies revealed that the wH domain in isolation was still able to bind to DNA, albeit with a lowered affinity and a loss of protection of the G-string when compared with the protection pattern generated by the full-length protein (Gaudier et al. 2007).

The second paper revealed the structure of the heterodimer of the ADP-bound forms of *Sulfolobus* Orc1-1 and Orc1-3 in complex with adjacent mini-ORB and C3 elements from *oriC2*. As in the *A. pernix* structure, both of the *Sulfolobus* proteins

have bipartite DNA-interaction surfaces, composed of wH domain and ISM. The two proteins abut one another on the DNA, burying about 360 Å² of surface in a protein-protein interface and generating an extensive positively-charged surface of about 2,500 Å that interacts with 28 base pairs of DNA. Despite this extensive interface, a total of only five bases are contacted specifically by the proteins (Dueber et al. 2007). Thus, the paucity of sequence specific contacts appears a general feature of Orc1/Cdc6-DNA interactions. As in *Aeropyrum*, the *Sulfolobus* complex reveals considerable protein-induced under-winding of the DNA. It seems possible therefore that in addition to the modest sequence-specific contacts, the binding of Orc1/Cdc6 proteins is also modulated by the innate deformability of its recognition sequence. If this is the case, the archaeal proteins may represent an evolutionary stepping stone between the tight, highly-sequence-dependent interactions of the bacterial initiator DnaA and the apparently much less sequence-dependent binding of ORC in most eukaryotes. There also appears to be a degree of malleability in the structures of the proteins themselves upon interaction with DNA. Examination of the disposition of *Aeropyrum* Orc1-1 and *Sulfolobus* Orc1-3 on DNA reveal that the ISM makes equivalent contacts with the minor groove of DNA. In contrast, in the *Sulfolobus* Orc1-1/Orc1-3-DNA structure, while the wH domains of both proteins make essentially equivalent interactions with DNA, the respective ISMs do not. More specifically, the interaction between Orc1-1 and Orc1-3 results in the ISM of Orc1-1 being repositioned into the adjacent major groove, altering the angle between AAA+ and wH domains in comparison with the disposition of these domains in Orc1-3. All the DNA bound structures of Orc1/Cdc6s are of the ADP-bound form of the proteins. AAA+ proteins typically function as higher order multimers with the ATP-binding site being found at the interface between protomers. Indeed, residues from both neighbours contribute to binding. The nucleotide is principally bound in a “cis-acting” cleft in one protomer but is additionally coordinated by “trans-acting” residues in the neighbour, the classic such residue is the arginine finger. The arginine finger coordinates the γ -phosphate of ATP and thus provides a means for receiving information from, and effecting conformational changes between, protomers during the nucleotide binding, hydrolysis and release cycle of the active site.

In the Orc1-1/Orc1-3-DNA structure, the *cis*-face of Orc1-1 points towards the *trans*-face of Orc1-3, however, the arginine finger of Orc1-3 points away from the bound ADP. Some degree of repositioning would therefore be required in order to allow the arginine finger to appropriately coordinate an ATP moiety bound by Orc1-1. Given the extensive nature of the protein-DNA contacts, it seems highly likely that any significant conformational alteration within the Orc1/Cdc6 will either remodel the protein-DNA interaction and/or possibly impact upon the protein-induced DNA deformation. Interestingly, if one superimposes the wH domain of the ADPNP bound form of Orc1-2 of *Aeropyrum* onto that of the DNA-bound Orc1-1, then the resultant predicted structure has the ISM some distance removed from the path of the DNA in the structure (Fig. 4.4). This raises the tantalizing possibility that the ATP-bound form of the protein may have its AAA+ domain disengaged from the DNA and thus potentially available for ATP-mediated contacts with the AAA+ domains of adjacent protomers.

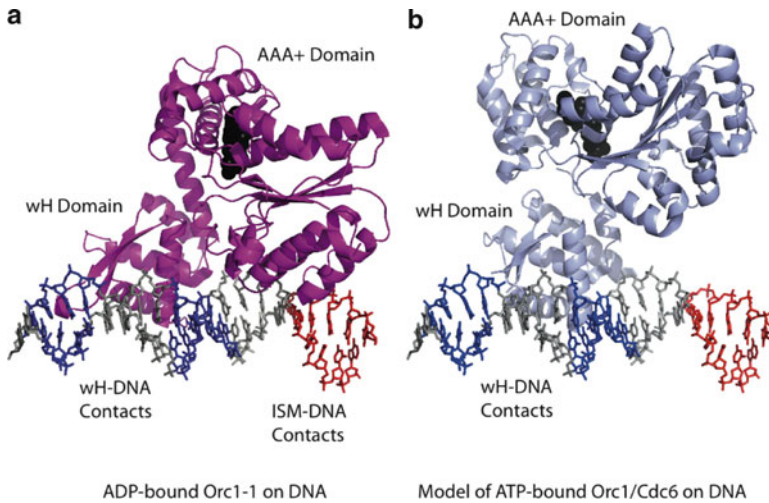


Fig. 4.4 Comparison of the disposition of the AAA+ domain in the known structure (**a**) of ADP-bound Orc1-1 (from *A. permix* Orc1-1 on DNA; PDB 2V1U) with a model (**b**) of how the ATP-bound form may interact with DNA. The model was generated by superimposing the structure of the ADPNP form of *A. permix* Orc1-2 (PDB 1WST) onto the DNA-bound structure of the ADP-form of Orc1-1. Colour coding of the DNA is as in Fig. 4.3; ADP and ADPNP are shown in *black*. The superimposition was generated using the wH domains as the initial point of alignment. The resultant model suggests that the AAA+ domain of the ATP-bound form may be some distance removed from the DNA and thus could potentially be available for additional protein-protein contacts. To date, no actual structure of the ATP-bound form of an Orc1/Cdc6 protein has been determined

A recent study investigated the rules of attraction between *Sulfolobus* Orc1/Cdc6s and origin DNA using a combination of biophysical and molecular-biological methodologies (Dueber et al. 2011). The analysis focused on the mini-ORB and C3 binding sites for Orc1-1 and Orc1-3 at *Sulfolobus oriC2*. The affinities of the isolated proteins for their cognate sites were 390 and 27 nM respectively. Orc1-1 showed a 12-fold lower affinity for a non-specific DNA oligonucleotide; Orc1-3 showed greater powers of discrimination, with a 280-fold difference in affinity. Mutation of conserved residues in either wing or helix of the initiators had dual impacts; the affinity of the mutant protein for DNA was significantly reduced and the ability of the protein to discriminate between specific and non-specific DNA sites was also impaired. A pair of conserved residues in the ISMs of Orc1-1 and Orc1-3 were also targeted for mutagenesis, these residues (G120 and L121 in Orc1-1 and G126 and I127 in Orc1-3) make non-sequence specific van der Waals contacts with DNA. Mutation of these residues had the anticipated effect of reducing the affinity of the initiators for their cognate sites. Surprisingly, the residues also proved important for determining the specificity of binding, despite the absence of direct contacts with the bases. These data suggest that the ISM plays a key role in reading an as yet unidentified aspect of the inherent geometry or deformability of the origin DNA. The affinity data for individual sites were complemented by footprinting

studies using the DNA conformation-sensitive reagent, copper phenanthroline. Although the resultant data were complex, some general principles could be gleaned. First, ISM mutations impacted on the DNA geometry, as revealed by altered hypersensitivity to the footprinting reagent. Second, some ISM mutations, most notably Orc1-1 (G120L, L121D) resulted in an extended region of protection, suggesting that impairing the ability of the ISM to interact with DNA actually facilitated the recruitment of a second protomer of the protein to an adjacent, presumably non-specific, site on the origin. This situation arising from mutation of the ISM is, of course, reminiscent of the model proposed above for an ATP-induced disengagement of the AAA+ domain from DNA (Fig. 4.4).

4.5 Beyond Binding Origins – What Do Orc1/Cdc6s do?

The Orc1/Cdc6s clearly bind to archaeal replication origins but how do they mediate replication initiation? Are they simply passive recruitment platforms for the replication machinery or do they actively mediate origin unwinding prior to helicase recruitment? The latter possibility would be analogous to the situation in bacteria where DnaA mediates localised DNA unwinding before the DnaB•DnaC (helicase•helicase loader) complex is recruited to the newly-exposed single-stranded DNA. However, there is little unambiguous data to support a role for archaeal Orc1/Cdc6s in mediating appropriate origin melting. Wigley and colleagues revealed that high concentrations of *A. pernix* Orc1-1 led to periodic sensitivity to nuclease P1 across the entire origin region *in vitro* (Grainge et al. 2006). However, it was not clear whether this was due to helical distortion or true melting of DNA. More recently, Ishino and colleagues have reported *P. furiosus* Orc1-1 mediated melting of DNA at the single *Pyrococcus* origin of replication *in vitro* as detected by nuclease P1 sensitivity assays (Matsunaga et al. 2010). Puzzlingly, however, this apparent melting was inhibited by ATP. Furthermore, the site of melting was 670 nt removed from the *in vivo* start site of replication, mapped previously by the same authors, raising questions regarding the physiological relevance of this observation.

It may not be too surprising that as yet there is no clinching proof for relevant origin melting by the archaeal initiators. Orc1/Cdc6 is not orthologous to bacterial DnaA and the organization of archaeal origins, with a distinct number of discrete DNA binding sites, clearly differs from the densely packed DnaA boxes in bacterial origins. Furthermore, recent studies in the orthologous eukaryotic system have provided strong support for ORC•Cdc6 mediating loading of MCM onto double stranded DNA rather than onto a pre-melted origin (Evrin et al. 2009; Remus et al. 2009). Perhaps the most parsimonious model for the archaeal system would be that Orc1/Cdc6 proteins do not lead directly to DNA melting in archaea either and that the MCM helicase is, as in eukarya, recruited to double stranded DNA. Melting could take place at a later stage during the activation of the MCM helicase (Bell 2011). Although number of laboratories have reported direct interactions between archaeal Orc1/Cdc6s and MCM, the mechanism of the putative loading reaction remains elusive to date.

Acknowledgements I thank Marek Drozd and Rachel Samson for their comments on this chapter. I would also like to thank Alessandro Costa for valuable discussions on the role of ATP in these proteins. Work in my lab is supported by the EPA Trust, Wellcome Trust and BBSRC.

References

- Bell SD (2011) DNA replication: archaeal oriGINS. *BMC Biol* 9:36
- Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem* 71:333–374
- Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P (2008) Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* 6:245–252
- Capaldi SA, Berger JM (2004) Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon *Methanothermobacter thermoautotrophicus*. *Nucleic Acids Res* 32:4821–4832
- Dueber ELC, Corn JE, Bell SD, Berger JM (2007) Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science* 317:1210–1213
- Dueber EC, Costa A, Corn JE, Bell SD, Berger JM (2011) Molecular determinants of origin discrimination by Orc1 initiators in archaea. *Nucleic Acids Res* 39:3621–3631
- Duggin IG, McCallum SA, Bell SD (2008) Chromosome replication dynamics in the archaeon *Sulfolobus acidocaldarius*. *Proc Natl Acad Sci USA* 105:16737–16742
- Edgell DR, Doolittle WF (1997) Archaea and the origin(s) of DNA replication proteins. *Cell* 89:995–998
- Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L, Hedlund BP, Brochier-Armanet C, Kunin V, Anderson I, Lapidus A, Goltsman E, Barry K, Koonin EV, Hugenholtz P, Kyrpides N, Wanner G, Richardson P, Keller M, Stetter KO (2008) A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc Natl Acad Sci USA* 105:8102–8107
- Erzberger JP, Berger JM (2006) Evolutionary relationships and structural mechanisms of AAA plus proteins. *Annu Rev Biophys Biomol Struct* 35:93–114
- Erzberger JP, Mott ML, Berger JM (2006) Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat Struct Mol Biol* 13:676–683
- Evrin C, Clarke P, Zech J, Lurz R, Sun JC, Uhle S, Li HL, 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 USA* 106:20240–20245
- Gaudier M, Schuwirth BS, Westcott SL, Wigley DB (2007) Structural basis of DNA replication origin recognition by an ORC protein. *Science* 317:1213–1216
- Grainge I, Gaudier M, Schuwirth BS, Westcott SL, Sandall J, Atanassova N, Wigley DB (2006) Biochemical analysis of a DNA replication origin in the archaeon *Aeropyrum pernix*. *J Mol Biol* 363:355–369
- Iyer LM, Leipe DD, Koonin EV, Aravind L (2004) Evolutionary history and higher order classification of AAA plus ATPases. *J Struct Biol* 146:11–31
- Jacob F, Cuzin F, Brenner S (1963) On regulation of DNA replication in bacteria. *Cold Spring Harb Symp Quant Biol* 28:329–348
- Liu JY, Smith CL, DeRyckere D, DeAngelis K, Martin GS, Berger JM (2000) Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. *Mol Cell* 6:637–648
- Lundgren M, Andersson A, Chen LM, Nilsson P, Bernander R (2004) Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci USA* 101:7046–7051
- Majernik AI, Chong JPJ (2008) A conserved mechanism for replication origin recognition and binding in archaea. *Biochem J* 409:511–518
- Matsunaga F, Takemura K, Akita M, Adachi A, Yamagami T, Ishino Y (2010) Localized melting of duplex DNA by Cdc6/Orc1 at the DNA replication origin in the hyperthermophilic archaeon *Pyrococcus furiosus*. *Extremophiles* 14:21–31

- Myllykallio H, Lopez P, Lopez-Garcia P, Heilig R, Saurin W, Zivanovic Y, Philippe H, Forterre P (2000) Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* 288:2212–2215
- Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, Allers T (2007) Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet* 3:729–743
- Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JFX (2009) Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139:719–730
- Robinson NP, Bell SD (2007) Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc Natl Acad Sci USA* 104:5806–5811
- Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, Bell SD (2004) Identification of two origins of replication in the single chromosome of the Archaeon *Sulfolobus solfataricus*. *Cell* 116:25–38
- Robinson NP, Blood KA, McCallum SA, Edwards PAW, Bell SD (2007) Sister chromatid junctions in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *EMBO J* 26:816–824
- She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CCY, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PMK, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, Van der Oost J (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci USA* 98:7835–7840
- Singleton MR, Morales R, Grainge I, Cook N, Isupov MN, Wigley DB (2004) Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. *J Mol Biol* 343:547–557

Chapter 5

Cdt1 and Geminin in DNA Replication Initiation

Christophe Caillat and Anastassis Perrakis

Abstract One of the mechanisms controlling the initiation of DNA replication is the dynamic interaction between Cdt1, which promotes assembly of the pre-replication license complex, and Geminin, which inhibits it. Specifically, Cdt1 cooperates with the cell cycle protein Cdc6 to promote loading of the minichromosome maintenance helicases (MCM) onto the chromatin-bound origin recognition complex (ORC), by directly interacting with the MCM complex, and by modulating histone acetylation and inducing chromatin unfolding. Geminin, on the other hand, prevents the loading of the MCM onto the ORC both by directly binding to Cdt1, and by modulating Cdt1 stability and activity. Protein levels of Geminin and Cdt1 are tightly regulated through the cell cycle, and the Cdt1-Geminin complex likely acts as a molecular switch that can enable or disable the firing of each origin of replication. In this review we summarize structural studies of Cdt1 and Geminin and subsequent insights into how this molecular switch may function to ensure DNA is faithfully replicated only once during S phase of each cell cycle.

Keywords Replication licensing • Cdt1 • Geminin

5.1 Cdt1 and Geminin: A Functional Preview

In eukaryotes, the control of the DNA replication license depends on the correct spatiotemporal assembly of the pre replication complex (pre-RC) on the DNA. First the origin recognition complex (ORC) binds to origins of replication. This complex then recruits Cdt1 and Cdc6, which in turn recruit the mini-chromosome maintenance

C. Caillat • A. Perrakis (✉)
Department of Biochemistry, Netherlands Cancer Institute,
Plesmanlaan 121, 1066CX Amsterdam, The Netherlands
e-mail: a.perrakis@nki.nl

(MCM) proteins. Together, these proteins form the functional pre-replication complex (pre-RC) that permits the loading of the DNA replication machinery (Blow and Dutta 2005; Stillman 2005) and subsequent replication.

The faithful replication of genomic DNA is ensured in part by the control of licensing by several overlapping pathways (Diffley 2010). Firstly, the formation of the pre-RC is limited in time to the late mitosis (M) and G1 phases of the cell cycle, to ensure that each origin of replication fires only once during S phase (Kearsey and Cotterill 2003). In addition, the activity of the ORC, Cdc6 and Cdt1 are tightly controlled to prevent re-replication in S and G2. Indeed over-expression of Cdc6 and Cdt1 was shown to induce re-replication (Blow and Dutta 2005; Machida et al. 2005). Several mechanisms also exist to control Cdt1 protein levels during the cell cycle to allow or prevent the loading of the MCM, and Cdt1 activity is controlled by cyclin-dependent kinases (Cdks). In yeast, Cdk-dependent regulation of transcription and proteolysis is necessary to prevent re-replication (Blow and Dutta 2005). In higher eukaryotes, the regulation by Cdks is also important, but additional mechanisms exist. Specifically, Cdt1 activity is regulated by ubiquitination and degradation through different pathways (Truong and Wu 2011) and by the binding of Cdt1 to Geminin (McGarry and Kirschner 1998; Wohlschlegel et al. 2000).

5.1.1 *The Multiple Faces of Geminin*

5.1.1.1 Geminin Functions in Replication Licensing

Geminin was first described as an inhibitor of DNA replication. It was shown to work by binding and sequestering Cdt1, thereby preventing MCM complex loading until the pre-RC needs to be assembled (McGarry and Kirschner 1998; Wohlschlegel et al. 2000). However, subsequent studies have revealed that the function of Geminin in replication licensing is far more complex.

In both yeast (Chen et al. 2007) and metazoans (Xouri et al. 2007), the recruitment of Cdt1 to chromatin appears to be a dynamic process crucial for MCM loading. In metazoans it is likely that this recruitment is subject to additional regulation by Geminin. Experiments in mammalian cells showed co-localization of Geminin and Cdt1 on chromatin (Kulartz and Knippers 2004) and recruitment of Geminin onto chromatin by Cdt1 in live cells (Xouri et al. 2007), suggesting that licensing inhibitory complexes are formed on chromatin. Indeed, Geminin is thought to have a role specifically in modulating Cdt1 interaction with the origins of replication on the DNA. Supporting this are experiments in *Xenopus* egg extracts, which showed that Geminin significantly stabilizes the binding of ORC, Cdc6 and Cdt1 on plasmid DNA (Waga and Zembutsu 2006). This apparent stabilization of Cdt1 by Geminin on the chromatin is likely caused by inactivated Cdt1 that cannot be released from the ORC-Cdc6 complex.

Interestingly, Geminin can also positively regulate Cdt1 by stabilizing low levels of Cdt1 during S phase, and higher levels in G2, mitosis (Ballabeni et al. 2004) and meiosis (Narasimhachar and Coue 2009). Also, the Cdt1-Geminin complex is still

able to license origins, at least under low concentrations of Geminin, suggesting its mechanism of action is more complicated than straightforward inhibition of Cdt1 (Lutzmann et al. 2006).

Histone acetylation mediated by Cdt1 and Geminin was recently shown to promote MCM loading and replication licensing. Cdt1 was found to promote the loading of MCM at the origins of replication by recruiting the histone acetylase HBO1, which acetylates histone H4 thereby inducing chromatin de-condensation (Miotto and Struhl 2008; Wong et al. 2010). Geminin was shown to inhibit HBO1-mediated histone acetylation in the context of an HBO1-Cdt1 complex (Miotto and Struhl 2010). In addition, Geminin association with the histone deacetylase HDAC11 (Wong et al. 2010) promotes the association of HDAC11 with Cdt1 in S phase, reducing histone acetylation and inhibiting licensing. Thus, Geminin appears to inhibit licensing by suppressing histone acetylation via Cdt1 in two synergistic manners: by repressing HBO1 acetylation and by promoting HDAC11 deacetylation.

5.1.1.2 Geminin in the Cell Cycle

Geminin and Cdt1 levels both oscillate strongly during the cell cycle: Cdt1 accumulates during G1, while Geminin accumulates during S, G2, and M phases (Nishitani et al. 2004). This robust characteristic reverse oscillation has recently been used as a marker for the cell cycle (Kulartz and Knippers 2004; Mechali and Lutzmann 2008; Sakaue-Sawano et al. 2008).

The synthesis of Geminin is transcriptionally regulated by the retinoblastoma tumour suppressor (RB)/E2F pathway, starting at the G1/S phase transition and continuing during S phase (Markey et al. 2004). Geminin levels persist through the S and G2 phases, and the protein is degraded only late in mitosis at the anaphase/metaphase transition, through anaphase promoting complex/cyclosome-mediated proteolysis. On the contrary, Cdt1 accumulates during G1 and is then ubiquitinated by two distinct E3 ubiquitin ligases, SCF-Skp2 and Cul4-Ddb1, during S and G2 (Nishitani et al. 2006), and subsequently degraded by proteolysis.

These profiles suggest that Geminin and Cdt1 may co-exist both late in mitosis, during the M/G1 transition when Cdt1 levels start accumulating, and then at the G1/S transition, as Geminin levels start to accumulate again. In human cells, Geminin stabilizes Cdt1 and promotes its accumulation at the G2/M transition (Ballabeni et al. 2004). In some cell lines, Cdt1 and Geminin have been shown to co-express only at the G1/S transition; Cdt1 is then degraded early in S phase, while chromatin-bound Geminin persists (Kulartz and Knippers 2004; Mechali and Lutzmann 2008; Sakaue-Sawano et al. 2008). Moreover, while silencing of Geminin has been shown to induce re-replication and cell cycle arrest in several cell lines (McGarry 2002; Melixetian et al. 2004; Zhu et al. 2004), deletion of Geminin does not affect the cell cycle in HeLa cells (Kulartz and Knippers 2004). Similarly, the cellular response to Geminin depletion by small interfering RNA (siRNA) has been shown to be cell type dependent, which could be used as a mechanism to target specific cancer cells (Zhu and Depamphilis 2009). This selectivity was proposed to

occur because Cdt1 inhibition by Geminin is the primary mechanism for preventing DNA re-replication in some cancer cells, whereas normal cells possess several partially redundant mechanisms. Thus Geminin depletion induces re-replication only in certain cancer cells, thereby triggering apoptosis. Further work is needed to fully characterize and uncover the function of these oscillating Geminin and Cdt1 levels over the cell cycle.

5.1.1.3 Geminin in Cell Differentiation

Besides its role in proliferation, Geminin also has a prominent role in cell differentiation (Seo and Kroll 2006). Intriguingly, in some cell lines Geminin is involved in the decision to direct cells to either divide or to differentiate. However the mechanisms involved in this are still unclear.

In *Xenopus* embryos, Geminin is required to restrain commitment and spatially restrict mesoderm, endoderm and non-neural ectoderm to their proper locations by regulating the expression of genes involved in early embryonic development (Lim et al. 2011). In hematopoietic stem cells, degradation of Geminin switches the cells from an undifferentiated state to a proliferative and differentiated state, and is directed by the RDCOXB4 complex (Hoxb4 bound to the ubiquitin ligase core component Roc1/Rbx1-Ddb1-Cul4a) (Ohno et al. 2010) or by the PcG (polycomb gene) complex 1 (Ohtsubo et al. 2008), both E3 ubiquitin ligases. Geminin can also regulate the acquisition of neural fate in embryonic stem cells (Yellajoshiyula et al. 2011). This function of Geminin in differentiation seems to be due to its ability to maintain chromatin, and more specifically neural genes, in an accessible and hyper-acetylated state, thereby promoting transcriptional activation.

Geminin can also interact with several homeodomain-containing regulators of DNA transcription such as Six3 (Del Bene et al. 2004) and Hox family members (Luo et al. 2004). The homeodomain is a DNA-binding domain found in hundreds of transcription factors that promote distinct developmental programs (Merabet et al. 2009). Geminin also interacts with the SWI/SNF remodelling complex catalytic subunit Brg1 (Seo et al. 2005), the PcG protein PRC1 (Luo et al. 2004) and with the co-repressor SMRT (Kim et al. 2006). More recently, Geminin has also been shown to interact with coiled-coil proteins through a coiled-coil interaction. Geminin interacts with the proteins ERNI and BERT, which are implicated in neural plate acquisition (Papanayotou et al. 2008). In addition Geminin interacts with the recently identified Geminin-related protein Idas (Pefani et al. 2011), which exhibits high levels of expression in the choroid plexus and the cortical hem of the developing mouse forebrain (Pefani et al. 2011).

As previously discussed, Cdt1 is implicated in the unfolding of chromatin through interaction with HBO1 and HDAC11 (Miotto and Struhl 2008; Wong et al. 2010) and binding of Geminin to Cdt1 inhibits HBO1 acetylase activity (Miotto and Struhl 2010). During development HBO1 functions as a transcriptional activator, which is indispensable for H3K14 acetylation and for the normal expression of essential genes regulating embryonic development (Kueh et al. 2011). Therefore, Geminin might also function in additional differentiation processes through its

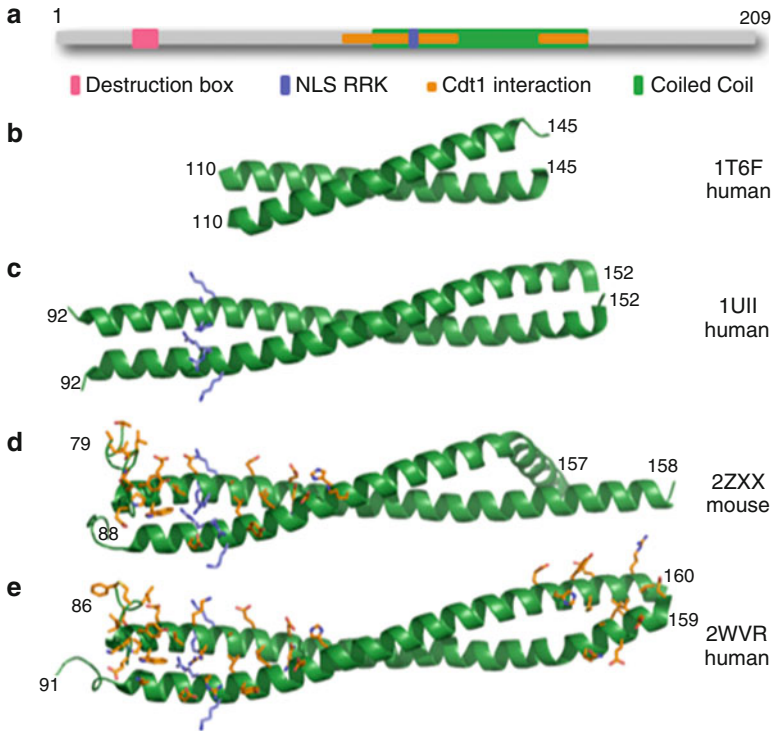


Fig. 5.1 Domains and motifs of Geminin **(a)** Domain structure of human Geminin. The coiled-coil domain, domains interacting with Cdt1, the NLS and the destruction box are shown. **(b–e)** The different structures of Geminin coiled-coils are shown as a cartoon representation. The residues of the NLS (*blue*) and the residues interacting with Cdt1 (*orange*) are shown in sticks

interaction with Cdt1 by modulating gene expression through histone deacetylation and chromatin compaction.

How can such a small protein (around 200 amino acids in humans) apparently interact with so many partners to mediate such diverse functions? The only interaction that has been analyzed in detail and characterized structurally is the Cdt1-Geminin interaction. This interaction is therefore known to be direct and it remains possible that some of the other interactions are actually indirect.

5.1.2 The Structure of Geminin

The analysis of deletion mutants of Geminin (Kroll et al. 1998; McGarry and Kirschner 1998; Benjamin et al. 2004) together with crystal structures (Saxena et al. 2004; Thepaut et al. 2004) have defined several domains within the Geminin protein (Fig. 5.1), roughly consisting of an N-terminal region (residues 1–95), a central

coiled-coil domain (residues 96–160) whose function and structure we will discuss, and a C-terminal domain (residues 160–209) with a not fully understood function, that binds the Brahma (Brm) catalytic subunit of the SWI/SNF chromatin remodeling complex, and is cleaved by caspase-3 at two sites during apoptosis, one of which removes the Brm interaction (Roukos et al. 2007).

5.1.2.1 The N-Terminal Domain

At the anaphase-metaphase transition, Geminin is degraded by the anaphase promoting complex/cyclosome (APC/C) (McGarry and Kirschner 1998). APC/C recognizes the destruction box (residues 23–31) and ubiquitinates Geminin, directing it to the proteasome for degradation.

Geminin is mainly localized in the nucleus (Kroll et al. 1998; McGarry and Kirschner 1998) and its known functions occur in the nucleus. While non-mammalian Geminin possesses a classic bipartite nuclear localisation signal (NLS) (Benjamin et al. 2004; Boos et al. 2006) located in the N-terminal domain, the mammalian Geminin NLS appears to be an arginine-arginine-lysine (RRK) sequence at residues 106–108 (Boos et al. 2006; Sakaue-Sawano et al. 2008) located in the coiled-coil domain. Interestingly, arginines 106 and 107 interact with Cdt1 in the Cdt1-Geminin complex, which suggests that this interaction could interfere with the nuclear localization of Geminin. The N-terminal domain of Geminin has been proposed to play a role in cell differentiation. A neutralizing domain, consisting of residues 38–90, has been identified as sufficient to induce uncommitted embryonic cells to differentiate into neurons (Kroll et al. 1998).

5.1.2.2 The Coiled-Coil Domain

The coiled-coil domain is one of the principal structural motifs involved in protein oligomerization. Predictions based on primary sequence analysis suggest that 2–10% of all the residues in proteins are coiled-coil regions (Moutevelis and Woolfson 2009; Wolf et al. 1997). Although these domains look structurally simple, they are highly versatile and are involved in many different functions. Coiled-coil domains are composed of two or more α -helical peptides, which interact in a specific manner to form a complex. They are typically composed of heptads, i.e. repeats of seven amino acids (with positions denoted *abcdefg*). The specificity and properties of the interaction between the two α -helices depends on the nature of the amino acids at each position. The amino acids at position *a* and *d* are mainly hydrophobic and the packing of the residues *a,d* from one α -helix on the residues *a,d* of the other α -helix constitute the core of the interface. The amino acids at position *e* and *g* are often charged residues and contribute to the specificity of the binding.

The crystal structure of the coiled-coil domain of Geminin alone (Saxena et al. 2004; Thepaut et al. 2004) and in complex with Cdt1 (De Marco et al. 2009;

Lee et al. 2004) has been determined by X-ray crystallography. Geminin forms a head-to-head homodimer in solution, typical for many coiled-coils, and also associates with itself *in vivo* (Saxena et al. 2004). The coiled-coil domain of Geminin consists of two α -helices interacting together through a left-handed superhelix, by regular interlocking side-chain interactions. The seven heptad repeats (residues 96–144) show some variation compared to the canonical coiled-coil sequence. This suggests that the packing of the Geminin:Geminin dimer is not ideal and does not form a stable complex. On the *a* and *d* positions, several residues (i.e. Ser96, Trp99, Ala103, Arg106, Ala113, Asn117, Lys127 and Asn138) are not ideal for stabilization. On the *e* and *g* positions, the side-chains of five hydrophobic residues are exposed to the solvent. Three of them (Val102, Ala109 and Leu114) are mediating hydrophobic interactions with Cdt1, suggesting that the Geminin homodimer is probably stabilized by the interaction with Cdt1. In the structure of the human Cdt1 in complex with Geminin (PDB 2WVR), residues 148–160 are extending the coiled-coil, by two additional heptads.

The non-ideal packing of the Geminin:Geminin homodimer raises the possibility that Geminin could form heterodimers with other partners of higher stability. This is strongly supported by two studies of coiled-coil containing proteins interacting with Geminin through coiled-coil interactions. Specifically, the coiled-coil domain of the protein ERNI, involved in establishing neural plate identity, can interact with itself and with the coiled-coil of Geminin (Papanayotou et al. 2008). Similarly, the coiled-coil of Idas can also interact with itself but preferentially with the coiled-coil of Geminin (Pefani et al. 2011). Interestingly, the binding of Idas to Geminin reduces the affinity of Geminin for Cdt1. Thus, the proposed heterodimerization of Geminin could modulate its interaction with other partners and thereby provide an additional mechanism for regulating Geminin function.

The Geminin coiled-coil domain was shown to interact directly with the homeodomain of the Hoxa11 protein using peptide arrays (Luo et al. 2004). However, it was also shown that the coiled-coil of Geminin is not sufficient to interact with Hoxa11 (Saxena et al. 2004) and that the N-terminal (1–70) or C-terminal (152–209) domains of Geminin are required. Del Bene et al. showed by two-hybrid analysis and pull-down assays that full-length Geminin is required to interact with the homeodomain containing transcription factor Six3 (Del Bene et al. 2004). We have so far been unable to detect an interaction between the recombinant homeodomain of Hox proteins and recombinant Geminin *in vitro*, using biophysical methods (unpublished results). The mechanism by which Geminin interacts with these transcription factors remains unknown.

5.1.3 The Structure of Cdt1

Cdt1 can be divided into three domains; an N-terminal domain (1–166) whose structure is unknown, and two winged helix domains (WHD) forming the middle (167–351) and C-terminal domains (352–546) (Fig. 5.2).

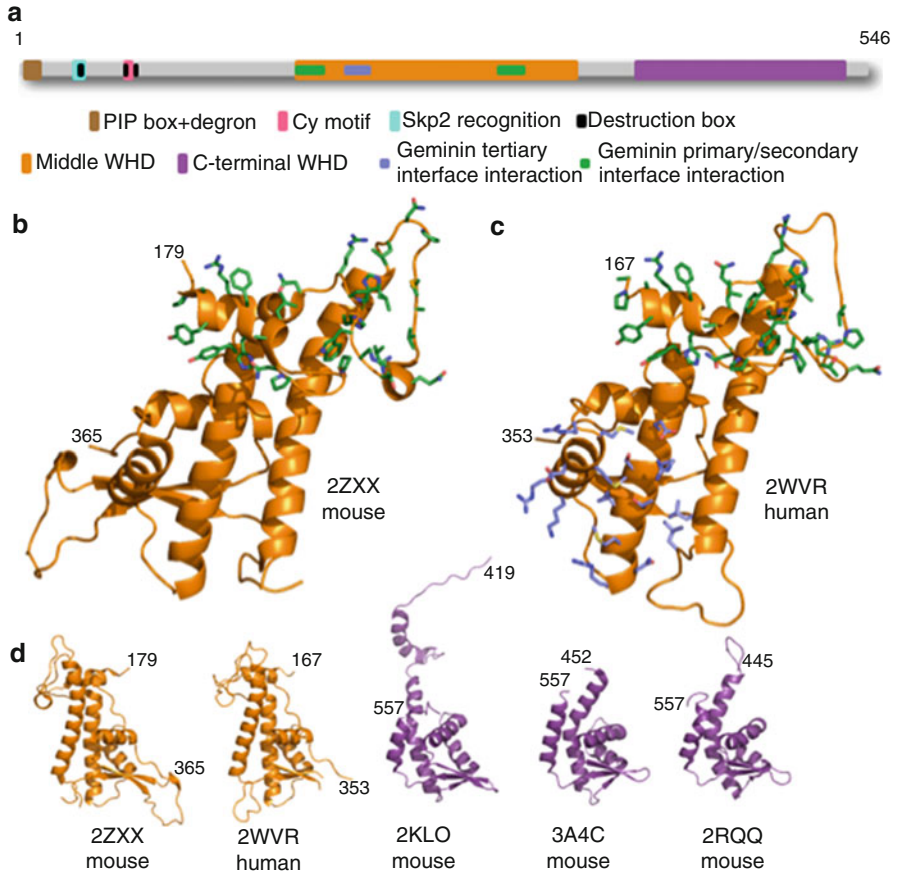


Fig. 5.2 Domains and motifs of Cdt1 (a) Domain structure of human Cdt1. The middle and C-terminal domains, the regions interacting with Geminin and the different regulation motifs of the N-terminal domain are represented. PCNA bound to chromatin bind the PIP box and subsequently Cdt2 recognize the degron motif which promotes the activity of the Cul4–Ddb1–Cdt2 E3 ligase in response to DNA damage. The cyclin A–Cdk binds the Cy motif and phosphorylates the Thr29 close to the Skp2 recognition site that promotes the binding of the SCF–Skp2 E3 ligase. The three destruction boxes recognized by the APC/C^{dh1} are also represented. (b) Structures of the human and (c) mouse middle domain of Cdt1. The residues interacting with Geminin are shown as sticks; residues implicated in the primary and secondary interaction in green and the residues implicated in the ternary interaction in blue. (d) The different structures of Cdt1 middle (orange) and C-terminal (purple) domains are shown as a cartoon representation

5.1.3.1 The N-Terminal Domain Is Highly Regulated

In metazoans the N-terminal domain of Cdt1 mediates interactions with several proteins implicated in different mechanism of regulation. This probably allows more sophisticated control of the regulation of licensing and also variability in the regulation of replication in different cell types.

The N-terminal domain of human Cdt1 is recognized by three distinct E3 ubiquitin ligases during the S and G2 phases (Truong and Wu 2011) to promote ubiquitin-mediated proteolysis. First, Cyclin A-Cdk binding (Liu et al. 2004; Sugimoto et al. 2004) to the Cy-motif RRL (68–70) of Cdt1, and subsequent phosphorylation of Thr29 close to the recognition site of Skp2 SPARPALR (31–38), induces binding of the SCF-Skp2 E3 ligase and ubiquitination of Cdt1 (Nishitani et al. 2006). Second, Cul4–Ddb1–Cdt2 targets Cdt1 during the S phase and in response to DNA damage (Havens and Walter 2009; Hu et al. 2004; Nishitani et al. 2006; Takeda et al. 2005). Degradation is dependent on Cdt1 binding to PCNA, which is involved in promoting DNA replication, through a consensus PCNA-interaction motif (PIP) located in the first 28 amino acids of Cdt1, and to subsequent Cdt2-mediated recognition of the degron motif adjacent to the PIP box (Arias and Walter 2006; Roukos et al. 2011). Finally, the APC/C^{Cdh1} is a third ubiquitin ligase recognizing three destruction boxes in the N-terminal domain of Cdt1 (Sugimoto et al. 2008).

Acetylation of the N-terminal domain of Cdt1 prevents ubiquitination by E3 ligases (Glozak and Seto 2009). This acetylation can be removed by the binding of HDAC11 on Cdt1. Given that Geminin binding to Cdt1 promotes the recruitment of HDAC11 (Wong et al. 2010), Geminin could be expected to enhance Cdt1 de-acetylation and thereby promote its degradation. Structural data on these complexes would help elucidate how this could be achieved. The N-terminal domain of Cdt1 also binds Cdc7 (Ballabeni et al. 2009), which seems to regulate the amount of Cdt1 bound to chromatin. Finally, Cdt1 has two NLS (Arentson et al. 2002) required for its nuclear import, which is probably mediated through an interaction with the import receptors importin α 1 and importin β 1 (Sugimoto et al. 2008).

5.1.3.2 The Structurally Conserved Winged Helix Domains

Winged helix domains (WHD) are primarily used to mediate DNA recognition but can also mediate protein-protein interactions (Gajiwala and Burley 2000). Many pre-RC proteins (Orc1, Orc2, Cdc6, Cdt1 and Mcm6) have a WHD, suggesting a common evolutionary origin (Khayrutdinov et al. 2009). The two WHDs of Cdt1 are found in the middle domain and the C-terminus. The structure of the middle domain of Cdt1 was determined by X-ray crystallography in a complex with the Geminin dimer (De Marco et al. 2009; Lee et al. 2004) and the C-terminal domain was recently determined by X-ray crystallography and solution NMR spectroscopy (Jee et al. 2010; Khayrutdinov et al. 2009). As these two domains share only 9% sequence identity, the C-terminal domain was not expected to be structurally similar to the middle domain ($Z=1.1$, RMSD of 3.4 Å for 91 C α atoms) (Khayrutdinov et al. 2009).

The middle domain of Cdt1 also contains the binding site for Geminin, although the interaction site is not located on the WHD fold itself but on the N-terminal and C-terminal extensions (Fig. 5.2). The strong interaction of Geminin with Cdt1 probably stabilizes the middle domain and affects its conformation, because when over-expressed alone it is highly unstable, but co-expression with Geminin stabilizes it. This middle WHD can also bind DNA (discussed in the next section). The C-terminal

WHD of Cdt1 does not bind to DNA but mediates interaction with the MCM through a C-terminal WHD domain in Mcm6 (Wei et al. 2010; Yanagi et al. 2002). Finally, Cdt1 can also interact with Mcm9, although the details of this interaction are unknown. Mcm9 seems to be a positive regulator of Cdt1 and prevents the recruitment of an excess of Geminin to Cdt1 (Lutzmann and Mechali 2008), suggesting that Mcm9 could interact with the middle domain of Cdt1.

5.1.3.3 The Recruitment of Cdt1 on Chromatin

In yeast, ORC and Cdc6 recruit Cdt1 and the MCM to origins of replication (Randell et al. 2006) via a direct interaction between Cdt1 and the Orc6 subunit (Chen et al. 2007). Experiments in *Xenopus* egg extracts suggest that also in metazoans, the loading of MCM by Cdt1 requires the presence of Cdc6 on chromatin (Tsuyama et al. 2005). However, there is no data on how Cdt1 is recruited to the ORC-Cdc6 complex. The analysis of the binding of Cdt1 on DNA shows that Cdt1 can bind DNA directly *in vitro*, in a sequence-, strand- and conformation-independent manner. It has been proposed that residues 1–293 of Cdt1 contain a DNA binding site that could be composed of two independent sites on the N-terminal and the middle domains (Yanagi et al. 2002). Fluorescence recovery after photobleaching (FRAP) experiments using deletion mutants of Cdt1 implicated the N-terminal (residues 1–140) and middle domains (residues 298–352) of Cdt1 in the recruitment of Cdt1 to chromatin in cells (Xouri et al. 2007). However, detailed information of Cdt1-DNA interactions are still missing, while interactions between Cdt1 and ORC subunits or with Cdc6 for the loading of the MCM have not been identified (see Chap. 2, this volume).

The middle WHD could contribute to the binding of Cdt1 on chromatin by binding to DNA. However, this would likely have to be in a non-canonical fashion, since in a canonical DNA-binding WHD, helix H3 recognizes the major groove of the DNA and has several basic residues (Gajiwala and Burley 2000), whereas in the middle WHD of Cdt1 the C-terminal loop of the domain folds back on the α -helix H3 (Fig. 5.3), which possesses several acidic residues.

The superposition of the middle domain of Cdt1 (PDB: 2WVR) with other WHDs in complexes with DNA, i.e. those from E2F (PDB: 1CF7), RTP (PDB: 2DPD), Orc1 (PDB: 2V1U) or RFX (PDB: 1DP7), shows that the sequences in the N-terminal and C-terminal of the WHD, which are stabilized by Geminin, would collide with the DNA. Since the structure of this Cdt1 domain is known only in a complex with Geminin, and since Geminin binding may induce conformational changes that stabilize this Cdt1 domain, as we previously discussed, DNA binding may only occur when the H3 helix is ‘released’ from the protection of Geminin, as also suggested by previous work (Yanagi et al. 2002). However, Geminin does not totally inhibit Cdt1 binding on chromatin as Geminin and Cdt1 have been shown to co-localize on chromatin (Gillespie et al. 2001; Kulartz and Knippers 2004; Lutzmann et al. 2006; Xouri et al. 2007). Structural data on the binding of Cdt1 to DNA would be needed to understand the mechanism of its recruitment to chromatin and how the binding of Geminin could affect it.

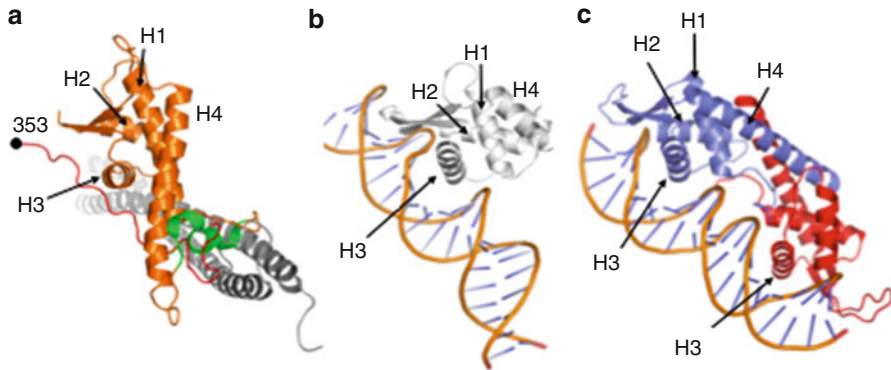


Fig. 5.3 Binding of the winged helix domains on DNA (a) Middle domain of Cdt1 (*orange*, with the Geminin interacting domains in *light green* and the C-terminal tail in *red*) in complex with Geminin (*grey*). The C-terminal tail of Cdt1 middle domain interacts with the H3 α -helix that normally contacts the DNA in canonical WHD domains (PDB: 2WVR). (b) WHD of ORC1 in complex with DNA (*silver*, PDB: 2V1U). (c) RTP dimer bound to DNA (PDB: 2DPD)

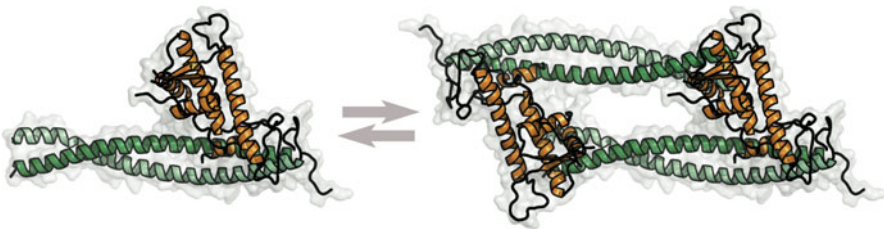


Fig. 5.4 The structures of the Cdt1-Geminin trimer and hexamer likely represent a licensing-permissive and a licensing-inhibitory complex and both exist in cells in different spatiotemporal contexts (PDB: 2WVR, (De Marco et al. 2009))

5.1.4 The Cdt1-Geminin Complex

The crystal structure of a truncated Cdt1-Geminin complex has been determined for the mouse (Lee et al. 2004) and human (De Marco et al. 2009) proteins (Fig. 5.3). In these structures the middle domain of Cdt1 (167–353) interacts with a dimer of Geminin (86–160) through tight primary and secondary interfaces with nano-molar affinity (Lee et al. 2004). This heterotrimeric complex can dimerize through a tertiary interface, forming a heterohexamer (De Marco et al. 2009).

5.1.4.1 The Primary and Secondary Interfaces

The primary and secondary interfaces are mainly hydrophobic and bury a surface of $\sim 1,240 \text{ \AA}^2$ (PDB: 2WVR), as shown by analyzing the interface with the interactive

tool PISA (Krissinel and Henrick 2007). The primary interface is composed of the two small α -helices located in the N-terminus of the middle WHD of Cdt1 (residues 167–182). These helices interact with the N-terminal regions of both monomers of the Geminin coiled-coil (residues 106–121). This is the most important interface for the interaction between Cdt1 and the Geminin dimer as multiple point mutations in Geminin at this interface reduce the affinity for Cdt1 by 2,000 fold and a construct of Cdt1 lacking these two N-terminal α -helices does not bind to Geminin (Lee et al. 2004). The secondary interface is composed of a loop located towards the C-terminus of the Cdt1 domain (residues 310–336) and a loop in the N-terminal of the coiled-coil of one of the Geminin monomers (residues 86–101). The contribution of this interface to the affinity of Geminin for Cdt1 is lower, and deletion of this loop on Cdt1 reduces the affinity for Geminin only about 15 fold.

5.1.4.2 The Tertiary Interface

De Marco et al. described a tertiary interface mediating the dimerization of the Cdt1-Geminin heterotrimer in a head to tail conformation, forming a heterohexamer. This interface between the C-terminal section of Geminin's coiled-coil (residues 145–160) and the middle domain of Cdt1 buries a surface of 650 Å² and is rich in charged residues, displaying eight hydrogen bonds and eight salt bridges. This ternary interaction seems to be a weak affinity interaction *in vitro*, and the complex likely exists in an equilibrium between a heterotrimer and a heterohexamer. The dimerization of the complex is dependent on the residues located at the ternary interface on both Cdt1 and Geminin. It is possible that the *in vivo* stoichiometry of the complex is affected by both the increase in local concentrations, caused by recruitment on chromatin (Lutzmann et al. 2006; Waga and Zembutsu 2006; Xouri et al. 2007), and/or by post-translational modifications, for example by Cdt1 phosphorylation, since Geminin was suggested to stabilize more efficiently a hyper-phosphorylated form of Cdt1 (Ballabeni et al. 2004).

5.1.4.3 Conformational Change of the N-Terminal Domain?

Geminin has been suggested to prevent the degradation of Cdt1, thereby stabilizing protein levels (Ballabeni et al. 2004; Narasimhachar and Coue 2009; Zhu and Depamphilis 2009). The SCF-Skp2 ubiquitin ligase induces proteolysis of Cdt1 at the G1/S phase transition, when the levels of Geminin are increasing. This suggests that the binding of Geminin on Cdt1 could prevent either Cdt1 phosphorylation by Cyclin A/Cdk or ubiquitination by SCF-Skp2. However, the currently known Geminin binding site on Cdt1 does not overlap with the recognition site of Cyclin A-CDK or SCF-Skp2. Interestingly though, pull-down assays have suggested that the 70–92 residue region of Geminin interacts with the N-terminal 100 residues of Cdt1 (Saxena et al. 2004). Thus, a possible mechanism to explain these observations could be that Geminin binding in the middle domain of Cdt1 induces a conformational change of the N-terminal region of Cdt1, thereby preventing its ubiquitination and degradation.

5.1.5 Models for a Cdt1-Geminin Molecular Switch

Geminin constructs containing the primary and secondary interface for interaction with Cdt1 (Geminin⁸²⁻¹⁴⁵) are unable to inhibit DNA synthesis in an *in vitro* DNA replication assay derived from *Xenopus* egg extracts (De Marco et al. 2009; Thepaut et al. 2004). The C-terminal end of the coiled-coil that is implicated in the formation of the hexameric complex seems to be crucial for the inhibition of DNA replication (De Marco et al. 2009; McGarry and Kirschner 1998). Geminin binding to Cdt1 does not inhibit Cdt1 association with chromatin (Waga and Zembutsu 2006), and the dynamic association of Cdt1 to chromatin recruits Geminin onto chromatin (Xouri et al. 2007). As the Cdt1-Geminin complex licenses chromatin but prevents re-replication in a concentration dependent manner, we have suggested that the changes in the oligomerization state of the Cdt1-Geminin complex works as a molecular switch at replication origins.

Several possible oligomerization states of Geminin and the Cdt1-Geminin complex have been described. SAXS data (Thepaut et al. 2004) and electron microscopy (Okorokov et al. 2004) data suggest that the coiled-coil of Geminin can form tetramers, where two homodimers of Geminin interact to form a tetramer. It has been proposed that the Cdt1-Geminin heterotrimer is recruited to chromatin and is able to load the MCM, but increasing concentrations of Geminin lead to the binding of additional Geminin on the Cdt1-Geminin heterotrimer resulting in an ‘inhibitory’ complex (Lutzmann et al. 2006). Crystallographic studies and SAXS experiments combined with *in vivo* data suggest the existence of a Cdt1-Geminin ‘permissive’ heterotrimer that dimerizes into an ‘inhibitory’ heterohexamer (De Marco et al. 2009).

The loading of the MCM is inhibited by Geminin when its concentration exceeds a threshold (McGarry and Kirschner 1998; Waga and Zembutsu 2006) suggesting a highly cooperative mechanism. Based on these data, a mathematical model of the regulation of Cdt1 by Geminin was developed (Ode et al. 2011). To take into account the ‘all-or-nothing’ inhibition, they proposed a model with inter-origin cooperativity, supported by experimentally measured distributions of inter-origin distances. In this model the Cdt1-Geminin complexes loaded on different origins of replication can interact with adjacent origins, forming complexes of higher oligomerization states. This reaction can propagate through the chromatin, inhibiting the origins of replication in a switch-like manner.

5.2 Conclusions

In higher eukaryotes, Cdt1 recruitment on the chromatin promotes chromatin unfolding and loading of the MCM complex, playing a role not only in the licensing of DNA replication but also in transcription and DNA repair. The binding of Geminin on Cdt1 regulates Cdt1 function by inhibiting MCM loading and affects Cdt1-mediated histone acetylation and deacetylation. The presence of Geminin in metazoans allows an additional layer of Cdt1 regulation on top of ubiquitin-mediated

degradation. The structural data collected on Cdt1 and Geminin have helped to improve the understanding of the mechanism of Cdt1 inhibition by Geminin. The Cdt1-Geminin complex can switch between a licensing ‘permissive’ to a licensing ‘inhibitory’ state depending on the concentration of Geminin and likely also on other post-translational modifications. Further studies on the recruitment of the Cdt1-Geminin complex to origins of replication, understanding of the cooperation with ORC and Cdc6 for loading the MCM, and consideration of origin co-operativity are required to understand the interplay of the Cdt1-Geminin complex that regulates the spatial and temporal loading of the MCM to license DNA replication.

References

- Arentson E, Faloon P, Seo J, Moon E, Studts JM, Fremont DH, Choi K (2002) Oncogenic potential of the DNA replication licensing protein CDT1. *Oncogene* 21:1150–1158
- Arias EE, Walter JC (2006) PCNA functions as a molecular platform to trigger Cdt1 destruction and prevent re-replication. *Nat Cell Biol* 8:84–90
- Ballabeni A, Melixetian M, Zamponi R, Masiero L, Marinoni F, Helin K (2004) Human geminin promotes pre-RC formation and DNA replication by stabilizing CDT1 in mitosis. *EMBO J* 23:3122–3132
- Ballabeni A, Zamponi R, Caprara G, Melixetian M, Bossi S, Masiero L, Helin K (2009) Human CDT1 associates with CDC7 and recruits CDC45 to chromatin during S phase. *J Biol Chem* 284:3028–3036
- Benjamin JM, Torke SJ, Demeler B, McGarry TJ (2004) Geminin has dimerization, Cdt1-binding, and destruction domains that are required for biological activity. *J Biol Chem* 279:45957–45968
- Blow JJ, Dutta A (2005) Preventing re-replication of chromosomal DNA. *Nat Rev Mol Cell Biol* 6:476–486
- Boos A, Lee A, Thompson DM, Kroll KL (2006) Subcellular translocation signals regulate Geminin activity during embryonic development. *Biol Cell* 98:363–375
- Chen S, de Vries MA, Bell SP (2007) Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading. *Genes Dev* 21:2897–2907
- De Marco V, Gillespie PJ, Li A, Karantzelis N, Christodoulou E, Klompmaker R, van Gerwen S, Fish A, Petoukhov MV, Iliou MS, Lygerou Z, Medema RH, Blow JJ, Svergun DI, Taraviras S, Perrakis A (2009) Quaternary structure of the human Cdt1-Geminin complex regulates DNA replication licensing. *Proc Natl Acad Sci USA* 106:19807–19812
- Del Bene F, Tessmar-Raible K, Wittbrodt J (2004) Direct interaction of geminin and Six3 in eye development. *Nature* 427:745–749
- Diffley JF (2010) The many faces of redundancy in DNA replication control. *Cold Spring Harb Symp Quant Biol* 75:135–142
- Gajiwala KS, Burley SK (2000) Winged helix proteins. *Curr Opin Struct Biol* 10:110–116
- Gillespie PJ, Li A, Blow JJ (2001) Reconstitution of licensed replication origins on *Xenopus* sperm nuclei using purified proteins. *BMC Biochem* 2:15
- Glozak MA, Seto E (2009) Acetylation/deacetylation modulates the stability of DNA replication licensing factor Cdt1. *J Biol Chem* 284:11446–11453
- Havens CG, Walter JC (2009) Docking of PIP box onto chromatin-bound PCNA creates a degron for the ubiquitin ligase CRL4^{Cdt1}. *Mol Cell* 35:93–104
- Hu J, McCall CM, Ohta T, Xiong Y (2004) Targeted ubiquitination of CDT1 by the DDB1-CUL4A-ROC1 ligase in response to DNA damage. *Nat Cell Biol* 6:1003–1009

- Jee J, Mizuno T, Kamada K, Tochio H, Chiba Y, Yanagi K, Yasuda G, Hiroaki H, Hanaoka F, Shirakawa M (2010) Structure and mutagenesis studies of the C-terminal region of licensing factor Cdt1 enable the identification of key residues for binding to replicative helicase Mcm proteins. *J Biol Chem* 285:15931–15940
- Kearsey SE, Cotterill S (2003) Enigmatic variations: divergent modes of regulating eukaryotic DNA replication. *Mol Cell* 12:1067–1075
- Khayrutdinov BI, Bae WJ, Yun YM, Lee JH, Tsuyama T, Kim JJ, Hwang E, Ryu KS, Cheong HK, Cheong C, Ko JS, Enomoto T, Karplus PA, Guntert P, Tada S, Jeon YH, Cho Y (2009) Structure of the Cdt1 C-terminal domain: conservation of the winged helix fold in replication licensing factors. *Protein Sci* 18:2252–2264
- Kim MY, Jeong BC, Lee JH, Kee HJ, Kook H, Kim NS, Kim YH, Kim JK, Ahn KY, Kim KK (2006) A repressor complex, AP4 transcription factor and geminin, negatively regulates expression of target genes in nonneuronal cells. *Proc Natl Acad Sci USA* 103:13074–13079
- Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372:774–797
- Kroll KL, Salic AN, Evans LM, Kirschner MW (1998) Geminin, a neuralizing molecule that demarcates the future neural plate at the onset of gastrulation. *Development* 125:3247–3258
- Kueh AJ, Dixon MP, Voss AK, Thomas T (2011) HBO1 is required for H3K14 acetylation and normal transcriptional activity during embryonic development. *Mol Cell Biol* 31:845–860
- Kulartz M, Knippers R (2004) The replicative regulator protein geminin on chromatin in the HeLa cell cycle. *J Biol Chem* 279:41686–41694
- Lee C, Hong B, Choi JM, Kim Y, Watanabe S, Ishimi Y, Enomoto T, Tada S, Cho Y (2004) Structural basis for inhibition of the replication licensing factor Cdt1 by geminin. *Nature* 430:913–917
- Lim JW, Hummert P, Mills JC, Kroll KL (2011) Geminin cooperates with Polycomb to restrain multi-lineage commitment in the early embryo. *Development* 138:33–44
- Liu E, Li X, Yan F, Zhao Q, Wu X (2004) Cyclin-dependent kinases phosphorylate human Cdt1 and induce its degradation. *J Biol Chem* 279:17283–17288
- Luo L, Yang X, Takihara Y, Knoetgen H, Kessel M (2004) The cell-cycle regulator geminin inhibits Hox function through direct and polycomb-mediated interactions. *Nature* 427:749–753
- Lutzmann M, Mechali M (2008) MCM9 binds Cdt1 and is required for the assembly of prereplication complexes. *Mol Cell* 31:190–200
- Lutzmann M, Maiorano D, Mechali M (2006) A Cdt1-geminin complex licenses chromatin for DNA replication and prevents rereplication during S phase in *Xenopus*. *EMBO J* 25:5764–5774
- Machida YJ, Hamlin JL, Dutta A (2005) Right place, right time, and only once: replication initiation in metazoans. *Cell* 123:13–24
- Markey M, Siddiqui H, Knudsen ES (2004) Geminin is targeted for repression by the retinoblastoma tumor suppressor pathway through intragenic E2F sites. *J Biol Chem* 279:29255–29262
- McGarry TJ (2002) Geminin deficiency causes a Chk1-dependent G2 arrest in *Xenopus*. *Mol Biol Cell* 13:3662–3671
- McGarry TJ, Kirschner MW (1998) Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell* 93:1043–1053
- Mechali M, Lutzmann M (2008) The cell cycle: now live and in color. *Cell* 132:341–343
- Melixetian M, Ballabeni A, Masiero L, Gasparini P, Zamponi R, Bartek J, Lukas J, Helin K (2004) Loss of Geminin induces rereplication in the presence of functional p53. *J Cell Biol* 165:473–482
- Merabet S, Hudry B, Saadaoui M, Graba Y (2009) Classification of sequence signatures: a guide to Hox protein function. *Bioessays* 31:500–511
- Miotto B, Struhl K (2008) HBO1 histone acetylase is a coactivator of the replication licensing factor Cdt1. *Genes Dev* 22:2633–2638
- Miotto B, Struhl K (2010) HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin. *Mol Cell* 37:57–66

- Moutevelis E, Woolfson DN (2009) A periodic table of coiled-coil protein structures. *J Mol Biol* 385:726–732
- Narasimhachar Y, Coue M (2009) Geminin stabilizes Cdt1 during meiosis in *Xenopus* oocytes. *J Biol Chem* 284:27235–27242
- Nishitani H, Lygerou Z, Nishimoto T (2004) Proteolysis of DNA replication licensing factor Cdt1 in S-phase is performed independently of geminin through its N-terminal region. *J Biol Chem* 279:30807–30816
- Nishitani H, Sugimoto N, Roukos V, Nakanishi Y, Saijo M, Obuse C, Tsurimoto T, Nakayama KI, Nakayama K, Fujita M, Lygerou Z, Nishimoto T (2006) Two E3 ubiquitin ligases, SCF-Skp2 and DDB1-Cul4, target human Cdt1 for proteolysis. *EMBO J* 25:1126–1136
- Ode KL, Fujimoto K, Kubota Y, Takisawa H (2011) Inter-origin cooperativity of geminin action establishes an all-or-none switch for replication origin licensing. *Genes Cells* 16:380–396
- Ohno Y, Yasunaga S, Ohtsubo M, Mori S, Tsumura M, Okada S, Ohta T, Ohtani K, Kobayashi M, Takihara Y (2010) Hoxb4 transduction down-regulates Geminin protein, providing hematopoietic stem and progenitor cells with proliferation potential. *Proc Natl Acad Sci USA* 107:21529–21534
- Ohtsubo M, Yasunaga S, Ohno Y, Tsumura M, Okada S, Ishikawa N, Shirao K, Kikuchi A, Nishitani H, Kobayashi M, Takihara Y (2008) Polycomb-group complex 1 acts as an E3 ubiquitin ligase for Geminin to sustain hematopoietic stem cell activity. *Proc Natl Acad Sci USA* 105:10396–10401
- Okorokov AL, Orlova EV, Kingsbury SR, Bagneris C, Gohlke U, Williams GH, Stoeber K (2004) Molecular structure of human geminin. *Nat Struct Mol Biol* 11:1021–1022
- Papanayotou C, Mey A, Birot AM, Saka Y, Boast S, Smith JC, Samarut J, Stern CD (2008) A mechanism regulating the onset of Sox2 expression in the embryonic neural plate. *PLoS Biol* 6:e2
- Pefani DE, Dimaki M, Spella M, Karantzelis N, Mitsiki E, Kyrousi C, Symeonidou IE, Perrakis A, Taraviras S, Lygerou Z (2011) Idas, a novel phylogenetically conserved geminin-related protein, binds to geminin and is required for cell cycle progression. *J Biol Chem* 286:23234–23246
- Randell JC, Bowers JL, Rodriguez HK, Bell SP (2006) Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol Cell* 21:29–39
- Roukos V, Iliou MS, Nishitani H, Gentzel M, Wilm M, Taraviras S, Lygerou Z (2007) Geminin cleavage during apoptosis by caspase-3 alters its binding ability to the SWI/SNF subunit Brahma. *J Biol Chem* 282:9346–9357
- Roukos V, Kinkhabwala A, Colombelli J, Kotsantis P, Taraviras S, Nishitani H, Stelzer E, Bastiaens P, Lygerou Z (2011) Dynamic recruitment of licensing factor Cdt1 to sites of DNA damage. *J Cell Sci* 124:422–434
- Sakaue-Sawano A, Ohtawa K, Hama H, Kawano M, Ogawa M, Miyawaki A (2008) Tracing the silhouette of individual cells in S/G2/M phases with fluorescence. *Chem Biol* 15:1243–1248
- Saxena S, Yuan P, Dhar SK, Senga T, Takeda D, Robinson H, Kornbluth S, Swaminathan K, Dutta A (2004) A dimerized coiled-coil domain and an adjoining part of geminin interact with two sites on Cdt1 for replication inhibition. *Mol Cell* 15:245–258
- Seo S, Kroll KL (2006) Geminin's double life: chromatin connections that regulate transcription at the transition from proliferation to differentiation. *Cell Cycle* 5:374–379
- Seo S, Herr A, Lim JW, Richardson GA, Richardson H, Kroll KL (2005) Geminin regulates neuronal differentiation by antagonizing Brg1 activity. *Genes Dev* 19:1723–1734
- Stillman B (2005) Origin recognition and the chromosome cycle. *FEBS Lett* 579:877–884
- Sugimoto N, Tatsumi Y, Tsurumi T, Matsukage A, Kiyono T, Nishitani H, Fujita M (2004) Cdt1 phosphorylation by cyclin A-dependent kinases negatively regulates its function without affecting geminin binding. *J Biol Chem* 279:19691–19697
- Sugimoto N, Kitabayashi I, Osano S, Tatsumi Y, Yugawa T, Narisawa-Saito M, Matsukage A, Kiyono T, Fujita M (2008) Identification of novel human Cdt1-binding proteins by a proteomics approach: proteolytic regulation by APC/C^{Cdh1}. *Mol Biol Cell* 19:1007–1021
- Takeda DY, Parvin JD, Dutta A (2005) Degradation of Cdt1 during S phase is Skp2-independent and is required for efficient progression of mammalian cells through S phase. *J Biol Chem* 280:23416–23423

- Thepaut M, Maiorano D, Guichou JF, Auge MT, Dumas C, Mechali M, Padilla A (2004) Crystal structure of the coiled-coil dimerization motif of geminin: structural and functional insights on DNA replication regulation. *J Mol Biol* 342:275–287
- Truong LN, Wu X (2011) Prevention of DNA re-replication in eukaryotic cells. *J Mol Cell Biol* 3:13–22
- Tsuyama T, Tada S, Watanabe S, Seki M, Enomoto T (2005) Licensing for DNA replication requires a strict sequential assembly of Cdc6 and Cdt1 onto chromatin in *Xenopus* egg extracts. *Nucleic Acids Res* 33:765–775
- Waga S, Zembutsu A (2006) Dynamics of DNA binding of replication initiation proteins during de novo formation of pre-replicative complexes in *Xenopus* egg extracts. *J Biol Chem* 281:10926–10934
- 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* 285:12469–12473
- Wohlschlegel JA, Dwyer BT, Dhar SK, Cvetic C, Walter JC, Dutta A (2000) Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. *Science* 290:2309–2312
- Wolf E, Kim PS, Berger B (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci* 6:1179–1189
- Wong PG, Glozak MA, Cao TV, Vaziri C, Seto E, Alexandrow M (2010) Chromatin unfolding by Cdt1 regulates MCM loading via opposing functions of HBO1 and HDAC11-geminin. *Cell Cycle* 9:4351–4363
- Xouri G, Squire A, Dimaki M, Geverts B, Verveer PJ, Taraviras S, Nishitani H, Houtsmuller AB, Bastiaens PI, Lygerou Z (2007) Cdt1 associates dynamically with chromatin throughout G1 and recruits Geminin onto chromatin. *EMBO J* 26:1303–1314
- Yanagi K, Mizuno T, You Z, Hanaoka F (2002) Mouse geminin inhibits not only Cdt1-MCM6 interactions but also a novel intrinsic Cdt1 DNA binding activity. *J Biol Chem* 277:40871–40880
- Yellajoshiyula D, Patterson ES, Elitt MS, Kroll KL (2011) Geminin promotes neural fate acquisition of embryonic stem cells by maintaining chromatin in an accessible and hyperacetylated state. *Proc Natl Acad Sci USA* 108:3294–3299
- Zhu W, Depamphilis ML (2009) Selective killing of cancer cells by suppression of geminin activity. *Cancer Res* 69:4870–4877
- Zhu W, Chen Y, Dutta A (2004) Rereplication by depletion of geminin is seen regardless of p53 status and activates a G2/M checkpoint. *Mol Cell Biol* 24:7140–7150

Chapter 6

MCM Structure and Mechanics: What We Have Learned from Archaeal MCM

Ian M. Slaymaker and Xiaojiang S. Chen

Abstract Minichromosome maintenance (MCM) complexes have been identified as the primary replicative helicases responsible for unwinding DNA for genome replication. This chapter discusses the current structural and functional understanding of MCMs and their role at origins of replication, which is based substantially on studies of MCM proteins complexes from archaeal genomes.

Keywords DNA replication • MCM • Hexameric and double hexameric helicase • Protein motor • Unwinding mechanism

6.1 Introduction

Eukaryotes and archaea employ a tightly regulated series of stepwise events to ensure a complete, high fidelity genome duplication event that occurs once, and only once, per cell cycle. Origins of replication spaced along chromosomes are engaged by pre-replication complexes (pre-RCs), which initiate DNA melting at the inception of S-phase. Two resulting replication forks originated from an origin travel in opposite directions, unwinding and copying DNA along the way.

The pre-RC is an ensemble complex consisting of the origin recognition complex (ORC), Cdc6, Cdt1 and the MCM helicase. MCM association with chromatin depends on the presence of all these factors (Donovan et al. 1997; Maiorano et al. 2000; Romanowski et al. 1996; Tanaka et al. 1997). Most components of the pre-RC are located specifically at potential origins and act locally to initiate firing.

I.M. Slaymaker • X.S. Chen (✉)
Molecular and Computational Biology, University of Southern California,
Los Angeles, CA 90089, USA
e-mail: Xiaojiang.chen@usc.edu

However, MCM complexes are unique in this respect, and are found broadly distributed along chromatin (Bailis and Forsburg 2003, 2004; Edwards et al. 2002b; Forsburg 2004; Kuipers et al. 2011; Liang et al. 1999; Pasion and Forsburg 2001; Tabancay and Forsburg 2006). Following the assembly of the pre-RC at the origin, additional factors, such as Cdc45 and GINS are incorporated into an intermediate pre-initiation complex (pre-IC) (Zou et al. 1997). The CDK and DDK protein kinases activate the Cdc45-MCM-GINS (CMG) complex, thought to be the core of the replication fork complex (Aparicio et al. 2006, 2009; Costa et al. 2011; Ives et al. 2010).

MCM proteins are AAA+(ATPases associated with diverse cellular activities) superfamily members that are known primarily as protein motors such as DNA helicases. MCMs unwind genomic double-stranded DNA (dsDNA) to expose single-stranded template during S-phase (Bailis and Forsburg 2004; Forsburg 2004; Pasion and Forsburg 2001; Tabancay and Forsburg 2006). Their function is not limited to unwinding DNA, however, as MCM has also been implicated in origin melting, genome repair and transcriptional regulation. Like all AAA+enzymes, MCM complexes hydrolyze ATP to fuel their substrate catalysis. Energy is transferred from the AAA+helicase motor to the central channel of the MCM complex to remodel the bound DNA substrate, splitting the DNA duplex ahead of the progressing replication fork (Labib et al. 2000). Cycles of ATP binding and hydrolysis drive MCM to translocate and unwind long stretches of DNA for the duplication of the entire genome.

Eukaryotic and archaeal MCM proteins likely evolved from a common ancestor with AAA+ core components very similar to modern eukaryotic replication machinery. From this common ancestor, eukaryotes evolved six MCM genes, the products of which form a ring shaped hetero-oligomeric MCM2-7 complex. Many archaea, however, evolved only a single MCM gene that produces a homo-oligomer with the same basic function as its hetero-hexameric counterpart. It should be noted that representatives of the recently characterized archaeal order *Methanococcales* have as many as eight MCM genes (Walters and Chong 2010). GINS, Cdc6, PCNA and ORC homologs are also found in several archaeal genomes, indicating an overall conserved system in both orders of life. For this reason, archaea are used as a model system to study replication.

Two archaeal MCM proteins, one from the thermophilic archaeon *Sulfolobus solfataricus* (ssoMCM) and another from *Methanothermobacter thermautotrophicus* (mtMCM), represent the best studied to date. MCM crystal structures from these two archaea provide the vast majority of structural information available. In the last decade, three MCM crystal structures have been solved. In order of publication these are: an N-terminal fragment of mtMCM (N-mtMCM) (Fletcher et al. 2003), a similar N-terminal fragment of ssoMCM (N-ssoMCM) (Liu et al. 2008) and a 4.3 Å near full-length ssoMCM (FL-ssoMCM) (Brewster et al. 2008). Additionally, an inactive MCM homolog from the archaeon *Methanopyrus kandleri* (mkaMcm2) was solved (Bae et al. 2009). FL-ssoMCM and the inactive truncated mkaMcm2 in the inactive deleted form crystallized as monomers, providing valuable new information about subunit domain organization.

Electron micrograph (EM) reconstructions provide low resolution maps of full-length MCM hexamer and double hexamer forms (Adachi et al. 1997; Costa et al. 2006a, b, 2008, 2011; Gomez-Llorente et al. 2005). With all of this information taken together, an understanding of how MCM functions at the core of the replisome begins to emerge.

6.2 Complex Organization: Hexamers and Double Hexamers

MCM complexes in eukaryotes and archaea assemble into a variety of oligomeric arrangements including hexamers, heptamers and some larger oligomers. This observation is no longer surprising, as many AAA+ proteins are found to form polymorphic oligomers, including hexameric and heptameric rings. Hexameric MCM rings are the most commonly observed oligomer and represent the active helicase. These hexameric helicases in archaeal and eukaryotic DNA replication commonly associate in a head-to-head double hexamer configuration.

MCMs (archaeal and eukaryotic) elute from size exclusion columns at a molecular weight congruent with hexamers or double hexamers (Brewster and Chen 2010; Brewster et al. 2008, 2010; Chong et al. 2000; Fletcher et al. 2003, 2005; Forsburg 2004; Gambus et al. 2011; Gomez-Llorente et al. 2005; McGeoch et al. 2005; Remus et al. 2009; Yu et al. 2002). Closed circular rings of MCM are clearly visualized by electron micrographs, exhibiting primarily hexameric and double hexameric rings (Fig. 6.1). It should be noted that heptamers and double heptamers are also found in relative abundance as examined by EM (Costa et al. 2006b, 2008; Gomez-Llorente et al. 2005). Double hexameric architecture was first shown by the crystal structure of N-mtMCM; N-ssoMCM was later crystallized as a single hexamer. Both display planar rings (Fig. 6.1a). The double hexamer architecture was recently shown to be a conserved eukaryotic Mcm2-7 architecture (Evrin et al. 2009; Gambus et al. 2011; Remus et al. 2009). These results strongly suggest that MCMs exist as a double hexamer at the replication origins at the initiation of replication.

The MCM double hexamer parallels that of SV40 large T antigen (LTag), another AAA+ hexameric helicase for viral replication in eukaryotic systems (Sclafani et al. 2004). Like MCM, LTag forms both hexamers and head-to-head double hexamers in solution and on dsDNA (Cuesta et al. 2010; Valle et al. 2006). Double hexamers clearly play an essential role for LTag, since *in vivo* DNA replication is not permitted when hexamer-hexamer interactions are disrupted. However, LTag hexamers do retain helicase activity *in vitro*, however, albeit at a 15-fold lower rate. These observations are echoed in studies of MCM: site-directed mutagenesis disrupting the double hexamerization of mtMCM causes between 5- and 13-fold less unwinding activity (depending on assay conditions) than wild-type (Fletcher et al. 2003, 2005; Sclafani et al. 2004).

Further evidence for the double hexamer as an active oligomer at the origin is found in the organization of origins within archaea. Three origins of replication identified in *S. solfataricus* (see Chap. 4, this volume) have oppositely facing ORC homolog

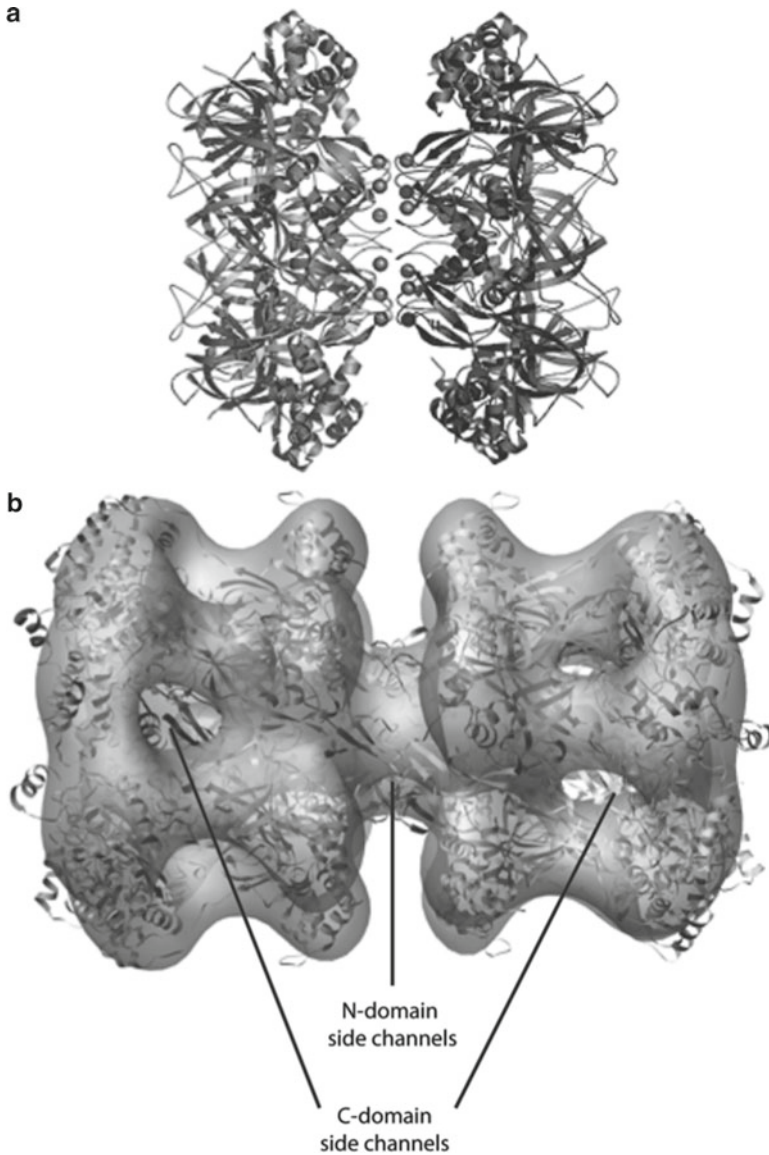


Fig. 6.1 The double hexamer architecture of two archaeal MCM complexes. **(a)** Double hexamer crystal structure of N-mtMCM (Adapted from Fletcher et al. (2003)). **(b)** ssoMCM full-length structure docked into an EM map with side channels labeled (Adapted from Brewster et al. (2008))

binding sites flanking an AT-rich duplex unwinding element (DUE) suggesting that hexamers are loaded in opposing directions i.e. head to head (Robinson et al. 2004). The length of the *Sulfolobus* DUE is ~65 base pairs, approximately the size of an MCM double hexamer (based on EM and crystal structure modeling), providing an

attractive model in which an MCM double hexamer is loaded directly onto dsDNA at the origin between two ORC proteins.

Research showing eukaryotic Mcm2-7 is loaded to origins of replication as double hexamers, provides convincing evidence for the biological significance of this oligomer (Evrin et al. 2009; Gambus et al. 2011; Remus et al. 2009). Mcm2-7 double hexamers reconstituted on dsDNA from purified yeast pre-RC components show that loading occurs in a head-to-head configuration similar to that of mtMCM (Remus et al. 2009). However, there was no indication that DNA had undergone any melting or unwinding in this study. If the *in vitro* pre-RC assembly thoroughly recapitulates *in vivo* MCM loading, this has significant implications for initiation and unwinding as it implies that origin melting occurs after MCM is loaded, not before or during the double hexamer loading/assembly at the origin. This is similar to the results obtained from the melting study for SV40 LTag system. Components involved in origin firing are not well understood and it is possible that the double hexamer contributes to origin melting in coordination with other pre-RC proteins, and once melted or at a certain stage of DNA replication (such as during elongation or approaching termination of replication), single hexamers carry out unwinding. Single molecule experiments demonstrated that *Xenopus* sister replisomes can function independently to replicate long stretches of DNA under the *in vitro* assay condition (Yardimci et al. 2010), which is consistent with the above hypothesis.

6.3 Helicase Activity

Several outstanding questions regarding MCM helicase mechanism during replication still remain to be addressed. First, how does MCM complex participate in melting origin DNA and generating viable replication forks? Secondly, what structural features are responsible for the physical separation of dsDNA into ssDNA, and where in the enzyme does this take place? Finally, how are the 6 subunits of an MCM hexamer coordinated to unwind DNA?

A strongly positively charged central channel runs straight from N- to C-terminal of MCM hexamers in MCM crystal structures (Fig. 6.2a), which is the primary binding site for DNA (Brewster et al. 2008; Fletcher et al. 2003; Liu et al. 2008). The length of the central channel in full-length ssoMCM hexamer is ~ 240 Å, sufficient for ~ 70 base pairs of straight B-form DNA (Brewster et al. 2008; Fletcher et al. 2003; Liu et al. 2008). The central channel diameter is sufficient to accommodate either ssDNA or dsDNA, thereby supporting a number of different unwinding models.

In addition to the primary central channel, side channels within the C domain are observed on the side-walls of the hexameric or double hexameric complexes (Figs. 6.1c and 6.2c). A side channel is found between each pair of monomers for a total of six side channels per hexamer (Brewster et al. 2008). The existence of positively charged side channels suggests a potential path through which DNA traverses during helicase activity (Fig. 6.2c). Additionally, within the double hexamer a

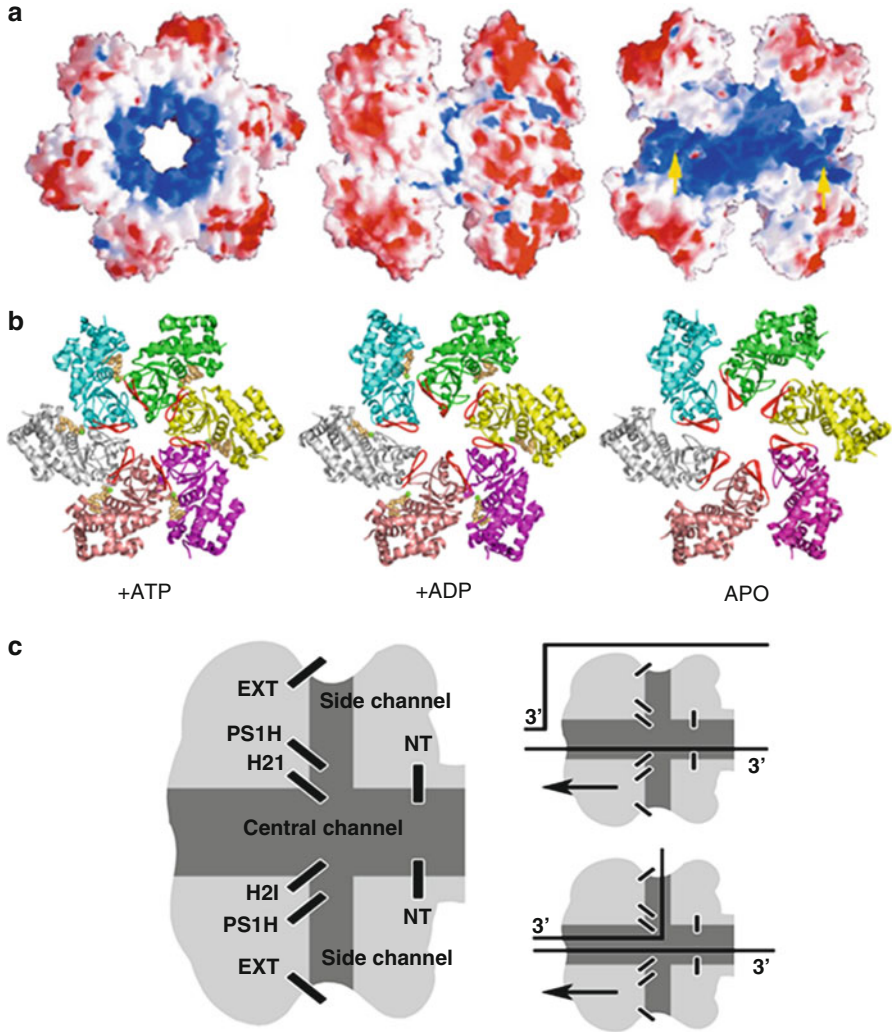


Fig. 6.2 The central channel features. (a) Electrostatics of N-mtMCM double hexamer structure showing positive central channel. *Yellow arrows* indicate the narrowest point (Adapted from Fletcher et al. (2003)). (b) LTag hexamer central channel narrowing as a result of nucleotide binding. Presensor-1 (PS1-hp) is colored *red* (Adapted from Brewster and Chen (2010)). (c) *Left*: cartoon of an MCM hexamer and the β -hairpins revealed by the crystal structure in the central channel. *Right*: *black lines* indicate possible pathways for DNA including or excluding side channels (Adapted from Brewster et al. (2008))

second set of N-terminal side channels is formed by the head-to-head interface of the double hexamer zinc binding domains (B-domain). It should be noted that the biological relevance for the obvious side channels has yet to be demonstrated.

Although MCM will bind nearly any DNA substrate, it can only unwind DNA with a forked end or a 3' overhang, and cannot unwind DNA with a 5' overhang or

blunt ended dsDNA (Bochman and Schwacha 2008; Ishimi 1997; Shin et al. 2003; You et al. 1999). This holds true for all isolated archaeal and eukaryotic MCM complexes to date. LTag requires identical substrate specifications for *in vitro* helicase activity in most cases, with one notable exception. If LTag is supplied with a specific origin sequence, it is capable of melting and unwinding long stretches (at least 1,000 bp) of blunt-end dsDNA from the middle portion of the dsDNA. Since origin DNA is presumably double-stranded prior to replication initiation, either MCM or another factor (or a combination of factors) must initially melt the duplex. Interestingly, if the N domain of ssoMCM is removed, the remaining C domain is capable of unwinding blunt-end dsDNA *in vitro* (Barry et al. 2007); however, the blunt-ended dsDNA tested was only 44 bp long, and it is unclear if much longer blunt-ended dsDNA can be unwound by this deletion mutant. It is possible that removing the N-terminal domain simulates an activation step or removes an inhibitory component for origin melting *in vivo*. However, no conditions have been found which permit wild-type MCM unwinding of blunt-end DNA *in vitro*. There is some *in vivo* evidence that MCM contributes to origin melting prior to S-phase. A *S. cerevisiae* *mcm5* mutant (*mcm5-bob1*) that bypasses a kinase checkpoint appears to melt DNA prior to S-phase (Geraghty et al. 2000). These observations raise the question: is MCM loaded onto dsDNA at origins and melted at a later stage, or does MCM load onto ssDNA that has been melted by other replication factors such as ORC and Cdc6? The *in vitro* reconstitution of a eukaryotic MCM double hexamer with no indication of origin melting (Remus et al. 2009), and failure to detect ssDNA in G1 arrested cells (Geraghty et al. 2000), seem to suggest the former.

Once origin DNA is melted, MCM assumes the role of helicase to unwind the replication forks. A number of unwinding models have been proposed to include the available data on MCM helicase activity. These are described below.

6.3.1 Steric Exclusion

The steric exclusion model proposes that the MCM helicase binds ssDNA and translocates away from the replication bubble, sterically excluding the opposite strand from the duplex. This model suggests that the active helicase acts as a single hexamer similar to DnaB family helicases found in prokaryotes and phages (Patel and Picha 2000). Evidence for the steric exclusion model is MCM's inability to unwind dsDNA without a 3' ssDNA overhang, suggesting that, at least *in vitro*, MCM threads onto ssDNA prior to unwinding.

If MCM is initially loaded at an origin as a double hexamer, this model implies the enzyme should split into two hexamers sometime after the replication initiation, with each hexamer reorganizing to encircle melted ssDNA strands and traveling in opposite directions along ssDNA. Support for this model comes from FRET data indicating that the 5' tail rapidly binds and unbinds the outer MCM surface during helicase activity (Rothenberg 2007).

The crystal structure of a related AAA+ papillomavirus E1 hexameric helicase provides evidence for the steric exclusion model for E1 helicase. In this structure ssDNA

is bound to the central channel, with each subunit making identical interactions with DNA (Enemark and Joshua-Tor 2006). In the steric exclusion model, this ssDNA in the central channel can be considered as the strand that would be translocating through the central channel, excluding the opposing strand. Central channel β -hairpins form a spiral staircase which tracks away from the ssDNA/dsDNA junction, pulling the ssDNA through the toroidal hexamer with cycles of ATP hydrolysis. There are several hairpins in the central channel that may play analogous roles (Brewster et al. 2008).

There are some inconsistencies with the steric exclusion model. The data showing that Mcm2-7 loads onto dsDNA double hexamer at origins (Remus et al. 2009) – and not ssDNA as this model would suggest – would require that MCM reorganizes once origin DNA is melted, switching from dsDNA bound to ssDNA bound. One possibility for this reorganization involves a Mcm2-7 “gate” between Mcm2 and Mcm5, which may open to allow ssDNA out of the central channel (Bochman and Schwacha 2010; Costa et al. 2011) (discussed in detail in Chap. 7, this volume). Similarly, archaeal MCMs are often observed to form broken rings. Displacement of the excluded DNA strand may also engage side channels to direct the newly unwound ssDNA away from MCM.

6.3.2 *Ploughshare*

The ploughshare model stipulates that a pinpointed force cleaves dsDNA into ssDNA within the central channel (Takahashi et al. 2005). MCM is loaded onto DNA in an inactive form (likely a double hexamer) by ORC and other replication factors and upon S-phase initiation, dsDNA is melted to ssDNA and the fork DNA interacts with MCM in such a way that a steric wedge, or ploughshare, is inserted at the ssDNA/dsDNA junction. By translocating along the genomic DNA, MCM pulls the ploughshare through the duplex dsDNA, cleaving it to ssDNA (Takahashi et al. 2005). If this model is accurate, a likely candidate for the ploughshare is the helix-2 insert (H2I, described below), which occupies and appears to dominate the central channel (Brewster et al. 2008) (Fig. 6.5).

6.3.3 *LTag Looping Model (or Strand Exclusion)*

Electron micrographs of LTag in the process of unwinding blunt-end origin containing dsDNA exhibit a curious particle shape suggesting that ssDNA is spooling out and away from the double hexamer as two loops (often referred to as ‘rabbit ears’) (Wessel et al. 1992). The crystal structure and EM structure of LTag –as well as that of archaeal MCM – reveal the presence of large side channels on the hexameric or double hexameric wall which may provide an avenue for ssDNA to spool away from the replication fork (Fig. 6.1b). A looping model has been proposed for LTag double-hexamer unwinding, in which dsDNA is pumped inside the double-hexamer,

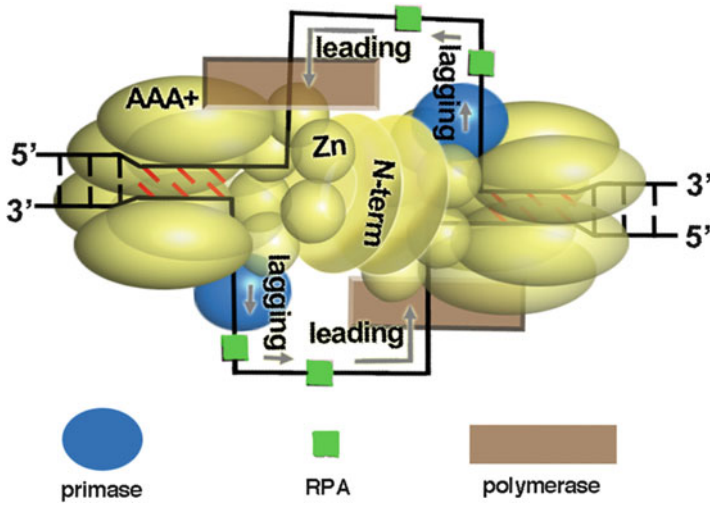


Fig. 6.3 The looping-model for bi-directional DNA unwinding by a double hexameric helicase, which may be in action at the initial origin melting and even the first phase of replication elongation. Two hexamers stay together through N-N interactions. The C-terminal AAA+ motor domains pump the dsDNA ahead of the fork into the double hexamer, extruding the separated ssDNA as loops. During replication, primase, polymerase, RPA, and other replication proteins dock on the side of the double hexamer near the side channels, capturing the emerging ssDNA loops as the template for the synthesis of the leading and lagging strands

and the separated ssDNAs inside each of the hexameric helicase are extruded through the side channels as loops (Fig. 6.3) (Gai et al. 2004, 2010; Li et al. 2003; Sclafani et al. 2004). The two growing DNA forks are held close together by the double-hexamer helicase while the ssDNA continues to be extruded from the side channel as loops, which can be captured and utilized by the primases and polymerases as template for daughter strand synthesis. The similarity of the double-hexamer architecture between LTag and MCM implies that MCM may function similarly by pulling or pumping dsDNA into the double hexamer which in turn extrudes ssDNA from its center.

Given LTag's predisposition for forming double hexamers, it is possible that single hexamers unwind DNA and the double hexamers found in 'rabbit ears' images are merely artifacts of *in vitro* unwinding. The persistence of double hexamers would present a problem for completing unwinding the circular genomes of archaea and SV40 virus because eventually the double hexamer would be unable to unwind the DNA when approaching replication termination. Alternatively, the double hexamer helicase may function at the origin melting stage, perhaps even for most of the elongation stage, and at termination, the double hexamer has to separate into less efficient single hexamers to wrap up replication of the entire genome.

Recent compelling *in vitro* evidence suggests that sister replisomes can split and travel in opposite directions, rather than pumping dsDNA through a central point of unwinding (Yardimci et al. 2010). Unlike *in vivo* conditions where the replication

machinery is believed to be anchored to the nuclear matrix, the replisome in this experiment was free in solution but the two dsDNA ends were anchored to prevent DNA from being pulled. Despite the anchoring, replication proceeded at normal rates, indicating that sister replisomes and MCM likely split and travel along DNA in this *in vitro* experiment. It is intriguing to find out at what stage(s) the double hexamer splits into two single hexamers in the form of functional replisomes after the origin melting and replication initiation *in vivo*.

6.3.4 Rotary Pump

The rotary pump model arose to explain the large abundance of MCM complexes on chromatin. This model suggests that MCM complexes translocate bidirectionally away from sites of loading and are anchored within replication factories (Laskey and Madine 2003). Once immobilized, multiple MCM complexes pump DNA through the central channel, rotating the dsDNA and introducing a negative twist that weakens the duplex. The key to this hypothesis is that two populations of MCM face opposite directions and untwist DNA in opposite directions, transferring the twist back to origins.

This model hinges on the observation that many MCM complexes are distributed along chromatin and speculation that they cooperatively unwind DNA. When MCM numbers are drastically reduced to only a single double hexamer per origin, replication occurs efficiently regardless (Edwards et al. 2002a). This indicates that only a subset of Mcm2-7 complexes bound to chromatin is essential for DNA unwinding and that this model may not be plausible.

6.4 Domains and Features of an MCM Subunit

Sequence similarities suggest all MCM proteins share a similar domain organization and can be separated into two major domains: the N-terminal domain (N domain) which is split into A, B and C subdomains, and the C-terminal domain (C domain) which contains the AAA+ core and at its extreme C-terminus, a small subdomain predicted to have a “winged helix” fold. The N- and C domains are connected by a flexible linker designated the N-C linker. Here we will discuss the important features of each domain, and how they contribute to the overall function of MCM complexes.

6.4.1 N Domain

MCM N domain sequences are poorly conserved between eukaryotic and archaeal MCMs, and even between MCM subunits within an organism. However, structure-based sequence alignments reveal a conservation of hydrophobic residues within

buried regions, and charged residues within the central channel (Fletcher et al. 2003). This suggests that although the primary sequence has mutated, the function and overall fold of the N domain remains consistent.

The function of the N domain is thought primarily to be regulatory: when this domain is deleted, the archaeal MCM (just the AAA+ motor and extreme C-terminal winged helix subdomain) is still capable of unwinding DNA (Barry et al. 2007, 2009). However, substrate specificity and processivity are lost, indicating that the N domain may be acting as a clamp to hold the AAA+ domain around dsDNA and prevent haphazard duplex unwinding. ssoMCM and mtMCM N domains (residues 1-268 and 2-286 respectively) have highly conserved structures and crystallized as hexamers with six-fold symmetry and a positively charged central channel (Fletcher et al. 2003; Liu et al. 2008). As previously discussed, only mtMCM crystallized as a head-to-head double hexamer.

Subdomain A is primarily composed of helices, forming a compact bundle which hangs off the outside of the hexamer (Fig. 6.4a) (Brewster et al. 2008; Fletcher et al. 2003). This subdomain is best known for its role in regulating MCM via kinases (Chen et al. 2005; Fletcher et al. 2003; Fletcher and Chen 2006; Geraghty et al. 2000; Hoang et al. 2007). A proline (Pro83) to leucine mutation within the A subdomain of yeast MCM5 bypasses a checkpoint mediated by Dbf4-dependent Cdc7 kinase (DDK), proposed to phosphorylate and promote proper assembly of the MCM complex (Fletcher and Chen 2006; Hoang et al. 2007). This proline aligns to a residue in mtMCM which mediates contact between subdomains A and C and promotes a rotation of subdomain A that is proposed to activate MCM (Chen et al. 2005; Fletcher et al. 2003; Fletcher and Chen 2006; Hoang et al. 2007). Mutation of the corresponding mtMCM proline (Pro62) to leucine causes a only slight shift of A subdomain due to the large leucine sidechain, however, helicase activity is decreased 14-fold confirming an conserved regulatory role for subdomain A (Fletcher et al. 2003; Fletcher and Chen 2006). This agrees with EM reconstructions suggesting this domain changes conformation (Chen et al. 2005; Fletcher and Chen 2006). The archaeal data suggest that the phosphorylation by DDK which targets Mcm4 just prior to replication initiation *in vivo* may induce a swing out of subdomain A in eukaryotic enzymes (Lei and Tye 2001).

Subdomain B is composed of three antiparallel β -sheets, which form a compact domain at the opposite protein end from the AAA+ motor (Fig. 6.4a). B subdomains of neighboring hexameric subunits are in close association with each other around the N-terminal end of the hexamer (Fletcher et al. 2003; Liu et al. 2008). Comparison between the N-ssoMCM and N-mtMCM hexamers revealed a rigid body “bowing in” of N-ssoMCM B subdomains, which narrows the central channel compared to N-mtMCM (Fletcher et al. 2003; Liu et al. 2008). This may indicate a flexibility of the B subdomain during helicase activity or promotion of oligomerization. The B subdomain coordinates a zinc at its tip using a $CX_2CX_nCX_2CX(C_4)$ motif that folds into an zinc binding domain (Fletcher et al. 2003). ssoMCM and mtMCM also engages a histidine to coordinate zinc, and in ssoMCM, this histidine replaces one of the cysteines (Fletcher et al. 2003; Liu et al. 2008). Mcm2-7 subunits have similar motifs capable of binding zinc within their N domains, suggesting a conserved function.

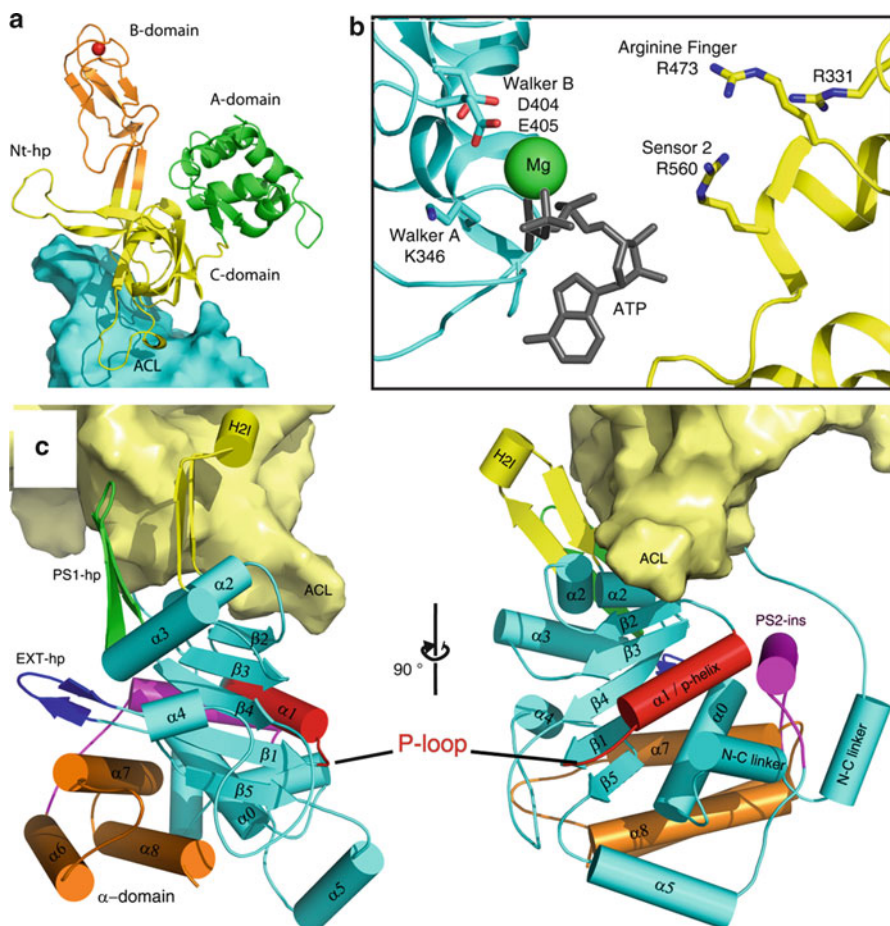


Fig. 6.4 Structural features of the N domain and at the nucleotide pocket of ssoMCM. (a) N domain features of ssoMCM (PDB: 3F9V). Blue solid represents C domain. (b) ATP pocket with modeled ATP (grey) and key residues labeled. (c) Detailed structural elements of the C domain of ssoMCM (PDB: 3F9V). The surface (cream color) represents the N domain and the ribbon diagram the C domain

Though little information is available for what this function is, the B-domain appears to be essential for double hexamerization (Fletcher et al. 2003). Mutation of His146 (found at the tip of the B subdomain) to alanine in ssoMCM severely affects DNA binding and helicase activity indicating an important, if unclear, role for the B subdomain (Fletcher et al. 2005; Gomez-Llorente et al. 2005; Sclafani et al. 2004). Clustering the B subdomain in hexamers may suggest that this mutation affects the enzymes ability to encircle DNA.

Subdomain C is the core of the N domain, and contains five antiparallel β -sheets that curl into a β -barrel like structure (Fig. 6.4a) (Brewster et al. 2008; Fletcher et al. 2003; Liu et al. 2008). The C subdomains harbor two particularly interesting

features: a positively charged β -hairpin which protrudes into the central channel (Nt-hp) and a loop known as the allosteric communication loop (ACL) (Fig. 6.4a, c) (Barry et al. 2009; Fletcher et al. 2003; Sakakibara et al. 2008). These features act in coordination to modulate the C domain through contacts with hairpins within the AAA+ core (Barry et al. 2009; Sakakibara et al. 2008). N domain crystal structures (both of mtMCM and ssoMCM) show that all six Nt-hps point into the central channel, contributing to the highly positive charge, and likely contacting DNA (Brewster et al. 2008; Fletcher et al. 2003; Liu et al. 2008). It is unlikely that this hairpin is essential for the physical splitting of the DNA duplex, as full-length MCM maintains helicase activity (although this is diminished) when it is deleted (Barry et al. 2009). These hairpins may track DNA through the central channel, in an analogous manner to those found in E1. Interestingly, the N domain fragment of mtMCM binds dsDNA only slightly less than full-length enzyme; however, when the two positively charged residues at the tip of the β -hairpin (Arg226 and Lys228 in mtMCM) are mutated to alanine, all DNA binding is abrogated (Fletcher et al. 2003). This indicates that the primary N domain feature interacting with DNA is likely the β -hairpin within the central channel. As mentioned previously, the B subdomain has some influence on DNA binding, but this may be via oligomeric interactions as opposed to DNA interactions. The C domain is connected to the AAA+ motor to via the N-C linker which extends from the β -barrel to the distal end of the C domain (Brewster et al. 2008). There is some indication that this linker is flexible, and permits AAA+ domain movement important for helicase activity (Brewster et al. 2008).

6.4.2 C Domain

The near full-length ssoMCM (FL-ssoMCM) and mkaMcm2 crystal structures revealed a typical organization of AAA+ features: the Walker A/B motif, P-loop and a five-parallel β -sheet core arranged in a canonical 51432 arrangement (Fig. 6.4c) (Bae et al. 2009; Brewster et al. 2008; Iyer et al. 2004). The C domain is often referred to as the motor domain, as it carries out the chemo-mechanical motion of MCM. Importantly, the C domain contains the ATP binding pocket.

6.4.2.1 ATP Binding Pocket

Two adjacent MCM monomers come together at the subunit interface to form a complete ATPase pocket (Fig. 6.4b) (Erzberger and Berger 2006; Iyer et al. 2004). Although neither a nucleotide-bound nor a multimer including the AAA+ domain structure has been solved, it is clear that a number of conserved residues contribute to the coordination and hydrolysis of ATP (Fig. 6.4b). Related AAA+ protein structures serve as templates for hypotheses regarding the organization of these residues within the MCM ATP pocket in lieu of a high resolution structure (Erzberger and

Berger 2006; Iyer et al. 2004). Like all AAA+ proteins, MCMs have a canonical P-loop motif bearing a conserved Walker A lysine (Fig. 6.4b). ATP is docked into the P-loop and a coordinated assembly of *cis* and *trans* subunits bind and catalyze the hydrolysis reaction. The Walker A lysine coordinates and binds the γ phosphate of ATP (Erzberger and Berger 2006; Iyer et al. 2004). The nearby Walker B motif coordinates Mg^{++} and water around the immobilized ATP molecule to catalyze hydrolysis of the γ phosphate, cleaving the ATP to ADP and phosphate. The Walker B glutamate in the sequence hhhhDE (where h is a hydrophobic residue) primes the water molecule for nucleophilic attack of the γ phosphate. The adjacent subunit projects a residue known as the arginine finger into the ATP pocket to stabilize and coordinate the associated nucleotide. If the arginine finger is mutated, MCM can no longer unwind DNA or hydrolyze ATP (Moreau et al. 2007). In the LTag ATP co-crystal structure (PDB: 1SVM) the arginine finger (R540) contacts with the γ phosphate of ATP (Gai et al. 2004). Once hydrolysis has occurred, R540 recedes from the LTag ATP pocket and away from the ADP (Gai et al. 2004; Li et al. 2003). Upon ATP binding within a LTag hexamer, a narrowing of the central channel (to 14 Å) ensues, and the next cycle of ATP hydrolysis again allows the arginine finger to recede from the ATP pocket. The global effect is an iris-like contraction and release mechanism, proposed to pump DNA and unwind duplex DNA within the central channel (Fig. 6.2b) (Gai et al. 2004). Hexamer models of MCM predict the arginine finger is positioned similarly, i.e. recessed from the ATP pocket as expected in the open-iris state (Brewster et al. 2008). It is possible that MCM utilizes ATP in a similar fashion, tightening the interface between monomers to pull itself along, or pull DNA through the central channel.

Additional highly conserved residues are essential for ATPase activity including sensor 1 and sensor 2 motifs. The sensor 1 motif is on $\beta 4$ of the AAA+ core sheets and bears an asparagine residue that coordinates the water molecule in conjunction with the Walker B glutamate (Fig. 6.4b) (Singleton et al. 2000; Story and Steitz 1992). The sensor 2 motif is on a bundle of helices ($\alpha 6$, $\alpha 7$ and $\alpha 8$) and contains a conserved arginine that coordinates and constrains the ATP within the pocket, usually interacting with the γ phosphate. Sensor 2 is atypically a *trans* residue in MCM proteins. Together all these residues transmit energy from hydrolyzed ATP to the central channel.

AAA+ proteins also transmit substrate information from the central channel to the ATP pocket (Mogni et al. 2009; Zhang and Wigley 2008). MCM's rate of hydrolysis notably increases with the addition of DNA (by 1.2-fold in ssoMCM) (Brewster et al. 2010; McGeoch et al. 2005). This indicates a direct line of communication from the DNA in the central channel to the ATPase active site. This occurs via a conserved AAA+ polar residue (T346 in ssoMCM) known as the "glutamate switch". The glutamate switch alters the position of the Walker B glutamate within the hhhhDE motif from an active conformation to an inactive conformation, triggered by substrate binding (Mogni et al. 2009; Zhang and Wigley 2008). When the glutamate switch is mutated, DNA no longer stimulates ATPase activity (Mogni et al. 2009; Zhang and Wigley 2008).

All C domain features discussed so far are common among AAA+ proteins, though specifically positioned in MCM (Erzberger and Berger 2006; Iyer et al. 2004).

Recently another essential conserved residue (R331 in ssoMCM) specific to MCMs was located on a hairpin (known as EXT-hp, see below) near the ATPase pocket (Brewster et al. 2008; Moreau et al. 2007). While the exact role of this residue remains unclear, it is near enough to the ATP pocket to physically influence the bound nucleotide (Brewster et al. 2010). This residue and the hairpin are further discussed below.

6.4.2.2 Hairpins, Helices and Inserts

MCM is a member of the Pre-sensor I superclade, shared with LTag, and the Pre-sensor II insertion clade, shared by the magnesium chelatase BChI. Proteins within these clades are characterized by a number of inserts and modifications to the basic AAA+core that sculpt the C domain into the helicase motor (Erzberger and Berger 2006; Iyer et al. 2004).

The PS1-hp is a long hairpin structure predicted to protrude into the central channel of the MCM hexamer (see Figs. 6.4c and 6.5) (Brewster et al. 2008; McGeoch et al. 2005; Moreau et al. 2007). A highly conserved lysine found at the tip of this hairpin is essential for MCMs helicase activity (McGeoch et al. 2005). Although unable to unwind DNA, the PS1-hp mutant MCM still binds dsDNA with wild-type affinity, suggesting the mutation uncouples DNA binding from unwinding and the PS1-hp is primarily involved in the unwinding mechanism (McGeoch et al. 2005). A homologous PS1-hp is found in the LTag central channel, bearing lysines and aromatic residues at its tip, all of which are essential for unwinding (Fig. 6.2b) (Gai et al. 2004). Importantly, ATP hydrolysis dramatically shifts the LTag hairpin 17 Å within the central channel during ATP hydrolysis, and accounts for the central channel iris narrowing (Gai et al. 2004). Although perhaps not utilizing the same mechanism as LTag, MCM likely harnesses a similar hairpin movement for unwinding. PS1-hp may have a specialized role in MCM, such as positioning DNA within the central channel, tracking DNA translocation similarly to E1, or prying the duplex open. PS1-hp may also work in conjunction with other features, such as the helix-2 insert (H2I), to cooperatively unwind DNA. Possibly, aspects of unwinding are delegated to each feature; for example, the H2I may disrupt the DNA duplex while the PS1-hp pulls the resulting strands apart.

The helix-2 insert (H2I) translates a portion of $\alpha 2$ into the central channel without interrupting the continuity of $\alpha 2$ hydrogen bonding, creating a hairpin structure with a piece of helix at its tip (Figs. 6.4c and 6.5) (Bae et al. 2009; Brewster et al. 2008; Erzberger and Berger 2006; Iyer et al. 2004). Like the PS1-hp, the H2I extends into the hexamer central channel where it likely interacts with substrate DNA (Brewster et al. 2008). H2I deletion completely abrogates helicase activity without compromising oligomerization and interestingly, significantly stimulates ssDNA and dsDNA binding (Jenkinson and Chong 2006). These mutational effects imply that the H2I is directly involved in unwinding the DNA duplex, but may also be forcing DNA into an energetically unfavorable conformation (Jenkinson and Chong 2006). In support of this, the H2I mutant, though having essentially wild-type basal

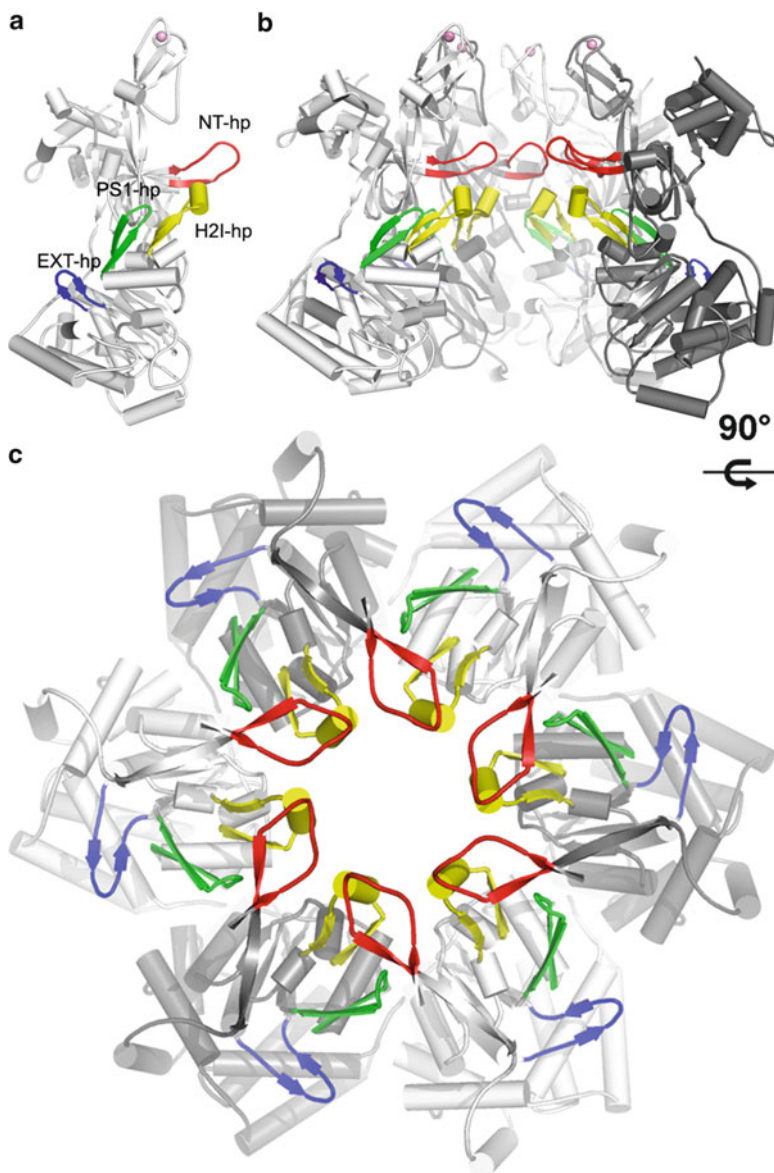


Fig. 6.5 The functionally important β -hairpins (in colors) of ssoMCM. (a) ssoMCM monomer (PDB: 3F9V) showing the β -hairpins from both N- and C domains. (b, c) The ssoMCM hexamer model showing the locations of the β -hairpins in the central channel viewing from the side and top of the hexamer (Adapted from Brewster and Chen (2010))

ATPase levels, has much higher level of DNA-stimulated hydrolysis, suggesting that the enzymes energetic load, normally overcome by nucleotide hydrolysis, is removed and the enzyme is essentially “spinning its wheels”. This implies a significant shift

of the H2I within the central channel occurs during ATP hydrolysis, similar to the PS1-hp in LTag. Evidence for this shift is found in the H2I deletion in mtMCM; when ATP is present, a tryptophan on the H2I is more solvent exposed than in the nucleotide-free enzyme, suggesting a conformational change during helicase activity (Jenkinson and Chong 2006). The mutational data for PS1-hp and the H2I indicate that both are cooperatively engaged during helicase activity and likely influence each other based on their positioning.

A hairpin predicted to be on the outer surface of the hexamer and below the side channels, is appropriately named the external hairpin (EXT-hp) (Figs. 6.4c and 6.5) (Brewster et al. 2008). This hairpin harbors the previously mentioned R331 residue involved in ATPase activity. A detailed analysis of each residue of the hairpin indicated that EXT-hp residues are also involved in DNA binding (Brewster et al. 2008, 2010). This could be interpreted as DNA threading through the side channels and engaging the EXT-hp. An alternate explanation is that mutations to the EXT-hp affect the ssoMCM subunit interface, which narrows or distorts the central channel, negatively influencing DNA binding.

Lastly, a long insertion (PS2-ins) between $\alpha 5$ and $\alpha 6$ reorganizes a canonical helical bundle domain (lid domain) found in PS1 superclade members (Fig. 6.4c) (Erzberger et al. 2006). The PS2-ins splits the canonical lid domain into an α helix bundle (α -domain consisting of $\alpha 6$, $\alpha 7$ and $\alpha 8$) and $\alpha 5$. $\alpha 5$ remains in its canonical location, but the helix bundle is repositioned to the opposite side of the subunit and away from the *cis* ATP pocket (Bae et al. 2009; Brewster et al. 2008). Importantly, this repositions the sensor 2 arginine on $\alpha 7$ into an atypical *trans* position (Moreau et al. 2007). The full-length ssoMCM crystal structure revealed that $\alpha 5$ is flanked by two long linkers, and predicted to associate with the adjacent monomers (Brewster et al. 2008). This may represent a key interface at the C terminal domain, and allow elasticity within the central channel.

The AAA+ core isolated from the N domain is still capable of unwinding DNA, albeit with much lower processivity and substrate specificity. Therefore this domain is considered to be the helicase motor that carries out the dynamic motion required for helicase activity. The PS1-hp and H2I appear to be intimately involved in the cleavage of duplex DNA and may function as the steric wedge in the ploughshare model, to redirect ssDNA in the steric exclusion model, or pull DNA through the central channel in the T-antigen model. Similar hairpins of LTag narrow and widen the central channel as ATP cycles through, making it likely that these hairpins are analogously involved in MCM dsDNA unwinding via force exerted by the H2I and the PS1-hp.

6.4.2.3 Winged Helix Domain

The wing helix domain is very small and is comprised of a few helices attached by a linker to $\alpha 8$. When this domain is removed from ssoMCM helicase activity is significantly stimulated, possibly indicating a regulatory role (Barry et al. 2007). This effect could be due to an influence on the attached $\alpha 6$, $\alpha 7$, $\alpha 8$ helix bundle which harbors the sensor 2 arginine. In eukaryotic MCM the winged helix domain

is responsible for interactions with Cdt1, a protein essential for Mcm2-7 loading (Khayrutdinov et al. 2009). An archaeal homolog of Cdt1 has not been discovered yet and it is curious that this domain is conserved between archaea and eukaryotes.

6.5 Inter- and Intra-Subunit Communication

The planar hexameric ring of MCM subunits comprises the active helicase unit. Within a hexamer, the interface between each pair of subunits harbors a single ATP active site, composed of *cis* and *trans* subunit contributions, for a total of six ATPase pockets per helicase unit. In the homohexameric archaeal helicase, each active site is equivalent, which is not the case in the Mcm2-7 complex, where each active site is unique (see Chap. 7, this volume, for details) (Bochman et al. 2008; Bochman and Schwacha 2008, 2010).

The coordination of hydrolysis within hexameric ring helicases fall into one of several types: concerted, sequential, probabilistic or semi-sequential (Crampton et al. 2006; Enemark and Joshua-Tor 2006; Gai et al. 2004; Martin et al. 2005; Singleton et al. 2000). LTag is thought to engage and hydrolyze ATP in each active site simultaneously, leading to the previously discussed iris-like contraction and relaxation cycles. This is observed in crystal structures containing nucleotides, in which all the ATP pockets are occupied (by ATP or ADP) or empty (Gai et al. 2004). Alternatively, ClpX subunits hydrolyze ATP probabilistically, suffering only a 1/6 reduction in enzyme efficiency from a single compromised active site (Martin et al. 2005). E1 and DnaB utilize a sequential mode of hydrolysis, in which each hydrolysis event is promoted by the neighboring ATPase active site (Donmez and Patel 2006; Enemark and Joshua-Tor 2006). Without DNA present, ssoMCM indicates a probabilistic mode of hydrolysis, meaning that hydrolysis cycles of a single subunit are not dictated by the neighboring subunits. However, during DNA unwinding MCM fits a model between sequential and random models, often termed semi-sequential (Moreau et al. 2007). This indicates that an individual subunit's ATP-dependent contribution to helicase activity is dependent on the conformation of other subunits in the complex. In short, the helicase can tolerate several inactive ATP pockets and still function.

A number of structural features at the subunit interface and within the central channel have been implicated in mediating subunit communication (Barry et al. 2009; Brewster et al. 2008; Moreau et al. 2007; Sakakibara et al. 2008). The ACL is so named for its role in regulating of the AAA+ domain via the N domain (Fig. 6.4a). The ACL function was initially recognized in mtMCM and subsequently validated in ssoMCM (Barry et al. 2009; Sakakibara et al. 2008). When deleted, the ACL mutant retains no helicase activity. This is curious since the AAA+ domain in isolation retains helicase activity. The missing ACL must somehow inhibit the helicase mechanism. It was subsequently shown that adding an Nt-hp deletion to the Δ ACL mutant restores helicase activity. The ACL is therefore proposed to play a role in properly positioning the Nt-hp within the central channel, and a mis-positioned

Nt-hp is implicated for helicase inhibition of the Δ ACL mutant (Barry et al. 2009). It seems likely then that the Nt-hp is performing a regulatory role through structural features identified as essential for helicase activity, possibly the PS1-hp and the H2I. The modeled hexamer of ssoMCM indicates that these features may in fact be in near each other in the central channel, offering support for this hypothesis. The N domain influences the H2I position in the central channel via an arginine that supports the ACL (R98 in mtMCM, R110 in ssoMCM). Mutation of this residue influences the solvent exposure of the H2I tryptophan similarly to ATP hydrolysis, indicative of a conformational change of H2I within the central channel. Although the data is limited, this may support the idea that the ACL regulates the central channel hairpin positioning (Barry et al. 2009; Sakakibara et al. 2008). Considered with the semi-sequential subunit communication, subunits may alternate between “on” and “off” conformations, inducing neighboring monomers to adopt an anti-conformation.

6.6 Higher-Order MCM Oligomers

Besides the hexamer and double hexamer, larger MCM oligomers have also been observed by EM, including heptamers, open rings, and filaments (Chen et al. 2005; Costa and Onesti 2009; Costa et al. 2006a, b, 2008; Gomez-Llorente et al. 2005; Yu et al. 2002). The largest of these oligomers are filaments, which were first observed in mtMCM EM preparations, and have been recognized *in vitro* for multiple archaeal species with varying length and dimensions (Chen et al. 2005; Costa et al. 2008). MCM filaments appear to depend on dsDNA, as purification that removes DNA also prevents filament formation (Costa et al. 2006b, 2008). This may offer an explanation for how the cell responds to exposed dsDNA in G1, and the sequence of events leading up to replication initiation. Chromosome fluorescent imaging shows that MCM is not focused at origins of replication but is instead liberally distributed along the entirety of the chromosome (Kuipers et al. 2011). MCM loading is clearly dependent on ORC, yet MCMs are found distal to ORC binding sites in large quantities (Edwards et al. 2002a; Takahashi et al. 2005). It is possible that ORC acts as a nucleation point for MCM filament extension. Recent work has shown that purified *Drosophila* Mcm2-7 fluctuates between a spiral “lock washer” state and a planar notched state, the second which has been observed in archaeal MCM (Costa et al. 2011). Inclusion of GINS and Cdc45 induce a planar notched state of Mcm2-7, and addition of ATP closes the gap, sealing Mcm2-7 into the closed hexamer. From the filament state, inducing a planar conformation may collapse the filament to a hexamer, effectively loading MCM to origin. This would suggest that the aforementioned “lock washer” state is an intermediate between the filament and the loaded hexamer. Alternatively, MCM filaments may function similarly to the bacterial initiator DnaA, proposed to wrap dsDNA around its outer surface as a filament, twisting the DNA in a way that transfers energy to a particular dsDNA region for duplex melting. Interestingly, mtMCM hexamers were reported to wrap dsDNA around their outer surface, inducing a 90° bend to the DNA (Costa et al. 2008).

6.7 Conclusions

Archaeal MCM research is in the process of revealing the intricate details of the MCM mechanism. Primary sequence conservation and similar complex organization with eukaryotic Mcm2-7 strongly suggests that archaeal MCM's mechanistic principals are transferable across orders of life. For this reason, archaeal enzymes will continue to be the focus of scientific interest and a source of significant replication research.

References

- Adachi Y, Usukura J, Yanagida M (1997) A globular complex formation by Nda1 and the other five members of the MCM protein family in fission yeast. *Genes Cells* 2:467–479
- Aparicio T, Ibarra A, Mendez J (2006) Cdc45-MCM-GINS, a new power player for DNA replication. *Cell Div* 1:18
- Aparicio T, Guillou E, Coloma J, Montoya G, Mendez J (2009) The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication. *Nucleic Acids Res* 37:2087–2095
- Bae B, Chen YH, Costa A, Onesti S, Brunzelle JS, Lin Y, Cann IK, Nair SK (2009) Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog. *Structure* 17:211–222
- Bailis JM, Forsburg SL (2003) It's all in the timing: linking S-phase to chromatin structure and chromosome dynamics. *Cell Cycle* 2:303–306
- Bailis JM, Forsburg SL (2004) MCM proteins: DNA damage, mutagenesis and repair. *Curr Opin Genet Dev* 14:17–21
- Barry ER, McGeoch AT, Kelman Z, Bell SD (2007) Archaeal MCM has separable processivity, substrate choice and helicase domains. *Nucleic Acids Res* 35:988–998
- Barry ER, Lovett JE, Costa A, Lea SM, Bell SD (2009) Intersubunit allosteric communication mediated by a conserved loop in the MCM helicase. *Proc Natl Acad Sci USA* 106:1051–1056
- Bochman ML, Schwacha A (2008) The Mcm2-7 complex has *in vitro* helicase activity. *Mol Cell* 31:287–293
- Bochman ML, Schwacha A (2010) The *Saccharomyces cerevisiae* Mcm6/2 and Mcm5/3 ATPase active sites contribute to the function of the putative Mcm2-7 'gate'. *Nucleic Acids Res* 38:6078–6088
- Bochman ML, Bell SP, Schwacha A (2008) Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability. *Mol Cell Biol* 28:5865–5873
- Brewster AS, Chen XS (2010) Insights into the MCM functional mechanism: lessons learned from the archaeal MCM complex. *Crit Rev Biochem Mol Biol* 45:243–256
- Brewster AS, Wang G, Yu X, Greenleaf WB, Carazo JM, Tjajadia M, Klein MG, Chen XS (2008) Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase. *Proc Natl Acad Sci USA* 105:20191–20196
- Brewster AS, Slaymaker IM, Afif SA, Chen XS (2010) Mutational analysis of an archaeal minichromosome maintenance protein exterior hairpin reveals critical residues for helicase activity and DNA binding. *BMC Mol Biol* 11:62
- Chen YJ, Yu X, Kasiviswanathan R, Shin JH, Kelman Z, Egelman EH (2005) Structural polymorphism of *Methanothermobacter thermoautotrophicus* MCM. *J Mol Biol* 346:389–394
- Chong JP, Hayashi MK, Simon MN, Xu RM, Stillman B (2000) A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. *Proc Natl Acad Sci USA* 97:1530–1535

- Costa A, Onesti S (2009) Structural biology of MCM helicases. *Crit Rev Biochem Mol Biol* 44:326–342
- Costa A, Pape T, van Heel M, Brick P, Patwardhan A, Onesti S (2006a) Structural basis of the *Methanothermobacter thermoautotrophicus* MCM helicase activity. *Nucleic Acids Res* 34:5829–5838
- Costa A, Pape T, van Heel M, Brick P, Patwardhan A, Onesti S (2006b) Structural studies of the archaeal MCM complex in different functional states. *J Struct Biol* 156:210–219
- Costa A, van Duinen G, Medagli B, Chong J, Sakakibara N, Kelman Z, Nair SK, Patwardhan A, Onesti S (2008) Cryo-electron microscopy reveals a novel DNA-binding site on the MCM helicase. *EMBO J* 27:2250–2258
- Costa A, Ilves I, Tamberg N, Petojevic T, Nogales E, Botchan MR, Berger JM (2011) The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol* 8:471–477
- Crampton DJ, Mukherjee S, Richardson CC (2006) DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. *Mol Cell* 21:165–174
- Cuesta I, Nunez-Ramirez R, Scheres SH, Gai D, Chen XS, Fanning E, Carazo JM (2010) Conformational rearrangements of SV40 large T antigen during early replication events. *J Mol Biol* 397:1276–1286
- Donmez I, Patel SS (2006) Mechanisms of a ring shaped helicase. *Nucleic Acids Res* 34:4216–4224
- Donovan S, Harwood J, Drury LS, Diffley JF (1997) Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proc Natl Acad Sci USA* 94:5611–5616
- Edwards MC, Tutter AV, Cvetcic C, Gilbert CH, Prokhorova TA, Walter JC (2002) MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in *Xenopus* egg extracts. *J Biol Chem* 277:33049–33057
- Enemark EJ, Joshua-Tor L (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* 442:270–275
- Erzberger JP, Berger JM (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct* 35:93–114
- Erzberger JP, Mott ML, Berger JM (2006) Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat Struct Mol Biol* 13:676–683
- 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 USA* 106:20240–20245
- Fletcher RJ, Chen XS (2006) Biochemical activities of the *BOB1* mutant in *Methanobacterium thermoautotrophicum* MCM. *Biochemistry* 45:462–467
- Fletcher RJ, Bishop BE, Leon RP, Sclafani RA, Ogata CM, Chen XS (2003) The structure and function of MCM from archaeal *M. thermoautotrophicum*. *Nat Struct Biol* 10:160–167
- Fletcher RJ, Shen J, Gomez-Llorente Y, Martin CS, Carazo JM, Chen XS (2005) Double hexamer disruption and biochemical activities of *Methanobacterium thermoautotrophicum* MCM. *J Biol Chem* 280:42405–42410
- Forsburg SL (2004) Eukaryotic MCM proteins: beyond replication initiation. *Microbiol Mol Biol Rev* 68:109–131
- Gai D, Zhao R, Li D, Finkielstein CV, Chen XS (2004) Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* 119:47–60
- Gai D, Chang YP, Chen XS (2010) Origin DNA melting and unwinding in DNA replication. *Curr Opin Struct Biol* 20:756–762
- Gambus A, Khoudoli GA, Jones RC, Blow JJ (2011) MCM2-7 form double hexamers at licensed origins in *xenopus* egg extract. *J Biol Chem* 286:11855–11864
- Geraghty DS, Ding M, Heintz NH, Pederson DS (2000) Premature structural changes at replication origins in a yeast minichromosome maintenance (MCM) mutant. *J Biol Chem* 275:18011–18021
- Gomez-Llorente Y, Fletcher RJ, Chen XS, Carazo JM, San Martin C (2005) Polymorphism and double hexamer structure in the archaeal minichromosome maintenance (MCM) helicase from *Methanobacterium thermoautotrophicum*. *J Biol Chem* 280:40909–40915

- Hoang ML, Leon RP, Pessoa-Brandao L, Hunt S, Raghuraman MK, Fangman WL, Brewer BJ, Sclafani RA (2007) Structural changes in Mcm5 protein bypass Cdc7-Dbf4 function and reduce replication origin efficiency in *Saccharomyces cerevisiae*. *Mol Cell Biol* 27:7594–7602
- Ilves I, Petojevic T, Pesavento JJ, Botchan MR (2010) Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell* 37:247–258
- Ishimi Y (1997) A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex. *J Biol Chem* 272:24508–24513
- Iyer LM, Leipe DD, Koonin EV, Aravind L (2004) Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol* 146:11–31
- Jenkinson ER, Chong JP (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 USA* 103:7613–7618
- Khayrutdinov BI, Bae WJ, Yun YM, Lee JH, Tsuyama T, Kim JJ, Hwang E, Ryu KS, Cheong HK, Cheong C, Ko JS, Enomoto T, Karplus PA, Guntert P, Tada S, Jeon YH, Cho Y (2009) Structure of the Cdt1 C-terminal domain: conservation of the winged helix fold in replication licensing factors. *Protein Sci* 18:2252–2264
- Kuipers MA, Stasevich TJ, Sasaki T, Wilson KA, Hazelwood KL, McNally JG, Davidson MW, Gilbert DM (2011) Highly stable loading of MCM proteins onto chromatin in living cells requires replication to unload. *J Cell Biol* 192:29–41
- Labib K, Tercero JA, Diffley JF (2000) Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science* 288:1643–1647
- Laskey RA, Madine MA (2003) A rotary pumping model for helicase function of MCM proteins at a distance from replication forks. *EMBO Rep* 4:26–30
- Lei M, Tye BK (2001) Initiating DNA synthesis: from recruiting to activating the MCM complex. *J Cell Sci* 114:1447–1454
- Li D, Zhao R, Lilyestrom W, Gai D, Zhang R, DeCaprio JA, Fanning E, Jochimiak A, Szakonyi G, Chen XS (2003) Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature* 423:512–518
- Liang DT, Hodson JA, Forsburg SL (1999) Reduced dosage of a single fission yeast MCM protein causes genetic instability and S-phase delay. *J Cell Sci* 112(Pt 4):559–567
- Liu W, Pucci B, Rossi M, Pisani FM, Ladenstein R (2008) Structural analysis of the *Sulfolobus solfataricus* MCM protein N-terminal domain. *Nucleic Acids Res* 36:3235–3243
- Maiorano D, Moreau J, Mechali M (2000) XCDT1 is required for the assembly of pre-replicative complexes in *Xenopus laevis*. *Nature* 404:622–625
- Martin A, Baker TA, Sauer RT (2005) Rebuilt AAA+ motors reveal operating principles for ATP-fuelled machines. *Nature* 437:1115–1120
- McGeoch AT, Trakselis MA, Laskey RA, Bell SD (2005) Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism. *Nat Struct Mol Biol* 12:756–762
- Mogni ME, Costa A, Ioannou C, Bell SD (2009) The glutamate switch is present in all seven clades of AAA+ protein. *Biochemistry* 48:8774–8775
- Moreau MJ, McGeoch AT, Lowe AR, Itzhaki LS, Bell SD (2007) ATPase site architecture and helicase mechanism of an archaeal MCM. *Mol Cell* 28:304–314
- Pasion SG, Forsburg SL (2001) Deconstructing a conserved protein family: the role of MCM proteins in eukaryotic DNA replication. *Genet Eng (N Y)* 23:129–155
- Pucci B, De Felice M, Rocco M, Esposito F, De Falco M, Esposito L, Rossi M, Pisani FM (2007) Modular organization of the *Sulfolobus solfataricus* mini-chromosome maintenance protein. *J Biol Chem* 282:12574–12582
- Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JF (2009) Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139:719–730
- Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, Bell SD (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* 116:25–38

- Romanowski P, Madine MA, Rowles A, Blow JJ, Laskey RA (1996) The *Xenopus* origin recognition complex is essential for DNA replication and MCM binding to chromatin. *Curr Biol* 6:1416–1425
- Sakakibara N, Kasiviswanathan R, Melamud E, Han M, Schwarz FP, Kelman Z (2008) Coupling of DNA binding and helicase activity is mediated by a conserved loop in the MCM protein. *Nucleic Acids Res* 36:1309–1320
- Scalfani RA, Fletcher RJ, Chen XS (2004) Two heads are better than one: regulation of DNA replication by hexameric helicases. *Genes Dev* 18:2039–2045
- Shin JH, Jiang Y, Grabowski B, Hurwitz J, Kelman Z (2003) Substrate requirements for duplex DNA translocation by the eukaryal and archaeal minichromosome maintenance helicases. *J Biol Chem* 278:49053–49062
- Singleton MR, Sawaya MR, Ellenberger T, Wigley DB (2000) Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* 101:589–600
- Story RM, Steitz TA (1992) Structure of the recA protein-ADP complex. *Nature* 355:374–376
- Tabancay AP Jr, Forsburg SL (2006) Eukaryotic DNA replication in a chromatin context. *Curr Top Dev Biol* 76:129–184
- Takahashi TS, Wigley DB, Walter JC (2005) Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. *Trends Biochem Sci* 30:437–444
- Tanaka T, Knapp D, Nasmyth K (1997) Loading of an MCM protein onto DNA replication origins is regulated by Cdc6p and CDKs. *Cell* 90:649–660
- Valle M, Chen XS, Donate LE, Fanning E, Carazo JM (2006) Structural basis for the cooperative assembly of large T antigen on the origin of replication. *J Mol Biol* 357:1295–1305
- Walters AD, Chong JP (2010) An archaeal order with multiple minichromosome maintenance genes. *Microbiology* 156:1405–1414
- Wessel R, Schweizer J, Stahl H (1992) Simian virus 40 T-antigen DNA helicase is a hexamer which forms a binary complex during bidirectional unwinding from the viral origin of DNA replication. *J Virol* 66:804–815
- Yardimci H, Loveland AB, Habuchi S, van Oijen AM, Walter JC (2010) Uncoupling of sister replisomes during eukaryotic DNA replication. *Mol Cell* 40:834–840
- You Z, Komamura Y, Ishimi Y (1999) Biochemical analysis of the intrinsic Mcm4-Mcm6-mcm7 DNA helicase activity. *Mol Cell Biol* 19:8003–8015
- Yu X, VanLoock MS, Poplawski A, Kelman Z, Xiang T, Tye BK, Egelman EH (2002) The *Methanobacterium thermoautotrophicum* MCM protein can form heptameric rings. *EMBO Rep* 3:792–797
- Zhang X, Wigley DB (2008) The ‘glutamate switch’ provides a link between ATPase activity and ligand binding in AAA+ proteins. *Nat Struct Mol Biol* 15:1223–1227
- Zou L, Mitchell J, Stillman B (1997) *CDC45*, a novel yeast gene that functions with the origin recognition complex and MCM proteins in initiation of DNA replication. *Mol Cell Biol* 17:553–563

Chapter 7

The Eukaryotic Mcm2-7 Replicative Helicase

Sriram Vijayraghavan and Anthony Schwacha

Abstract In eukaryotes, the Mcm2-7 complex forms the core of the replicative helicase – the molecular motor that uses ATP binding and hydrolysis to fuel the unwinding of double-stranded DNA at the replication fork. Although it is a toroidal hexameric helicase superficially resembling better-studied homohexameric helicases from prokaryotes and viruses, Mcm2-7 is the only known helicase formed from six unique and essential subunits. Recent biochemical and structural analyses of both Mcm2-7 and a higher-order complex containing additional activator proteins (the CMG complex) shed light on the reason behind this unique subunit assembly: whereas only a limited number of specific ATPase active sites are needed for DNA unwinding, one particular ATPase active site has evolved to form a reversible discontinuity (gate) in the toroidal complex. The activation of Mcm2-7 helicase during S-phase requires physical association of the accessory proteins Cdc45 and GINS; structural data suggest that these accessory factors activate DNA unwinding through closure of the Mcm2-7 gate. Moreover, studies capitalizing on advances in the biochemical reconstitution of eukaryotic DNA replication demonstrate that Mcm2-7 loads onto origins during initiation as a double hexamer, yet does not act as a double-stranded DNA pump during elongation.

Keywords Mcm2-7 • DNA replication • Gate model • AAA⁺ proteins • GINS • Cdc45 • Cell cycle

S. Vijayraghavan • A. Schwacha (✉)
Department of Biological Sciences, University of Pittsburgh,
Pittsburgh, PA 15260, USA
e-mail: schwacha@pitt.edu

7.1 Introduction

Although DNA is usually double-stranded, DNA polymerases universally require single-stranded DNA as a substrate. In most organisms, double-stranded DNA (dsDNA) is unwound into single-stranded DNA (ssDNA) by the replicative helicase, a molecular motor that uses ATP binding and hydrolysis to separate two DNA strands ahead of the polymerases at the replication fork. In eukaryotes, the catalytic core of this activity has been identified as Mcm2-7, a heterohexameric helicase that forms a toroidal complex uniquely generated from six distinct and essential subunits termed Mcm2 through Mcm7. This review focuses on recent structural and biochemical advances of this helicase; for additional background on Mcm2-7, the reader is directed to several earlier reviews (Bochman and Schwacha 2009; Forsburg 2004; Takahashi et al. 2005).

Despite the lack of a high-resolution Mcm2-7 structure, bioinformatics (Koonin 1993) and the crystal structure of the simpler but related homohexameric archaeal Mcm complex (this volume, Chap. 6; Brewster et al. 2008; Fletcher et al. 2003) confirm that Mcm2-7 is a AAA⁺ ATPase (ATPases associated with a variety of cellular activities; reviewed in Ogura and Wilkinson 2001). ATPase active sites in AAA⁺ proteins are formed at dimer interfaces. One subunit contributes specific *cis* motifs to the active site (e.g. Walker A and B), while the other subunit contributes specific *trans* motifs (e.g. arginine fingers) (Fig. 7.1a) (Erzberger and Berger 2006; Iyer et al. 2004; Koonin 1993). As each subunit participates in two different ATPase active sites, the study of isolated Mcm ATPase active site dimers has uncovered the basic subunit architecture of Mcm2-7 (Bochman et al. 2008; Costa et al. 2011; Davey et al. 2003). The six component ATPase active sites are named by the subunits that form them, and include the Mcm5/3, Mcm3/7, Mcm7/4, Mcm4/6, Mcm6/2, and Mcm2/5 active sites (Fig. 7.1b). Electron microscopy of the complex confirms that like most other replicative helicases and many AAA⁺ proteins, Mcm2-7 is toroidal with an open central channel (Biswas and Tsodikov 2008; Bochman and Schwacha 2007; Costa et al. 2011; Enemark and Joshua-Tor 2006; Gai et al. 2004; Pape et al. 2003; Patel and Picha 2000). Mcm2-7 presumably interacts with DNA within the central channel, and the observation that the dimensions of this central channel are sufficient to accommodate either ssDNA or dsDNA has inspired a variety of DNA unwinding models for Mcm2-7 (discussed below). Although the manner in which ATP binding and hydrolysis are coupled to DNA unwinding is unknown, sequence homology between Mcm2-7 and better studied AAA⁺ helicases suggests the involvement of several positively charged β -hairpin ‘fingers’ that protrude from each subunit into the central channel (Fig. 7.1c). Two particular fingers (pre-sensor 1 (PS1) hairpin and the helix-2 insert (H2I) hairpin) are contained within the C-terminal AAA⁺ domain, and are likely involved in coupling ATP binding and hydrolysis to DNA manipulation (Brewster et al. 2008; Jenkinson and Chong 2006) (Fig. 7.1c). In particular, studies with related helicases (Brewster et al. 2010; Gai et al. 2004) indicate that the PS1 hairpin undergoes a striking (~ 17 Å) conformational change in the presence of ATP (Gai et al. 2004).

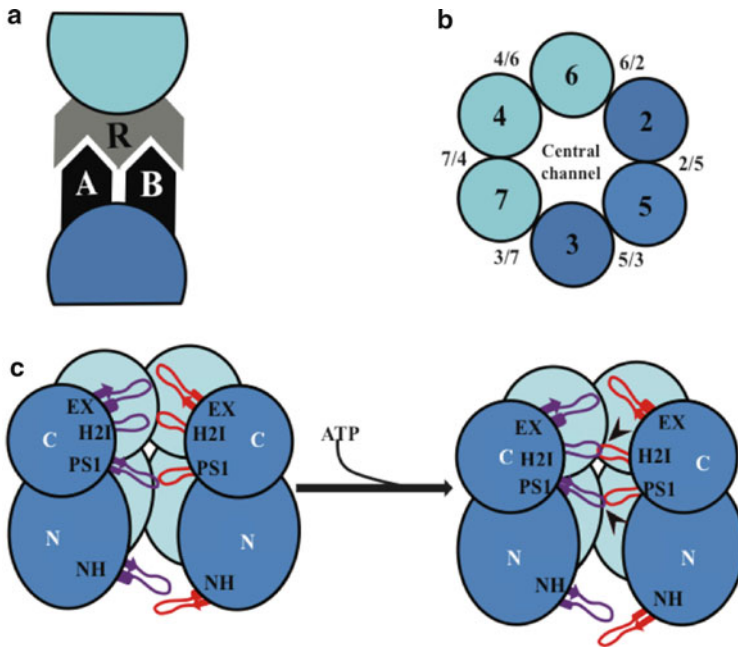


Fig. 7.1 Subunit organization of Mcm2-7. (a) AAA+ active sites form in *trans*. (b) Subunit architecture of Mcm2-7. (c) Cross-sectional view through the Mcm complex showing the distribution of β -hairpins around the central channel and the conformational rearrangement of H2I and PS1 following ATP binding (*arrowheads*, discussed in Brewster et al. 2008; Jenkinson and Chong 2006). Mcm subunits are depicted with N- and C-terminal domains. *EX* external hairpins, *PS1* presensor-1 hairpin, *H2I* helix-2 insert hairpin, *NH* N-terminal hairpin, *A* Walker A, *B* Walker B, and *R* arginine finger

As both the molecular motor driving replication fork progression and a major regulatory focus, Mcm2-7 interacts with many additional proteins. Although a detailed understanding of these accessory factors is lacking, sufficient evidence exists to classify these interactions into the groups listed below; the known genetic and physical interactions of these proteins with Mcm2-7 are summarized in Table 7.1.

- Proteins involved with loading Mcm2-7 onto replication origins, including ORC (origin recognition complex), Cdc6 and particularly Cdt1, which binds Mcm2-7 through its C-terminal domain (Ferenbach et al. 2005; Forsburg 2004; Wei et al. 2010).
- Proteins that regulate Mcm2-7 helicase activity – likely in addition to various kinases (e.g., Cdc7/Dbf4, CDK and Mec1) (Bruck and Kaplan 2009; Chuang et al. 2009; Liku et al. 2005; Randell et al. 2010; Sheu and Stillman 2010), this group includes Cdc45 and the GINS complex which are the presumed activators of Mcm2-7 (discussed below). In addition, various replication checkpoint proteins (Mrc1/Tof1/Csm3) that are likely involved in the inactivation of Mcm2-7 in the presence of collapsed replication forks or other types of DNA damage

Table 7.1 Summary of Mcm2-7 interactions

Interacting factors	Mcm2-7 subunit involved	Type ^a	Assay method ^b	Selected references
<i>Initiation factors</i>				
Orc1-6	2,4,5,7	G	SPA	(Loo et al. 1995)
Cdc6	2,5,7	P	TH,IP,BRC	(Randell et al. 2006; Seki and Diffley 2000)
Cdt1	2,6	P	IP,AC,EM	(Kawasaki et al. 2006; Remus et al. 2009)
Mcm10	2,6,7	G,P	TH,AC,SPA	(Homesley et al. 2000; Im et al. 2009)
<i>Elongation factors</i>				
GINS (Psf1-3, Sld5)	3,5	P	EM,IP,AC	(Costa et al. 2011; Gambus et al. 2006; Remus et al. 2009)
Cdc45	2,5	G, P	SPA,EM, IP,AC	(Costa et al. 2011; Ilves et al. 2010; Im et al. 2009)
Ctf4	4	G, P	AC,SPA	(Gambus et al. 2009)
Mrc1	2,4,6	G, P	IP,SPA	(Lou et al. 2008; Nedelcheva et al. 2005)
RPA1-3	3,4,5,6,7	P	IP	(Nakaya et al. 2010)
<i>Regulatory kinases</i>				
CDK (Cdc28)	3	P	BRC	(Liku et al. 2005; Tanaka et al. 2007a)
DDK (Cdc7/ Dbf4)	2-7	G,P	SPA,BA,IP	(Hardy et al. 1997; Jares et al. 2000; Sheu and Stillman 2010)
Mec1	2,4,6	G,P	BRC,SPA	(Lee et al. 2010; Randell et al. 2010)
<i>Checkpoint factors</i>				
Mrc1	6	P	IP	(Komata et al. 2009)
Tof1	2,4,5	P	AC,MS,IP	(Nedelcheva et al. 2005; Numata et al. 2010)
Csm3	7	P	IP	(Numata et al. 2010)
Rad17	7	P	TH,IP	(Tsao et al. 2004)
<i>Chromatin factors</i>				
FACT (Spt16,Pob3)	2,4,6,7	P	MS,IP,GF	(Tan et al. 2010)
HBO1	2	P	TH	(Burke et al. 2001)
HistoneH3	2	P	AC	(Ishimi et al. 1996)
MCM-BP	7	P	AC,DC	(Nishiyama et al. 2011)
<i>Sister chromatid cohesion</i>				
Cohesin (Smc1,Smc3)	6,7	P	MS,IP	(Ryu et al. 2006)
<i>Cell cycle regulators</i>				
Rb	3,6,7	G, P	TH,IP	(Mukherjee et al. 2010; Numata et al. 2010; Sterner et al. 1998)

^ainteraction abbreviations: *P* physical, *G* genetic

^bassay method abbreviations: *AC* affinity capture, *BRC* biochemical reconstitution, *DC* density centrifugation, *EM* electron microscopy, *GF* gel filtration, *IP* immunoprecipitation, *MS* mass spectrometry, *SPA* synthetic phenotype/mutant analysis, *TH* two-hybrid analysis

(Komata et al. 2009; Lopes et al. 2001; Lou et al. 2008; Nedelcheva et al. 2005; Numata et al. 2010) can be placed in this class.

- Proteins that physically link Mcm2-7 to the rest of the replication fork, including Mcm10 and Ctf4 (which link Mcm2-7 to Pol α /primase) (Gambus et al. 2009; Ricke and Bielinsky 2004) and Mrc1 (that links Mcm2-7 to the leading strand DNA polymerase – Pol ϵ) (Lou et al. 2008).
- Proteins that provide additional functionality to the replication fork, including the chromatin remodeling complex FACT (Tan et al. 2010; Wittmeyer et al. 1999), and proteins involved in sister chromatid cohesion (Guillou et al. 2010; Ryu et al. 2006).

Although Mcm2-7 superficially resembles other well-characterized homohexameric helicases, biochemical and structural analysis of the complex has proved largely intractable until recently. This impasse was chiefly due to the difficulty in reconstituting biologically-relevant initiation and elongation reactions *in vitro*, an inability to observe DNA unwinding from purified Mcm2-7 *in vitro*, and a paucity of knowledge concerning the function of numerous essential replication factors that interact with Mcm2-7 but lack obvious orthologs in better-studied prokaryotic and viral replication systems. Although significant technical problems remain, recent advances in these areas (detailed below) have begun to crack the ‘Mcm problem’.

7.2 The ‘Mcm Problem’ and Nonequivalent ATPase Active Sites

The biochemical reconstitution of Mcm2-7 helicase activity hinged upon understanding why particular subunits within the complex inhibit helicase activity under typical *in vitro* conditions. Ongoing genomic analysis has revealed that all currently annotated eukaryotic genomes contain identifiable orthologs of each of the six corresponding *MCM* genes, which further underscores the importance of a heterohexameric organization (Iyer et al. 2004; Liu et al. 2009). Because replicative helicases from bacteria, viruses and archaea are homohexamers in which each ATPase active site is apparently capable of equal biochemical activity, the heterohexameric nature of Mcm2-7 begs a question – *why are six different subunits required?*

From initial biochemical investigation of Mcm2-7, it was inferred that individual Mcm subunits were functionally distinct. While the heterohexameric complex (Mcm2-7) lacked *in vitro* helicase activity, a specific hexameric Mcm subcomplex (Mcm467) was competent to unwind DNA (Bochman and Schwacha 2007; Davey et al. 2003; Ishimi 1997; Lee and Hurwitz 2000). Consistent with a potential inhibitory role for the Mcm2, Mcm3, and Mcm5 subunits, addition of either the Mcm5/3 dimer or Mcm2 monomer inhibits the DNA unwinding activity of Mcm467 (Sato et al. 2000). In the case of Mcm2, inhibition requires a functional ATPase active site, because Mcm2 containing a mutation in the Walker A motif (*mcm2KA*) binds to Mcm467 but does not inhibit its DNA unwinding activity (Stead et al. 2009). Moreover, further analysis of Mcm467 indicates that helicase activity only depends upon the

Mcm7/4 ATPase active site, as a dimer of Mcm4 and Mcm7 can unwind DNA without Mcm6 (Kanter et al. 2008). These results present a paradox, because considerable *in vivo* data indicate that all six subunits are coordinately required for both initiation and elongation (Aparicio et al. 1997; Kubota et al. 2003; Labib et al. 2000; Leon et al. 2008).

Mutational analysis verifies the non-equivalence of Mcm2-7 ATPase active sites. Studies of ATP hydrolysis and ATP-dependent ssDNA binding of *S. cerevisiae* Mcm2-7 complexes containing mutations in individual ATPase active site motifs (Walker A, Walker B, and arginine finger), indicates that both biochemical activities depend upon the Mcm3/7 and Mcm7/4 active sites. In contrast, the remaining sites (largely involving Mcm2, Mcm3 and Mcm5) contribute little to either activity (reviewed in Bochman and Schwacha 2009). In addition, analysis of Mcm dimers corresponding to individual ATPase active sites (see below) indicates that the Mcm3/7 and Mcm7/4 active sites account for the bulk ATP hydrolysis activity of Mcm2-7 complex (Bochman et al. 2008; Davey et al. 2003). Although the relative functional contribution of the individual Mcm ATPase active sites may be somewhat different in metazoans (Ilves et al. 2010), results for *S. cerevisiae* divide the Mcm subunits into two functional groups: those needed for DNA unwinding (Mcm4, Mcm7, and possibly Mcm6), and those that possibly serve an essential regulatory role (Mcm2, Mcm3 and Mcm5).

The subunit architecture of Mcm2-7 further underscores the nonequivalence of ATPase active sites. *S. cerevisiae* Mcm2-7 active site dimers can be formed and studied either by reconstitution from purified subunits (Davey et al. 2003) or by purification of co-expressed subunits from baculovirus-infected insect cell lines (Bochman et al. 2008). Since most Mcm subunits are present in two different dimer combinations, the dimer association data defines the order of Mcm subunits within the intact heterohexamer (Fig. 7.1b). Because only five out of the six possible dimer combinations could be isolated, a literal interpretation of the data would suggest that the Mcm complex is linear, with a subunit order of Mcm5-3-7-4-6-2. Conversely, electron microscopy of Mcm2-7 (Sato et al. 2000), as well as analogy to better-studied AAA⁺ proteins (Hanson and Whiteheart 2005; Pape et al. 2003), indicates that Mcm2-7 forms a toroidal structure. To accommodate this data, Mcm2 and Mcm5 must be adjacent to each other in the complex. However, because a stable Mcm2/5 dimer could not be isolated, the dimer interface between them must be substantially weaker than those of the other Mcm active sites. The predicted juxtaposition of Mcm2 and Mcm5 within the *Drosophila* Mcm2-7 has been subsequently confirmed by electron microscopy (Costa et al. 2011).

7.3 Discovery of Mcm2-7 Helicase Activity and the Mcm2/5 Gate

The conundrum that Mcm467 had helicase activity, while Mcm2-7 did not, hindered significant biochemical advances in this area for nearly a decade. Two groups solved this problem using different approaches, and in the process they discovered a likely function for the Mcm regulatory subunits.

7.3.1 Differences in Circular ssDNA Binding Between Mcm2-7 and Mcm467

To search for clues that explain the difference in helicase activity between Mcm2-7 and Mcm467, their ATP-dependent ssDNA binding activities were compared. Analysis of recombinant Mcm complexes purified from baculovirus-infected insect cells indicated that they have similar affinities for *linear* ss- or dsDNA (Bochman and Schwacha 2007). However, these two complexes differ in their ability to bind *circular* ssDNA. Whereas hexameric Mcm467 (dimer of Mcm467 trimers) binds circular ssDNA poorly (Bochman and Schwacha 2008; Ma et al. 2010), Mcm2-7 binds to circular ssDNA depending upon the order of ATP addition to the reaction (Bochman and Schwacha 2007, 2008). If Mcm2-7 and circular ssDNA are pre-incubated together, subsequent addition of ATP results in a high extent of circular ssDNA binding. In contrast, if Mcm2-7 and ATP are pre-incubated together, followed by addition of circular ssDNA, Mcm2-7 binds circular ssDNA poorly. The intrinsic DNA binding properties of Mcm2-7 do not vary under these conditions, as linear DNA substrates generated from the circular substrate bind Mcm2-7 to the same extent regardless of the order of ATP addition. Because hexameric helicases in general bind DNA within the central channel (Patel and Picha 2000) and the archaeal Mcm complex has been specifically shown to bind in this manner (Fletcher et al. 2003; Liu et al. 2008; McGeoch et al. 2005), a topological model was proposed. Mcm467 was postulated to have a closed toroidal structure that prevents the passage of circular DNA into the central channel, while Mcm2-7 has an accessible central channel that could be closed by an ATP-dependent conformational change.

Mutational analysis of Mcm2-7 demonstrated that the Mcm2/5 ATPase active site was responsible for the differences in circular ssDNA binding. Except for certain mutations in the Mcm3/7 or Mcm7/4 active sites, complexes containing mutations in conserved ATPase active site motifs in the remaining active sites only lead to minor defects in linear ssDNA binding (Bochman and Schwacha 2007, 2010). Mcm2-7 complexes containing mutations in the remaining active sites were tested for circular ssDNA binding. Although some small effects were observed for certain mutations in the Mcm6/2 and Mcm5/3 active sites (Bochman and Schwacha 2010), two independent mutations in the 2/5 active site – a Walker A mutation in *MCM5* (*mcm5KA*) and an arginine finger mutation in *MCM2* (*mcm2RA*) – abolished the ATP pre-incubation effect on circular ssDNA binding (Bochman and Schwacha 2007, 2010). Their activities in this assay were, however, completely opposite – the *mcm5KA* mutation causes Mcm2-7 to bind circular ssDNA in both the presence and absence of ATP pre-incubation, while the *mcm2RA* mutation blocks the ability of Mcm2-7 to bind circular ssDNA under either condition. Two additional points should be noted: firstly, in a hexameric helicase such as the SV40 large T-antigen, oligomerization into a hexamer requires ATP (Wang and Prives 1991) and secondly, although specific AAA⁺ proteins can vary, Walker A mutations tend to block both ATP binding and hydrolysis, while arginine finger mutations tend to affect hydrolysis more than ATP binding (Hanson and Whiteheart 2005). Taken together, these results directly imply that ATP binding to the Mcm2/5 active site functions as a

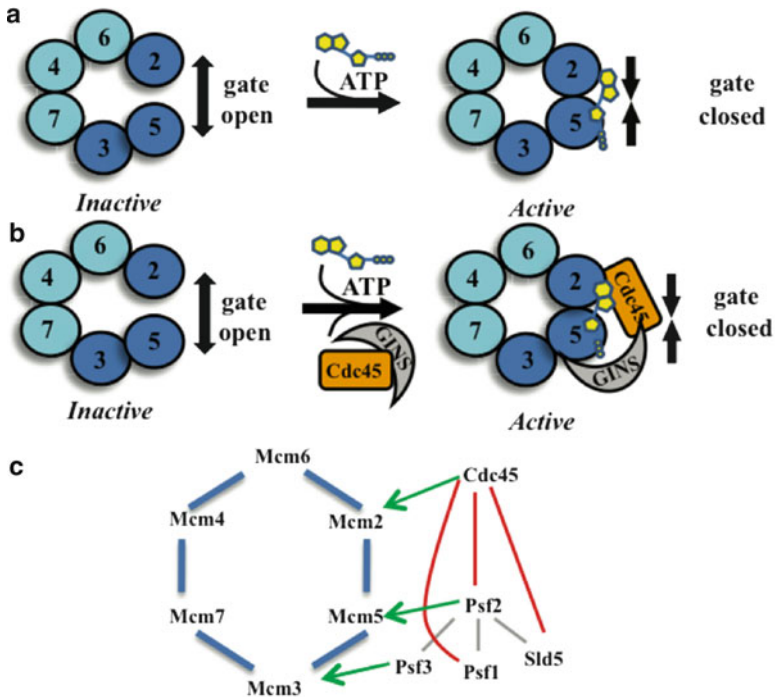


Fig. 7.2 The gate model for Mcm2-7 complex activation. (a) Original model of the Mcm2/5 gate. (b) Current gate model showing the interaction of Cdc45 and GINS with Mcm2 and Mcm5 to switch the complex from the open to closed form. (c) Summary of physical interactions among Mcm2-7, Cdc45 and members of the GINS complex (see text for details)

‘switch’ governing the transition between open and closed conformations of the Mcm2-7 complex (Fig. 7.2a).

7.3.2 An *In Vitro* Condition That ‘Closes’ Mcm2-7 Stimulates Its Helicase Activity

Hexameric helicases are believed to unwind dsDNA as a closed toroid to facilitate a wave of conformational changes around the ring that drives DNA unwinding (Erzberger and Berger 2006). This suggests that perhaps the reason why Mcm2-7 lacks DNA unwinding activity *in vitro* is because it is predominantly in an open conformation, and thus unable to propagate the necessary conformational changes. Although the simple experiment to pre-incubate Mcm2-7 with ATP prior to addition of a DNA helicase substrate failed to stimulate Mcm2-7 helicase activity (Bochman and Schwacha 2008), other possibilities were considered. One possibility was that ions present in the buffer system were preferentially stabilizing the open conformation. Biochemical analysis ultimately identified that either acetate or glutamate

support the helicase activity of the *S. cerevisiae* complex, while the use of chloride in the initial system poisoned this activity (Bochman and Schwacha 2008). Interestingly, pre-incubation of Mcm2-7 with glutamate preferentially reduces circular ssDNA binding relative to linear ssDNA binding (Bochman and Schwacha 2008), mimicking the pre-incubation effect with ATP and consistent with the notion that the closed form of Mcm2-7 is required for helicase activity.

7.3.3 *The Mcm2/5 ‘Gate’ Model – The Open Conformation and DNA Unwinding Are Mutually Exclusive*

The reconstitution of Mcm2-7 helicase activity confirmed its identity as the eukaryotic replicative helicase, and provided an important glimpse into the function of the Mcm regulatory sites. Combined with the subunit association studies indicating that the Mcm2/5 active site is structurally unstable (above), these results suggest that the Mcm2/5 site is indeed the site of the discontinuity or ‘gate’ in the toroid. The correlation between the closed conformation and helicase activity implies that the status of the gate (Mcm2/5 active site) regulates DNA unwinding in some manner. Because DNA replication necessitates the loading and unloading of Mcm2-7 from chromosomes, as well as the stimulation or repression of its helicase activity under different circumstances, the gate could be expected to have an important *in vivo* mechanistic and regulatory impact on Mcm2-7 (see below).

7.4 The CMG Complex

An alternative approach to reconstitute the *in vitro* helicase activity of Mcm2-7 was to identify replication factors that stimulate this activity. Although, in principle, any protein involved in either initiation or elongation could assume this role, interest soon centered on two specific candidates: Cdc45 (Hennessy et al. 1991; Moir et al. 1982; Zou et al. 1997) and the heterotetrameric GINS complex (Japanese for go-ichi-ni-san or 5-1-2-3, corresponding to Sld5, Psf1, Psf2, and Psf3, see Chap. 8) (Takayama et al. 2003). Neither replication factor has an apparent ortholog in prokaryotic or viral systems, consistent with the possibility that they might be required for the proposed unique role of activating the replicative helicase in eukaryotic DNA replication. Several critical experiments demonstrated the importance of these two factors in fork progression; similar to studies of Mcm2-7 (Labib et al. 2000), inactivation of either Cdc45 or members of the GINS complex (Sld5 and Psf1) during S-phase blocks ongoing fork progression (Takayama et al. 2003; Tercero et al. 2000). Likewise, immunodepletion of Cdc45 from the *Xenopus in vitro* replication system also blocks elongation (Pacek and Walter 2004). Finally, using appropriate affinity tags, a very large multi-protein complex was isolated from yeast, containing Cdc45 and the GINS complex in combination with Mcm2-7 and

other proteins (Gambus et al. 2006). Although this ‘replisome progression complex’ (Broderick and Nasheuer 2009; Liu et al. 2009) seemed a promising target for biochemical investigation, its poor yield precluded this possibility.

7.4.1 *Discovery of the CMG Complex*

Drosophila embryonic extracts proved to be a more fruitful source of higher-order complexes containing Mcm2-7. Following gentle purification of Cdc45 using both conventional and immunoaffinity methods, a complex was isolated that additionally contained Mcm2-7 and GINS; this was subsequently named the CMG complex (Cdc45/Mcm2-7/GINS) (Moyer et al. 2006). This complex has robust DNA unwinding activity, suggesting a strong stimulatory role for Cdc45 and GINS on Mcm2-7 (Moyer et al. 2006). However, because this system is poorly suited for generating mutant Mcm2-7 complexes, some doubt remained over whether the observed helicase activity was indeed Mcm2-7-specific.

To overcome this problem, a recombinant baculovirus system was developed that co-expressed all of the eleven CMG subunits in sufficient abundance to allow their detailed biochemical analysis (Ilves et al. 2010). To dispel the remaining doubts concerning Mcm2-7 dependence of the observed DNA unwinding, various mutant CMG complexes were generated and shown to lack DNA unwinding activity. This recombinant system also allowed the production of both Mcm2-7 and CMG for biochemical comparison. Although the *Drosophila* Mcm2-7 complex in isolation had limited helicase activity in the presence of glutamate, the CMG complex had much better activity. This observation validated the initial belief that these two auxiliary factors were, in fact, Mcm2-7 activators. Combined with the demonstration of a greatly elevated degree of ATP hydrolysis from the CMG complex, it could be surmised that that Cdc45 and GINS activate Mcm2-7 by stimulating its ATPase activity.

7.4.2 *CMG Structure – Cdc45 and GINS Close the Mcm2/5 Gate*

A comparative structural analysis of Mcm2-7 and the CMG complex by transmission electron microscopy and image averaging uncovered the potential basis of Mcm2-7 activation by Cdc45 and GINS (Costa et al. 2011). In the absence of GINS and Cdc45, Mcm2-7 preparations contain two subpopulations in likely equilibrium: approximately 28% are in a planar quasi-symmetric (notched) form; and 72% are present as a non-planar spiral (lock washer) form (Costa et al. 2011). Three-dimensional reconstruction to approximately 35 Å resolution revealed that both subpopulations contain a gap in the ring structure. Treatment of the Mcm2-7 preparation with the ATP transition state analog ADP.BeF₃ causes a small shift in the population from the lock washer form to the notched form (36% notched and ~65%

lock washer), weakly consistent with the previous prediction that the binding of ATP to the Mcm2/5 active site could either close the gate (increase the planar form) (Fig. 7.2a) or perhaps slow the interconversion of the two forms (Bochman and Schwacha 2007). Through addition of large epitope tags, the subunit architecture of Mcm2-7 that was previously deduced in *S. cerevisiae* was confirmed. Moreover, the gap in both subpopulations was shown to be flanked by Mcm2 and Mcm5 (Costa et al. 2011), consistent with the prior biochemical prediction of the Mcm2/5 gate (Bochman and Schwacha 2007, 2008).

Further structural analysis of the CMG complex demonstrated that Cdc45 and GINS bridge the Mcm2/5 gate (Costa et al. 2011) (Fig. 7.2b). The CMG complex structure reconstructed to a resolution of 28 Å consists of a uniform population of notched molecules. However in contrast to Mcm2-7, addition of ATP analogs uniformly and significantly narrows the Mcm2/5 gap within the population. These results suggest that in the presence of ATP, Cdc45 and GINS work together as a 'latch' to shut the Mcm2/5 gate (Fig. 7.2b) and, presumably activate Mcm2-7 helicase activity.

Analysis of the CMG complex containing epitope-tagged GINS subunits combined with available structural information on GINS (Chang et al. 2007; Choi et al. 2007) defined the physical contacts among Cdc45, Mcm2-7 and GINS (Costa et al. 2011). Although a prior study indicated that the GINS complex physically interacts with Mcm4 (Ilves et al. 2010), structural analysis of the CMG complex indicates that the Psf2 and Psf3 subunits of GINS make a series of intricate physical contacts with the N-terminal domains of both Mcm3 and Mcm5 (Fig. 7.2c). In the presence of ADP.BeF₃, conformational changes within Mcm2-7 extend these contacts to the C-terminal AAA⁺ domains of Mcm3 and Mcm5. Cdc45 has a similar series of intricate physical interactions with GINS (largely through the Sld5 and Psf2 subunits), Mcm2, and Mcm5 that position it to bridge the Mcm2/5 interface (Fig. 7.2c). The individual physical interactions among Mcm2-7, GINS and Cdc45 are likely to be weak, consistent with the previous observation that a simple addition of GINS and Cdc45 to Mcm2-7 does not reconstitute *in vitro* helicase activity (Moyer et al. 2006). Furthermore, these results may explain why additional essential replication factors (Sld2 and Sld3) are required *in vivo* for the loading of GINS and Cdc45 onto replication forks (Kamimura et al. 2001; Muramatsu et al. 2010).

7.4.3 Possible Regulation of the Mcm2/5 Gate

The surprisingly good correlation between biochemical (Bochman and Schwacha 2007, 2008) and structural data (Costa et al. 2011) from two independent groups puts the Mcm2/5 gate model on a firm footing, and simultaneously provides a compelling and testable model for the functional role of the Mcm2-7 regulatory subunits (Mcm2, 3, and 5), GINS, and Cdc45. Although no data currently validates the *in vivo* relevance of the Mcm2/5 gate, obvious roles for the gate in helicase loading during initiation, activation of helicase activity in G1/S, temporary deactivation of helicase activity during the checkpoint response, and unloading the complex during termination can

be readily imagined (reviewed in Bochman and Schwacha 2009). For any of these potential roles, there must exist some currently unknown way to regulate the opening or closing of the gate. Given that gate closure apparently requires ATP, Cdc45 and GINS, its regulation must focus on one of these attributes.

At least two general models to regulate the gate could be envisioned – one would be regulation of interactions among members of the CMG complex. Presumably, this happens in S-phase, as the loading of Cdc45 and GINS onto Mcm2-7 at the G1/S transition requires cyclin-dependent kinase (CDK) phosphorylation of Sld2 and Sld3, a critical step that ties DNA replication to cell cycle progression (Tanaka et al. 2007b; Zegerman and Diffley 2007). The unique ternary assembly of the CMG complex might facilitate such regulation, because disruption of any of the multiple weak interactions might be expected to disassemble the entire CMG complex and inactivate Mcm2-7. Alternatively, analogous to well-studied small GTPases that are regulated by proteins that affect nucleotide exchange (GEFs) or catalytic turnover (GAPs) (reviewed in (Bos et al. 2007)), the status of the Mcm2/5 gate could be similarly controlled. Although the ATP hydrolysis activity of Mcm2-7 is stimulated in combination with Cdc45 and GINS (Ilves et al. 2010), it is unclear if this is due to a GAP-like activity intrinsic to these activator proteins or simply because Cdc45 and GINS cause a conformational change in Mcm2-7 that reconstitutes a functional Mcm2/5 active site.

7.5 How Does Mcm2-7 Unwind DNA?

Despite the availability of high resolution structural data, there is still no consensus on the mechanism of DNA unwinding by hexameric helicases: compare, for example, unwinding models of SV40 large T-antigen (Gai et al. 2004) with those of bovine papillomavirus E1 (Enemark and Joshua-Tor 2006). To account for a plethora of *in vivo* and *in vitro* observations, various unwinding models have been proposed for Mcm2-7 (Takahashi et al. 2005). In general, these models can be classified as those in which Mcm2-7 interacts in its central channel either with dsDNA or with ssDNA. Even though most hexameric helicases are capable of sterically causing strand displacement by binding and translocating along a complementary strand *in vitro*, there is some debate if these *in vitro* observations reflect the true *in vivo* mechanism. Although a definitive conclusion has yet to be reached, several investigations using conditions that more closely reflect the *in vivo* state have considerably sharpened our understanding on how Mcm2-7 is loaded onto DNA during initiation and how it unwinds DNA during elongation.

7.5.1 Mcm2-7 Loads as Double Hexamers onto dsDNA

Recent biochemical reconstitution of early stages of replication initiation using purified components (origin DNA, ORC, Cdc6, and Cdt1) from either *S.cerevisiae* or *Xenopus* egg extracts, has facilitated studies of Mcm2-7 in complex with DNA

(Evrin et al. 2009; Gambus et al. 2011; Remus et al. 2009). Electron microscopy of these complexes demonstrates that Mcm2-7 binds origins of replication as double hexamers (Evrin et al. 2009; Remus et al. 2009). Furthermore, by additionally visualizing both protein and DNA, the Mcm2-7 double hexamers were shown to encircle dsDNA and not ssDNA (Evrin et al. 2009). Although this result was unexpected because the bacterial initiator protein DnaA melts origin DNA to a single stranded form prior to loading of the replicative helicase DnaB (Bramhill and Kornberg 1988), ORC binding has never been observed to lead to measurable DNA unwinding (Remu et al. 2009). These Mcm2-7–DNA complexes are resistant to dissociation by high salt, suggesting a topological association of Mcm2-7 with DNA, as has been extensively observed *in vivo* (Donovan et al. 1997). Consistent with a topological mode of DNA binding, the binding half life of Mcm2-7 on linear DNA is much shorter than on the corresponding circular substrate (half-life of 10 min vs. 60 min, respectively, see Remus et al. 2009).

Mcm2-7 double hexamers assemble in a head-to-head configuration, with the C-termini containing AAA⁺ ATPase domains directed outward (Remus et al. 2009). The archaeal MCM from *Sulfolobus solfataricus* adopts a similar orientation on forked DNA substrates, with the C-terminal directed towards the ssDNA/dsDNA junction (McGeoch et al. 2005). The observed double hexamers appear to be the physiologically relevant intermediates during initiation. Double hexamers have been reconstituted on DNA both in solution as well as in a potentially more concentrated state (DNA immobilized on beads), lending credence to the fact that double hexamerization is not an artifact resulting from forced interactions (Remus et al. 2009). Furthermore, the consistent observation of only double-hexameric Mcm2-7 complexes on DNA strongly suggests that Mcm2-7 hexamers are loaded concurrently rather than sequentially. Moreover, the failure to detect DNA distortion within the pre-replicative complex (pre-RC) supports the idea that Mcm2-7 loads onto dsDNA and that origin melting occurs downstream of Mcm2-7 loading.

7.5.2 *Single-Molecule Studies Eliminate the dsDNA Pump Model for Elongation*

At least *in vitro*, Mcm2-7 (and probably the CMG complex) likely unwinds DNA by tracking along a single DNA strand while sterically displacing the complementary strand (Bochman and Schwacha 2008; Costa et al. 2011). In contrast, Mcm2-7 double hexamers could unwind DNA via a dsDNA ‘pump’ mechanism. In this model, both hexamers remain functionally and physically coupled while pumping flanking dsDNA into the complex and releasing them between the two hexamers in a single-stranded form (Takahashi et al. 2005) (Fig. 7.3). SV40 large T-antigen is believed to unwind DNA in this manner, and the characteristic ‘rabbit ear’ DNA products have been observed using electron microscopy (Wessel et al. 1992).

Recent single-molecule approaches to study DNA replication using immobilized DNA and cell-free extracts from *Xenopus* oocytes hold much promise in addressing

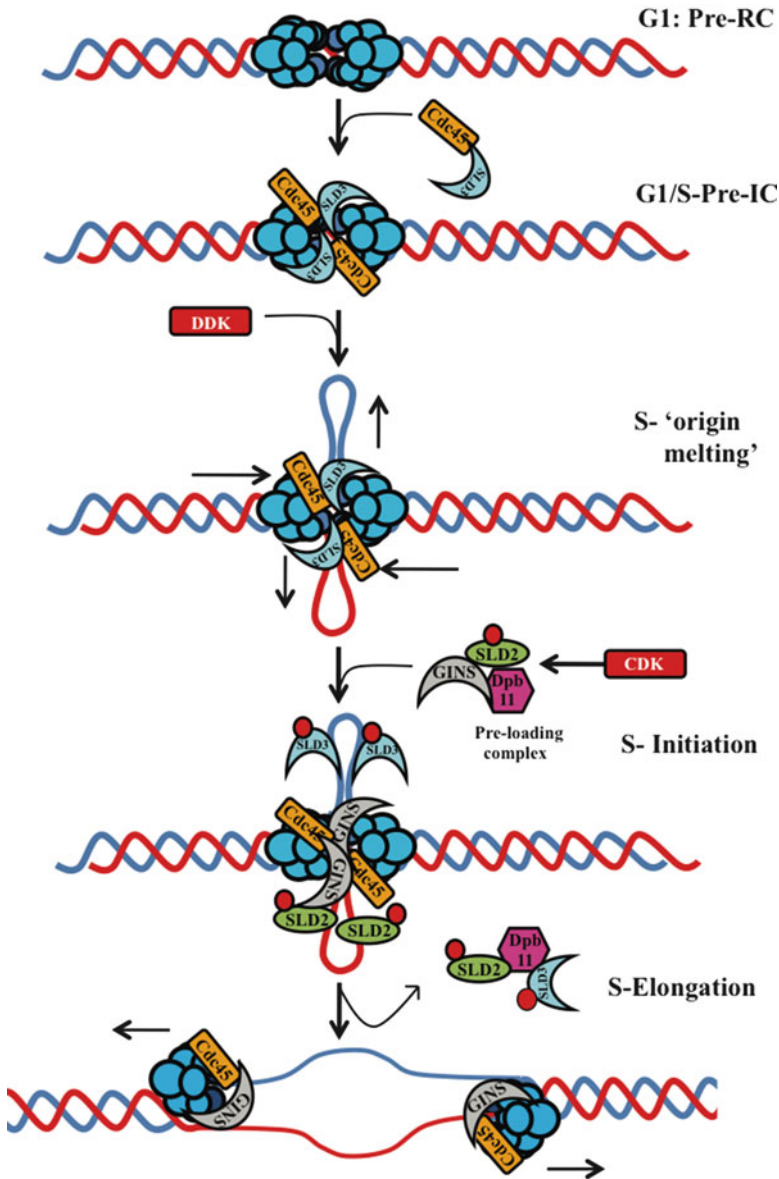


Fig. 7.3 Summary model of CMG loading and activation during DNA replication. *Red circles* phosphorylation, Mcm2-7 head-to-head double hexamers– *blue* (depicted with N-termini facing each other and C-termini facing away from each other, see text for details)

these questions (Yardimci et al. 2010). Although complete eukaryotic DNA replication has yet to be reconstituted using purified components, an efficient *in vitro* system using *Xenopus* egg extracts and exogenous DNA has been well-characterized and validated as a reliable model for *in vivo* eukaryotic DNA replication (Blow and

Laskey 1986; Masuda et al. 2003; Newport 1987; Prokhorova and Blow 2000). This system has been recently adopted for single-molecule analysis. Using total internal reflection fluorescence (TIRF) microscopy to visualize DNA replication by both an increase in DNA thickness (following staining with fluorescent DNA-specific dyes), as well as localization of newly incorporated UTP-digoxigenin (following addition of fluorescently-labeled anti-DIG antibodies), the bidirectional incorporation of nucleotides into tethered lambda DNA was visualized and verified as being Mcm-dependent (Yardimci et al. 2010).

Additional experiments using the single-molecule approach strongly argue against the dsDNA pump model for elongation. This model proposes that dsDNA is bidirectionally pumped into the helicase and predicts that unwinding (and DNA replication) should stall if both ends of an actively replicating substrate are tethered to prevent DNA movement into the pump (see Fig. 1A in Yardimci et al. 2010). DNA substrates were designed in which either one or both ends were modified with biotin. Upon addition to an avidin-coated flow cell, one end of the DNA binds the avidin coating, while flow within the cell quickly stretches the DNA to 90% of its estimated contour length. DNA containing the second biotin then binds the avidin coating to generate a stretched DNA molecule with both ends tethered. Upon addition of the *in vitro* *Xenopus* DNA replication system followed by extensive UTP-digoxigenin incorporation (replication over >60% of the DNA), no effects on the rate or extent of DNA replication were observed between singly and doubly-tethered lambda DNA molecules (Yardimci et al. 2010). These results demonstrate that eukaryotic DNA replication does not require coupling between sister replication forks and disfavor the obligatory use of a dsDNA pump during elongation.

7.6 Speculative Model for Mcm2-7 Function

In combination, the above investigations suggest that Mcm2-7 loads as dsDNA-bound double hexamers during initiation but is likely active as an ssDNA-bound CMG complex during elongation. If this interpretation is correct, two mechanistic problems must be solved: firstly, one of the two hexamer types should melt the double-stranded origin to generate ssDNA and secondly, the dsDNA-bound double hexamers will require remodeling to release one of the two strands to the external surface of the complex to form ssDNA-bound single hexamers (Fig. 7.3). Recent studies suggest that Sld2 and Sld3 may be involved in one or both scenarios.

Sld2 and Sld3 serve to load GINS and Cdc45 onto the pre-RC (Kamimura et al. 2001; Muramatsu et al. 2010). CDK-dependent phosphorylation of Sld2 and Sld3 facilitates their association with Dpb11, and is the central event that couples DNA replication to S-phase progression (Zegerman and Diffley 2007). *In vivo*, unphosphorylated Sld3 assists origin loading of Cdc45 in G1 (Aparicio et al. 1999). At the G1/S transition, phosphorylation of Sld2 facilitates its binding to Dpb11, which further associates with GINS and Pol ϵ to form the 'pre-loading' complex (pre-LC) (Muramatsu et al. 2010). The pre-LC then binds phosphorylated

Sld3 and loads GINS and Pol ϵ onto the nascent replication fork. Sld2, Sld3 and Dpb11 are subsequently released from the replication origin (Kanemaki and Labib 2006; Masumoto et al. 2000). Although initially thought to be yeast-specific features of DNA replication, metazoan orthologs of Sld2, such as RecQ4 (Matsuno et al. 2006), and Sld3, such as Treslin/Ticrr (Sanchez-Pulido et al. 2010), have been recently identified.

In addition to their loading activity, both Sld2 and Sld3 have an origin-specific ssDNA binding activity (Bruck and Kaplan 2011b). GST pull-down experiments indicate that Sld3 preferentially binds one specific strand of the budding yeast replication origins *ARS1* and *ARS305* ($K_d = 53$ nM) but has much lesser affinity for either dsDNA or the opposite strand ($K_d > 200$ nM) (Bruck and Kaplan 2011b). Similarly, Sld2 preferentially binds the complementary single-stranded origin region (Kanter and Kaplan 2010). Although these proteins may passively assist origin unwinding by binding ssDNA, experiments with Sld3 suggest a more complex scenario. Sld3 physically associates *in vitro* with both Cdc45 and Mcm2-7 (to generate the Cdc45-Mcm2-7-Sld3 (CMS) complex) in the absence of CDK phosphorylation (Bruck and Kaplan 2011a; Kamimura et al. 2001). Furthermore, Sld3 and GINS bind Mcm2-7 and Cdc45 in a mutually exclusive manner, indicating that Sld3 must dissociate from the pre-RC before GINS can load to form the CMG complex (Bruck and Kaplan 2011a). *In vitro*, the association between Sld3 and ssDNA destabilizes its association with Mcm2-7 and facilitates CMG formation (Bruck and Kaplan 2011b). These observations may suggest several things: firstly, as GINS loading requires a partially single-stranded origin to assist the removal of Sld3, the origin may be unwound prior to CMG assembly, and secondly, as Sld3 and GINS bind both Mcm2-7 and Cdc45 in a competitive manner, Sld3 may form a complex with Mcm2-7 and Cdc45 that is the structural or perhaps functional alternative to the CMG complex (Bruck and Kaplan 2011a).

Precedents in other systems indicate that double hexameric helicases can melt dsDNA with a pump-like activity (Enemark and Joshua-Tor 2006; Wessel et al. 1992). For example, the AAA⁺ viral helicase E1 has two different modes of DNA unwinding. While it unwinds DNA by steric exclusion of the unbound strand as a single hexamer (Enemark and Joshua-Tor 2006), the double hexameric form is specifically needed for initial origin dsDNA unwinding (Schuck and Stenlund 2005). Although the Mcm2-7 single-molecule experiments are inconsistent with a dsDNA pump activity during elongation, they lack sufficient resolution to preclude the possibility of limited dsDNA pump activity during initiation (Yardimci et al. 2010). Under appropriate conditions (e.g., phosphorylation by regulatory kinases, or binding of additional replication factors), perhaps Mcm2-7 has a similar dsDNA pumping activity (Fig. 7.3). A specific candidate for DNA unwinding would be the CMS complex. If Sld3 can substitute for GINS to form the CMS complex (the CMS complex has been observed *in vitro*, see Bruck and Kaplan 2011a), perhaps Sld3 positions Cdc45 over the Mcm2/5 gate to generate an alternative Mcm2-7 higher-order complex that specializes in origin unwinding. Alternatively, perhaps Sld3 serves to stimulate specific Mcm ATPase active sites that are not directly involved in DNA unwinding during elongation.

Moreover, the apparent ssDNA-binding properties of Sld2 and Sld3 may assist the remodeling of the Mcm2-7 double hexamers into the CMG complex. Although both Sld2 and Sld3 bind single-stranded origin DNA, they prefer *opposite* strands (Bruck and Kaplan 2011b; Kanter and Kaplan 2010). In a dsDNA pumping model, origin DNA would be extruded between the double hexamers, forming a ‘rabbit-eared’ structure in which the two complementary origin strands separate (Wessel et al. 1992). The binding of Sld2 and Sld3 to specific and opposite origin strands may assist strand passage between individual hexamers (possibly through the Mcm2/5 gate) to the outside of Mcm2-7. Additionally, the force of extruding ssDNA as ‘rabbit ears’ may destabilize the double hexamers, perhaps resulting in their separation into two bidirectionally moving single hexamers, and/or providing mechanical stress to assist the predicted DNA remodeling required for CMG formation (Fig. 7.3).

Acknowledgements We express our gratitude to N. Simon, M. Bochman and E. Tsai for their critical reading and insightful comments on the manuscript. This work is supported by IRO1GM83985 grant from NIH.

Note added in proof Using the *Xenopus* DNA replication system and single-molecule experiments utilizing strand-specific roadblocks, Fu et al. (Cell 146, 931-941, 2011) recently demonstrated that the CMG helicase can track along single-stranded DNA in a 3′-5′ direction on the leading strand while selectively excluding the lagging strand template. While a dsDNA pump would effectively stall regardless of the placement of the obstruction on either strand, a strand-specific translocation mechanism would selectively allow bypass of an obstruction one of the two strands. By presenting strong experimental evidence in favor of the latter scenario, Fu et al. validate the steric exclusion model of DNA unwinding by the Mcm2-7 helicase.

References

- Aparicio OM, Weinstein DM, Bell SP (1997) Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell* 91:59–69
- Aparicio OM, Stout AM, Bell SP (1999) Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. *Proc Natl Acad Sci USA* 96:9130–9135
- Biswas T, Tsodikov OV (2008) Hexameric ring structure of the N-terminal domain of *Mycobacterium tuberculosis* DnaB helicase. *FEBS J* 275:3064–3071
- Blow JJ, Laskey RA (1986) Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of *Xenopus* eggs. *Cell* 47:577–587
- Bochman ML, 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* 282:33795–33804
- Bochman ML, Schwacha A (2008) The Mcm2-7 complex has *in vitro* helicase activity. *Mol Cell* 31:287–293
- Bochman ML, Schwacha A (2009) The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiol Mol Biol Rev* 73:652–683
- Bochman ML, Schwacha A (2010) The *Saccharomyces cerevisiae* Mcm6/2 and Mcm5/3 ATPase active sites contribute to the function of the putative Mcm2-7 ‘gate’. *Nucleic Acids Res* 38:6078–6088

- Bochman ML, Bell SP, Schwacha A (2008) Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability. *Mol Cell Biol* 28:5865–5873
- Bos JL, Rehmann H, Wittinghofer A (2007) GEFs and GAPs: critical elements in the control of small G proteins. *Cell* 129:865–877
- Bramhill D, Kornberg A (1988) Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. *Cell* 52:743–755
- Brewster AS, Wang G, Yu X, Greenleaf WB, Carazo JM, Tjajadia M, Klein MG, Chen XS (2008) Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase. *Proc Natl Acad Sci USA* 105:20191–20196
- Brewster AS, Slaymaker IM, Afif SA, Chen XS (2010) Mutational analysis of an archaeal minichromosome maintenance protein exterior hairpin reveals critical residues for helicase activity and DNA binding. *BMC Mol Biol* 11:62
- Broderick R, Nasheuer HP (2009) Regulation of Cdc45 in the cell cycle and after DNA damage. *Biochem Soc Trans* 37:926–930
- Bruck I, Kaplan D (2009) Dbf4-Cdc7 phosphorylation of Mcm2 is required for cell growth. *J Biol Chem* 284:28823–28831
- Bruck I, Kaplan DL (2011a) GINS and Sld3 compete with one another for Mcm2-7 and Cdc45 binding. *J Biol Chem* 286:14157–14167
- Bruck I, Kaplan DL (2011b) Origin single-stranded DNA releases Sld3 from Mcm2-7, allowing GINS to bind Mcm2-7. *J Biol Chem* 286:18602–18613
- Burke TW, Cook JG, Asano M, Nevins JR (2001) Replication factors MCM2 and ORC1 interact with the histone acetyltransferase HBO1. *J Biol Chem* 276:15397–15408
- Chang YP, Wang G, Bermudez V, Hurwitz J, Chen XS (2007) Crystal structure of the GINS complex and functional insights into its role in DNA replication. *Proc Natl Acad Sci USA* 104:12685–12690
- Choi JM, Lim HS, Kim JJ, Song OK, Cho Y (2007) Crystal structure of the human GINS complex. *Genes Dev* 21:1316–1321
- Chuang L-C, Teixeira LK, Ja W, Henze M, Yates JR, Méndez J, Reed SI (2009) Phosphorylation of Mcm2 by Cdc7 promotes pre-replication complex assembly during cell-cycle re-entry. *Mol Cell* 35:206–216
- Costa A, Ilves I, Tamberg N, Petojevic T, Nogales E, Botchan MR, Berger JM (2011) The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol* 18:471–477
- Davey MJ, Indiani C, O'Donnell M (2003) Reconstitution of the Mcm2-7p heterohexamer, subunit arrangement, and ATP site architecture. *J Biol Chem* 278:4491–4499
- Donovan S, Harwood J, Drury LS, Diffley JF (1997) Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proc Natl Acad Sci USA* 94:5611–5616
- Enemark EJ, Joshua-Tor L (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* 442:270–275
- Erzberger JP, Berger JM (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 USA* 106:20240–20245
- Ferenbach A, Li A, Brito-Martins M, Blow JJ (2005) Functional domains of the *Xenopus* replication licensing factor Cdt1. *Nucleic Acids Res* 33:316–324
- Fletcher RJ, Bishop BE, Leon RP, Sclafani RA, Ogata CM, Chen XS (2003) The structure and function of MCM from archaeal *M. thermoautotrophicum*. *Nat Struct Biol* 10:160–167
- Forsburg SL (2004) Eukaryotic MCM proteins: beyond replication initiation. *Microbiol Mol Biol Rev* 68:109–131
- Gai D, Zhao R, Li D, Finkielstein CV, Chen XS (2004) Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* 119:47–60
- Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, Labib K (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol* 8:358–366

- Gambus A, van Deursen F, Polychronopoulos D, Foltman M, Jones RC, Edmondson RD, 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
- Gambus A, Khoudoli GA, Jones RC, Blow JJ (2011) MCM2-7 form double hexamers at licensed origins in *Xenopus* egg extract. *J Biol Chem* 286:11855–11864
- Guillou E, Ibarra A, Coulon V, Casado-Vela J, Rico D, Casal I, Schwob E, Losada A, Mendez J (2010) Cohesin organizes chromatin loops at DNA replication factories. *Genes Dev* 24:2812–2822
- Hanson PI, Whiteheart SW (2005) AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* 6:519–529
- Hardy CF, Dryga O, Seematter S, Pahl PM, Sclafani RA (1997) *mcm5/cdc46-bob1* bypasses the requirement for the S phase activator Cdc7p. *Proc Natl Acad Sci USA* 94:3151–3155
- Hennessy KM, Lee A, Chen E, Botstein D (1991) A group of interacting yeast DNA replication genes. *Genes Dev* 5:958–969
- Homesley L, Lei M, Kawasaki Y, Sawyer S, Christensen T, Tye BK (2000) Mcm10 and the MCM2-7 complex interact to initiate DNA synthesis and to release replication factors from origins. *Genes Dev* 14:913–926
- Ilves I, Petojevic T, Pesavento JJ, Botchan MR (2010) Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell* 37:247–258
- Im JS, Ki SH, Farina A, Jung DS, Hurwitz J, Lee JK (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 USA* 106:15628–15632
- Ishimi Y (1997) A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex. *J Biol Chem* 272:24508–24513
- Ishimi Y, Ichinose S, Omori A, Sato K, Kimura H (1996) Binding of human minichromosome maintenance proteins with histone H3. *J Biol Chem* 271:24115–24122
- Iyer LM, Leipe DD, Koonin EV, Aravind L (2004) Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol* 146:11–31
- Jares P, Donaldson A, Blow JJ (2000) The Cdc7/Dbf4 protein kinase: target of the S phase checkpoint? *EMBO Rep* 1:319–322
- Jenkinson ER, Chong JP (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 USA* 103:7613–7618
- Kamimura Y, Tak YS, Sugino A, Araki H (2001) Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in *Saccharomyces cerevisiae*. *EMBO J* 20:2097–2107
- Kanemaki M, Labib K (2006) Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. *EMBO J* 25:1753–1763
- Kanter DM, Kaplan DL (2010) Sld2 binds to origin single-stranded DNA and stimulates DNA annealing. *Nucleic Acids Res* 39:2580–2592
- Kanter DM, Bruck I, Kaplan DL (2008) Mcm subunits can assemble into two different active unwinding complexes. *J Biol Chem* 283:31172–31182
- Kawasaki Y, Kim HD, Kojima A, Seki T, Sugino A (2006) Reconstitution of *Saccharomyces cerevisiae* prereplicative complex assembly *in vitro*. *Genes Cells* 11:745–756
- 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* 29:5008–5019
- Koonin EV (1993) A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Res* 21:2541–2547
- 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* 17:1141–1152
- Labib K, Tercero JA, Diffley JF (2000) Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science* 288:1643–1647

- Lee JK, Hurwitz J (2000) Isolation and characterization of various complexes of the minichromosome maintenance proteins of *Schizosaccharomyces pombe*. *J Biol Chem* 275:18871–18878
- Lee C, Liachko I, Bouten R, Kelman Z, Tye BK (2010) Alternative mechanisms for coordinating polymerase alpha and MCM helicase. *Mol Cell Biol* 30:423–435
- Leon RP, Tecklenburg M, Sclafani RA (2008) Functional conservation of beta-hairpin DNA binding domains in the Mcm protein of *Methanobacterium thermoautotrophicum* and the Mcm5 protein of *Saccharomyces cerevisiae*. *Genetics* 179:1757–1768
- Liku ME, Nguyen VQ, Rosales AW, Irie K, Li JJ (2005) CDK phosphorylation of a novel NLS-NES module distributed between two subunits of the Mcm2-7 complex prevents chromosomal rereplication. *Mol Biol Cell* 16:5026–5039
- Liu W, Pucci B, Rossi M, Pisani FM, Ladenstein R (2008) Structural analysis of the *Sulfolobus solfataricus* MCM protein N-terminal domain. *Nucleic Acids Res* 36:3235–3243
- Liu Y, Richards TA, Aves SJ (2009) Ancient diversification of eukaryotic MCM DNA replication proteins. *BMC Evol Biol* 9:60
- Loo S, Fox CA, Rine J, Kobayashi R, Stillman B, Bell S (1995) The origin recognition complex in silencing, cell cycle progression, and DNA replication. *Mol Biol Cell* 6:741–756
- Lopes M, Cotta-Ramusino C, Pelliccioli A, Liberi G, Plevani P, Muzi-Falconi M, Newlon CS, Foiani M (2001) The DNA replication checkpoint response stabilizes stalled replication forks. *Nature* 412:557–561
- Lou H, Komata M, Katou Y, Guan Z, Reis CC, Budd M, Shirahige K, Campbell JL (2008) Mrc1 and DNA polymerase ϵ function together in linking DNA replication and the S phase checkpoint. *Mol Cell* 32:106–117
- Ma X, Stead BE, Rezvanpour A, Davey MJ (2010) The effects of oligomerization on *Saccharomyces cerevisiae* Mcm4/6/7 function. *BMC Biochem* 11:37
- Masuda T, Mimura S, Takisawa H (2003) CDK- and Cdc45-dependent priming of the MCM complex on chromatin during S-phase in *Xenopus* egg extracts: possible activation of MCM helicase by association with Cdc45. *Genes Cells* 8:145–161
- Masumoto H, Sugino A, Araki H (2000) Dpb11 controls the association between DNA polymerases α and ϵ and the autonomously replicating sequence region of budding yeast. *Mol Cell Biol* 20:2809–2817
- Matsuno K, Kumano M, Kubota Y, Hashimoto Y, Takisawa H (2006) The N-terminal noncatalytic region of *Xenopus* RecQ4 is required for chromatin binding of DNA polymerase alpha in the initiation of DNA replication. *Mol Cell Biol* 26:4843–4852
- McGeoch AT, Trakselis MA, Laskey RA, Bell SD (2005) Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism. *Nat Struct Mol Biol* 12:756–762
- Moir D, Stewart SE, Osmond BC, Botstein D (1982) Cold-sensitive cell-division-cycle mutants of yeast: isolation, properties, and pseudoreversion studies. *Genetics* 100:547–563
- Moyer SE, Lewis PW, Botchan MR (2006) Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci USA* 103:10236–10241
- Mukherjee P, Winter SL, Alexandrow MG (2010) Cell cycle arrest by transforming growth factor beta1 near G1/S is mediated by acute abrogation of prereplication complex activation involving an Rb-MCM interaction. *Mol Cell Biol* 30:845–856
- Muramatsu S, Hirai K, Tak YS, Kamimura Y, Araki H (2010) CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol ϵ , and GINS in budding yeast. *Genes Dev* 24:602–612
- Nakaya R, Takaya J, Onuki T, Moritani M, Nozaki N, Ishimi Y (2010) Identification of proteins that may directly interact with human RPA. *J Biochem* 148:539–547
- Nedelcheva MN, Roguev A, Dolapchiev LB, Shevchenko A, Taskov HB, Stewart AF, Stoynev SS (2005) Uncoupling of unwinding from DNA synthesis implies regulation of MCM helicase by Tof1/Mrc1/Csm3 checkpoint complex. *J Mol Biol* 347:509–521
- Newport J (1987) Nuclear reconstitution *in vitro*: stages of assembly around protein-free DNA. *Cell* 48:205–217

- Nishiyama A, Frappier L, Mechali M (2011) MCM-BP regulates unloading of the MCM2-7 helicase in late S phase. *Genes Dev* 25:165–175
- Numata Y, Ishihara S, Hasegawa N, Nozaki N, Ishimi Y (2010) Interaction of human MCM2-7 proteins with TIM, TIPIN and Rb. *J Biochem* 147:917–927
- Ogura T, Wilkinson AJ (2001) AAA+ superfamily ATPases: common structure – diverse function. *Genes Cells* 6:575–597
- Pacek M, Walter JC (2004) A requirement for MCM7 and Cdc45 in chromosome unwinding during eukaryotic DNA replication. *EMBO J* 23:3667–3676
- Pape T, Meka H, Chen S, Vicentini G, van Heel M, Onesti S (2003) Hexameric ring structure of the full-length archaeal MCM protein complex. *EMBO Rep* 4:1079–1083
- Patel SS, Picha KM (2000) Structure and function of hexameric helicases. *Annu Rev Biochem* 69:651–697
- Prokhorova TA, Blow JJ (2000) Sequential MCM/P1 subcomplex assembly is required to form a heterohexameric with replication licensing activity. *J Biol Chem* 275:2491–2498
- Randell JC, Bowers JL, Rodriguez HK, Bell SP (2006) Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol Cell* 21:29–39
- Randell JC, Fan A, Chan C, Francis LI, Heller RC, Galani K, Bell SP (2010) Mec1 is one of multiple kinases that prime the Mcm2-7 helicase for phosphorylation by Cdc7. *Mol Cell* 40:353–363
- Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JF (2009) Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139:719–730
- Ricke RM, Bielinsky AK (2004) Mcm10 regulates the stability and chromatin association of DNA polymerase α . *Mol Cell* 16:173–185
- Ryu MJ, Kim BJ, Lee JW, Lee MW, Choi HK, Kim ST (2006) Direct interaction between cohesin complex and DNA replication machinery. *Biochem Biophys Res Commun* 341:770–775
- Sanchez-Pulido L, Diffley JF, Ponting CP (2010) Homology explains the functional similarities of Treslin/Ticrr and Sld3. *Curr Biol* 20:R509–510
- Sato M, Gotow T, You Z, Komamura-Kohno Y, Uchiyama Y, Yabuta N, Nojima H, Ishimi Y (2000) Electron microscopic observation and single-stranded DNA binding activity of the Mcm4,6,7 complex. *J Mol Biol* 300:421–431
- Schuck S, Stenlund A (2005) Assembly of a double hexameric helicase. *Mol Cell* 20:377–389
- Seki T, Diffley JF (2000) Stepwise assembly of initiation proteins at budding yeast replication origins *in vitro*. *Proc Natl Acad Sci USA* 97:14115–14120
- Sheu YJ, Stillman B (2010) The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature* 463:113–117
- Stead BE, Sorbara CD, Brandl CJ, Davey MJ (2009) ATP binding and hydrolysis by Mcm2 regulate DNA binding by Mcm complexes. *J Mol Biol* 391:301–313
- Sterner JM, Dew-Knight S, Musahl C, Kornbluth S, Horowitz JM (1998) Negative regulation of DNA replication by the retinoblastoma protein is mediated by its association with MCM7. *Mol Cell Biol* 18:2748–2757
- Takahashi TS, Wigley DB, Walter JC (2005) Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. *Trends Biochem Sci* 30:437–444
- 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* 17:1153–1165
- Tan BC, Liu H, Lin CL, Lee SC (2010) Functional cooperation between FACT and MCM is coordinated with cell cycle and differential complex formation. *J Biomed Sci* 17:11
- Tanaka S, Tak YS, Araki H (2007a) The role of CDK in the initiation step of DNA replication in eukaryotes. *Cell Div* 2:16
- Tanaka S, Umemori T, Hirai K, Muramatsu S, Kamimura Y, Araki H (2007b) CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature* 445:328–332
- Tercero JA, Labib K, Diffley JF (2000) DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p. *EMBO J* 19:2082–2093

- Tsao CC, Geisen C, Abraham RT (2004) Interaction between human MCM7 and Rad17 proteins is required for replication checkpoint signaling. *EMBO J* 23:4660–4669
- Wang EH, Prives C (1991) ATP induces the assembly of polyoma large tumor antigen into hexamers. *Virology* 184:399–403
- 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* 285:12469–12473
- Wessel R, Schweizer J, Stahl H (1992) Simian virus 40 T-antigen DNA helicase is a hexamer which forms a binary complex during bidirectional unwinding from the viral origin of DNA replication. *J Virol* 66:804–815
- Wittmeyer J, Joss L, Formosa T (1999) Spt16 and Pob3 of *Saccharomyces cerevisiae* form an essential, abundant heterodimer that is nuclear, chromatin-associated, and copurifies with DNA polymerase alpha. *Biochemistry* 38:8961–8971
- Yardimci H, Loveland AB, Habuchi S, van Oijen AM, Walter JC (2010) Uncoupling of sister replisomes during eukaryotic DNA replication. *Mol Cell* 40:834–840
- Zegerman P, Diffley JF (2007) Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* 445:281–285
- Zou L, Mitchell J, Stillman B (1997) *CDC45*, a novel yeast gene that functions with the origin recognition complex and MCM proteins in initiation of DNA replication. *Mol Cell Biol* 17:553–563

Chapter 8

The GINS Complex: Structure and Function

Katsuhiko Kamada

Abstract Eukaryotic chromosomal DNA replication is controlled by a highly ordered series of steps involving multiple proteins at replication origins. The eukaryotic GINS complex is essential for the establishment of DNA replication forks and replisome progression. GINS is one of the core components of the eukaryotic replicative helicase, the CMG (Cdc45-MCM-GINS) complex, which unwinds duplex DNA ahead of the moving replication fork. Eukaryotic GINS also links with other key proteins at the fork to maintain an active replisome progression complex. Archaeal GINS homologues play a central role in chromosome replication by associating with other replisome components. This chapter focuses on the molecular events related with DNA replication initiation, and summarizes our current understanding of the function, structure and evolution of the GINS complex in eukaryotes and archaea.

Keywords CMGcomplex • Crystal structure • DNA replication • Replication fork

8.1 Introduction

The initiation of chromosomal DNA replication in eukaryotic cells is a highly regulated process that requires the coordinated actions of a large number of essential and optional protein factors (Bell and Dutta 2002; Diffley 2004). The early events of chromosome replication have been well studied in the budding yeast *Saccharomyces cerevisiae*. During the early stages of the cell cycle (G1), the loading of the six subunits of the conserved protein complex known as ORC (origin recognition complex)

K. Kamada (✉)

Chromosome Dynamics Laboratory, RIKEN Advanced Science Institute,
2-1 Hirosawa, Wako, Saitama 351-0198, Japan
e-mail: kamadak@riken.jp

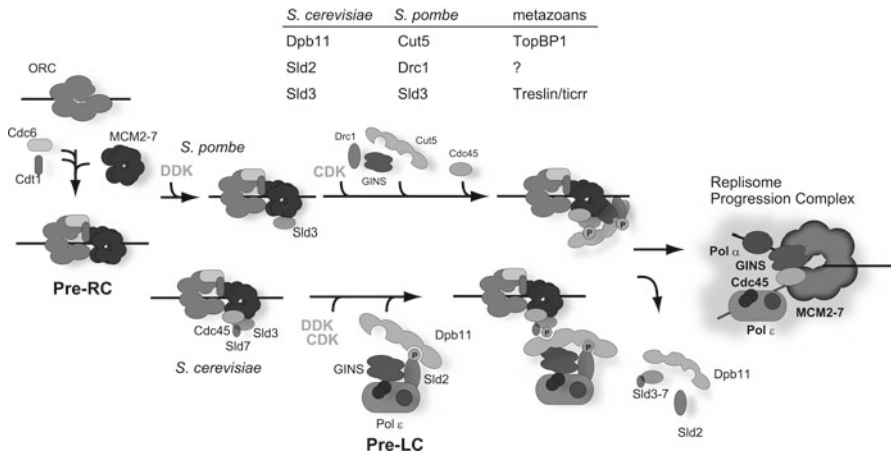


Fig. 8.1 Models of the ordered assembly of replication initiation in yeast. In the G1 phase, Cdc6 binds to ORC at origin DNA, and the loadings of inactive MCM complexes and Cdt1 form the pre-RC. Pre-RC assembly at this stage is said to license the origin. In *S. pombe*, at the onset of S phase, Sld3 binds to replication origins in a manner dependent on DDK activity. Depending on Sld3 and CDK, GINS and Cut5 then bind to the origin in a mutually dependent manner. The loading of Cdc45 depends on the other factors. In *S. cerevisiae*, Sld3 is complexed with Sld7 and associates with Cdc45 at early replication origins. Phosphorylation of Sld3 by S-CDK allows, by way of Dpb11 binding, the recruitment of the pre-LC to the CDK-dependent assembled complex. The pre-LC contains Dpb11, Sld2, Pol ϵ and the GINS tetramer. At this time, Dpb11, Sld2 and GINS are recruited in a mutually dependent manner. By the combined actions of kinases, an active MCM complex is formed with Cdc45 and GINS, namely the CMG complex, which most likely functions as the replicative helicase. Thus, the replisome progression complex (RPC) is assembled with other protein factors around the single CMG complex

onto the DNA origin or prospective sites for the initiation of replication is a prerequisite for the assembly of the pre-replicative complex (pre-RC) (Fig. 8.1). ORC functions essentially as a platform for binding Cdc6 and Cdt1. These two proteins subsequently facilitate the loading of the MCM (mini-chromosome maintenance) 2-7 complex onto the DNA to form the pre-RC. Pre-RC assembly, also known as replication licensing, occurs only during G1, when cyclin dependent kinase (CDK) activity is low, ensuring that replication occurs once, and only once, per cell cycle. Therefore, the MCM complex just loaded onto the origin or the isolated complex is inactive as a DNA helicase. During G1/S transition, the recruitment of other replication factors such as Cdc45 and GINS, and the combined action of two S-phase promoting kinases, the S-phase cyclin-dependent (S-CDK) and Cdc7-Dbf4 (DDK) kinases, lead to the assembly of an active DNA helicase complex at replication origins. This activation results in the unwinding of DNA and the subsequent recruitment of the replication machineries (DNA polymerases and the primase complex). Once the replisome is activated, the replication forks move bidirectionally from the origin DNA into the surrounding sequences. ORC stays at the origin in yeast cells, whereas MCM, Cdc45, GINS and DNA polymerases move with the fork. Meanwhile, numerous additional protein factors are recruited around the Cdc45-MCM-GINS core

(Moyer et al. 2006), forming a larger macromolecular assembly known as the RPC (replisome progression complex) (Gambus et al. 2006).

The present review summarizes the currently available information on the function, structure and evolution of the recently identified tetrameric GINS complex. The GINS complex consists of four proteins (Sld5-Psf1-Psf2-Psf3) and was named for the Japanese 'go-ichi-ni-san', which means 5-1-2-3 (Takayama et al. 2003). The complex is essential for the initiation and elongation stages of chromosome replication. The function of the GINS complex has been investigated by a variety of methods in many species, and the three-dimensional structure of the human GINS complex has been solved by X-ray crystallography. Archaeal GINS homologues have also been identified, providing insight into its evolutionary origins.

8.2 Discovery of GINS

The regulation of the four subunit DNA polymerase ϵ (Pol ϵ) complex, which functions in leading strand synthesis, was originally investigated by Araki and co-workers, who performed genetic screens in the budding yeast (Araki et al. 1995). Increased expression of *dpb11* was found to suppress a mutation in the *dpb2* gene, which encodes the second largest subunit of Pol ϵ . Dpb11 is required for the loading of Pol ϵ onto DNA during chromosomal DNA replication. However, Dpb11 is displaced from the origin after initiation and does not normally move with DNA replication forks (Masumoto et al. 2000). Studies have focused on screening to isolate synthetic lethal mutants in combination with the temperature sensitive *dpb11-1* allele (Kamimura et al. 1998). Isolated mutants were categorized into distinct *sld* (synthetic lethality with *dpb11-1*) groups, and named *sld1* (or *dpb3*: Dpb3 is the third largest subunit of Pol ϵ), *sld2*, *sld3* and *sld4* (or *cdc45*) (Kamimura et al. 1998, 2001; Masumoto et al. 2002). Of these groups, *sld5* encoded an essential protein for cell growth (Takayama et al. 2003). Using one of the thermosensitive mutants, sequential multicopy suppressor screening and an immunoprecipitation assay identified three additional interacting factors, namely Psf1 (Partner of Sld Five 1), Psf2 and Psf3. These four proteins are stoichiometric components of a complex that is required for cell growth and chromosomal DNA replication.

Labib and co-workers also independently isolated these genes from a collection of budding yeast strains that was systematically screened for genes essential for cell viability but of unknown function (Kanemaki et al. 2003). These authors utilized a method based on the conditional and rapid degradation of the target protein *in vivo* (Labib et al. 2000). Strains were generated to express each target protein with a 'heat-inducible degron' cassette at its N-terminus (Dohmen et al. 1994). Under higher temperature conditions, the unfolded degron is recognized by the Ubr1 protein, ubiquitinated and subsequently degraded by the proteasome. In the original report, three of the four subunits of the GINS complex were identified as Cdc105, Cdc101 and Cdc102, which correspond to Sld5, Psf1 and Psf2, respectively (Kanemaki et al. 2003). These proteins are required for the establishment and

normal progression of DNA replication forks. The fourth factor, Cdc103 or Psf3, was later isolated using a biochemical approach in yeast cell extracts.

Each of the four GINS proteins is essential for cell viability in budding yeast. ChIP (chromatin immunoprecipitation) experiments have shown that the GINS proteins associate with origin sequences in early S-phase before moving with the replication fork as replication continues. In addition, dense-isotope substitution experiments demonstrated that the GINS complex is required for both establishment and progression of replication, indicating that the GINS complex is not only a passenger at the fork but an active constituent (Kanemaki et al. 2003).

Takisawa and co-workers identified sequence homologues of the four yeast GINS proteins and first characterized a higher eukaryotic GINS complex (Kubota et al. 2003). Antibodies raised against Sld5 co-depleted Psf1, Psf2 and Psf3 from *Xenopus* egg extracts, indicating that the four proteins coexist as a complex. Biochemical analysis showed that the recombinant complex had an expected molecular mass of approximately 100 kDa.

8.3 GINS Functions

8.3.1 *Replication Initiation in the Budding Yeast*

Identification of CDK and DDK targets is important for the understanding of how replication initiation is controlled. In the budding yeast, Sld2 and Sld3 are the minimal set of S-phase CDK substrates required for the initiation of DNA replication. The CDK-mediated phosphorylation of Thr600 and Ser622 in Sld3 (the functional orthologue in yeast of Treslin/Ticrr in humans) is essential for cell viability, and provides a binding site for another essential replication factor, Dpb11 (the functional orthologue of TopBP1 in humans) (Tanaka et al. 2007; Zegerman and Diffley 2007). The Dpb11 protein possesses two sets of tandemly aligned BRCT (BRCA1 C-terminal) domains, and the N-terminal pair is responsible for binding to phosphorylated Sld3. Sld3 appears to form a complex with a stabilization factor, Sld7 (Tanaka et al. 2011), and also associates with an essential factor, Cdc45, which is recruited to the pre-RC through its interaction with MCM. Interestingly, a mutant form of Cdc45 (known as JET) can partially bypass the requirement for CDK (Tanaka et al. 2007), probably by adopting a conformation that promotes the Sld3-Dpb11 interaction. Another essential substrate, Sld2, has multiple CDK phosphorylation motifs and contains a region that is partially homologous to human RecQL4, although unlike Sld2, RecQL4 appears to function after Cdc45 and GINS have been loaded onto replication origins (Matsuno et al. 2006; Sangrithi et al. 2005). In particular, the phosphorylation of Thr84 is crucial to create a binding site for Dpb11 through the C-terminal pair of BRCT domains (Masumoto et al. 2002; Tak et al. 2006). Combination of a phospho-mutant form of Sld2 Thr84 with the Sld3-Dpb11 fusion bypasses the requirement for S-CDK in replication initiation (Zegerman and Diffley 2007).

Similar results were obtained when the phospho-mimicking Sld2 mutant was combined with cells expressing JET and overproducing Dpb11, showing that Dpb11 bridges phosphorylated Sld2 and Sld3 (Tanaka et al. 2007) (Fig. 8.1).

Origin association of GINS is dependent on the function of Sld3 and Dpb11. In *sld3-5* or *dpb11-26* mutant cells, ChIP assays showed the absence of GINS binding to the replication origin (Takayama et al. 2003), suggesting that this event occurs downstream of Sld3 and Dpb11 function. However, GINS function is required for the association of Dpb11 with origins, indicating that origin binding of GINS and Dpb11 are mutually dependent events.

In addition to the interactions among these factors, CDK promotes the formation of a fragile complex called the preloading complex (pre-LC), which contains Pole, GINS, Sld2 and Dpb11, in a DDK-independent manner (Muramatsu et al. 2010) (Fig. 8.1). This complex is detected only in formaldehyde-fixed cell extracts, and is somewhat heterogeneously assembled. However, its presence supports the function of Dpb11 as a molecular bridge between Cdc45-Sld3 and Sld2 to recruit the pre-LC components to the origin. Therefore, the pre-LC acts as a carrier of GINS to the pre-RC to efficiently activate subsequent steps. Thus, CDK might promote the initiation of DNA replication by enhancing GINS recruitment to origins through the formation of the pre-LC. Once DNA replication starts, Sld2, Sld3 and Dpb11 do not move with the replication fork, whereas Cdc45 and GINS act as major components at the replication fork (Kanemaki and Labib 2006). In this scenario, the pre-LC must be disrupted for the initiation of DNA replication to occur.

8.3.2 Replication Initiation in the Fission Yeast

Several conditional lethal mutants of GINS isolated in the fission yeast *Schizosaccharomyces pombe* have been used for the analysis of replication initiation. A temperature-sensitive mutant, *psf2-209*, was initially isolated that was unable to re-replicate chromosomal DNA in the absence of cyclin B (Gomez et al. 2005). When shifted to the restrictive temperature, the mutant displays a somewhat heterogeneous phenotype with nuclear abnormalities and a delay in S-phase progression. The *psf2-209* cells require the presence of a functional DNA damage checkpoint for viability even at the permissive temperature. The *psf3-1* mutant, which was isolated by targeted screening, causes an early S-phase block when shifted to the restrictive temperature (Yabuuchi et al. 2006). This mutation can be suppressed by overproduction of other subunits and by Sld2, and is synthetically lethal at the normal permissive temperature together with mutant alleles of *cut5* and *cdc20*, which encode the fission yeast Dpb11 homologue and the largest subunit of Pol ϵ respectively, as well as with *sld3* and *cdc45*.

As in the budding yeast, the ChIP assay has been used to show that GINS associates with origin DNA in the early S-phase. The loading of GINS and Cut5 onto DNA depends on Sld3, whereas Sld3 binds to origins independently (Fig. 8.1). In contrast, the recruitment of Cdc45 requires Sld3 and GINS, but not *vice versa*.

These results indicate that origin loading of Sld3 is the most upstream reaction required for assembly at the beginning of S-phase (Yabuuchi et al. 2006). The recruitment of GINS and Cut5 to origins is mutually dependent on Sld3, and Cdc45 is the last factor recruited. The unwinding of DNA at the replication origin thus occurs after the assembly of these factors.

The biochemical function of GINS in the fission yeast was studied through the fusion of GINS subunits to a steroid hormone-binding domain (HBD) (Pai et al. 2009). In this system, the 90 kDa heat shock protein Hsp90 binds to the HBD in the absence of β -estradiol and can inactivate the tagged protein through structural interference, but binding of β -estradiol to the HBD causes Hsp90 displacement, which can reactivate the fusion protein (Picard 2000). Fusing the β -estradiol HBD to Psf1 and Psf2 produces cells that require the constant presence of β -estradiol for cell viability. Removal of β -estradiol causes rapid cell cycle arrest in early S-phase, and in the case of Psf2, nuclear delocalization of Psf3. Inactivation of GINS has distinct effects on the replication origin association and chromatin binding of two of the replicative DNA polymerases. Inactivation of Psf1 leads to loss of chromatin bindings of Pol ϵ and Cdc45. In contrast, chromatin association of the catalytic subunit of DNA polymerase α (Pol α) is not affected, suggesting that GINS functions in a pathway that involves Cdc45 and is necessary for Pol ϵ chromatin binding.

8.3.3 Replication Initiation in Higher Eukaryotes

The *Xenopus laevis* cell-free system is a powerful tool for the biochemical study of eukaryotic chromosomes (Chong et al. 1997). Extracts prepared from *Xenopus* eggs are able to replicate exogenously added sperm DNA while preserving the replication licensing system of the cell cycle. In Sld5-depleted *Xenopus* extracts, incorporation of labeled nucleotides into chromosomal DNA is not detected, indicating that Sld5 is required for efficient replication initiation (Kubota et al. 2003). Further analysis demonstrated that the GINS complex binds to chromatin in a manner dependent on replication licensing and S-phase CDK. As seen in budding yeast, the binding of GINS and Cdc45 to chromatin is interdependent. Lack of chromatin binding of the two factors leads to a failure in the loading of the replicative polymerases Pol α and Pol ϵ . Cut5 (TopBP1 in humans) is also required for chromatin loading of GINS and Cdc45.

Analyses of endogenous human GINS have shown that the four components of GINS are expressed during mitotic cycles, degraded shortly after the cells are driven into G0, and resynthesized upon re-entry into the cell cycle (Aparicio et al. 2009). The complex is present in abundance, even in untransformed cell lines, showing an average of 100,000 GINS complexes per cell. The number of molecules present varies in different cell types, with significantly higher concentrations seen in immortalized cells. While the level of GINS is constant throughout the cell cycle, a significant fraction of human GINS associates with chromatin specifically during S-phase. This chromatin-associated material co-purifies with MCM and Cdc45 in

extracts prepared from nuclease-treated S-phase chromatin, consistent with the existence of CMG complexes in human cells.

In human cells, evidence points to the possible involvement of GINS in the Fanconi anemia (FA) tumor suppressor pathway, which plays an important role in the maintenance of genomic stability in eukaryotic cells (Tumini et al. 2011). Psf2 physically interacts with the FANCF protein, which is a component of the FA multiprotein nuclear core complex, as confirmed by yeast two-hybrid assays and co-immunoprecipitation experiments. GINS is required to efficiently load components of the FA complex onto the chromosome at the G1/S transition. However, depletion of Psf2 is not sufficient to inhibit the monoubiquitylation of FANCD2 or its localization to nuclear foci following DNA damage, suggesting that involvement of the GINS complex is limited to recruiting or stabilizing the complex.

8.3.4 *GINS in the Replication Progression Complex*

Using tandem affinity column purification, Gambus et al. identified a specific set of proteins from budding yeast cells that interacted with GINS (Gambus et al. 2006). In addition to Cdc45 and MCM, GINS binds to essential initiation and elongation factors by forming large replisome progression complexes (RPCs), which are assembled during initiation and disassembled at the end of S phase. Isolated RPC components include Mrc1 (checkpoint mediator for fork stalling), the Tof1-Csm3 complex (pausing factor for replication forks at protein-DNA barriers), the histone chaperone FACT, type I topoisomerase (thought to release positive supercoils at the fork), and Ctf4 and MCM10, which are known to bind Pol α . The GINS complex may be a part of larger complexes that include the MCM helicase and other proteins at the replication fork.

Other evidence for the involvement of GINS in the progression complex comes from studies of fork pausing. The *Xenopus* cell-free system allows replication initiation to occur independently of the DNA sequence. Pacek et al. synthesized plasmid DNA containing a biotin group at a single site in which the resulting biotin-streptavidin pairing halts replication fork progression, and detected the presence of individual target proteins by CHIP (Pacek et al. 2006). This approach allowed the identification of all three replicative DNA polymerases (α , δ and ϵ), and MCM2-7, Cdc45, GINS, and MCM10 as components of the replisome. In the presence of the DNA polymerase inhibitor aphidicolin, which causes uncoupling of a highly processive DNA helicase from the stalled replisome, only Cdc45, GINS, and MCM2-7 were enriched at the pause site. It is highly likely that the MCM, Cdc45 and GINS proteins form a large molecular machine that functions as a replicative helicase.

A different version of the RPC, containing Pol α , was purified under conditions that induced the displacement of MCM10; coupling of Pol α to the RPC requires the Ctf4 protein (Gambus et al. 2009). In the budding yeast, Ctf4 was originally identified as a gene product whose mutated form showed decreased chromosome transmission fidelity during mitosis (Spencer et al. 1990), and is also known to bind

Pol α (Miles and Formosa 1992). GINS and Ctf4 were shown to form a stable complex *in vitro* and the Sld5 subunit was found to be responsible for their interaction (Tanaka et al. 2009), indicating that the interaction between these two components in the RPC is crucial for linking the replicative helicase MCM2-7 to Pol α . However, the subsequent contribution of MCM10 and Ctf4 at the replication fork as a tether between Pol α and the MCM2-7 helicase was not addressed.

Several lines of evidence show that the GINS complex communicates with polymerase complexes. Using surface plasmon resonance and co-immunoprecipitation, a recombinant human GINS was shown to bind to a purified dimeric primase complex (De Falco et al. 2007), although the dissociation constant was in the low micromolar range and the interaction has yet to be observed in cells. Moreover, a two-order molar excess of the GINS complex resulted in a ten-fold increase in DNA synthesis by the Pol α -primase complex *in vitro*, but did not stimulate the activity of primase itself.

Other studies have indicated that GINS also functions as an accessory module for Pol ϵ (Takayama et al. 2003). This is originally based on the results of two-hybrid assays showing an interaction between Psf1 and the Dpb2 subunit of Pol ϵ . The *Xenopus* GINS complex also binds to Pol ϵ *in vitro* and stimulates its activity (Shikata et al. 2006). How these *in vitro* observations might reflect the role of GINS *in vivo* has not been determined. It remains to be explored whether GINS works multi-functionally during the progression of the replication fork.

Many studies have reported snapshots of the progression of the DNA replication fork in the budding yeast genome. Sekedat et al. monitored the dynamic progression of the GINS complex during the cell cycle using whole genome time-resolved chromatin immunoprecipitation combined with microarray analysis (Sekedat et al. 2010). Previous data were interpreted to indicate that the dynamics of fork progression are strongly affected by local chromatin structures or macromolecular protein architecture. Interestingly, they showed that GINS spreads bidirectionally and symmetrically from active replication origins independent of genome location, with a highly uniform rate of 1.6 ± 0.3 kb/min, revealing that replication fork dynamics in yeast are simpler and more uniform than previously envisaged. Direct measurement of a fork specific associated protein allowed simulations for progression of the RPC and estimation of origin firing efficiencies.

8.4 Structure of GINS

8.4.1 Overall Structure

The structure of the human GINS heterotetramer has been solved at 2.3–3.0 Å resolution by three groups using similar crystallization conditions (PDB codes: 2E9X, 2EHO and 2Q9Q) (Kamada et al. 2007; Choi et al. 2007; Chang et al. 2007). Although the four full-length subunits of the complex form a stable 1:1:1:1 heterotetrameric complex when they are expressed simultaneously in *Escherichia coli*,

mass spectrometry studies revealed that the C-terminal region of Psf1 is sensitive to various proteases. Two of the three groups solved the structure of a GINS complex containing a C-terminally truncated version of Psf1 (Choi et al. 2007; Kamada et al. 2007). The third group used the full-length Psf1 but the resulting electron density map of the C-terminal region was disordered (Chang et al. 2007).

The structure of the GINS heterotetramer resembles a trapezoid with dimensions of ~ 60 Å deep, ~ 65 Å high, and ~ 110 Å wide (Fig. 8.2a). Each of the subunits occupies roughly a quarter of the trapezoid. An Sld5 and Psf1 heterodimer forms the top layer, and Psf2 and Psf3 the bottom layer, creating a pseudo-dyad axis in the center of the molecule. Sld5 and Psf2 are arranged roughly along the axis, sharing a horizontal interface (Fig. 8.2a). Psf1 and Psf3 show a similar orientation. There are few diagonal contacts between Sld5 and Psf3 and between Psf1 and Psf2. At the center of the complex is a narrow cavity (Fig. 8.2b). Calculation of the electrostatic potential of the molecular surface showed that it is highly acidic.

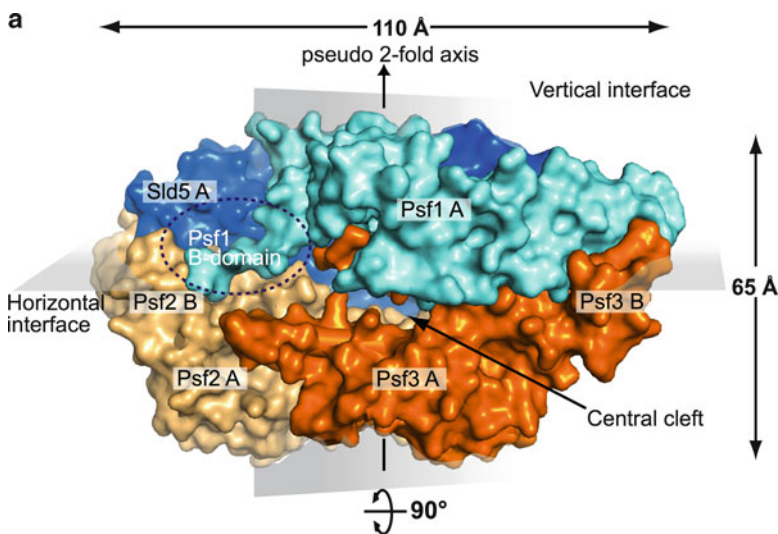


Fig. 8.2 Overall structure and subunit arrangement of the tetrameric GINS complex. (a) Molecular surface representation of the heterotetramer, viewed perpendicular to the pseudo-two fold axis of the complex. Two major (vertical and horizontal) subunit-interfaces are roughly shown by *grey planes*. (b) The molecular surface cut open along the horizontal interface, is viewed from the upper direction of the pseudo-two fold axis of (a), to show an internal part of the central cleft. A flat vertical interface between the A-domains of Psf2 and Psf3 is seen clearly. The width of the flat entrance of the cleft is about 22 Å, while the bottom of the cleft is almost closed by the linker region of the Sld5 B-domain, with a 6 Å pore remaining. (c) Cartoon representation viewed from the opposite side of that seen in (a). Mutation sites of the GINS subunits found in yeasts (*Sc*, *S. cerevisiae*; *Sp*, *S. pombe*) are indicated by *arrows* (Gomez et al. 2005; Takayama et al. 2003; Yabuuchi et al. 2006). All the sites are located in subunit interfaces or the internal core of subunits

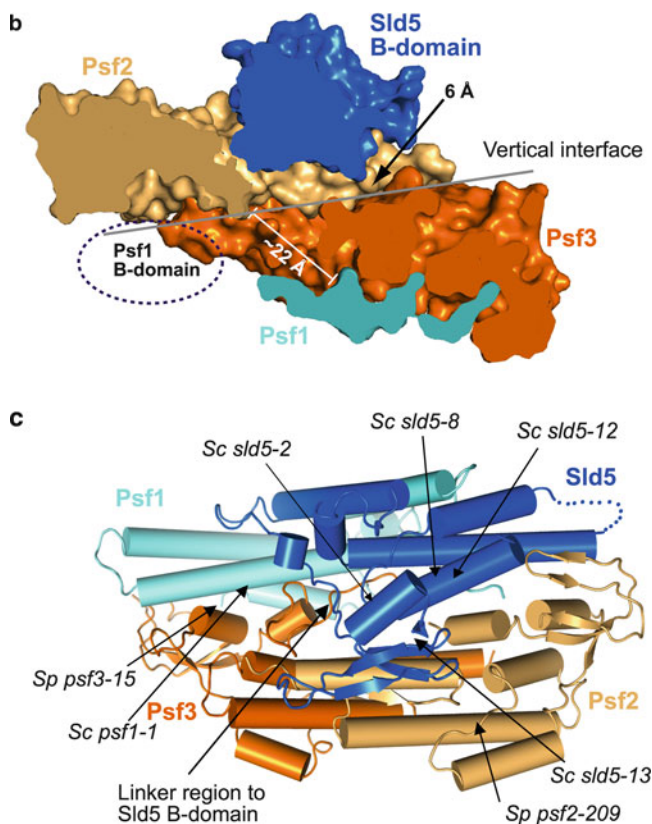


Fig. 8.2 (continued)

8.4.2 Two Structural Domains in All Subunits

Each subunit is composed of two distinct structural domains: an α -helix-rich (A)-domain and a β -strand-rich (B)-domain (Fig. 8.3a). As predicted by sequence analysis, the two structural domains occur in the order A-B in Sld5 and Psf1, but in the reverse order in Psf2 and Psf3 (Fig. 8.3b). Superposition of the individual domains in the subunits shows that each domain shares a common fold, suggesting that the truncated C-terminal region of Psf1 that is lacking in the crystal structures has a similar folding to the B-domain. All the A-domains contain five α -helices. The fourth element of Psf2 exceptionally adopts a β -strand conformation to interact with the Sld5 subunit (Fig. 8.3b). The B-domain has an extended N-terminal region containing a one α -helical turn, followed by a globular core comprised of an α -helix and two small antiparallel β -sheets.

Notably, structure-based sequence alignment reveals that the A-domain of each subunit contains two perfectly conserved residues, namely arginine in the third helix and glutamate in the fifth helix (Fig. 8.3b) (Kamada et al. 2007). These two residues are located in an internal portion of the A-domain and form a bidentate hydrogen bond

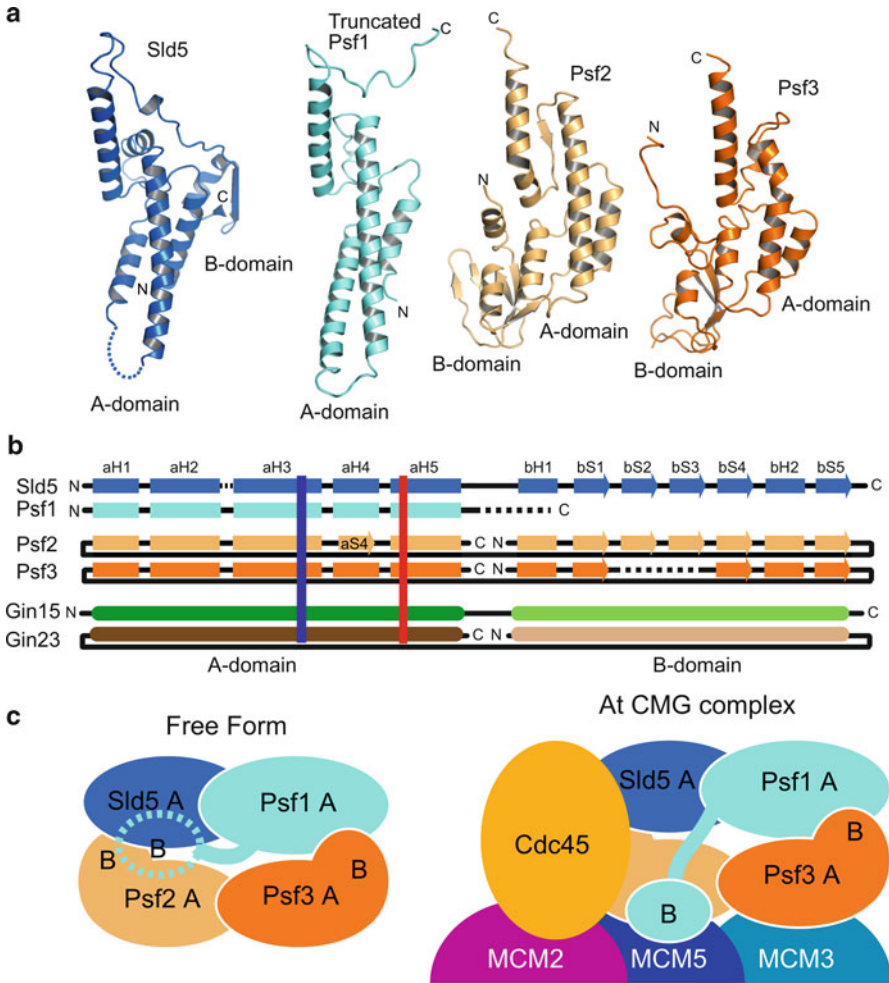


Fig. 8.3 Structures and organization of subunits, and a functional model of the complex. (a) Ribbon drawings of the four subunits of the solved structure. In Sld5, two A- and B-domains are almost separated with about 20 residues of a linker region. In contrast, the two domains of Psf2 and Psf3 are tightly bound to each other by hydrophobic interactions. (b) Schematic diagrams of secondary structural elements and domains of eukaryotic and archaeal GINS subunits. The order of the two domains has been inverted in the course of evolution creating two subfamilies (A-B and B-A types). The locations of the arginine and glutamate residues within the A-domain, which are perfectly conserved, are shown by vertical blue and red bars, respectively, forming hydrogen bonds with each other. (c) In solution, the B-domain of Psf1 is not involved in the formation of the tetrameric complex. GINS contributes to the formation of the CMG complex through in the interaction of the Psf1 C-terminal B-domain and the peripheral surface of the tetrameric complex

within the hydrophobic environment. The pairing of these residues neutralizes their opposing electrostatic forces and creates a hydrophobic plane, which contributes to the stability of the interior protein structure. The conservation of these residues in all the subunits can be considered a structural vestige, and indicates that GINS

subunits evolved from a common ancestral gene module. In fact, the hydrogen bond network followed by these residues is crucial to the fundamental structure of the A-domain. Mutation of a conserved arginine residue of Psf2 in fission yeast (R133K; Arg124 in human) causes the temperature-sensitive phenotype of the *psf2-209* allele (Gomez et al. 2005). This mutation could disturb the formation of the bidentate hydrogen bond and destabilize the subunit structure (Fig. 8.2c).

Two vertical and horizontal interfaces significantly contribute to heterotetramer formation (Fig. 8.2a). The vertical interface is created by heterodimerization of the A-domains of Sld5 and Psf1 and the A-domains of Psf2 and Psf3. The third and fifth helices of each pair of subunits are related by the pseudo-dyad axis and form a four-helix bundle (Fig. 8.2b). In contrast, the B-domains of Psf2 and Psf3 create the horizontal interface of the top and bottom layers through their interactions with the first and third helices of the Sld5 and Psf1 A-domains. At this interface, the arginine residues of Sld5 or Psf1 plug into hydrophobic patches on the B-domain surface of Psf2 or Psf3, resulting in the stabilization of the horizontal interface. The effects of disrupting this interface can be seen with the budding yeast *psf1-1* mutant allele (Takayama et al. 2003). Substitution of Arg84 in yeast Psf1 (Arg74 in human) with glycine confers lethality at non-permissive temperatures and severely disturbs complex formation (Fig. 8.2c).

The two B domains of Sld5 and Psf1 play functionally distinct roles. The C-terminal B-domain of Sld5 is embedded between the Sld5 and Psf2 A-domains and helps stabilize the tetrameric complex (Fig. 8.2b). The C-terminus of this domain is buried inside completely, and mutations or additional tags at/near this domain destabilize the complex (Fig. 8.2c). Deletion of the Sld5 B-domain also severely weakens the horizontal interface between the top and bottom layers of the complex without affecting the vertical interfaces between Sld5 and Psf1 or between Psf2 and Psf3 (Kamada et al. 2007). In contrast, the C-terminal region of Psf1 is dispensable for tetramer formation. Biochemical experiments indicate that this domain does not interact stably with a tetrameric complex composed of the B-domain-less Psf1 and full-length Sld5, Psf2 and Psf3, suggesting that the Psf1 B-domain is simply tethered by its linker region to the tetrameric core complex in solution (Kamada et al. 2007).

8.4.3 Functional Interface of the GINS Complex

Addition of recombinant human GINS to *Xenopus* egg extracts immunodepleted of endogenous GINS restores full replication activity only when full-length Psf1 is present; complexes containing C-terminally truncated Psf1 do not replicate sperm chromatin, even with the addition in *trans* of the recombinant Psf1 B-domain (Kamada et al. 2007). Moreover, the linker region tethering the Psf1 B-domain is crucial for replication activity. In contrast, an unstable mutant complex lacking the Sld5 B-domain shows weak but nonetheless detectable activity. This non-stoichiometric complex is the result of the weakening of horizontal-interface interactions. This unstable mutant

is quantitatively equivalent to thermosensitive mutants previously isolated from yeast (Takayama et al. 2003), indicating that the presence of the Psf1 B-domain is more critical than the stability of the overall complex. Chromatin binding experiments also show that this linker peptide region to the Psf1 B-domain is important for the initiation process, consistent with previous observations of the interdependent chromatin binding of GINS and Cdc45 (Kubota et al. 2003). The Psf1 B-domain is not stably anchored on the surface of the tetrameric core complex, and the conserved hydrophobic residues of the linker to the B-domain are functionally important. Sequence conservation in the GINS complex maps to the surface where the B-domain of Psf1 is most likely adjacent. Consistent with this explanation, results of a complementation assay using point mutations on the surface of the yeast GINS complex proposed that several patches critical for cell viability are on the similar peripheral surface of Sld5 and Psf2 (Choi et al. 2007). It is possible that the Psf1 B-domain is a functional interface for association with Cdc45, the MCM complex or both on chromatin (Fig. 8.3c). It seems reasonable that conformational changes in the linker peptide would allow association with protein(s) that target the GINS complex.

8.4.4 GINS and the CMG Complex

The GINS complex has been isolated biochemically with Cdc45 and MCM2-7; together these are referred to as the CMG (Cdc45-MCM-GINS) complex. Botchan and co-workers purified the complex from *Drosophila* embryo extracts by following Cdc45-containing fractions (Moyer et al. 2006). They found a stable complex (of predicted molecular weight of 711 kDa) containing Cdc45, the MCM2-7 hexamer and GINS that was associated with ATP-dependent DNA helicase activity. The amount of the complex was highly limited, containing only ~5% of the total Cdc45 and GINS proteins and ~1% of the total MCM proteins in the extracts. The authors also reconstituted *Drosophila* MCM2-7 and CMG complexes using the baculovirus expression system, and tested them for helicase activity (Ilves et al. 2010), showing that the recombinant CMG complex possesses significantly enhanced helicase activity on a circular DNA template, and that the hexameric MCM motor does not function as a simple summation of six equally contributing ATPase domains. A recent single-particle electron microscopy (EM) study produced low resolution (~30 Å) structures of the *Drosophila* MCM and CMG complexes, revealing their overall architecture and providing insights into subunit-subunit contacts (Costa et al. 2011). GINS and Cdc45 are located on the surface of the MCM ring and bridge a gap between MCM2 and MCM5 (discussed in detail in Chap. 7, this volume). GINS makes extensive contacts to Cdc45 via the A-domain of Psf2, and is anchored to the N-terminal domains of MCM5 and MCM3 via contacts with the underneath of the A-domains of Psf2 and Psf3, respectively (Fig. 8.3c). Judged by density comparison depending on the nucleotide status of CMG, the B-domain of Psf1 is likely involved in binding to AAA+ domain of MCM5, which is consistent with biochemical results discussed above (Sect. 8.4.3).

8.4.5 EM Images and DNA Clamping Action

Prior to the publication of the crystal structures of the human GINS complex, two EM studies of the GINS structure were reported. In the original study, the image of a rotary shadowed recombinant *Xenopus* GINS complex appears as a ring-shaped structure with a central hole with a diameter of ~ 40 Å (Kubota et al. 2003). In the second study, single-particle EM with three-dimensional reconstruction at 33 Å resolution showed recombinant human GINS as a C-shaped complex (Boskovic et al. 2007). However, high-resolution crystallographic data show a tapered central cleft (Fig. 8.2b), which is visually different from the images of the complex in the EM studies. This may reflect the limitations of visualization technique of EM for macromolecular analysis. The previous impressions of a ring-like structure can be accounted for by the low density distribution of electrons at the center of the complex.

The impact of EM images led to the speculation that GINS has a DNA clamp-like function similar to that of PCNA (proliferating cell nuclear antigen, see Chap. 15) (Kubota et al. 2003). Crystallographic studies revealed that the inner space is small and that its surface has high negative electrostatic potential with low sequence conservation, which makes the cleft unsuitable for the accommodation of even ssDNA. Because the subunit interfaces of the complex consist mostly of tight hydrophobic interactions, rearrangement of the subunit orientation to make the central cleft wider is impossible without huge energy input. On the other hand, it has been proposed that unplugging the Psf3 N-terminal peptide from the complex would allow the cleft to become a wider cavity (Chang et al. 2007). However, the peptide region is stably bound to the Sld5 subunit inside the cavity by hydrophobic interactions with low crystallographic temperature factors. DNA binding activity of the N-terminal peptide-less complex has not been shown conclusively to date.

The DNA binding ability of the tetrameric GINS complex is somewhat controversial. Boskovic et al. used electrophoretic mobility-shift assays (EMSA) to show that a recombinant human GINS complex preferentially binds to single stranded DNA or molecules containing stretches of ssDNA than to a probe consisting of only double stranded DNA (Boskovic et al. 2007). To date, this remains the only evidence for robust GINS-DNA interactions. In contrast, *Pyrococcus furiosus* GINS does not have DNA binding activity (Yoshimochi et al. 2008). Moreover, a PfuMCM band shift is not observed in the presence of PfuGINS, even at micromolar concentrations, indicating that PfuGINS cannot assist the loading of PfuMCM onto DNA. Ilves et al. have observed binding affinity to forked DNA substrates using recombinant *Drosophila* GINS containing full-length or a C-terminally truncated version of Psf3 (Ilves et al. 2010). The affinity of either complex for DNA was considerably weaker compared to that of human GINS. Nonetheless, deletion of the C-terminal basic region of Psf3 alters the DNA binding and helicase activity of CMG, suggesting that it may indirectly enhance the DNA binding properties of the CMG complex.

8.5 Archaeal GINS

8.5.1 Structure and Evolution

The lack of detectable sequence conservation among the subunits of eukaryotic GINS led to the belief that the four subunits were unrelated. Koonin and colleagues first reported archaeal GINS homologues and proposed that the four eukaryotic subunits diverged from a common ancestor (Makarova et al. 2005). Archaea encode two GINS proteins (Fig. 8.3b), which appear to have been derived from each other by swapping of the two small modules (termed A- and B-domains in the crystal structure). This suggests that the last common ancestor of archaea and eukaryotes may have contained two GINS paralogues. Presumably, the four genes encoding the eukaryotic GINS subunits evolved from one ancestral gene via at least two rounds of gene duplication and permutation after an initial round of duplication to produce the reversal of the two domains. Further analyses have demonstrated that all archaeal organisms encode an A-B type protein (variously termed Gins15 or Gins51) related to Sld5 and Psf1 (MacNeill 2010; Yoshimochi et al. 2008). By contrast, the second gene that encodes a B-A type protein (Gins23) related to Psf2 and Psf3 is found only in crenarchaeal species. Most euryarchaeal organisms lack Gins23, indicating that known archaeal Gins23 proteins are already highly divergent. The evolutionary branching point is difficult to identify, implying that Gins15, which is predominantly present in archaea, is functionally more important than Gins23 despite that fact that Gins23 was the original protein that was lost later in the course of evolution.

Co-expression of archaeal Gins15 and Gins23 proteins results in formation of a tetrameric complex comprising a dimer of dimers (Marinsek et al. 2006; Yoshimochi et al. 2008). Recently the crystal structure of the 2:2 type complex from *Thermococcus kodakaraensis* has been reported (Oyama et al. 2011). As expected, this complex adopts a similar architecture to that of the human GINS complex (Fig. 8.4a). The structure preserves the top and bottom layers, composed of Gins15 and Gins23 homodimers. The horizontal interface between the two layers is basically created by the B-domains through interactions with their A-domains, although the B-domain of Gins15 is not involved in stable tetramer formation (Oyama et al. 2011). Considering the roles of the B-domain in eukaryotes, the B-domains of Gins15 may play a role, like B-domain of Psf1, in the archaeal replisome.

In some archaeal genomes, the gene encoding Gins15 is found just downstream to the gene encoding PriS, which is the small subunit of primase, the enzyme responsible for RNA primer synthesis. Swiatek and MacNeill reported the structural similarity of archaeal GINS to PriS, and provided an interesting explanation of domain derivation of archaeal GINS by evolutionary gene transfer (Swiatek and MacNeill 2010). Multiple sequence alignments indicated that the C-terminal domain (CTD) of PriS is conserved in most archaeal lineages, but not in *Thermococcales* and eukarya. A crystal structure of PriS demonstrates that the PriS-CTD seems to be a structural platform for a neighboring zinc-binding domain (Lao-Sirieix et al. 2005)

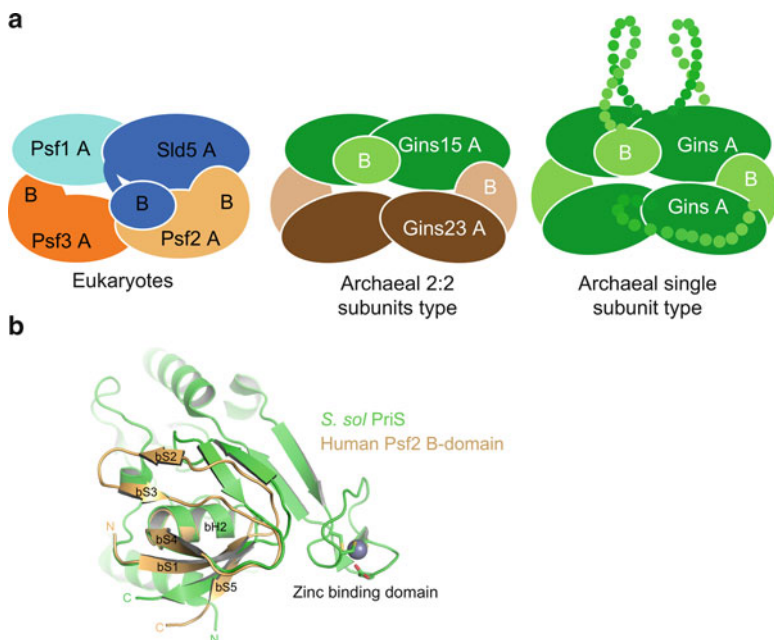


Fig. 8.4 Structure of the archaeal GINS complex. **(a)** Comparison of subunit organization between eukaryotic GINS and two types of archaeal GINS complexes. The archaeal 2:2 subunits complex adopts a similar orientation to that seen in the eukaryote complex. The single subunit type is a hypothetical model with interdomain loops by gradient dots (Oyama et al. 2011). The drawing colors are based on Fig. 8.3b. **(b)** Superimposition of the B-domain of the human Psf2 subunit and the small subunit of the *Sulfolobus solfataricus* primase (PriS, see PDB: 1ZT2) (Lao-Sirieix et al. 2005). The C-terminal domain of PriS has a similar structure to the B-domain seen in GINS subunits. In some archaeal species, the ORF of PriS is often found in a region upstream of that of Gins15, resulting in a line of B-A-B order on the chromosome DNA region

(Fig. 8.4b), while the precise role of the domain is uncertain. The domain of PriS was found to have a similar folding pattern as the B-domain of GINS subunits despite limited sequence similarities (Fig. 8.4b), suggesting that the similarity between the two domains is an evolutionary vestige resulting from the tandem duplication of the Gins15 gene (A-B order). First, linear duplication of a Gins15 ORF could have occurred in the last common archaeo-eukaryotic ancestor. Then, deletion of Gins15 A-domain sequences would have resulted in the fusion of the Primase domain and the remaining B-domain, generating an ORF encoding a PriS protein in the last common archaeal ancestor. However, another possibility might be considered that tandem duplication of a Gins23 gene (B-A order) could have generated PriS-CTD and Gins15, and the subsequent horizontal transfer of a Gins23 ORF by duplication would have resulted in its emergence in archaeal ancestors at a later stage of evolution. More concrete lines of evidence supporting the archaeal gene context are required to draw a definitive conclusion.

8.5.2 *Biological Functions of Archaeal GINS*

The archaeal DNA replication machinery is most likely a simplified version of the eukaryotic system. Therefore, the study of archaeal DNA replication systems should help elucidate eukaryotic replication mechanisms.

Biochemical characterizations of several distinct archaeal GINS complexes have been reported so far. In the crenarchaeote *Sulfolobus solfataricus* (Marinsek et al. 2006), the Gins23 protein interacts with the non-catalytic N-terminal domain of MCM helicase in a yeast two-hybrid system. This interaction was also detected by co-immunoprecipitation experiments in cell extracts of *S. solfataricus*. Gins23 also interacts in the two-hybrid and pull-down assays with both subunits of primase, as seen in human GINS (De Falco et al. 2007). However, addition of the SsoGINS complex does not affect the activity of primase or MCM helicase. Further column purification of Gins23 from the cell extracts identified two interacting factors: the other GINS subunit Gins15 and a protein with similarity to the DNA-binding domain of the bacterial RecJ homologue (RecJdbh). *E. coli* RecJ, which has exonuclease activity on ssDNA, plays an important role in the processing of stalled replication forks (Courcelle et al. 2003). A recent paper also showed that a similar factor is isolated with the ssDNA-specific exonuclease activity in *T. kodakaraensis* (GINS-associated nuclease: GAN) and its specific interaction with Gins15 enhances the nuclease activity (Li et al. 2011). It is an attractive and likely situation that GINS coordinates and regulates such additional factors with MCM at the archaeal replication fork.

The biochemical properties of GINS from the hyperthermophilic archaeon *P. furiosus* have also been investigated (Yoshimochi et al. 2008). Similar to the situation in *S. solfataricus*, recombinant *P. furiosus* Gins15 and Gins23 also form a tetrameric GINS complex at a 2:2 M ratio. Gins23 interacts with the MCM helicase in a yeast two-hybrid assay, and the two proteins are detected as an immunocomplex from *P. furiosus* extracts. Moreover, two-hybrid analysis also showed that Gins15 interacts with the *P. furiosus* Orc1/Cdc6 homologue and CHIP studies using anti-Gins23 antibody demonstrated that the PfuGINS complex associates more dominantly with the *P. furiosus* replication origin during the exponential growth phase compared with the stationary phase. As seen in *S. solfataricus*, purified PfuMCM and PfuGINS cannot form a stable complex even in the presence of DNA. Interestingly, recombinant PfuGINS can stimulate MCM ATPase and helicase activities when present in excess over PfuMCM. The helicase activity of PfuMCM is weaker compared with MCM proteins from other archaeal organisms. Therefore, full activity for archaeal MCM helicase might require other auxiliary factors like GAN.

The GINS complex from the euryarchaeote *Thermoplasma acidophilum* has also been characterized (Ogino et al. 2011). *T. acidophilum* encodes a Gins15-type subunit only. Interestingly, this protein exists as homotetramer and also physically interacts with its cognate MCM. How can this GINS form the tetrameric structure without the Gins23 subunit? EM studies of the complex answer this question

(Oyama et al. 2011). The Gins15 subunit contains a long intervening loop structure between the A and B domains, which is potentially intrinsically disordered. Spatial domain relocation of the single subunit by the loop could allow the creation of different domain configurations mimicking Gins23 to form the tetrameric complex (Fig. 8.4a).

8.6 Conclusions and Prospects

Since the GINS complex belatedly appeared on the stage of the replication field in 2003, GINS has been recognized as one of the most ancient component of the replisome in eukaryotes and archaea. GINS undeniably plays a central role as the chromosome replicative helicase, the CMG complex, and is fundamentally important for both the establishment and progression at DNA replication forks. Although, to date, the enzyme activity of the GINS tetrameric complex has not been reported, GINS and Cdc45 seem to play as a gate keeper for the activity of the replicative helicase. Therefore, further investigation of the function of this complex is important for understanding of the eukaryotic replication process.

In the initiation step, the heterotetrameric GINS complex is integrated with Cdc45 to allow MCM2-7 to be an active form of helicase by a certain mechanism. This process not only depends on structural protein factors, but also on the local concentration of S-phase protein kinases. Does GINS also participate in a phosphorylation event? Combinatory proteomics characterized human Psf2 as a protein phosphorylated by the ATM and ATR protein kinases, two major signal amplifiers for the DNA-damage response (Matsuoka et al. 2007). In Psf2, two sites of phosphorylation are listed: Thr180 and Ser182 in the extreme C-terminal region. Considering the low sequence conservation of this region, the effect of these phosphorylation sites on GINS function is not clear. It would be of interest to assess whether phosphor-mimic mutations have an effect on cell viability under DNA damage induced conditions. In contrast, whether GINS acts as an acceptor of phosphopeptides like the BRCT domain remains to be investigated. Crystal structures revealed the presence of several sulfate molecules with weak density on the surface of the complex (Choi et al. 2007; Kamada et al. 2007). These compounds are used as a precipitant for crystallization, and their binding to protein surfaces sometimes mimics phosphate binding. It would be interesting to reconsider the possibility of phosphopeptide binding in replication initiation.

In the elongation stage, GINS functions as a part of the CMG complex that unwinds DNA during S-phase. How GINS and Cdc45 mechanistically contribute to the function of CMG? Incorporation of GINS with Cdc45 to MCM enables efficient loading to circular DNA, suggesting that these auxiliary factors create a gate on the MCM2-7 ring (see Chap. 7). In fact, GINS partially contributes to the DNA binding affinity of the CMG complex (Ilves et al. 2010). EM studies of the CMG complex also suggest that these factors direct ssDNA to the inside channel of the MCM2-7 ring (Costa et al. 2011). In correlation with the change in overall structure, the

helicase activity of MCM2-7 in the CMG complex is functionally asymmetric, based on mutations of the Walker A motif of individual subunits (Ilves et al. 2010). In circular homohexameric helicases, crystallographically distinct NTP-binding sites share several different configurations that allosterically affect the DNA-binding loops, allowing for ssDNA movement within a subunit (Enemark and Joshua-Tor 2008). In the CMG complex, such a rotary translocation mechanism might not be used; instead, repetitive actions mainly driven by specific NTP-binding subunits might apply. It is possible that the asymmetric structure of the CMG caused by incorporation of GINS and Cdc45 facilitates biased engagement of the loops coupled with the specific subunits that effectively utilize energy from ATP hydrolysis. Consequently, coordinated escort of ssDNA by the motivated loop might be achieved in the central channel of the CMG complex. The inherent function of GINS is veiled by the large structure of the MCM helicase. Determination of the high resolution structure of the CMG complex bound to DNA is a high priority and would provide an important insight for understanding of the mechanism of action of the replisome core.

In archaea, primitive forms of the GINS complex are employed in the replication machinery. The structure of an archaeal GINS complex not only represents the fundamental architecture but also provides interesting aspects for domain configurations of a single-subunit form of archaeal GINS. It is highly intriguing how an associated factor like GAN and RecJdbh are structurally involved in the replication assembly. Unraveling archaeal replisome components with simplified structures is important to understand the architecture of replication machineries in higher eukaryotes.

Acknowledgments I am grateful to T. Hirano of the Hirano laboratory for critical reading of this manuscript. Funding was obtained through a Grant-in-Aid for Scientific Research(C) (21570125).

References

- Aparicio T, Guillou E, Coloma J, Montoya G, Mendez J (2009) The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication. *Nucleic Acids Res* 37:2087–2095
- Araki H, Leem SH, Phongdara A, Sugino A (1995) Dpb11, which interacts with DNA polymerase II (ϵ) in *Saccharomyces cerevisiae*, has a dual role in S-phase progression and at a cell cycle checkpoint. *Proc Natl Acad Sci USA* 92:11791–11795
- Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem* 71:333–374
- Boskovic J, Coloma J, Aparicio T, Zhou M, Robinson CV, Mendez J, Montoya G (2007) Molecular architecture of the human GINS complex. *EMBO Rep* 8:678–684
- Chang YP, Wang G, Bermudez V, Hurwitz J, Chen XS (2007) Crystal structure of the GINS complex and functional insights into its role in DNA replication. *Proc Natl Acad Sci USA* 104:12685–12690
- Choi JM, Lim HS, Kim JJ, Song OK, Cho Y (2007) Crystal structure of the human GINS complex. *Genes Dev* 21:1316–1321
- Chong JP, Thommes P, Rowles A, Mahubani HM, Blow JJ (1997) Characterization of the *Xenopus* replication licensing system. *Methods Enzymol* 283:549–564
- Costa A, Ilves I, Tamberg N, Petojevic T, Nogales E, Botchan MR, Berger JM (2011) The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol* 18:471–477

- Courcelle J, Donaldson JR, Chow KH, Courcelle CT (2003) DNA damage-induced replication fork regression and processing in *Escherichia coli*. *Science* 299:1064–1067
- De Falco M, Ferrari E, De Felice M, Rossi M, Hubscher U, Pisani FM (2007) The human GINS complex binds to and specifically stimulates human DNA polymerase α -primase. *EMBO Rep* 8:99–103
- Diffley JF (2004) Regulation of early events in chromosome replication. *Curr Biol* 14:R778–R786
- Dohmen RJ, Wu P, Varshavsky A (1994) Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science* 263:1273–1276
- Enemark EJ, Joshua-Tor L (2008) On helicases and other motor proteins. *Curr Opin Struct Biol* 18:243–257
- Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, Labib K (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol* 8:358–366
- Gambus A, van Deursen F, Polychronopoulos D, Foltman M, Jones RC, Edmondson RD, Calzada A, Labib K (2009) A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase α within the eukaryotic replisome. *EMBO J* 28:2992–3004
- Gomez EB, Angeles VT, Forsburg SL (2005) A screen for *Schizosaccharomyces pombe* mutants defective in rereplication identifies new alleles of *rad4+*, *cut9+* and *psf2+*. *Genetics* 169:77–89
- Ilves I, Petojevic T, Pesavento JJ, Botchan MR (2010) Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell* 37:247–258
- 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* 14:388–396
- Kamimura Y, Masumoto H, Sugino A, Araki H (1998) Sld2, which interacts with Dpb11 in *Saccharomyces cerevisiae*, is required for chromosomal DNA replication. *Mol Cell Biol* 18:6102–6109
- Kamimura Y, Tak YS, Sugino A, Araki H (2001) Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in *Saccharomyces cerevisiae*. *EMBO J* 20:2097–2107
- Kanemaki M, Labib K (2006) Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. *EMBO J* 25:1753–1763
- Kanemaki M, Sanchez-Diaz A, Gambus A, Labib K (2003) Functional proteomic identification of DNA replication proteins by induced proteolysis *in vivo*. *Nature* 423:720–724
- 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* 17:1141–1152
- Labib K, Tercero JA, Diffley JF (2000) Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science* 288:1643–1647
- Lao-Sirieix SH, Nookala RK, Roversi P, Bell SD, Pellegrini L (2005) Structure of the heterodimeric core primase. *Nat Struct Mol Biol* 12:1137–1144
- Li Z, Pan M, Santangelo TJ, Chemnitz W, Yuan W, Edwards JL, Hurwitz J, Reeve JN, Kelman Z (2011) A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res* 39:6114–6123
- MacNeill SA (2010) Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochem J* 425:489–500
- Makarova KS, Wolf YI, Mekhedov SL, Mirkin BG, Koonin EV (2005) Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res* 33:4626–4638
- Marinsek N, Barry ER, Makarova KS, Dionne I, Koonin EV, Bell SD (2006) GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep* 7:539–545
- Masumoto H, Sugino A, Araki H (2000) Dpb11 controls the association between DNA polymerases α and ϵ and the autonomously replicating sequence region of budding yeast. *Mol Cell Biol* 20:2809–2817

- Masumoto H, Muramatsu S, Kamimura Y, Araki H (2002) S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. *Nature* 415:651–655
- Matsuno K, Kumano M, Kubota Y, Hashimoto Y, Takisawa H (2006) The N-terminal noncatalytic region of *Xenopus* RecQ4 is required for chromatin binding of DNA polymerase alpha in the initiation of DNA replication. *Mol Cell Biol* 26:4843–4852
- Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER 3rd, Hurov KE, Luo J, Bakalarski CE, Zhao Z, Solimini N, Lerenthal Y, Shiloh Y, Gygi SP, Elledge SJ (2007) ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 316:1160–1166
- Miles J, Formosa T (1992) Evidence that POB1, a *Saccharomyces cerevisiae* protein that binds to DNA polymerase alpha, acts in DNA metabolism *in vivo*. *Mol Cell Biol* 12:5724–5735
- Moyer SE, Lewis PW, Botchan MR (2006) Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci USA* 103:10236–10241
- Muramatsu S, Hirai K, Tak YS, Kamimura Y, Araki H (2010) CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol ϵ , and GINS in budding yeast. *Genes Dev* 24:602–612
- Ogino H, Ishino S, Mayanagi K, Haugland GT, Birkeland NK, Yamagishi A, Ishino Y (2011) The GINS complex from the thermophilic archaeon, *Thermoplasma acidophilum* may function as a homotetramer in DNA replication. *Extremophiles* 15:529–539
- 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
- Pacek M, Tutter AV, Kubota Y, Takisawa H, Walter JC (2006) Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol Cell* 21:581–587
- Pai CC, Garcia I, Wang SW, Cotterill S, Macneill SA, Kearsley SE (2009) GINS inactivation phenotypes reveal two pathways for chromatin association of replicative α and ϵ DNA polymerases in fission yeast. *Mol Biol Cell* 20:1213–1222
- Picard D (2000) Posttranslational regulation of proteins by fusions to steroid-binding domains. *Methods Enzymol* 327:385–401
- Sangrithi MN, Bernal JA, Madine M, Philpott A, Lee J, Dunphy WG, Venkitaraman AR (2005) Initiation of DNA replication requires the RECQL4 protein mutated in Rothmund-Thomson syndrome. *Cell* 121:887–898
- Sekedat MD, Fenyo D, Rogers RS, Tackett AJ, Aitchison JD, Chait BT (2010) GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Mol Syst Biol* 6:353
- Shikata K, Sasa-Masuda T, Okuno Y, Waga S, Sugino A (2006) The DNA polymerase activity of Pol ϵ holoenzyme is required for rapid and efficient chromosomal DNA replication in *Xenopus* egg extracts. *BMC Biochem* 7:21
- Spencer F, Gerring SL, Connelly C, Hieter P (1990) Mitotic chromosome transmission fidelity mutants in *Saccharomyces cerevisiae*. *Genetics* 124:237–249
- Swiatek A, MacNeill SA (2010) The archaeo-eukaryotic GINS proteins and the archaeal primase catalytic subunit PriS share a common domain. *Biol Direct* 5:17
- Tak YS, 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* 25:1987–1996
- 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* 17:1153–1165
- 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* 445:328–332
- Tanaka H, Katou Y, Yagura M, Saitoh K, Itoh T, Araki H, Bando M, Shirahige K (2009) Ctf4 coordinates the progression of helicase and DNA polymerase alpha. *Genes Cells* 14:807–820

- Tanaka T, Umemori T, Endo S, Muramatsu S, Kanemaki M, Kamimura Y, Obuse C, Araki H (2011) Sld7, an Sld3-associated protein required for efficient chromosomal DNA replication in budding yeast. *EMBO J* 30:2019–2030
- Tumini E, Plevani P, Muzi-Falconi M, Marini F (2011) Physical and functional crosstalk between Fanconi anemia core components and the GINS replication complex. *DNA Repair (Amst)* 10:149–158
- Yabuuchi H, Yamada Y, Uchida T, Sunathvanichkul T, Nakagawa T, Masukata H (2006) Ordered assembly of Sld3, GINS and Cdc45 is distinctly regulated by DDK and CDK for activation of replication origins. *EMBO J* 25:4663–4674
- 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* 283:1601–1609
- Zegerman P, Diffley JF (2007) Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* 445:281–285

Chapter 9

The Pol α -Primase Complex

Luca Pellegrini

Abstract Initiation of DNA synthesis in eukaryotic replication depends on the Pol α -primase complex, a multi-protein complex endowed with polymerase and primase activity. The Pol α -primase complex assembles the RNA-DNA primers required by the processive Pol δ and Pol ϵ for bulk DNA synthesis on the lagging and leading strand, respectively. During primer synthesis, the primase subunits synthesise *de novo* an oligomer of 7–12 ribonucleotides in length, which undergoes limited extension with deoxyribonucleotides by Pol α . Despite its central importance to DNA replication, little is known about the mechanism of primer synthesis by the Pol α -primase complex, which comprises the steps of initiation, ‘counting’ and hand-off of the RNA primer by the primase to Pol α , followed by primer extension with dNTPs and completion of the RNA-DNA hybrid primer. Recent biochemical and structural work has started to provide some insight into the molecular basis of initiation of DNA synthesis. Important advances include the structural characterisation of the evolutionarily related archaeal primase, the elucidation of the mechanism of interaction between Pol α and its B subunit and the observation that the regulatory subunit of the primase contains an iron-sulfur cluster domain that is essential for primer synthesis.

Keywords Pol α • Primase • primer

L. Pellegrini (✉)
Department of Biochemistry, University of Cambridge,
80 Tennis Court Road, Cambridge CB2 1GA, UK
e-mail: lp212@cam.ac.uk

9.1 Introduction

In the semi-conservative model of DNA replication each parental strand acts as a template for the synthesis of a complementary DNA chain. DNA polymerases are responsible for synthesising the novel strands of nucleic acid accurately and efficiently. However, polymerisation depends on the presence of an existing oligonucleotide primer annealed to the template that is extended by addition of deoxynucleotides to the 3' hydroxyl group of the primer.

In order to be able to begin duplication of their genome, all organisms have evolved a specialised DNA-dependent RNA polymerase, termed primase, which is endowed with the unique ability to initiate DNA synthesis *de novo* (Frick and Richardson 2001; Kuchta and Stengel 2010). Primases can assemble from the ribonucleotide pool of the cell the short RNA primers that are utilised by the DNA polymerase for processive synthesis. The antiparallel arrangement of the parental strands and the obligate 5'–3' direction of synthesis of DNA polymerase imply that DNA synthesis needs to be primed repeatedly on the lagging strand. Thus, primase activity is constantly required at the replication fork.

Priming DNA synthesis is a universal requirement of DNA replication and consequently primases are present in all kingdoms of life. Certain important features of primase activity, such as the tendency to initiate with a purine base and to synthesise RNA primers of between 5 and 10 nucleotides, appear to be universal. However, prokaryotic and eukaryotic primases differ radically in structural organisation and mechanism of primer synthesis. Whereas bacterial primase is a single polypeptide, in eukaryotic cells the primase is a heterodimer of catalytic or small subunit (PriS) and regulatory or large subunit (PriL). In addition, in eukaryotic cells the primase is normally associated in a specific, constitutive complex with DNA polymerase α (Pol α) and its B subunit (Muzi-Falconi et al. 2003). The resulting heterotetramer forms the Pol α -primase complex, the multi-subunit protein assembly that initiates DNA synthesis in eukaryotic replication. In the complex, the RNA primer synthesised by the primase undergoes limited extension with deoxynucleotides by Pol α . The resulting RNA-DNA primer is then utilised for processive synthesis by Pol ϵ and Pol δ on the leading and lagging strands, respectively.

The first reports describing the isolation of primase activity and the biochemical characterisation of the Pol α -primase complex in cell extracts of eukaryotic organisms date to almost 30 years ago (Conaway and Lehman 1982; Hubscher 1983; Kaufmann and Falk 1982; Plevani et al. 1984; Yagura et al. 1982). Despite extensive experimental efforts, our understanding of the molecular steps of RNA primer initiation, completion and limited elongation with dNTPs by the Pol α -primase complex is still surprisingly limited. In this chapter, I will review the current state of knowledge concerning the structural information currently available for the Pol α -primase complex and describe how the present data inform our understanding of the mechanism of primer synthesis in eukaryotic replication.

9.2 Primase

9.2.1 Prim Fold of the Catalytic Subunit

The eukaryotic primase is a heterodimer of catalytic or small subunit (PriS) and regulatory or large (PriL) subunit. Most of our knowledge of its three-dimensional architecture comes from crystallographic analysis of evolutionarily related archaeal primases, which show clear sequence similarity and subunit organisation with their eukaryotic counterparts.

To date, two crystal structures for the isolated catalytic subunit of archaeal primases have been reported (Augustin et al. 2001; Ito et al. 2003) (Fig. 9.1a). The crystallographic analysis showed that the core structural elements of the catalytic subunit fold in a novel DNA-dependent RNA polymerase domain, the ‘prim’ fold, that has at its core two flanged beta sheets accommodating the active site and surrounded by alpha helices on the outside. Overall, the structure of the catalytic

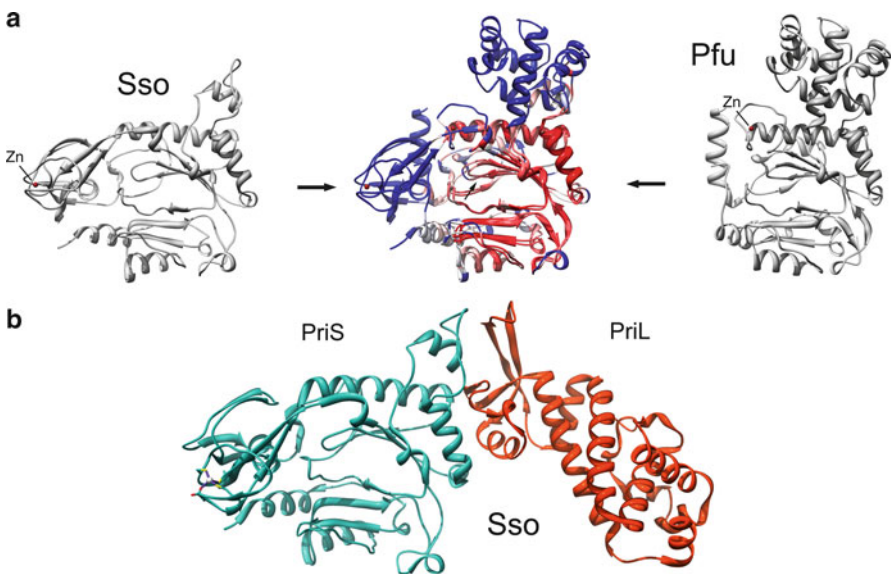


Fig. 9.1 Structure of the archaeal/eukaryotic primase. (a) Ribbon diagrams of the catalytic subunits of archaeal primases from *Sulfolobus solfataricus* (Sso, PDB 1zt2) and *Pyrococcus furiosus* (Pfu, PDB 1 g71) are shown in grey. In the middle of the panel, a superposition of the two structures is shown, coloured according to structure similarity: parts of the structure of the catalytic subunits that are similar are drawn in red, whereas divergent regions are drawn in blue. The colouring scheme highlights the elements of secondary structure that constitute the ‘prim’ fold. (b) Ribbon representation of the crystal structure of the *Sulfolobus solfataricus* primase in its heterodimeric form. The catalytic subunit (PriS) is drawn in green and the large subunit (PriL) in orange. The PriL polypeptide present in the crystal is a C-terminally truncated version that lacks the Fe-S cluster domain (see Fig. 9.2)

subunit has a flat, slab-like appearance with a very exposed active site, entirely different from the classic 'right-hand' fold of DNA polymerases. The structure of the catalytic subunit is completed by a smaller, species-specific domain that can occupy different positions on the rim of the core 'prim' fold (Fig. 9.1a).

Although the archaeal/eukaryotic primase catalytic subunit fold is novel, the mechanism of nucleotide polymerization is likely to be similar to the general enzymatic mechanism of nucleic acid synthesis. Thus, a triad of aspartate residues have been identified that are invariant across archaeal and eukaryotic primases, and which are deemed to be necessary for catalytic activity (Augustin et al. 2001; Ito et al. 2003) (Fig. 9.2a). It therefore appears that convergent evolution has driven primases to adopt the same mechanism of catalysis, involving two divalent metal ions, proposed for other DNA and RNA polymerases (Yang et al. 2006). The position of the active site was confirmed by diffusion of uridine triphosphate (UTP) in the *Pyrococcus horikoshii* primase crystals (Ito et al. 2003) (Fig. 9.2a).

A highly conserved feature of the catalytic subunit is the presence of a zinc-binding motif situated on the same side of the catalytic subunit as the active site; its putative functional role remains uncertain but its proximity to the active site suggests an involvement in the mechanism of catalysis (Fig. 9.1a). Intriguingly, bacterial primases also contain a zinc-binding domain that has been implicated in binding single-stranded DNA (Corn et al. 2005). It is tempting to speculate that the zinc-binding motif in archaeal and eukaryotic primases might fulfill comparable functional roles. However, in contrast to what observed in the archaeal primase, the zinc-binding motif of the bacterial primase is located in a separate domain of the polypeptide chain.

The prim fold is probably of ancient evolutionary derivation as it has been found in a number of different genetic contexts (Iyer et al. 2005), including multifunctional enzymes with combined primase and helicase activities, encoded in the genome of the bacterium *Bacillus cereus* (McGeoch and Bell 2005) and in a plasmid of the archaeon *Sulfolobus islandicus* (Lipps et al. 2003), as well as in certain multi-functional bacterial enzymes active in non-homologous end joining pathway of DNA repair in mycobacteria (Brissett et al. 2007).

9.2.2 The Archaeal/Eukaryotic Primase Is an Iron-Sulfur Protein

The archaeal/eukaryotic primase is a constitutive heterodimer of two subunits, PriS and PriL. The role of the non-catalytic polypeptide in primase activity has remained mysterious for a long time, however it is known that PriL performs an essential role in the priming reaction (Zerbe and Kuchta 2002). Indeed, PriL is an essential gene in yeast (Foiiani et al. 1989).

Recent crystallographic and biochemical analysis of the archaeal/eukaryotic primase has made important progress in understanding the role of PriL in RNA primer synthesis. The structure of a truncated form of the heterodimeric primase from the thermophilic archaeon *S. solfataricus*, lacking the conserved C-terminal portion of

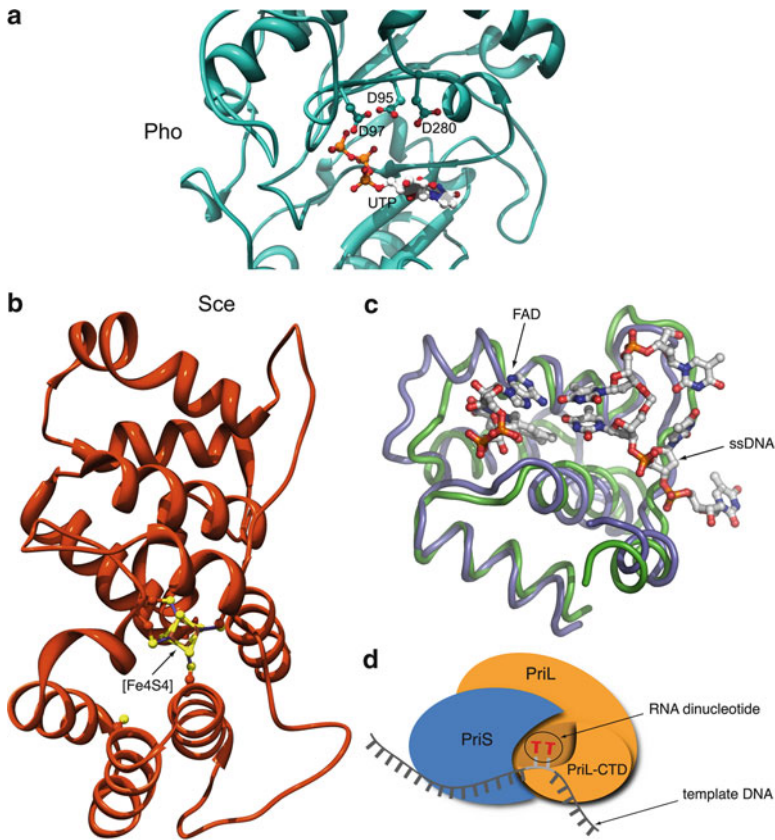


Fig. 9.2 Active site architecture of the archaeal/eukaryotic primase. **(a)** Close-up view of the archaeal primase from *Pyrococcus horikoshii* (PDB 1v34) with a molecule of UTP bound in the active site. The side chains of the aspartate residues forming the putative catalytic triad are also shown. **(b)** Structure of the conserved C-terminal domain of the large subunit of the yeast primase. The [4Fe-4S] cofactor and the cysteine ligands are also shown in ball-and-stick representation. **(c)** Superimposition of the yeast PriL-CTD on the active site of cryptochrome Cry3 from *A. thaliana*, bound to FAD and ssDNA. The PriL-CTD is drawn as a thin tube in blue, the DNA photolyase in green. FAD and ssDNA are shown as sticks, coloured according to element type. **(d)** A possible model for the essential role of the PriL-CTD in initiation of RNA primer synthesis. In the model, the PriL-CTD is an integral part of the active site and assists the catalytic subunit PriS in binding the two initial ribonucleotides at the initiation site on the template DNA

PriL, illuminated the relative special arrangement of the two subunits (Lao-Sirieix et al. 2005) (Fig. 9.1b): PriL is a mainly helical polypeptide that extends from the narrow edge of the slab-shaped catalytic subunit and is better understood as a structural ‘arm’ that serves to position its conserved C-terminal domain (PriL-CTD) in the correct position for catalysis. Indeed, the PriL-CTD is essential in the initiation reaction of the heterodimeric primase, but not for its polymerase activity, pointing to a crucial function of the PriL-CTD in the initial step of di-nucleotide synthesis.

Surprisingly, it was found that PriL-CTD contains a [4Fe-4S] cofactor (Klinge et al. 2007; Weiner et al. 2007). The iron-sulfur centre appears to have a structural role in maintaining the C-terminal sequence of PriL in its correct three-dimensional shape (Agarkar et al. 2011; Sauguet et al. 2010; Vaithiyalingam et al. 2010) (Fig. 9.2b). At this stage, functional roles of the Fe-S cofactor depending on its redox status remain speculative but cannot be ruled out.

The molecular mechanism underlying PriL-CTD's involvement in primer synthesis remains unclear, but an intriguing clue as to how this might happen comes from the unexpected structural similarity between PriL-CTD and the active site of DNA photolyase/cryptochrome family of DNA repair enzymes (Sauguet et al. 2010) (Fig. 9.2c). The mode of binding of single-stranded (ss) DNA and flavin adenine dinucleotide (FAD) observed in the co-crystal structure of the DASH cryptochrome 3 from *Arabidopsis thaliana* (Pokorny et al. 2008) suggests that the PriL-CTD could adopt a similar mode of interaction with the template DNA and ribonucleotides during *de novo* RNA synthesis. In particular, the spacial relationship between FAD and the extruded, cross-linked pyrimidine dimer observed in the active site of DNA photolyase suggests a possible arrangement for the pairing of the first dinucleotide of the RNA primer onto template DNA during the initiation step catalysed by the primase. Thus, the PriL-CTD might participate in RNA primer synthesis by assisting the catalytic subunit PriS in the simultaneous binding of the two initial RNA nucleotides and by promoting dinucleotide base-pairing with template DNA at the initiation site (Fig. 9.2d).

9.3 DNA Polymerase α

9.3.1 Catalytic Activity

In contrast to prokaryotic and bacteriophage replication, where the primer is composed exclusively of RNA, in eukaryotic replication the primer is a hybrid RNA-DNA molecule. The polymerase responsible for synthesising the deoxy-nucleotide portion of the primer is DNA polymerase α (Pol α), a member of the B-family of DNA polymerases and the first DNA polymerase to be detected in mammalian cells. The role of Pol α in priming synthesis is to extend the RNA primer synthesised by the primase with deoxy-nucleotides, in order to assemble an RNA-DNA oligonucleotide of about 20–25 nucleotides. According to current models, transfer of the 3'-hydroxyl of the RNA primer between active sites of the primase and Pol α takes place via an intra-molecular hand-off (Copeland and Wang 1993; Eki et al. 1991; Kuchta et al. 1990; Sheaff et al. 1994). However, the nature of the molecular switch that regulates the transfer is still unknown. The mature RNA-DNA primer is then utilised by DNA Polymerase δ and ϵ for processive synthesis on the lagging and leading strands, respectively (Stillman 2008) (see Chaps. 12 and 13). Upon completion of lagging strand synthesis and joining of the Okazaki fragments, the primer is excised by the combined action of Pol δ and PCNA, which displace the 5' end of the

downstream fragment, and the enzymatic activities of the Fen1 nuclease and the Dna2 nuclease-helicase, which excise 5' flaps of ssDNA (Burgers 2009).

A common feature of all three replicative polymerases – α , δ and ϵ – is the presence of a conserved region that extends past the polymerase fold to a cysteine-rich C-terminal region (CTD) that is necessary for DNA replication and cell viability. A conserved pattern of eight cysteine residues indicated that the CTD of the polymerase likely binds two metal ions, such as Zn^{2+} . Biochemical evidence implicates this region of the polymerase in the interaction with the primase (Mizuno et al. 1999).

9.3.2 Structure of the B Subunit and Its Interaction with Pol α

The three major replicative DNA polymerases – Pols α , δ and ϵ – share unifying features of their subunit organisation that reveal a clear evolutionary relationship (Fig. 9.3a) (Johansson and MacNeill 2010) (see also Chap. 12 and 13). Of their different cohorts of accessory subunits, only the so-called B subunit is present in all three polymerase assemblies and is clearly conserved in eukaryotic organisms (Aravind and Koonin 1998; Makiniemi et al. 1999). Interestingly, an orthologue of the B subunit has also been found in archaeal organisms as the single accessory polypeptide of a replicative polymerase.

Reflecting their high degree of conservation, the catalytic and B subunits are the only indispensable polymerase components. A large body of experimental evidence has highlighted the functional importance of the CTD interaction with the B subunit (Dua et al. 1998; Mizuno et al. 1999; Sanchez Garcia et al. 2004). Thus, a heterodimer of catalytic and B subunit represents the conserved functional core of the three replicative polymerases.

Recent crystallographic evidence has demonstrated the mode of interaction between the CTD of yeast Pol α and its B subunit (Klinge et al. 2009) (Fig. 9.3b). The B subunit fold derives from the intimate association of an N-terminal oligonucleotide/oligosaccharide (OB) domain with an inactive C-terminal phosphoesterase domain. The CTD of yeast Pol α adopts an elongated bilobal shape reminiscent of an asymmetrically proportioned saddle. Each lobe contains a zinc-binding module: the lobe with the four N-terminal cysteine ligands (Zn-1) is larger and includes additional secondary structure elements as well as irregular coil structure; the lobe with the four C-terminal cysteine ligands is smaller and formed entirely by the zinc-binding module (Zn-2). The two lobes are connected by a three-helix bundle that represents the central portion or 'backbone' of the saddle-shaped CTD.

The two zinc-binding motifs bear a clear structural relationship to each other; in both cases, metal-binding results from the 'handshake' interaction of two β ribbons, each providing a pair of cysteine ligands for the tetrahedral coordination of the zinc atom. However, the crystal structure indicates that the two metal-binding motifs take on distinct functional roles. The Zn-2 motif is an integral part of the Pol α CTD – B subunit interface, whereas the Zn-1 motif is removed from the interface

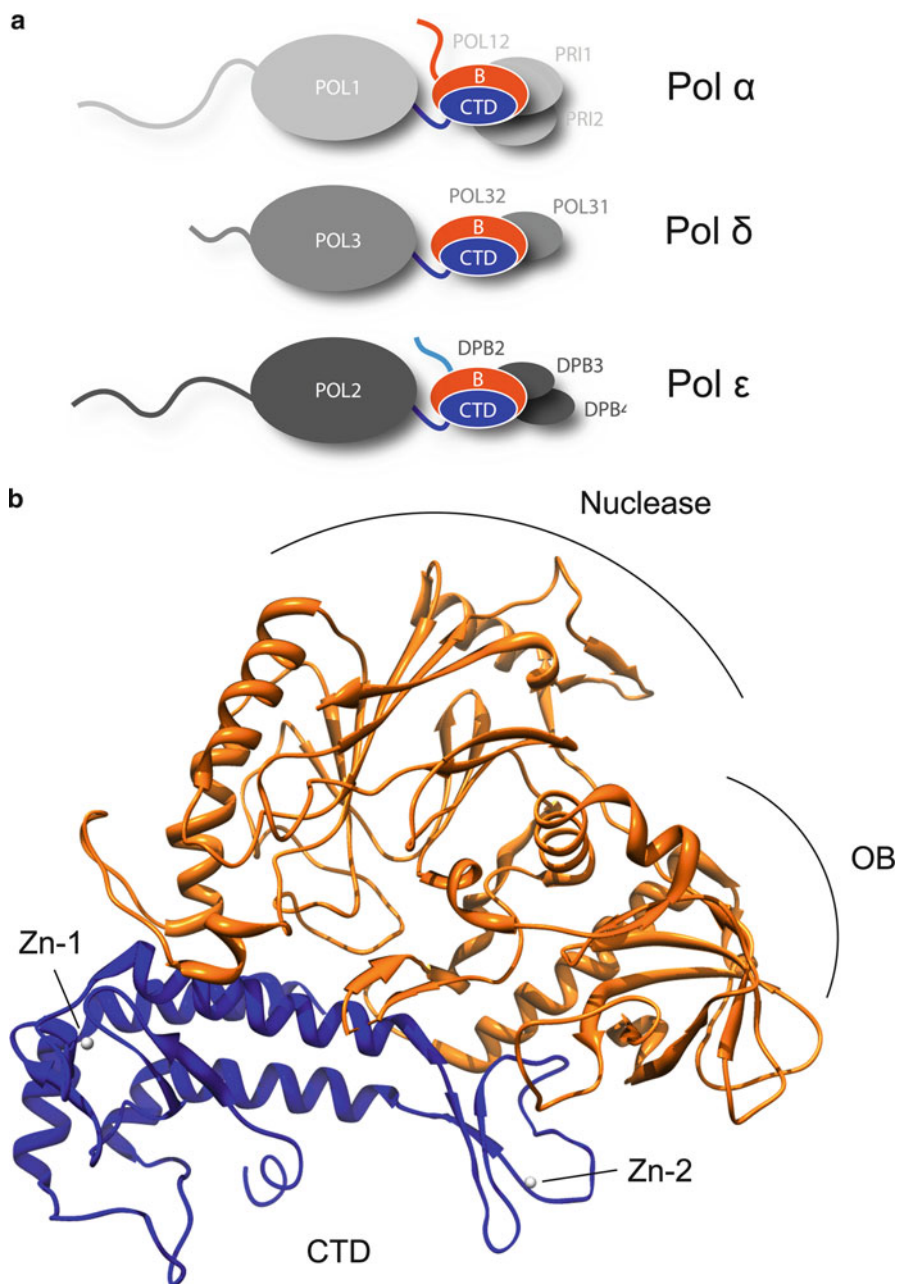


Fig. 9.3 Structure of the yeast Pol α CTD – B subunit complex. **(a)** Cartoons of the three multi-subunit replicative DNA polymerases α , δ and ϵ , highlighting the conserved B subunit (blue) and Pol CTDs (orange). **(b)** Crystal structure of the Pol α CTD – B complex. The B subunit is shown as a orange ribbon, and the Pol α CTD as blue ribbon. The position and extend of the oligonucleotide/oligosaccharide (OB) and inactive nuclease domains are shown, as well as the location of the two zinc atoms

and more exposed to solvent, possibly poised for interactions with other protein factors or DNA.

The large surface area (4,500 Å²) buried at the interface between the Pol α CTD and the B subunit indicates a tight association between the two polypeptides, supporting the notion that the B subunit performs a function that requires its constitutive association to Pol α . The presence of a phosphodiesterase domain juxtaposed to an OB domain suggests that the evolutionary ancestor of the B subunit might have possessed an enzymatic function, perhaps required alongside the polymerase activity of Pol α . In support of this hypothesis, some nuclease activity appears to have been retained by the archaeal orthologue of the eukaryotic B subunit (Jokela et al. 2004; Shen et al. 2004). Thus, it appears that, in the course of evolution of the replication machinery, the B subunit has morphed from a nuclease into a protein scaffold element responsible for mediating multiple and concomitant protein-protein interactions within the eukaryotic replisome (Baranovskiy et al. 2008; Klinge et al. 2009).

9.4 Towards a Concerted Mechanism for Primer Synthesis by the Pol α -Primase Complex

As already mentioned, the nucleic acid primer that begins synthesis in eukaryotic replication is a hybrid RNA-DNA molecule. Virtually nothing is known about the molecular switch that coordinates the hand-off of the nascent primer between active sites of primase and Pol α . However, given the high frequency of initiation events at the replication fork and the necessity to synchronize priming of synthesis with fork progression and processive synthesis of bulk DNA, it seems reasonable to assume that a coordinated, intra-molecular mechanism of primer transfer between primase and Pol α must exist. Indeed, experimental evidence exists in support of such a model (Copeland and Wang 1993; Eki et al. 1991; Kuchta et al. 1990; Sheaff et al. 1994).

The physical association of the primase and polymerase polypeptides provides a physical basis for their functional coupling. Recent electron microscopy reconstructions of the 3D-architecture of the Pol α - B subunit (Klinge et al. 2009) and of the Pol α -primase complex (Nunez-Ramirez et al. 2011) shows that it is organised as a dumbbell-shaped particle with flexibly connected lobes (Fig. 9.4a, b). The catalytic domain of Pol α resides in one lobe, whereas the other lobe is formed by the interaction of the Pol α CTD, the B subunit and the two primase subunits. The high degree of mobility observed in the relative orientation of the two catalytic lobes might be functionally relevant for the hand-off mechanism of the RNA primer between primase and Pol α (Fig. 9.4c). Thus, it is possible to envisage the process of RNA-DNA primer synthesis as requiring a series of transitions between specific conformational conformations of the Pol α -primase, leading in turn to RNA primer synthesis, transfer to and extension with dNTPs by Pol α (Fig. 9.4c). Confirmation of such mechanism will require the structural characterization of the Pol α -primase complex in its key intermediate states during primer synthesis.

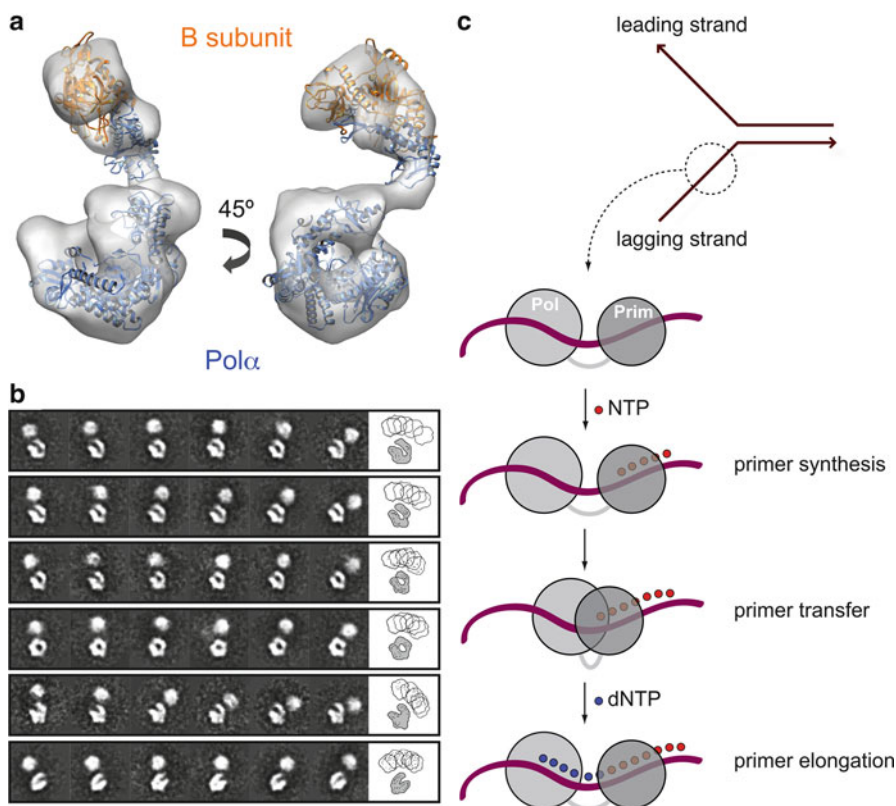


Fig. 9.4 Structure of the accessory B subunit and its interaction with Pol α . **(a)** Ribbon model of the Pol α CTD – B complex structure is fitted into the EM reconstruction of the Pol α – B complex. The catalytic domain of the archaeal polymerase from *T. gorgonarius* (PDB 2VWJ) has also been fitted in the EM density. Two rotated views of the EM reconstruction are shown, represented as white transparent density. The Pol α CTD and the catalytic domain of the archaeal polymerase are shown in *blue* and the B subunit in *orange*. **(b)** Collection of 2D reference-free averages of the Pol α -primase. Each row shows a set of averages where the Pol lobe is oriented according to the same view. The conformational flexibility of the primosome is highlighted by a cartoon outline showing the superposition of all averages in a row. Scale bar represents 15 nm. **(c)** Hypothetical diagram of the steps of RNA–DNA primer synthesis, based on a conformational rearrangement of the Pol and Prim lobes during primer transfer

9.5 Outlook

Elucidating in detail the molecular mechanisms by which the Pol α -primase complex initiates nucleic acid synthesis is essential to our comprehension of DNA replication. Despite its central role in the process of genomic duplication, how the complex assembles the RNA–DNA oligonucleotides that prime DNA synthesis remains surprisingly obscure. The evidence discussed here represents a hopeful

sign that rapid progress will be made in providing a structural basis for the enzymatic activity of the of Pol α -primase complex. Indeed, crystallographic models of archaeal or eukaryotic derivation for almost all the proteins or protein domains of the Pol α -primase complex are now available. The challenge for the future will be to integrate our current structural knowledge in a complete picture of the necessary enzymatic steps underlying RNA–DNA primer synthesis in eukaryotic replication.

References

- Agarkar VB, Babayeva ND, Pavlov YI, Tahirov TH (2011) Crystal structure of the C-terminal domain of human DNA primase large subunit: implications for the mechanism of the primase-polymerase α switch. *Cell Cycle* 10(6):926–931
- Aravind L, Koonin EV (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res* 26(16):3746–3752
- Augustin MA, Huber R, Kaiser JT (2001) Crystal structure of a DNA-dependent RNA polymerase (DNA primase). *Nat Struct Biol* 8(1):57–61
- Baranovskiy AG, Babayeva ND, Liston VG, Rogozin IB, Koonin EV, Pavlov YI, Vassilyev DG, Tahirov TH (2008) X-ray structure of the complex of regulatory subunits of human DNA polymerase δ . *Cell Cycle* 7(19):3026–3036
- Brissett NC, Pitcher RS, Juarez R, Picher AJ, Green AJ, Dafforn TR, Fox GC, Blanco L, Doherty AJ (2007) Structure of a NHEJ polymerase-mediated DNA synaptic complex. *Science* 318(5849):456–459
- Burgers PM (2009) Polymerase dynamics at the eukaryotic DNA replication fork. *J Biol Chem* 284(7):4041–4045
- Conaway RC, Lehman IR (1982) A DNA primase activity associated with DNA polymerase alpha from *Drosophila melanogaster* embryos. *Proc Natl Acad Sci USA* 79(8):2523–2527
- Copeland WC, Wang TS (1993) Enzymatic characterization of the individual mammalian primase subunits reveals a biphasic mechanism for initiation of DNA replication. *J Biol Chem* 268(35):26179–26189
- Corn JE, Pease PJ, Hura GL, Berger JM (2005) Crosstalk between primase subunits can act to regulate primer synthesis in *trans*. *Mol Cell* 20(3):391
- Dua R, Levy DL, Campbell JL (1998) Role of the putative zinc finger domain of *Saccharomyces cerevisiae* DNA polymerase ϵ in DNA replication and the S/M checkpoint pathway. *J Biol Chem* 273(45):30046–30055
- Eki T, Enomoto T, Masutani C, Miyajima A, Takada R, Murakami Y, Ohno T, Hanaoka F, Ui M (1991) Mouse DNA primase plays the principal role in determination of permissiveness for polyomavirus DNA replication. *J Virol* 65(9):4874–4881
- Foiani M, Santocanale C, Plevani P, Lucchini G (1989) A single essential gene, *PR12*, encodes the large subunit of DNA primase in *Saccharomyces cerevisiae*. *Mol Cell Biol* 9(7):3081–3087
- Frick DN, Richardson CC (2001) DNA primases. *Annu Rev Biochem* 70:39–80
- Hubscher U (1983) The mammalian primase is part of a high molecular weight DNA polymerase α polypeptide. *EMBO J* 2(1):133–136
- Ito N, Nureki O, Shirouzu M, Yokoyama S, Hanaoka F (2003) Crystal structure of the *Pyrococcus horikoshii* DNA primase-UTP complex: implications for the mechanism of primer synthesis. *Genes Cells* 8(12):913–923
- Iyer LM, Koonin EV, Leippe DD, Aravind L (2005) Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res* 33(12):3875–3896

- Johansson E, MacNeill SA (2010) The eukaryotic replicative DNA polymerases take shape. *Trends Biochem Sci* 35:339–347
- Jokela M, Eskelinen A, Pospiech H, Rouvinen J, Syväoja JE (2004) Characterization of the 3' exonuclease subunit DP1 of *Methanococcus jannaschii* replicative DNA polymerase D. *Nucleic Acids Res* 32(8):2430–2440
- Kaufmann G, Falk HH (1982) An oligoribonucleotide polymerase from SV40-infected cells with properties of a primase. *Nucleic Acids Res* 10(7):2309–2321
- Klinge S, Hirst J, Maman JD, Krude T, Pellegrini L (2007) An iron-sulfur domain of the eukaryotic primase is essential for RNA primer synthesis. *Nat Struct Mol Biol* 14(9):875–877
- Klinge S, Nunez-Ramirez R, Llorca O, Pellegrini L (2009) 3D architecture of DNA Pol α reveals the functional core of multi-subunit replicative polymerases. *EMBO J* 28(13):1978–1987
- Kuchta RD, Stengel G (2010) Mechanism and evolution of DNA primases. *Biochim Biophys Acta* 1804(5):1180–1189
- Kuchta RD, Reid B, Chang LM (1990) DNA primase. Processivity and the primase to polymerase α activity switch. *J Biol Chem* 265(27):16158–16165
- Lao-Sirieix SH, Nookala RK, Roversi P, Bell SD, Pellegrini L (2005) Structure of the heterodimeric core primase. *Nat Struct Mol Biol* 12(12):1137–1144
- Lipps G, Rother S, Hart C, Krauss G (2003) A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. *EMBO J* 22(10):2516–2525
- Makiniemi M, Pospiech H, Kilpelainen S, Jokela M, Vihinen M, Syvaaja JE (1999) A novel family of DNA-polymerase-associated B subunits. *Trends Biochem Sci* 24(1):14–16
- McGeoch AT, Bell SD (2005) Eukaryotic/archaeal primase and MCM proteins encoded in a bacteriophage genome. *Cell* 120(2):167–168
- Mizuno T, Yamagishi K, Miyazawa H, Hanaoka F (1999) Molecular architecture of the mouse DNA polymerase α -primase complex. *Mol Cell Biol* 19(11):7886–7896
- Muzi-Falconi M, Giannattasio M, Foiani M, Plevani P (2003) The DNA polymerase α -primase complex: multiple functions and interactions. *ScientificWorldJournal* 3:21–33
- Nunez-Ramirez R, Klinge S, Sauguet L, Melero R, Recuero-Checa MA, Kilkenny M, Perera RL, Garcia-Alvarez B, Hall RJ, Nogales E, Pellegrini L, Llorca O (2011) Flexible tethering of primase and DNA Pol α in the eukaryotic primosome. *Nucleic Acids Res* 39(18):8187–8199
- Plevani P, Badaracco G, Augl C, Chang LM (1984) DNA polymerase I and DNA primase complex in yeast. *J Biol Chem* 259(12):7532–7539
- Pokorny R, Klar T, Hennecke U, Carell T, Batschauer A, Essen LO (2008) Recognition and repair of UV lesions in loop structures of duplex DNA by DASH-type cryptochrome. *Proc Natl Acad Sci USA* 105(52):21023–21027
- Sanchez Garcia J, Ciuffo LF, Yang X, Kearsley SE, MacNeill SA (2004) The C-terminal zinc finger of the catalytic subunit of DNA polymerase δ is responsible for direct interaction with the B-subunit. *Nucleic Acids Res* 32(10):3005–3016
- Sauguet L, Klinge S, Perera RL, Maman JD, Pellegrini L (2010) Shared active site architecture between the large subunit of eukaryotic primase and DNA photolyase. *PLoS One* 5(4):e10083
- Sheaff RJ, Kuchta RD, Ilesley D (1994) Calf thymus DNA polymerase α -primase: “communication” and primer-template movement between the two active sites. *Biochemistry* 33(8):2247–2254
- Shen Y, Tang XÁ, Yokoyama H, Matsui E, Matsui I (2004) A 21-amino acid peptide from the cysteine cluster II of the family D DNA polymerase from *Pyrococcus horikoshii* stimulates its nuclease activity which is Mre11-like and prefers manganese ion as the cofactor. *Nucleic Acids Res* 32(1):158–168
- Stillman B (2008) DNA polymerases at the replication fork in eukaryotes. *Mol Cell* 30(3):259–260
- Vaithiyalingam S, Warren EM, Eichman BF, Chazin WJ (2010) Insights into eukaryotic DNA priming from the structure and functional interactions of the 4Fe-4S cluster domain of human DNA primase. *Proc Natl Acad Sci* 107(31):13684–13689
- Weiner BE, Huang H, Dattilo BM, Nilges MJ, Fanning E, Chazin WJ (2007) An iron-sulfur cluster in the C-terminal domain of the p58 subunit of human DNA primase. *J Biol Chem* 282(46):33444–33451

- Yagura T, Koza T, Seno T (1982) Mouse DNA replicase. DNA polymerase associated with a novel RNA polymerase activity to synthesize initiator RNA of strict size. *J Biol Chem* 257(18):11121–11127
- Yang W, Lee JY, Nowotny M (2006) Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol Cell* 22(1):5
- Zerbe LK, Kuchta RD (2002) The p58 subunit of human DNA primase is important for primer initiation, elongation, and counting. *Biochemistry* 41(15):4891–4900

Chapter 10

The Structure and Function of Replication Protein A in DNA Replication

Aishwarya Prakash and Gloria E.O. Borgstahl

Abstract In all organisms from bacteria and archaea to eukarya, single-stranded DNA binding proteins play an essential role in most, if not all, nuclear metabolism involving single-stranded DNA (ssDNA). Replication protein A (RPA), the major eukaryotic ssDNA binding protein, has two important roles in DNA metabolism: (1) in binding ssDNA to protect it and to keep it unfolded, and (2) in coordinating the assembly and disassembly of numerous proteins and protein complexes during processes such as DNA replication. Since its discovery as a vital player in the process of replication, RPA's roles in recombination and DNA repair quickly became evident. This chapter summarizes the current understanding of RPA's roles in replication by reviewing the available structural data, DNA-binding properties, interactions with various replication proteins, and interactions with DNA repair proteins when DNA replication is stalled.

Keywords Replication protein A • DNA replication • Single-stranded DNA binding protein • OB-fold • Protein-protein interaction • G-quadruplex

A. Prakash

Department of Microbiology and Molecular Genetics, The Markey Center for Molecular Genetics, University of Vermont, Given Medical Building, 89 Beaumont Avenue, Burlington, VT 05405, USA

G.E.O. Borgstahl (✉)

The Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, 987696 Nebraska Medical Center, Omaha, NE 68198-7696, USA
e-mail: gborgstahl@unmc.edu

10.1 Introduction

DNA replication is a cleverly orchestrated, fundamental process occurring within cells that allows organisms to duplicate the vast amounts of genetic information carried within DNA. This process occurs during the S-phase of the cell cycle and must be completed for healthy cells to divide. Replication of eukaryotic chromosomes is initiated at replication origins. These origins, ~30–100 kb apart and scattered along each chromosome, serve to recruit several proteins that constitute the replisome. The replisome, an enormous multiprotein-DNA complex, comprises proteins that unwind the DNA-double helix, stabilize ssDNA regions generated during the initial steps, and copy the DNA with accuracy and speed.

Briefly, replication begins at the origins upon binding of the origin recognition complex and proceeds bidirectionally in a semi-discontinuous manner (Campbell 1986; Wold 2010). The double-stranded DNA (dsDNA) is melted and unwound by a DNA helicase after which the ssDNA regions produced are coated rapidly by RPA (Oakley and Patrick 2010; Wold 1997). RPA is in abundance in cells and its binding protects ssDNA. It is thought to unfold DNA secondary structures, and keep them from reforming, before the DNA is replicated. During the initiation of replication, RPA functions to recruit the DNA polymerase α -primase complex (Pol α -primase) to the replication origins (see Chap. 9). Pol α -primase lays down an RNA-DNA primer to initiate leading and lagging strand synthesis, after which the leading strand is extended continuously. The clamp loader, replication factor C (RFC), assembles the sliding clamp, proliferating cell nuclear antigen (PCNA), at the end of the primer which then displaces Pol α -primase (see Chaps. 14 and 15). During the elongation phase RPA is believed to play a role in stimulating DNA polymerase δ (Pol δ) and DNA polymerase ϵ (Pol ϵ) which carry out highly processive DNA synthesis (see Chaps. 12 and 13). The lagging strand is constructed in a similar fashion to the leading strand but in the opposite 3'–5' direction and as a series of short Okazaki fragments, each of which is synthesized 5'–3'. When Pol δ approaches the RNA primer of the downstream Okazaki fragment, ribonuclease (RNase) H1 removes all but the last RNA nucleotide of the DNA primer. RPA is involved in the recruitment of the Dna2 endonuclease, which cleaves the RPA bound primers and RPA is therefore thought to play a role in Okazaki fragment processing (Bae et al. 2001, 2003; MacNeill 2001). Following this, the flap endonuclease 1 (FEN1) exonuclease complex (Chap. 16) removes the last RNA nucleotide and the gap is filled in by Pol δ . DNA ligase joins the Okazaki fragment to the growing strand (Wold 2010; Kunkel and Burgers 2008) (Chap. 17). DNA replication is regarded as a tightly regulated process that involves the coordinated action of numerous factors that function to copy the DNA efficiently with minimal error, in order to maintain genomic stability.

In the 1980s, when the molecular biology of DNA replication was still in its infancy, scientists relied on *in vitro* reconstitution analyses to study this process. Due to its simplistic genome organization, Simian Virus 40 (SV40) virus replication was used as a model system. T-antigen, a virally encoded protein, plays a central role in

the binding and unwinding of the viral DNA during the initial stages of replication. This protein in addition to six others was necessary for proper replication (Weinberg et al. 1990). The T-antigen requires ATP and another cellular protein to successfully perform its unwinding functions. This cellular protein was determined to be RPA (Wold and Kelly 1988). RPA has been studied extensively since its discovery and is thought to be the primary eukaryotic ssDNA binding protein that is involved in several facets of DNA metabolism including replication, recombination, and repair (Binz et al. 2004; Bochkarev and Bochkareva 2004; Broderick et al. 2010; Fanning et al. 2006; Iftode et al. 1999; MacNeill 2001; Mer et al. 2000a; Oakley and Patrick 2010; Sakaguchi et al. 2009; Turchi et al. 1999; Wold 1997; Zou et al. 2006).

RPA is an abundant protein in cells: in humans it is the most abundant ssDNA binding protein with 5×10^4 to 2.4×10^5 molecules of RPA per cell (Kenny et al. 1990; Seroussi and Lavi 1993). RPA is essential for cell survival and there is a constant level of RPA protein during the cell cycle (Din et al. 1990). Down regulation of RPA with small-interfering RNA (siRNA) results in prolonged S-phase during the cell cycle, accumulation of DNA strand breaks, G2/M arrest, and cell death (Haring et al. 2008). Since the primary function of RPA is to bind any naked ssDNA generated during cellular processes, it is not surprising that cells cannot survive without it. RPA accumulates at sites of replication, called replication foci, in the nucleus just prior to the initiation of replication and remains localized during the DNA synthesis phase with 10–50 RPA molecules per replicating strand in the replication fork (Seroussi and Lavi 1993).

RPA is heterotrimeric in nature: in humans the three subunits are named RPA1, RPA2 and RPA3 in decreasing order of size where RPA1 is 70 kDa, RPA2 is 32 kDa and RPA3 is 14 kDa (Fig. 10.1a). Each of these subunits have folded domains called “oligonucleotide binding folds” (OB-folds) (Fig. 10.1b, c). RPA1 contains four OB-folds (F, A, B, and C – see Fig. 10.1) that are separated by intrinsically disordered linkers. OB-folds A, B, and C bind DNA, whereas OB-fold F is a protein interaction domain. RPA2 has a disordered N-terminus, OB-fold D at its center and a winged-helix-loop-helix (wHLH) protein interaction domain at its C-terminus connected by a disordered linker. RPA3 is composed of OB-fold E. The OB-folds are conserved in structure, with more structural homology than sequence homology (Figs. 10.1c and 10.2). Each of these domains is involved in specialized functions that involve ssDNA binding, recognition of damaged DNA and noncanonical DNA, protein-protein interactions, inter-subunit interactions, and post-translational modifications such as phosphorylation (Iftode et al. 1999; Oakley and Patrick 2010; Wold 1997).

10.2 Evolution of RPA

Single-strand DNA binding (SSB) proteins are essential in mediating several aspects of DNA metabolism. These proteins have been identified in organisms from prokaryotes to eukaryotes, and in archaea (Chedin et al. 1998). The bacterial SSB is

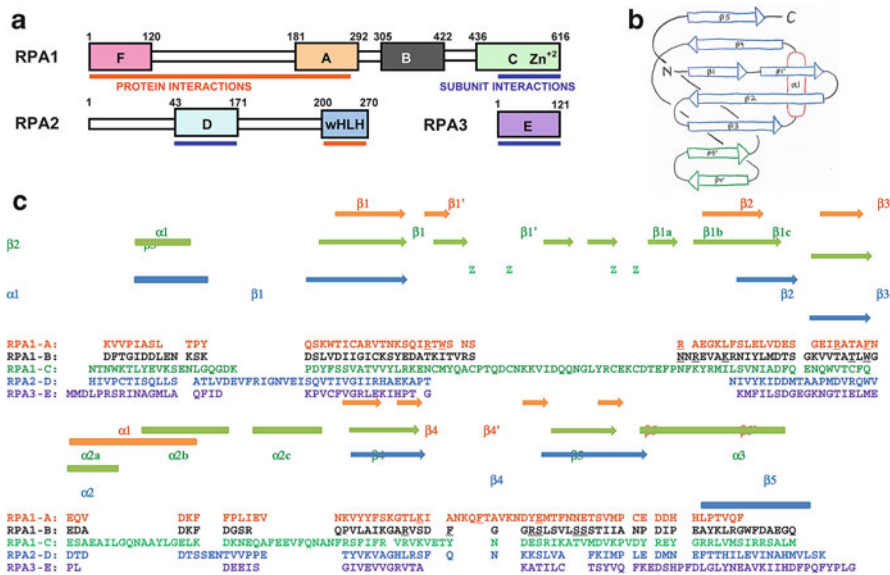


Fig. 10.1 The domain structure of RPA. (a) Schematic drawing of the three subunits (RPA1=70 kDa, RPA2=32 kDa, and RPA3=14 kDa), the folded domains (*thick colored rectangles*) and flexible linkers defined by limited proteolysis and NMR studies (Brosey et al. 2009; Gomes et al. 1996; Gomes and Wold 1995, 1996) (*thin white rectangles*). The OB-folds are labeled A–F, individually colored, and this color code is used in all figures in this chapter. The zinc finger on RPA1 is indicated (Zn²⁺). The winged-helix-loop-helix domain is labeled wHLH. (b) General topology of the OB-folds (Bochkarev et al. 1997). The β -strands are indicated by *arrows* and the α -helix by an oval. The *blue* β -strands correspond to those that comprise the OB-fold. The L12 loop lies between β_1' and β_2 and the L45 loop lies between β_4' and β_5' . (c) Sequence and secondary structure alignment of domains A–E based on structure (Bochkarev et al. 1997, 1999; Bochkareva et al. 2002). *Orange* secondary structure elements represent domains A and B, *green* elements represent domain C, and *blue* elements represent domains D and E. Lower case *z* indicates the Cys residues in domain C that bind zinc. Residues in RPA1-A and RPA1-B that bind ssDNA are *underlined*

expressed from one gene and functions as a homotetramer. In contrast, eukaryotic RPA is expressed from three separate genes forming three subunits and functions as a heterotrimer. Despite these differences, the two proteins are structurally similar suggesting they originated from a common ancestor before evolving into the proteins they are today. In the archaeal homologs features like the zinc finger motif (present at the C-terminal domain of RPA1) developed, which is not present in the bacterial homologs, representing a link between prokaryotic (archaeal) and eukaryotic proteins.

In addition to the three canonical subunits of RPA, a homolog of human RPA2 (with 47% amino acid identity and 63% similarity) called RPA4 was discovered (Kemp et al. 2010). RPA4 readily forms an alternative heterotrimeric complex with RPA1 and RPA3, called aRPA and is expressed in all human tissues, albeit at different levels. This alternate form of RPA failed to support replication in the *in vitro*

SV40 replication system indicating that the roles for this protein are significantly different than canonical RPA (Haring et al. 2010). RPA4 has been implicated in the initial steps of nucleotide excision repair (NER) and also during the Rad51 dependent strand exchange step of homologous recombination (HR), indicating a role for this protein in DNA repair (Mason et al. 2010).

10.3 RPA Structure

RPA is a dynamic complex in solution and not surprisingly, the quest for a full-length crystal structure of RPA has resembled that for the Holy Grail. However, over the past few decades, several groups have reported either NMR or crystal structures of various domains and truncated subunits of RPA (Fig. 10.2). These structures have enabled researchers to piece together some of the structural basis for the numerous essential interactions of RPA with DNA and various interacting proteins.

The structures of RPA2-wHLH and RPA1-F protein interaction domains have been determined. An NMR structure of the C-terminal region of RPA2 comprising residues 172–270, revealed a wHLH domain formed by a right-handed three-helix bundle and three short anti-parallel β -strands (Fig. 10.2a) (Mer et al. 2000b). This wHLH domain is an important protein-protein interaction domain in DNA repair, mediating interactions with, for example, XPA, UNG2 and RAD52 (Jackson et al. 2002; Mer et al. 2000b; Stigger et al. 1998). The N-terminal RPA1-F domain, encompassing residues 8–108 was also studied with NMR; it forms a five-stranded β -barrel which is capped on both ends by a short helix (Jacobs et al. 1999) (Fig. 10.2b). This region was shown to associate with various proteins including p53, VP16, Gal4 and XPG (Bochkareva et al. 2005; He et al. 1993, 1995) and is very important in DNA replication. Residues 109–168 form an unstructured flexible linker to RPA1-A. The β -barrel contains two loops on one side that form a basic cleft containing one lysine and five arginine residues which extend from one end of the β -barrel. This basic cleft was proposed to form a binding surface for the acidic motifs of transcriptional activators, repair proteins and replication proteins.

X-ray crystallography was used to study the structure of the primary DNA binding domains of RPA1: A (residues 180–290) and B (residues 300–420). RPA1-AB are arranged in a tandem orientation and connected with an extended, flexible inter-domain linker (Fig. 10.2c). The crystal structure revealed that each domain contains an OB-fold structure with an N-terminal extension with RPA1-B also having a C-terminal helix (Bochkareva et al. 2001). Without ssDNA bound, the flexible linker between RPA1-A and RPA1-B can adopt multiple conformations. RPA1-AB was co-crystallized with a poly dC₈ oligonucleotide (Fig. 10.2d) (Bochkarev et al. 1997; Pfuetzner et al. 1997). This structure clearly showed that both OB-folds contain ssDNA binding sites. Upon binding ssDNA the OB-folds reorient, the interdomain linker is stabilized and the binding surfaces coalign to tightly bind the oligonucleotide (Fig. 10.3a, b). The oligonucleotide crosses on the β -strands on both OB-folds and between the loops L12 and L45 (Fig. 10.2d) and these loops

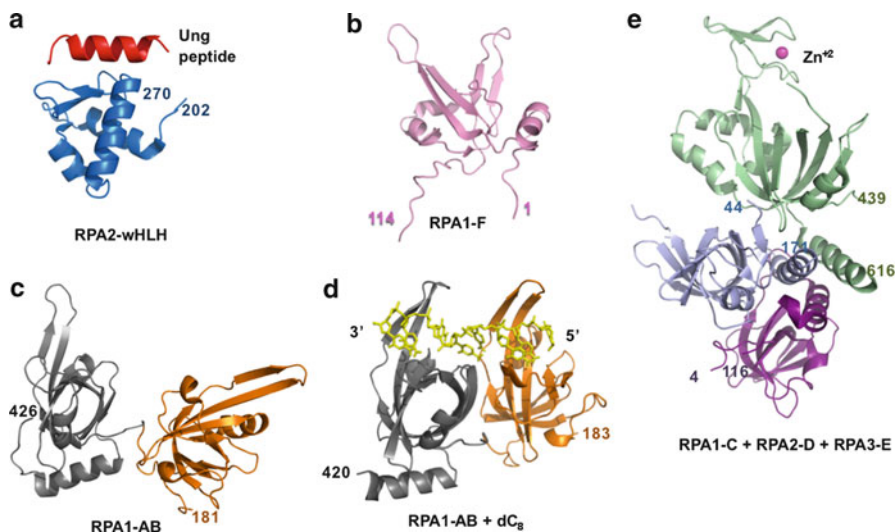


Fig. 10.2 X-ray and NMR structures of RPA domains. (a) Domain RPA2-wHLH with Ung peptide bound (Mer et al. 2000b); (b) Domain RPA1-F; (c, d) Major ssDNA binding domains RPA1-AB with and without ssDNA bound (Bochkarev et al. 1997; Bochkareva et al. 2001); (e) Trimerization core containing domains RPA1-C, RPA2-D and RPA3-E (Bochkareva et al. 2002). Domains colored as in Fig. 10.1. PDB IDs used were 1L1O, 1JMC, 1FGU, 1EWI and 1DPU, respectively

significantly change their conformation when DNA is bound (Fig. 10.3a, b). The cytosine bases are tucked into the binding cleft and the phosphodiester backbone interacts with basic patches on the surface (Fig. 10.3b). The loops move to a closed conformation by folding around the oligonucleotide and securely hold it in the depths of the charge and shape compatible binding cleft. Each OB-fold makes contacts with three nucleotides and there are two nucleotides between the OB-folds (Fig. 10.2d). RPA1-A makes more extensive contacts with the ssDNA than RPA1-B. Thus we have a structural description for how RPA's primary DNA binding domains bind pyrimidine-rich ssDNA with high affinity.

The crystal structure of the RPA2/3 core was solved (Bochkarev et al. 1999). This construct included only the central region of RPA2 (residues 43–171) and full-length RPA3 which were resistant to limited proteolytic digestion (Fig. 10.4a). This structure revealed that both RPA2 core and RPA3 contain canonical OB-fold structures with an N-terminal extension and a C-terminal helix. The heterodimer interface is mediated by the C-terminal helices on both subunits through a helix-helix interaction, while a higher order (dimer of dimers) interaction is mediated by a four-helix bundle (Fig. 10.4a). This helix bundle was proposed to play a role in trimerization of the full-length protein.

Full-length RPA2/3 was solved in several crystal forms (Deng et al. 2007). In these crystals the N-terminus (residues 1–42) and C-terminal wHLH domain (residues 175–270) were disordered with very weak electron density. The ordered OB-fold

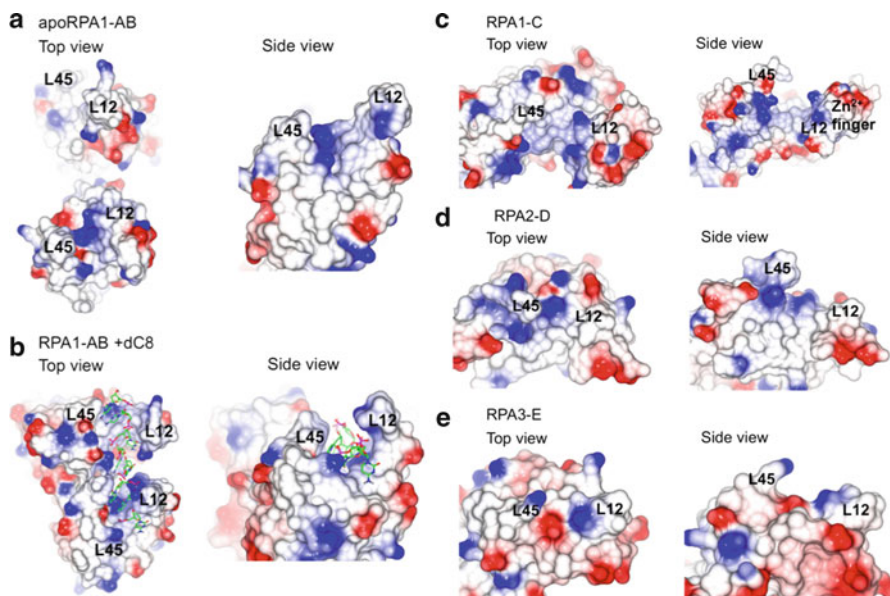


Fig. 10.3 Electrostatic surfaces of RPA's DNA binding domains. (a, b) Major ssDNA binding domains RPA1-AB with and without ssDNA bound. In the side view, the A domain is in the foreground. In the top view, the A domain is at the bottom; (c) RPA1-C domain; (d) RPA2-D domain; (e) RPA3-E. Surface figures were created using ccp4mg with -0.5 V (red) to $+0.5$ V (blue) (Potterton et al. 2004)

regions were very similar to the previously solved RPA2/3 core crystal structure. However, the higher order quaternary structures formed between the heterodimers were significantly different. The four-helix bundle previously thought to be important for forming the heterotrimer only occurred when the deletion construct was crystallized. The full-length RPA2/3 crystals contained different dimer-dimer interfaces (Fig. 10.3b, c). These differences in quaternary structure may reflect the actual structural repertoire of the RPA heterotrimer and the relative locations of the RPA2/3 $\alpha 2$ helices may represent alternate locations for the RPA1 $\alpha 3$ helix when the trimer forms.

The crystal structure of the RPA trimerization core included the C-terminus of RPA1 (residues 436–616), the core of RPA2 (residues 45–171) and RPA3 (Fig. 10.2e) (Bochkareva et al. 2002). All three domains comprise an OB-fold structure flanked by a C-terminal α -helix and are structurally similar (Fig. 10.1c). The hydrophobic interactions present between these helices form a three-helix bundle and mediate the trimerization of the domains (Fig. 10.4d). Also, a six-helix bundle forms between trimers in the crystal lattice (Fig. 10.4e). The individual RPA2 and RPA3 core regions in this structure are identical to those found in the dimer core structure. The binding surface of RPA2 and RPA3 is much shallower than seen with RPA1-AB (Fig. 10.3d, e), which corresponds with their weaker binding to ssDNA. Also, the floor of the binding cleft on RPA2-D is positively charged, and for RPA3-E the floor

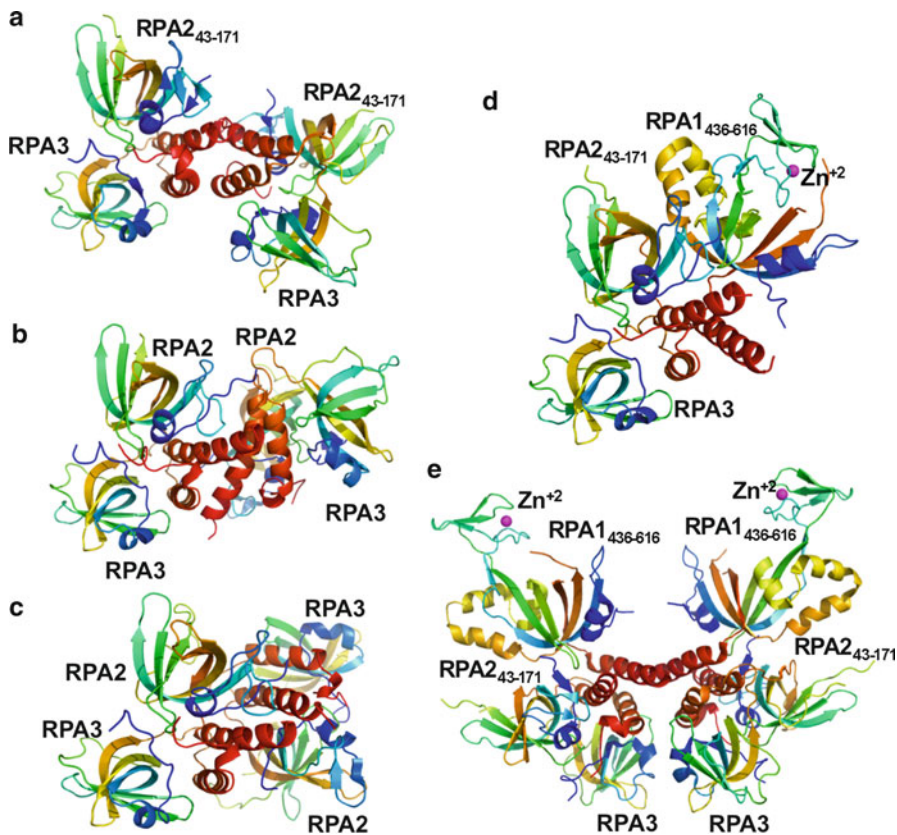


Fig. 10.4 Four-helix bundle quaternary structures formed by the RPA2/3 heterodimer and by the RPA trimer core. (a) Dimer core (Bochkarev et al. 1999); (b) Orthorhombic; and (c) Hexagonal crystal forms of full length dimer (Deng et al. 2007); (d) Three helical bundle formed by one trimer core; (e) Six helical bundle formed in the crystal lattice of the trimer core (Bochkareva et al. 2002). Domains colored by sequence number with N-terminal residues shown in *blue* and C-terminal residues *red*. For parts **a**, **b** and **c**, PDB IDs 1QUQ, 2PI2 and 2PQA were used, respectively. For parts **d** and **e**, PDB ID 1L1O was used

is negatively-charged with positively-charged loops. These observations imply that these OB-folds bind ssDNA differently than RPA1-AB.

RPA1-C contains a zinc-finger motif that sets it apart from the other OB-folds (Bochkareva et al. 2000). This zinc finger is present between strands $\beta 1$ and $\beta 2$, and the zinc ion present in the structure is coordinated by four cysteine residues (Cys 481, 486, 500 and 503; Fig. 10.1c). This is similar to transcription factors involved in DNA binding that have also been shown to possess zinc finger domains (Krishna et al. 2003). The binding cleft is deep and the surface charges for RPA1-C are very similar to RPA1-AB (Fig. 10.3c), implying that it will bind ssDNA in a similar conformation. Unfortunately, at this time we have no structural data on how RPA1-C, RPA2-D and RPA3-E interact with ssDNA.

Small angle x-ray scattering (SAXS) has been used to study the structural dynamics of the N-terminal half of RPA1 (including domains RPA1-F, -A and -B) when bound to ssDNA (Pretto et al. 2010). Consistent with previous reports, SAXS data indicate that binding of ssDNA to RPA1-FAB reduces the interdomain flexibility between RPA1-A and -B but has no effect on RPA1-F which is available for protein interactions. These data support a model where RPA1-F remains structurally independent of RPA1-AB when RPA is bound to ssDNA, thereby allowing RPA to form critical protein-protein interactions.

Structural studies on various domain constructs of RPA have provided a wealth of structural knowledge that has helped considerably in understanding the complex functions of RPA. Unfortunately, we still do not have a complete structural model for RPA. Methods have been developed using mass spectrometry to follow the reactivity of amino acids to proteolytic and chemical modification to test theoretical models of RPA built using the available domain structures (Nuss et al. 2006). These reactivities have been employed to construct and test a complete model for the structure of RPA (Nuss et al. 2009). This RPA structural model contains stable domains and highly flexible non-domain regions. The overall structure is discoidal, and its surface is predominantly negatively charged with neutral and positive patches coinciding with ssDNA or protein binding sites. This leaves one face of the structure largely negative for interaction with basic protein molecules. The DNA binding OB-folds (A, B, C, and D) are exposed to solvent and, with the exception of OB-fold D, they are on the periphery of the complex. This structure is consistent with ssDNA binding simultaneously to domains A–D. Most of the protein binding sites on RPA are also exposed and accessible to protein ligands. Four relatively long (>20 amino acids) regions of the RPA primary structure are coiled or intrinsically disorganized as judged by primary structure analysis. This model is helpful in understanding RPA function but is still limited in the understanding of full-length RPA because of the flexible nature of the protein.

Full-length heterotrimeric RPA was analyzed using NMR and gave rich insight into the folding and structural dynamics of this multidomain, flexible protein (Brosey et al. 2009). The NMR spectra on the RPA trimer contained over 350 of the 550 expected signals domains F, A, B, wHLH and the N-terminus of RPA2. The signals were nearly identical in position on the spectra as those from the isolated domains. This indicates these domains are structurally independent from each other in the absence of DNA. Signals from RPA-CDE core were absent in the spectra from the full-length protein, indicating it had a slow rate of tumbling due to the drag caused by the attachment of the five other domains. Experiments conducted in the presence of DNA confirmed that the basic RPA1-F domain and the acidic RPA2-wHLH domain played no role in binding to ssDNA and remained available for binding to other protein factors. Upon binding to DNA, a structural rearrangement and alignment of RPA1-AB with RPA-CDE was observed. Changes were also seen in the NMR signals of the N-terminal region of RPA2 reflecting remodeling of this region. This last observation may explain how the N-terminus of RPA2 with ssDNA bound is more accessible to kinase activity during DNA repair processes than the free form (Fotedar and Roberts 1992). We look forward to more NMR experiments on intact RPA and the full-length crystal structure in the future.

10.4 Interactions of RPA with Single-Stranded DNA

The DNA binding properties of RPA have been extensively studied and reveal several important features. Importantly, its ssDNA binding function protects DNA from nucleases and aids in unfolding any secondary structures that DNA forms which may disrupt DNA processing. RPA binds ssDNA with a much greater affinity when compared with dsDNA or RNA, and binds ssDNA with low cooperativity and a 5'→3' molecular polarity. In fact, RPA binds ssDNA over 1,000-fold better than dsDNA with an association constant in the range of 10^9 – 10^{11} M⁻¹ (Kim et al. 1992, 1994). The binding of RPA also appears to be sequence dependent, as it prefers to bind polypyrimidine sequences over polypurine sequences. One can summarize the order of RPA binding to nucleic acids in order of decreasing affinity as follows: polypyrimidine > mixed ssDNA > polypurine ssDNA >> damaged dsDNA > dsDNA ≡ RNA. RPA-ssDNA binding depends on two important factors, the length of the ssDNA sequence and salt conditions used in the assay. Shorter ssDNA sequences have lower binding constants for RPA with association constants ranging from 10^7 to 10^9 M⁻¹. The binding of RPA to pyrimidine-rich sequences is so tight that salt concentrations >1.5 M are necessary to weaken its interaction with the ssDNA for comparative studies (Kim et al. 1992; Wold 1997).

Several lines of evidence indicate that the binding of RPA to ssDNA causes a significant change in its conformation. Limited proteolysis experiments revealed that without ssDNA present, RPA1 and RPA2 are degraded within minutes (Gomes et al. 1996). RPA3 was resistant to proteolysis in these experiments. When a polypyrimidine oligonucleotide (dT₃₀) was present, RPA1 and RPA2 became more resistant to degradation and the domain structure of RPA, used in the many structural studies already discussed, was revealed. Additionally, electron microscopy (EM) images of the RPA-ssDNA complex indicate that the complex can adopt three different molecular shapes: globular, elongated, or contracted depending on the salt concentrations present in the reactions (Treuner et al. 1996). Using scanning-transmission electron microscopy (STEM), RPA was shown to adopt different conformations upon DNA binding (Blackwell et al. 1996). These complexes were observed as either an 8 nt mode which is more compact and globular or a 30 nt elongated binding mode. These early observations were then incorporated into the models for ssDNA binding described next.

The versatility of RPA's numerous possible interactions with ssDNA comes from the multiple DNA binding domains of RPA (Fanning et al. 2006; Sakaguchi et al. 2009). RPA1-A and -B are known as the primary DNA binding domains. These domains bind DNA with 10–50-fold lower affinity when compared to full-length RPA depending on the length and nature of the DNA sequence. RPA1-C and RPA2-D have some, albeit weak, DNA binding activity on the order of 10^{-5} – 10^{-6} M⁻¹. Based on the numerous ssDNA interactions performed with RPA, a sequential model was proposed for DNA binding by RPA (Bastin-Shanower and Brill 2001; Bochkarev and Bochkareva 2004; Fanning et al. 2006). In these models, RPA1-A recognizes ssDNA first and this is followed by the binding of RPA1-B. Together RPA1-A and -B bind a footprint of 8 nt. This is followed by the binding of RPA1-C

which is then involved in binding a 12–23 nt segment of DNA. The binding of RPA2-D covers a length of 25–30 nt which is the most characterized, well-known footprint of RPA. This sequential model, updated with the recent data about RPA1-F and RPA2-wHLH being protein binding domains, is summarized in Fig. 10.5.

In the absence of the primary DNA binding domains, a construct containing only the trimer core RPA-CDE was capable of recognizing a primer-template junction (Dickson et al. 2009; Kolpashchikov et al. 2000a, b; Pestryakov et al. 2003, 2004, 2007; Pestryakov and Lavrik 2008; Weissbart et al. 2004). RPA3-E plays a vital role in the recognition of this primer-template junction since the same core with RPA3 deleted could not properly recognize the 3' end of the primer-template junction. There is evidence for RPA3-E interacting with ssDNA molecules bound to trimeric RPA (Pestryakov et al. 2007); the polarity of this interaction is on the 3' side of the oligonucleotide (Salas et al. 2009). Despite RPA's traditional preference for pyrimidine-rich sequences, more light has been shed on the interaction of RPA with biologically-relevant mixed ssDNA sequences (Deng et al. 2009) and the binding preferences of individual domains (Prakash et al. 2011b). Additionally, RPA is now known to bind non-canonical ssDNA sequences capable of forming complex secondary structures (Fan et al. 2009; Salas et al. 2006; Wu et al. 2008). These secondary structures pose a difficult challenge for DNA replication and the involvement of RPA in conquering them appears to be important, as described below.

10.5 DNA Structure and Requirement for RPA

The versatile nature of DNA and its ability to form stable secondary structures has intrigued scientists for a long time (see Mirkin 2008, for review). Some of these structures include DNA hairpins, cruciforms, triple-helical DNA, *i*-motif and G-quadruplex structures (Fig. 10.6). The formation and stabilization of these secondary structures *in vivo* has sparked the interest of researchers all over the world for decades because of their potential role in stalling replication thereby leading to disease progression (Voineagu et al. 2009; Wells 2007). The section below discusses some of these DNA arrangements in a disease-relevant context emphasizing the requirement for proteins like RPA to help unfold these structures and/or to signal a stress-response.

DNA hairpins are formed when ssDNA bends back onto itself forming duplex DNA and terminating in a loop (Voineagu et al. 2008). These are commonly formed by inverted repeat sequences. The stability of these hairpins is dependent upon the GC content of the sequence. The most common sequences that have the capability to form hairpins are trinucleotide repeats (TNRs), where the trinucleotide (for example, CNG or GAA, where N is any nucleotide) sequence is repeated multiple times (Lahue and Slater 2003). There are now over 20 known neurological disorders that involve TNRs, including Huntington's disease, Fragile X syndrome, and myotonic dystrophy (Cummings and Zoghbi 2000; Mirkin 2006, 2007). TNRs can occur in non-coding sequences as well as within coding sequences. NMR structural studies of these repeat

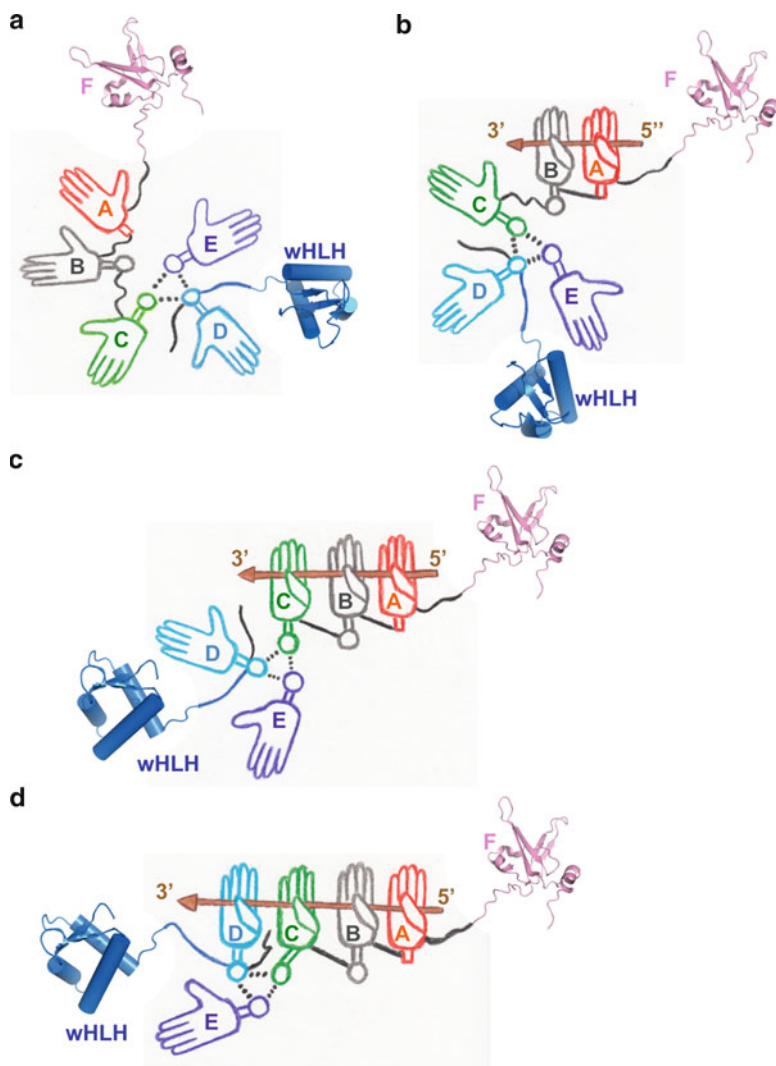
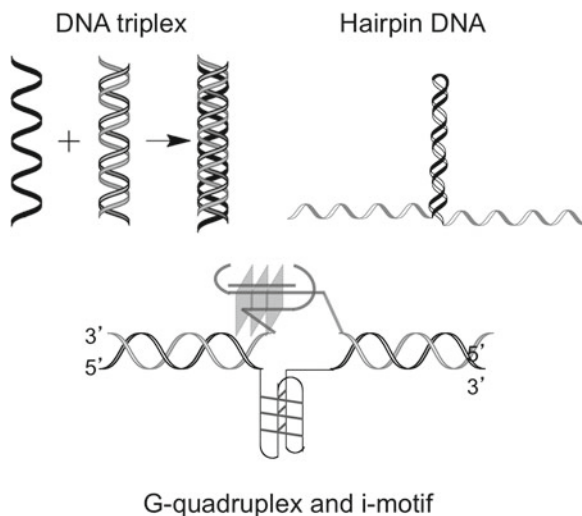


Fig. 10.5 Sequential binding model for RPA. (a) Unbound RPA in globular conformation. (b) Binding of 8–10 nt by RPA1-A and RPA1-B; (c) Binding of 13–15 nt by RPA1-A, RPA1-B and RPA1-C; (d) The 30 nt binding mode with all four DNA binding OB-fold domains. Domains RPA1-F and RPA2-wHLH are involved in protein-protein interactions. This model was created combining information from what is known about the flexible regions of RPA, the order of DNA binding, which domains primarily bind ssDNA (A–D) and which are involved in protein-protein interactions (F and wHLH) and speculation that various helical bundles might form the heterotrimer quaternary interface (Figs. 10.3 and 10.4) (Bastin-Shanower and Brill 2001; Bochkarev and Bochkareva 2004; Bochkareva et al. 2002; Brosey et al. 2009; Deng et al. 2007; Fanning et al. 2006; Gomes et al. 1996)

Fig. 10.6 RPA must bind ssDNA secondary structures and keep them from forming during DNA replication. Schematic representation of the structures of a DNA triplex (*upper left*), a DNA hairpin (*upper right*) and a G-quadruplex opposite an *i*-motif structure (*lower part*)



sequences reveal and confirm the formation of hairpin and mismatched DNA duplex structures (Mariappan et al. 1998). In such instances when the replicative polymerase encounters a stable secondary structure that was not unwound by a helicase or DNA binding protein, it skips over the region resulting in a loss of genetic information, genome instability and disease progression.

Another non-canonical DNA secondary structure that appears to have an effect on process like replication and transcription in the formation of triple helical DNA, often called triplex DNA (Bissler 2007). These structures are formed when a third strand of DNA binds to the major groove of a double-stranded DNA, using Hoogsteen base-pairing. Triplexes can form intermolecularly where the third strand originates from a second DNA molecule or from a triplex forming oligonucleotide (TFO), whereas in the case of intramolecular triplexes, also commonly referred to as H-DNA, the third triplex forming strand originates from a region within the same DNA molecule. TFOs are being exploited as therapeutic agents to target specific genes because of their ability to bind duplex DNA with high-affinity (Jain et al. 2008). Through the years, it has been noted that H-DNA structures can be formed by triple repeat structures. In the case of Friedreich's ataxia, an expansion of the intronic sequence d(GAA)_n forms a triplex structure that halts DNA polymerization *in vitro* (Mirkin 1999). Thus, triplex DNA structures also pose a challenge for DNA replication.

The knowledge that G-rich regions in DNA form non-B DNA secondary structures like G-quadruplexes (often called tetraplexes or G4 DNA) has been known for a long time; contrary to their being a nuisance, these sequences have potentially important roles in regulating cellular metabolism (Dai et al. 2010; Gellert et al. 1962; Huppert 2008). *In vitro*, G-rich sequences can form a variety of G-quadruplex structures. Four planar guanine residues interact via Hoogsteen hydrogen bonds to form a G-quartet (Huppert and Balasubramanian 2007; Patel et al. 2007). A G-quadruplex results from the stacking of two or more G-quartets. The formation

of G-quadruplexes also depends on the presence of monovalent cations such as sodium or potassium ions. The precise ion preference depends on the sequence and nature of the G-quadruplexes (Marathias and Bolton 1999). G-quadruplexes form at the telomere and in the promoter regions of proto-oncogenes, such as c-MYC, VEGF, c-KIT and Bcl-2 (Eddy and Maizels 2006, 2009; Patel et al. 2007). An *i*-motif can form on the strand opposite the G-quadruplex (Dai et al. 2010). These locations indicate that the occurrence of G-quadruplexes and *i*-motifs might be regulatory and play a role in the formation and progression of many cancers.

10.6 RPA Binding to Non-canonical DNA Structures

As discussed previously, hairpin structures present the replication machinery with a challenge and a roadblock if not properly unwound or melted. *In vitro*, RPA was shown to bind preferentially to hairpin structures with a 3' protruding end. However, in this study, RPA did not significantly melt or unfold the hairpin structures (de Laat et al. 1998). However, as will be discussed in later sections, RPA also serves to recruit other DNA binding proteins such as helicases that enable successful unwinding of DNA. Thus RPA binding could be a crucial initial first step in binding the ssDNA regions generated by hairpin structures in DNA and further aiding in the unfolding of these structures through the recruitment of other proteins. In contrast to *E. coli* and T4 ssDNA binding proteins, RPA was shown to melt a DNA triplex containing a pyrimidine third strand annealed to duplex DNA (Wu et al. 2008). In the same study, cellular analyses using HeLa cells indicated that depletion of RPA caused an increase in triplex DNA content. This emphasizes a physiological role for RPA in binding and unfolding such secondary structures.

Compared with the above-mentioned secondary structures, a significantly greater number of studies were performed with RPA binding to G-quadruplex DNA. Some of these studies are summarized here. Native gel electrophoresis, cross-linking, and fluorescence resonance energy transfer experiments indicate that RPA can bind and unfold a 21-mer telomeric G-quadruplex sequence (Salas et al. 2006). Most recently, studies employing CD (circular dichroism) indicate that RPA can bind and melt intramolecular G-quadruplex structures (Fan et al. 2009). In fact, it was demonstrated that RPA could bind a purine-rich, G-quadruplex forming sequence with a similar affinity as the complementary pyrimidine-rich sequence. Interestingly, the above studies showing RPA unfolding G-quadruplexes were all done in the presence of Na⁺ ions. It was subsequently shown that K⁺ (and a porphyrin drug) can stabilize G-quadruplex forming sequence from RPA unfolding (Prakash et al. 2011a). G-quadruplex forming sequences can induce instability during leading-strand replication when cells are treated with a G-quadruplex stabilizing drug or in the absence of the G-quadruplex unwinding Pif1 helicase (Lopes et al. 2011). It is possible that RPA may have a role in these types of errors in DNA replication.

RPA helps prevent the accumulation of telomeric DNA in cells employing alternative lengthening of telomeres, supports telomerase activity in yeast, restores

human telomerase activity *in vitro*, and causes telomere shortening in human cancer cells (Grudic et al. 2007; Kobayashi et al. 2010). The human Dna2 protein possesses both helicase and nuclease activities during lagging strand DNA replication and it specifically binds to telomeric regions that have the propensity to form G-quadruplexes (Masuda-Sasa et al. 2008). Although the helicase activity of Dna2 is effective in unwinding G-quadruplex DNA, this secondary structure causes attenuation of nuclease activity. The presence of RPA bound to the G-quadruplex DNA restores the nuclease activity of Dna2, thus emphasizing the requirement for RPA during telomere biogenesis.

The diverse nature of RPA binding to ssDNA has been explored by several groups. However, so far the data are limited since most studies on RPA, and its domains, have been performed using primarily poly-pyrimidine ssDNA sequences. The specific ssDNA sequences preferred by the DNA binding OB-fold domains of RPA were studied using SELEX (Systematic Evolution of Ligands by EXponential enrichment) methods (Prakash et al. 2011b). Not surprisingly, SELEX with full-length RPA revealed no specific sequence preference. The most interesting SELEX result was obtained with RPA-CDE which selected a 20-mer G-rich sequence that formed an intramolecular G-quadruplex. Fluorescence polarization (FP) binding studies to verify and understand the SELEX results were conducted where the selected G-quadruplex, a TC-rich complement of the G-quadruplex, a polyA and a polyG sequence were tested using five different RPA constructs: (i) full length RPA, (ii) RPA1-AB, (iii) RPA-CDE-core, (iv) RPA-DE, and (v) RPA1-C. These extensive FP binding studies indicate that domains RPA1-A, -B and -C of contribute to the “universal binder” functions of RPA. The similarities of their binding surfaces support this observation (Fig. 10.3). Binding affinity, with the RPA-C construct indicated that this construct binds to TC-rich and G-rich sequences alike with a binding constant $\sim 3 \mu\text{M}$. Most importantly RPA2-D and RPA3-E appear to contribute to a more specialized function for binding preferentially to G-rich sequences. CD studies showed that full length RPA and RPA-CDE core bind and unfold the G-quadruplex. RPA-DE on the other hand stabilized the G-quadruplex secondary structure. Note RPA2-D is unique in that it features positive charge on the floor of the binding cleft and a model for how RPA2-D could bind a folded G-quadruplex was built (see Fig. 8(e) in Prakash et al. 2011b). Taken together, it is likely that RPA-DE can recognize the G-quadruplex fold and in the context of the RPA heterotrimer, the G-quadruplex becomes unfolded. Also RPA-DE might recognize DNA secondary structures, such as G-quadruplexes or DNA hairpins and then recruit DNA helicases, like Dna2, to help unwind and unfold these structures for proper DNA replication.

10.7 RPA Binding to Damaged DNA

DNA is constantly being subjected to assault by either exogenous or endogenous factors that cause damage. Some exogenous agents include ultra-violet (UV) light, ionizing radiation (IR), toxic chemicals, and chemotherapeutic drugs. Endogenous

agents include reactive oxygen species (ROS), free radicals, secondary structures formed within DNA, and others. When the replicative machinery encounters lesions in the DNA caused by one or more of these factors, stalling occurs, causing replication arrest which further leads to a cascade of events to take place that signal the damage is present so that it is either repaired or bypassed (Hyrien 2000). The binding of RPA to damaged DNA has also been studied extensively. The first such report indicated RPAs interaction with DNA damage adducts and crosslinks, mediated by cisplatin (Clugston et al. 1992). In another study involving cisplatin induced DNA damage, RPA was seen to bind the damaged duplex DNA with a 10–50-fold increase in affinity over undamaged duplex DNA (Patrick and Turchi 1998). Conflicting reports exist in the literature as to whether RPA prefers to bind to the damaged DNA strand or the undamaged strand *in vitro* (Hermanson-Miller and Turchi 2002; Schweizer et al. 1999). Further, the binding of RPA has been studied with UV-induced damaged DNA where RPA bound preferentially to the 6-4- photoproduct thus formed. Therefore, not only does RPA have functions in binding and coating ssDNA regions formed during replication, but also binds to sites of DNA damage that can occur as part of the process. The binding and interactions of proteins involved during replication, either upon replication stalling or during normal replication, to RPA will be discussed in the next section.

10.8 Role in Recruiting Proteins to the Replication Fork

While RPA is binding ssDNA, it also helps coordinates DNA replication by binding to other replicative proteins at the appropriate place and time. RPA's primary replicative protein interaction domain appears to be the N-terminal RPA1-F domain (Figs. 10.1a and 10.5). Large T-antigen, some of the helicases, replication factor C (RFC), Dna2 and Pol α -primase all interact with RPA1-F (Fanning et al. 2006). The C-terminal RPA2-wHLH domain has been shown to also be important in binding T-antigen and proteins involved in processing stalled replication forks. RPA binds ssDNA at the replication fork immediately after the initiation of replication and then these interactions between RPA and other proteins are essential for forming an active DNA replication fork indicating that RPA is a proteinaceous glue of sorts.

Human RPA was originally recognized as a component necessary for SV40 DNA replication *in vitro* (Fairman and Stillman 1988). The interaction between RPA and the SV40 large T-antigen was shown to be essential for primosome assembly (Melendy and Stillman 1993). More specifically, interaction with the SV40 large T-antigen is mediated by both the RPA2-wHLH and the RPA1-F domains (Han et al. 1999; Taneja et al. 2007) and both of these domains were absolutely required for successful DNA replication. Large T-antigen residues 164–249, located within the DNA binding domain, are responsible for mediating this interaction with RPA (Weisshart et al. 1998). Large T-antigen actively loads RPA onto nascent ssDNA after initiation. NMR analyses indicated that the T-antigen, RPA1-F and a short 8-mer oligonucleotide can form a stable ternary complex (Jiang et al. 2006). This complex was disrupted by increasing the length of the DNA bound to RPA,

thereby indicating a conformational change within the protein that is required for loading onto the DNA (Arunkumar et al. 2005). Thus, T-antigen protein interactions with RPA, plus RPA's conformational change upon binding ssDNA (Fig. 10.5) load RPA on to ssDNA during initiation.

In addition, during the initial stages of replication RPA also forms a stable complex with DNA Pol α -primase and as well as with T-antigen (Dornreiter et al. 1992). The primase domain of Pol α -primase and RPA1-F domain mediate this interaction (Braun et al. 1997). RPA was shown to stimulate Pol α -primase activity and reduce misincorporation by this polymerase, thereby increasing its processivity.

During the process of replication, Pol α -primase is replaced by a switching mechanism where RFC, the eukaryotic clamp loader (see Chap. 14), binds to the 3' end of the nascent DNA and loads PCNA and Pol δ (Waga and Stillman 1994). This switch occurs in the presence of RPA where RPA1-F binds Rfc4, one of the five subunits of RFC (Kim and Brill 2001). Thus, RPA participates in loading PCNA through an RFC protein-protein complex.

Furthermore, during the elongation stage of DNA replication RPA stimulates the action of Pol δ and Pol ϵ , an activity that could be the result of RPA's interaction with PCNA. Pol δ is one of the replicative polymerases which functions mainly in lagging strand synthesis (McElhinny et al. 2008) (see Chap. 12). This polymerase competes with RFC for RPA, resulting in displacement of RFC from the 3' end (Yuzhakov et al. 1999). During the processing of Okazaki fragments, the Dna2 helicase/endonuclease aids in removing the RNA primers of these fragments. RPA plays a role in the stimulation of Dna2 endonuclease activity mediated by direct protein-protein interactions at the N-terminal domains of Dna2 and RPA1-F (Bae et al. 2001, 2003).

Another group of proteins that interact with RPA during the replication process are the RecQ family of helicases. The Werner syndrome protein (WRN), a member of this class of helicases that localizes to sites of stalled replication, directly interacts with RPA and the Mre11 complex upon replication arrest (Constantinou et al. 2000). In contrast to human RPA, *E. coli* SSB and bacteriophage T4 gene 32 protein (gp32) failed to stimulate WRN helicase unwinding of long DNA duplexes, indicating a specific interaction between WRN and RPA (Brosh et al. 1999). The interaction of WRN and RPA is substantially increased at stalled replication forks (Machwe et al. 2011). Similarly, a Bloom syndrome helicase (BLM) interacts with RPA using its N-terminal acidic domain. The basic N-terminal RPA1-F domain interacts with both the WRN and BLM helicases (Doherty et al. 2005). This interaction stimulates the helicases' ability to unwind long DNA substrates. These results suggest that the critical interactions between RPA and WRN or BLM helicases play an important role in the mechanism of RPA stimulated DNA unwinding during replication.

As indicated by the above examples, protein-protein interactions mediated by RPA are essential for successful replication. However, when the replication fork stalls, a DNA damage response (DDR) ensues involving the recruitment of repair proteins, several of which require an initial interaction with RPA. RPA is involved in cell cycle checkpoint signaling in addition to the DDR. Signaling from a stalled replication fork involves proteins that are sensors, mediators, transducers or effectors. Sensor proteins like ATM/ATR, the 9-1-1 complex and the MRN complex sense the

damage and through mediator proteins such as 53BP1, TopBP1, claspin, etc., mediate and recruit proteins that aid in restoring the replication fork (Sogo et al. 2002; Zou and Elledge 2003). RPA is required for the recruitment of the ATR kinase to sites of DNA damage and for ATR-mediated CHK1 phosphorylation and activation *in vivo*. The N-terminal region of RPA1 also stimulates the binding of ATR interacting protein (ATRIP) to ssDNA *in vitro* enabling the ATR-ATRIP complex to associate with DNA. The N-terminal region of RPA1 binds ATRIP, Rad9 and Mre11. Binding of RPA to Rad9 promotes ATR signaling (Xu et al. 2008). The Rad9 protein is part of the 9-1-1 (Rad9-Rad1-Hus1) clamp protein complex that plays a key role in cellular response to DNA damage (Kemp and Sancar 2009). The pro-apoptotic BH3-interaction death domain agonist (BID) associates with RPA1-F and stimulates the recruitment/stabilization of ATR-ATRIP to the DNA damage sensor complex (Liu et al. 2011). The Rad17 protein aids in loading the clamp complex onto the DNA via an RPA-mediated interaction. Further phosphorylation of the Rad17 protein activates the downstream cell cycle check-point to mediate DNA repair or alternatively leads to apoptosis (Gottifredi and Prives 2005). It has also been indicated that RPA-coated ssDNA recruits the protein Cut5 which facilitates the binding of the sensor protein ATR, Pol α -primase and Rad1 to damaged DNA (Parrilla-Castellar and Karnitz 2003). Another protein SMARCAL1 localizes to stalled replication forks via an interaction with the RPA2-wHLH domain. Silencing of SMARCAL1 causes an increase in RPA binding to chromatin (Bansbach et al. 2009). From all the above examples, it is evident that these proteins that are necessary for successful replication require an interaction with RPA.

It is noteworthy to mention here that RPA itself is phosphorylated in a cell cycle dependent manner and is hyperphosphorylated in response to DNA damage (Oakley et al. 2001; Oakley and Patrick 2010). Studies on RPA phosphorylation have been primarily focused on the N-terminal region of RPA32 because this domain is conserved in higher eukaryotes and up to ten phosphorylation sites have been noted on RPA32 (Ser4, Ser8, Ser11-13, Thr21, Ser23, Ser29, Ser33 and Thr98). The kinases that are known to phosphorylate RPA are ATM, ATR and DNA-protein kinase (DNA-PK). Although, the phosphorylation of RPA does not directly impact the process of DNA replication, some studies report an inhibitory effect (Vassin et al. 2004). It has been shown that RPA mediates recombination-based repair during replication stress (Sleeth et al. 2007). RPA's interaction with RAD52 in this repair pathway involves RPA1 and RPA2-wHLH domains (Jackson et al. 2002) and is activated by phosphorylation (Deng et al. 2009). So it can be surmised that the phosphorylation of RPA forms a link between signaling from a stalled replication fork to the initiation of DNA repair, mediated via extensive protein-protein interactions.

10.9 Concluding Remarks: Future Research on RPA

Despite the vast knowledge of RPA gained over the past three decades, RPA still poses an enigma to scientists interested in facets of DNA metabolism involving replication, recombination and repair. Although various aspects of RPA's binding to

DNA have been elucidated, mechanistically, the recruitment of RPA to ssDNA regions is still largely unknown. Thus, a key question that still remains is, how does RPA sense ssDNA regions? Does it remain loosely bound to DNA at all times in a “*cis*-fashion” or are there other signals that lead to a “*trans*” recruitment of RPA. The spatial-temporal regulation of RPA binding to DNA within a cell remains a mystery. Other related questions that influence our thinking about RPA include: What is the mechanism by which RPA is released from DNA so it can be handed-off to the next protein? Does phosphorylation of RPA play a role in facilitating the release of RPA from DNA by causing a conformational change in the protein? How does RPA recognize, bind and relax secondary structures formed in ssDNA regions? What is the global organization of RPA domains during all of its different functional states? How are these changes in architecture used to drive function? How do the various interaction domains serve as exchange points for different proteins and drive transitions in the DNA processing machinery?

Since the interactions between RPA and DNA are crucial in several different pathways, it is fathomable that disrupting this interaction could have disastrous deleterious effects on a cell. However, in the case of rapidly dividing cancer cells, targeting this interaction with small molecule inhibitors might enhance the efficacy of DNA damaging agents currently in use as chemotherapeutics. Recent studies by the Turchi lab have indicated that small molecule inhibitors *in vitro* can target the OB-folds of RPA. One such compound prevented cell cycle progression, induced cytotoxicity, and increased the efficacy of chemotherapeutic damaging agents (Anciano Granadillo et al. 2010; Shuck and Turchi 2010). In addition, through high-throughput screening, small molecule inhibitors of the N-terminal protein-protein interaction domain of RPA1 were discovered. Such novel compounds that disrupt RPA’s interactions with other proteins also possess further therapeutic potential (Glanzer et al. 2011). The knowledge of the full-length structure of RPA would aid in a more complete understanding of the protein and perhaps assist in the design of more potent small molecule inhibitors. In addition to being targeted by chemotherapeutic drugs, RPA has also been shown to be a prognostic indicator for patients with astrocytomas (Kanakis et al. 2011).

Until recently, RPA was thought to be the sole SSB involved in several processes involving DNA metabolism, however two novel proteins human SSB (hSSB1 and 2) were recently discovered to participate in DNA-damage signal transduction. These proteins are more closely related to the archaeal SSB in terms of domain structure. The relationship between RPA and the two hSSB proteins has not been completely teased out, although the roles for hSSB1 in DSB repair have been well documented (Richard et al. 2008).

For proper cellular function, it is apparent that the DNA within the cell has to be properly replicated and protected. Disturbing the peaceful equilibrium in the cell by DNA damaging agents can lead to replication stress, errors in replication, genomic instability, disease progression and/or cell-death. RPA is one of the key players in maintaining genomic integrity by its involvement in not only the complex replication process but also in the interrelated DNA-repair processes. Future experimental work on this complex protein is necessary and will help define how RPA performs its numerous roles in the cell.

Acknowledgements This work was supported by the American Cancer Society [RSG-02-162-01-GMC], NCI Eppley Cancer Center Support Grant [P30CA036727] and the Nebraska Department of Health and Human Services grants [2011-05 & 2012-04]. Aishwarya Prakash was supported by a University of Nebraska Medical Center graduate fellowship and Presidential graduate fellowship and would also like to thank Dr. Sylvie Doublie for her support [NIH/NCI P01CA098993].

References

- Anciano Granadillo VJ, Earley JN, Shuck SC, Georgiadis MM, Fitch RW, Turchi JJ (2010) Targeting the OB-folds of replication protein A with small molecules. *J Nucleic Acids* 2010:304035
- Arunkumar AI, Klimovich V, Jiang X, Ott RD, Mizoue L, Fanning E, Chazin WJ (2005) Insights into hRPA32 C-terminal domain-mediated assembly of the simian virus 40 replisome. *Nat Struct Mol Biol* 12:332–339
- Bae SH, Bae KH, Kim JA, Seo YS (2001) RPA governs endonuclease switching during processing of Okazaki fragments in eukaryotes. *Nature* 412:456–461
- Bae KH, Kim HS, Bae SH, Kang HY, Brill S, Seo YS (2003) Bimodal interaction between replication-protein A and Dna2 is critical for Dna2 function both *in vivo* and *in vitro*. *Nucleic Acids Res* 31:3006–3015
- Bansbach CE, Betous R, Lovejoy CA, Glick GG, Cortez D (2009) The annealing helicase SMARCAL1 maintains genome integrity at stalled replication forks. *Genes Dev* 23:2405–2414
- Bastin-Shanower SA, Brill SJ (2001) Functional analysis of the four DNA binding domains of replication protein A. The role of RPA2 in ssDNA binding. *J Biol Chem* 276:36446–36453
- Binz SK, Sheehan AM, Wold MS (2004) Replication protein A phosphorylation and the cellular response to DNA damage. *DNA Repair (Amst)* 3:1015–1024
- Bissler JJ (2007) Triplex DNA and human disease. *Front Biosci* 12:4536–4546
- Blackwell LJ, Borowiec JA, Mastrangelo IA (1996) Single-stranded-DNA binding alters human replication protein A structure and facilitates interaction with DNA-dependent protein kinase. *Mol Cell Biol* 16:4798–4807
- Bochkarev A, Bochkareva E (2004) From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. *Curr Opin Struct Biol* 14:36–42
- Bochkarev A, Pfuetzner RA, Edwards AM, Frappier L (1997) Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* 385:176–181
- Bochkarev A, Bochkareva E, Frappier L, Edwards AM (1999) The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. *EMBO J* 18:4498–4504
- Bochkareva E, Korolev S, Bochkarev A (2000) The role for zinc in replication protein A. *J Biol Chem* 275:27332–27338
- Bochkareva E, Belegu V, Korolev S, Bochkarev A (2001) Structure of the major single-stranded DNA-binding domain of replication protein A suggests a dynamic mechanism for DNA binding. *EMBO J* 20:612–618
- Bochkareva E, Korolev S, Lees-Miller SP, Bochkarev A (2002) Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J* 21:1855–1863
- Bochkareva E, Kaustov L, Ayed A, Yi GS, Lu Y, Pineda-Lucena A, Liao JC, Okorokov AL, Milner J, Arrowsmith CH, Bochkarev A (2005) Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. *Proc Natl Acad Sci USA* 102:15412–15417
- Braun KA, Lao Y, He Z, Ingles CJ, Wold MS (1997) Role of protein-protein interactions in the function of replication protein A (RPA): RPA modulates the activity of DNA polymerase α by multiple mechanisms. *Biochemistry* 36:8443–8454

- Broderick S, Rehm K, Concannon C, Nasheuer HP (2010) Eukaryotic single-stranded DNA binding proteins: central factors in genome stability. *Subcell Biochem* 50:143–163
- Brose CA, Chagot ME, Ehrhardt M, Pretto DI, Weiner BE, Chazin WJ (2009) NMR analysis of the architecture and functional remodeling of a modular multidomain protein, RPA. *J Am Chem Soc* 131:6346–6347
- Brosh RM Jr, Orren DK, Nehlin JO, Ravn PH, Kenny MK, Machwe A, Bohr VA (1999) Functional and physical interaction between WRN helicase and human replication protein A. *J Biol Chem* 274:18341–18350
- Campbell JL (1986) Eukaryotic DNA replication. *Annu Rev Biochem* 55:733–771
- Chedin F, Seitz EM, Kowalczykowski SC (1998) Novel homologs of replication protein A in archaea: implications for the evolution of ssDNA-binding proteins. *Trends Biochem Sci* 23:273–277
- Clugston CK, McLaughlin K, Kenny MK, Brown R (1992) Binding of human single-stranded DNA binding protein to DNA damaged by the anticancer drug *cis*-diamminedichloroplatinum (II). *Cancer Res* 52:6375–6379
- Constantinou A, Tarsounas M, Karow JK, Brosh RM, Bohr VA, Hickson ID, West SC (2000) Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest. *EMBO Rep* 1:80–84
- Cummings CJ, Zoghbi HY (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum Mol Genet* 9:909–916
- Dai J, Hatzakis E, Hurley LH, Yang D (2010) i-motif structures formed in the human c-MYC promoter are highly dynamic—insights into sequence redundancy and i-motif stability. *PLoS One* 5:e11647
- de Laat WL, Appeldoorn E, Sugasawa K, Weterings E, Jaspers NG, Hoeijmakers JH (1998) DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes Dev* 12:2598–2609
- Deng X, Habel JE, Kabaleswaran V, Snell EH, Wold MS, Borgstahl GE (2007) Structure of the full-length human RPA14/32 complex gives insights into the mechanism of DNA binding and complex formation. *J Mol Biol* 374:865–876
- Deng X, Prakash A, Dhar K, Baia GS, Kolar C, Oakley GG, Borgstahl GE (2009) Human replication protein A-Rad52-single-stranded DNA complex: stoichiometry and evidence for strand transfer regulation by phosphorylation. *Biochemistry* 48:6633–6643
- Dickson AM, Krasikova Y, Pestryakov P, Lavrik O, Wold MS (2009) Essential functions of the 32 kDa subunit of yeast replication protein A. *Nucleic Acids Res* 37:2313–2326
- Din S, Brill SJ, Fairman MP, Stillman B (1990) Cell-cycle-regulated phosphorylation of DNA replication factor A from human and yeast cells. *Genes Dev* 4:968–977
- Doherty KM, Sommers JA, Gray MD, Lee JW, von Kobbe C, Thoma NH, Kureekattil RP, Kenny MK, Brosh RM Jr (2005) Physical and functional mapping of the replication protein A interaction domain of the Werner and Bloom syndrome helicases. *J Biol Chem* 280:29494–29505
- Dornreiter I, Erdile LF, Gilbert IU, von Winkler D, Kelly TJ, Fanning E (1992) Interaction of DNA polymerase α -primase with cellular replication protein A and SV40 T antigen. *EMBO J* 11:769–776
- Eddy J, Maizels N (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res* 34:3887–3896
- Eddy J, Maizels N (2009) Selection for the G4 DNA motif at the 5' end of human genes. *Mol Carcinog* 48:319–325
- Fairman MP, Stillman B (1988) Cellular factors required for multiple stages of SV40 DNA replication *in vitro*. *EMBO J* 7:1211–1218
- Fan JH, Bochkareva E, Bochkarev A, Gray DM (2009) Circular dichroism spectra and electrophoretic mobility shift assays show that human replication protein A binds and melts intramolecular G-quadruplex structures. *Biochemistry* 48:1099–1111
- Fanning E, Klimovich V, Nager AR (2006) A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res* 34:4126–4137

- Fotedar R, Roberts JM (1992) Cell cycle regulated phosphorylation of RPA-32 occurs within the replication initiation complex. *EMBO J* 11:2177–2187
- Gellert M, Lipsett MN, Davies DR (1962) Helix formation by guanylic acid. *Proc Natl Acad Sci USA* 48:2013–2018
- Glanzer JG, Liu S, Oakley GG (2011) Small molecule inhibitor of the RPA70 N-terminal protein interaction domain discovered using *in silico* and *in vitro* methods. *Bioorg Med Chem* 19:2589–2595
- Gomes XV, Wold MS (1995) Structural analysis of human replication protein A. Mapping functional domains of the 70-kDa subunit. *J Biol Chem* 270:4534–4543
- Gomes XV, Wold MS (1996) Functional domains of the 70-kilodalton subunit of human replication protein A. *Biochemistry* 35:10558–10568
- Gomes XV, Henricksen LA, Wold MS (1996) Proteolytic mapping of human replication protein A: evidence for multiple structural domains and a conformational change upon interaction with single-stranded DNA. *Biochemistry* 35:5586–5595
- Gottifredi V, Prives C (2005) The S phase checkpoint: when the crowd meets at the fork. *Semin Cell Dev Biol* 16:355–368
- Grudic A, Jul-Larsen A, Haring SJ, Wold MS, Lonning PE, Bjerkvig R, Boe SO (2007) Replication protein A prevents accumulation of single-stranded telomeric DNA in cells that use alternative lengthening of telomeres. *Nucleic Acids Res* 35:7267–7278
- Han Y, Loo YM, Militello KT, Melendy T (1999) Interactions of the papovavirus DNA replication initiator proteins, bovine papillomavirus type 1 E1 and simian virus 40 large T antigen, with human replication protein A. *J Virol* 73:4899–4907
- Haring SJ, Mason AC, Binz SK, Wold MS (2008) Cellular functions of human RPA1. Multiple roles of domains in replication, repair, and checkpoints. *J Biol Chem* 283:19095–19111
- Haring SJ, Humphreys TD, Wold MS (2010) A naturally occurring human RPA subunit homolog does not support DNA replication or cell-cycle progression. *Nucleic Acids Res* 38:846–858
- He Z, Brinton BT, Greenblatt J, Hassell JA, Ingles CJ (1993) The transactivator proteins VP16 and GAL4 bind replication factor A. *Cell* 73:1223–1232
- He Z, Henricksen LA, Wold MS, Ingles CJ (1995) RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature* 374:566–569
- Hermanson-Miller IL, Turchi JJ (2002) Strand-specific binding of RPA and XPA to damaged duplex DNA. *Biochemistry* 41:2402–2408
- Huppert JL (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem Soc Rev* 37:1375–1384
- Huppert JL, Balasubramanian S (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* 35:406–413
- Hyrien O (2000) Mechanisms and consequences of replication fork arrest. *Biochimie* 82:5–17
- Iftode C, Daniely Y, Borowiec JA (1999) Replication protein A (RPA): the eukaryotic SSB. *Crit Rev Biochem Mol Biol* 34:141–180
- Jackson D, Dhar K, Wahl JK, Wold MS, Borgstahl GE (2002) Analysis of the human replication protein A:Rad52 complex: evidence for crosstalk between RPA32, RPA70, Rad52 and DNA. *J Mol Biol* 321:133–148
- Jacobs DM, Lipton AS, Isern NG, Daughdrill GW, Lowry DF, Gomes X, Wold MS (1999) Human replication protein A: global fold of the N-terminal RPA-70 domain reveals a basic cleft and flexible C-terminal linker. *J Biomol NMR* 14:321–331
- Jain A, Wang G, Vasquez KM (2008) DNA triple helices: biological consequences and therapeutic potential. *Biochimie* 90:1117–1130
- Jiang X, Klimovich V, Arunkumar AI, Hysinger EB, Wang Y, Ott RD, Guler GD, Weiner B, Chazin WJ, Fanning E (2006) Structural mechanism of RPA loading on DNA during activation of a simple pre-replication complex. *EMBO J* 25:5516–5526
- Kanakis D, Levidou G, Gakiopoulou H, Eftichiadis C, Thymara I, Fragkou P, Trigka EA, Boviatsis E, Patsouris E, Korkolopoulou P (2011) Replication protein A: a reliable biologic marker of prognostic and therapeutic value in human astrocytic tumors. *Hum Pathol* 10:1545–1553
- Kemp M, Sancar A (2009) DNA distress: just ring 9-1-1. *Curr Biol* 19:R733–R734

- Kemp MG, Mason AC, Carreira A, Reardon JT, Haring SJ, Borgstahl GE, Kowalczykowski SC, Sancar A, Wold MS (2010) An alternative form of replication protein A expressed in normal human tissues supports DNA repair. *J Biol Chem* 285:4788–4797
- Kenny MK, Schlegel U, Furneaux H, Hurwitz J (1990) The role of human single-stranded DNA binding protein and its individual subunits in simian virus 40 DNA replication. *J Biol Chem* 265:7693–7700
- Kim HS, Brill SJ (2001) Rfc4 interacts with Rpa1 and is required for both DNA replication and DNA damage checkpoints in *Saccharomyces cerevisiae*. *Mol Cell Biol* 21:3725–3737
- Kim C, Snyder RO, Wold MS (1992) Binding properties of replication protein A from human and yeast cells. *Mol Cell Biol* 12:3050–3059
- Kim C, Paulus BF, Wold MS (1994) Interactions of human replication protein A with oligonucleotides. *Biochemistry* 33:14197–14206
- Kobayashi Y, Sato K, Kibe T, Seimiya H, Nakamura A, Yukawa M, Tsuchiya E, Ueno M (2010) Expression of mutant RPA in human cancer cells causes telomere shortening. *Biosci Biotechnol Biochem* 74:382–385
- Kolpashchikov DM, Ivanova TM, Boghachev VS, Nasheuer HP, Weisshart K, Favre A, Pestryakov PE, Lavrik OI (2000a) Synthesis of base-substituted dUTP analogues carrying a photoreactive group and their application to study human replication protein A. *Bioconjug Chem* 11:445–451
- Kolpashchikov DM, Pestryakov PE, Wlassoff WA, Khodyreva SN, Lavrik OI (2000b) Study of interaction of human replication factor A with DNA using new photoreactive analogs of dTTP. *Biochemistry (Mosc)* 65:160–163
- Krishna SS, Majumdar I, Grishin NV (2003) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res* 31:532–550
- Kunkel TA, Burgers PM (2008) Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol* 18:521–527
- Lahue RS, Slater DL (2003) DNA repair and trinucleotide repeat instability. *Front Biosci* 8:s653–s665
- Liu Y, Vaithiyalingam S, Shi Q, Chazin WJ, Zinkel SS (2011) BID binds to replication protein A and stimulates ATR function following replicative stress. *Mol Cell Biol* 31:4298–4309
- Lopes J, Piazza A, Bermejo R, Kriegsman B, Colosio A, Teulade-Fichou MP, Foiani M, Nicolas A (2011) G-quadruplex-induced instability during leading-strand replication. *EMBO J* 30:4033–4046
- Machwe A, Lozada E, Wold MS, Li GM, Orren DK (2011) Molecular cooperation between the Werner syndrome protein and replication protein A in relation to replication fork blockage. *J Biol Chem* 286:3497–3508
- MacNeill SA (2001) DNA replication: partners in the Okazaki two-step. *Curr Biol* 11:R842–R844
- Marathias VM, Bolton PH (1999) Determinants of DNA quadruplex structural type: sequence and potassium binding. *Biochemistry* 38:4355–4364
- Mariappan SV, Silks LA 3rd, Chen X, Springer PA, Wu R, Moyzis RK, Bradbury EM, Garcia AE, Gupta G (1998) Solution structures of the Huntington's disease DNA triplets, (CAG)_n. *J Biomol Struct Dyn* 15:723–744
- Mason AC, Roy R, Simmons DT, Wold MS (2010) Functions of alternative replication protein A in initiation and elongation. *Biochemistry* 49:5919–5928
- Masuda-Sasa T, Polaczek P, Peng XP, Chen L, Campbell JL (2008) Processing of G4 DNA by Dna2 helicase/nuclease and replication protein A (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition. *J Biol Chem* 283:24359–24373
- McElhinny AS, Li JL, Wu L (2008) Mastermind-like transcriptional co-activators: emerging roles in regulating cross talk among multiple signaling pathways. *Oncogene* 27:5138–5147
- Melendy T, Stillman B (1993) An interaction between replication protein A and SV40 T antigen appears essential for primosome assembly during SV40 DNA replication. *J Biol Chem* 268:3389–3395
- Mer G, Bochkarev A, Chazin WJ, Edwards AM (2000a) Three-dimensional structure and function of replication protein A. *Cold Spring Harb Symp Quant Biol* 65:193–200

- Mer G, Bochkarev A, Gupta R, Bochkareva E, Frappier L, Ingles CJ, Edwards AM, Chazin WJ (2000b) Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. *Cell* 103:449–456
- Mirkin S (1999) Structure and biology of H DNA. In: Malvy C, Harel-Bellan A (eds) Triple helix forming oligonucleotides. Kluwer Academic, Norwell, pp 193–222
- Mirkin SM (2006) DNA structures, repeat expansions and human hereditary disorders. *Curr Opin Struct Biol* 16:351–358
- Mirkin SM (2007) Expandable DNA repeats and human disease. *Nature* 447:932–940
- Mirkin SM (2008) Discovery of alternative DNA structures: a heroic decade (1979–1989). *Front Biosci* 13:1064–1071
- Nuss JE, Sweeney DJ, Alter GM (2006) Reactivity-based analysis of domain structures in native replication protein A. *Biochemistry* 45:9804–9818
- Nuss JE, Sweeney DJ, Alter GM (2009) Prediction of and experimental support for the three-dimensional structure of replication protein A. *Biochemistry* 48:7892–7905
- Oakley GG, Patrick SM (2010) Replication protein A: directing traffic at the intersection of replication and repair. *Front Biosci* 15:883–900
- Oakley GG, Loberg LI, Yao J, Risinger MA, Yunker RL, Zernik-Kobak M, Khanna KK, Lavin MF, Carty MP, Dixon K (2001) UV-induced hyperphosphorylation of replication protein A depends on DNA replication and expression of ATM protein. *Mol Biol Cell* 12:1199–1213
- Parrilla-Castellar ER, Karnitz LM (2003) Cut5 is required for the binding of ATR and DNA polymerase α to genotoxin-damaged chromatin. *J Biol Chem* 278:45507–45511
- Patel DJ, Phan AT, Kuryavyy V (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res* 35:7429–7455
- Patrick SM, Turchi JJ (1998) Human replication protein A preferentially binds cisplatin-damaged duplex DNA *in vitro*. *Biochemistry* 37:8808–8815
- Pestryakov PE, Lavrik OI (2008) Mechanisms of single-stranded DNA-binding protein functioning in cellular DNA metabolism. *Biochemistry (Mosc)* 73:1388–1404
- Pestryakov PE, Weisshart K, Schlott B, Khodyreva SN, Kremmer E, Grosse F, Lavrik OI, Nasheuer HP (2003) Human replication protein A. The C-terminal RPA70 and the central RPA32 domains are involved in the interactions with the 3'-end of a primer-template DNA. *J Biol Chem* 278:17515–17524
- Pestryakov PE, Khlimankov DY, Bochkareva E, Bochkarev A, Lavrik OI (2004) Human replication protein A (RPA) binds a primer-template junction in the absence of its major ssDNA-binding domains. *Nucleic Acids Res* 32:1894–1903
- Pestryakov PE, Krasikova YS, Petrusheva IO, Khodyreva SN, Lavrik OI (2007) The role of p14 subunit of replication protein A in binding to single-stranded DNA. *Dokl Biochem Biophys* 412:4–7
- Pfuetzner RA, Bochkarev A, Frappier L, Edwards AM (1997) Replication protein A. Characterization and crystallization of the DNA binding domain. *J Biol Chem* 272:430–434
- Potterton L, McNicholas S, Krissinel E, Gruber J, Cowtan K, Emsley P, Murshudov GN, Cohen S, Perrakis A, Noble M (2004) Developments in the CCP4 molecular-graphics project. *Acta Crystallogr D Biol Crystallogr* 60:2288–2294
- Prakash A, Kieken F, Marky LA, Borgstahl GE (2011a) Stabilization of a G-quadruplex from unfolding by replication protein A using potassium and the porphyrin TMPyP4. *J Nucleic Acids* 2011:529828
- Prakash A, Natarajan A, Marky LA, Ouellette MM, Borgstahl GE (2011b) Identification of the DNA-binding domains of human replication protein A that recognize G-quadruplex DNA. *J Nucleic Acids* 2011:896947
- Preto DI, Tsutakawa S, Brosey CA, Castillo A, Chagot ME, Smith JA, Tainer JA, Chazin WJ (2010) Structural dynamics and single-stranded DNA binding activity of the three N-terminal domains of the large subunit of replication protein A from small angle X-ray scattering. *Biochemistry* 49:2880–2889
- Richard DJ, Bolderson E, Cubeddu L, Wadsworth RI, Savage K, Sharma GG, Nicolette ML, Tsvetanov S, McIlwraith MJ, Pandita RK, Takeda S, Hay RT, Gautier J, West SC, Paull TT,

- Pandita TK, White MF, Khanna KK (2008) Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature* 453:677–681
- Sakaguchi K, Ishibashi T, Uchiyama Y, Iwabata K (2009) The multi-replication protein A (RPA) system—a new perspective. *FEBS J* 276:943–963
- Salas TR, Petrusseva I, Lavrik O, Bourdoncle A, Mergny JL, Favre A, Saintome C (2006) Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic Acids Res* 34:4857–4865
- Salas TR, Petrusseva I, Lavrik O, Saintome C (2009) Evidence for direct contact between the RPA3 subunit of the human replication protein A and single-stranded DNA. *Nucleic Acids Res* 37:38–46
- Schweizer U, Hey T, Lipps G, Krauss G (1999) Photocrosslinking locates a binding site for the large subunit of human replication protein A to the damaged strand of cisplatin-modified DNA. *Nucleic Acids Res* 27:3183–3189
- Seroussi E, Lavi S (1993) Replication protein A is the major single-stranded DNA binding protein detected in mammalian cell extracts by gel retardation assays and UV cross-linking of long and short single-stranded DNA molecules. *J Biol Chem* 268:7147–7154
- Shuck SC, Turchi JJ (2010) Targeted inhibition of replication protein A reveals cytotoxic activity, synergy with chemotherapeutic DNA-damaging agents, and insight into cellular function. *Cancer Res* 70:3189–3198
- Sleeth KM, Sorensen CS, Issaeva N, Dziegielewska J, Bartek J, Helleday T (2007) RPA mediates recombination repair during replication stress and is displaced from DNA by checkpoint signalling in human cells. *J Mol Biol* 373:38–47
- Sogo JM, Lopes M, Foisani M (2002) Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science* 297:599–602
- Stigger E, Drissi R, Lee SH (1998) Functional analysis of human replication protein A in nucleotide excision repair. *J Biol Chem* 273:9337–9343
- Taneja P, Boche I, Hartmann H, Nasheuer HP, Grosse F, Fanning E, Weisshart K (2007) Different activities of the largest subunit of replication protein A cooperate during SV40 DNA replication. *FEBS Lett* 581:3973–3978
- Treuner K, Ramsperger U, Knippers R (1996) Replication protein A induces the unwinding of long double-stranded DNA regions. *J Mol Biol* 259:104–112
- Turchi JJ, Henkels KM, Hermanson IL, Patrick SM (1999) Interactions of mammalian proteins with cisplatin-damaged DNA. *J Inorg Biochem* 77:83–87
- Vassin VM, Wold MS, Borowiec JA (2004) Replication protein A (RPA) phosphorylation prevents RPA association with replication centers. *Mol Cell Biol* 24:1930–1943
- Voineagu I, Narayanan V, Lobachev KS, Mirkin SM (2008) Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc Natl Acad Sci USA* 105:9936–9941
- Voineagu I, Freudenreich CH, Mirkin SM (2009) Checkpoint responses to unusual structures formed by DNA repeats. *Mol Carcinog* 48:309–318
- Waga S, Stillman B (1994) Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication *in vitro*. *Nature* 369:207–212
- Weinberg DH, Collins KL, Simanek P, Russo A, Wold MS, Virshup DM, Kelly TJ (1990) Reconstitution of simian virus 40 DNA replication with purified proteins. *Proc Natl Acad Sci USA* 87:8692–8696
- Weisshart K, Taneja P, Fanning E (1998) The replication protein A binding site in simian virus 40 (SV40) T antigen and its role in the initial steps of SV40 DNA replication. *J Virol* 72:9771–9781
- Weisshart K, Pestryakov P, Smith RW, Hartmann H, Kremmer E, Lavrik O, Nasheuer HP (2004) Coordinated regulation of replication protein A activities by its subunits p14 and p32. *J Biol Chem* 279:35368–35376
- Wells RD (2007) Non-B DNA conformations, mutagenesis and disease. *Trends Biochem Sci* 32:271–278
- Wold MS (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem* 66:61–92

- Wold MS (2010) Eukaryotic replication protein A. Encyclopedia of life science. Wiley, Chichester
- Wold MS, Kelly T (1988) Purification and characterization of replication protein A, a cellular protein required for *in vitro* replication of simian virus 40 DNA. Proc Natl Acad Sci USA 85:2523–2527
- Wu Y, Rawtani N, Thazhathveetil AK, Kenny MK, Seidman MM, Brosh RM Jr (2008) Human replication protein A melts a DNA triple helix structure in a potent and specific manner. Biochemistry 47:5068–5077
- Xu X, Vaithiyalingam S, Glick GG, Mordes DA, Chazin WJ, Cortez D (2008) The basic cleft of RPA70N binds multiple checkpoint proteins, including RAD9, to regulate ATR signaling. Mol Cell Biol 28:7345–7353
- Yuzhakov A, Kelman Z, Hurwitz J, O'Donnell M (1999) Multiple competition reactions for RPA order the assembly of the DNA polymerase δ holoenzyme. EMBO J 18:6189–6199
- Zou L, Elledge SJ (2003) Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science 300:1542–1548
- Zou Y, Liu Y, Wu X, Shell SM (2006) Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses. J Cell Physiol 208:267–273

Chapter 11

Structural Biology of Replication Initiation Factor Mcm10

Wenyue Du, Melissa E. Stauffer, and Brandt F. Eichman

Abstract Minichromosome maintenance protein 10 (Mcm10) is a non-enzymatic replication factor required for proper assembly of the eukaryotic replication fork. Mcm10 interacts with single-stranded and double-stranded DNA, DNA polymerase α and Mcm2-7, and is important for activation of the pre-replicative complex and recruitment of subsequent proteins to the origin at the onset of S-phase. In addition, Mcm10 has recently been implicated in coordination of helicase and polymerase activities during replication fork progression. The nature of Mcm10's involvement in these activities, whether direct or indirect, remains unknown. However, recent biochemical and structural characterization of Mcm10 from multiple organisms has provided insights into how Mcm10 utilizes a modular architecture to act as a replisome scaffold, which helps to define possible roles in origin DNA melting, Pol α recruitment and coordination of enzymatic activities during elongation.

Keywords DNA binding • DNA replication • Mcm10 • Minichromosome maintenance • Protein-protein interaction

W. Du • B.F. Eichman (✉)
Departments of Biological Sciences and Biochemistry,
Center for Structural Biology, Vanderbilt University,
Nashville, TN 37232, USA
e-mail: wenyue.du@vanderbilt.edu; brandt.eichman@vanderbilt.edu

M.E. Stauffer
Scientific Editing Solutions, Walworth, WI 53184, USA
e-mail: melstauff@yahoo.com

11.1 Replication Initiation

DNA replication can be divided into three primary stages: initiation, elongation and termination (Bell and Dutta 2002; Garg et al. 2005). Initiation commences during the G1-phase of the cell cycle, during which the replisome – the protein complex responsible for DNA unwinding and synthesis at an active replication fork – begins to assemble at origins of replication (Fig. 11.1). Initiation begins with origin licensing, in which the origin recognition complex (ORC), coupled with Cdc6 and Cdt1, loads the minichromosome maintenance (Mcm) proteins Mcm2-7 onto DNA as a head-to-head double hexamer (Remus et al. 2009). This marks the formation of the pre-replicative complex (pre-RC), which remains inactive in G1-phase.

The transition to S-phase is accompanied by origin activation. Mcm10 is one of the first proteins loaded onto chromatin at the onset of S-phase and it is essential for the subsequent recruitment of other replisome proteins (Wohlschlegel et al. 2002). At this point, two phosphorylation events take place to activate the pre-RC. In yeast, cyclin-dependent kinase (CDK) phosphorylates Sld2 and Sld3 and facilitates their binding to Dpb11 (Tanaka et al. 2007; Zegerman and Diffley 2007) and Dbf4-dependent kinase (DDK), composed of Cdc7 and Dbf4, directly phosphorylates Mcm2 and Mcm4 (Lei et al. 1997; Sheu and Stillman 2006). Mcm10 is important for both of these events. It has been shown to stimulate Mcm2-7 phosphorylation by DDK and may, in fact, recruit DDK to the pre-RC (Lee et al. 2003). In addition, the human RecQ4 helicase contains a Sld2-like sequence that is both a phosphorylation target of CDK and a binding site for Mcm10, suggesting that phosphorylation may act as a switch for RecQ4 activity by modulating its interaction with Mcm10 (Xu et al. 2009). These phosphorylation events enable the subsequent loading of two helicase cofactors, Cdc45 and the GINS complex, to form the pre-initiation complex (pre-IC) with the help of Mcm10 and other factors, including And-1/Ctf4 (Im et al. 2009; Tanaka and Nasmyth 1998; Wohlschlegel et al. 2002; Zou and Stillman 2000). Cdc45, GINS, and Mcm2-7 form the CMG complex, which is considered to be the active form of the replicative helicase (see Chaps. 6–8, this volume; Costa et al. 2011; Gambus et al. 2006; Ilves et al. 2010; Moyer et al. 2006; Pacek et al. 2006). Denaturation of origin DNA into single strands forms two bidirectional replication forks and is marked by recruitment of single-stranded DNA (ssDNA) binding protein replication protein A (RPA, see Chap. 10, this volume).

The initiation phase concludes upon recruitment of the DNA synthesis machinery to the emerging replication fork. Fork firing requires DNA polymerase α (Pol α)-primase to initiate DNA synthesis by generating RNA primers and short stretches of DNA on both leading and lagging strands (see Chap. 9, this volume). Mcm10 and And-1/Ctf4 have been implicated in loading Pol α onto chromatin, as well as physical coupling of Pol α and Mcm2-7 (Gambus et al. 2009; Im et al. 2009; Lee et al. 2010; Ricke and Bielinsky 2004; Zhu et al. 2007). Elongation proceeds through processive DNA synthesis by replicative DNA polymerases δ and ϵ , which require the sliding clamp, proliferating cell nuclear antigen (PCNA), and the clamp loader, replication factor C (RFC) (see Chaps. 12–15, this volume). Fork progression requires concerted DNA unwinding and synthesis through coordination of activities among the CMG complex and polymerases α , δ , and ϵ .

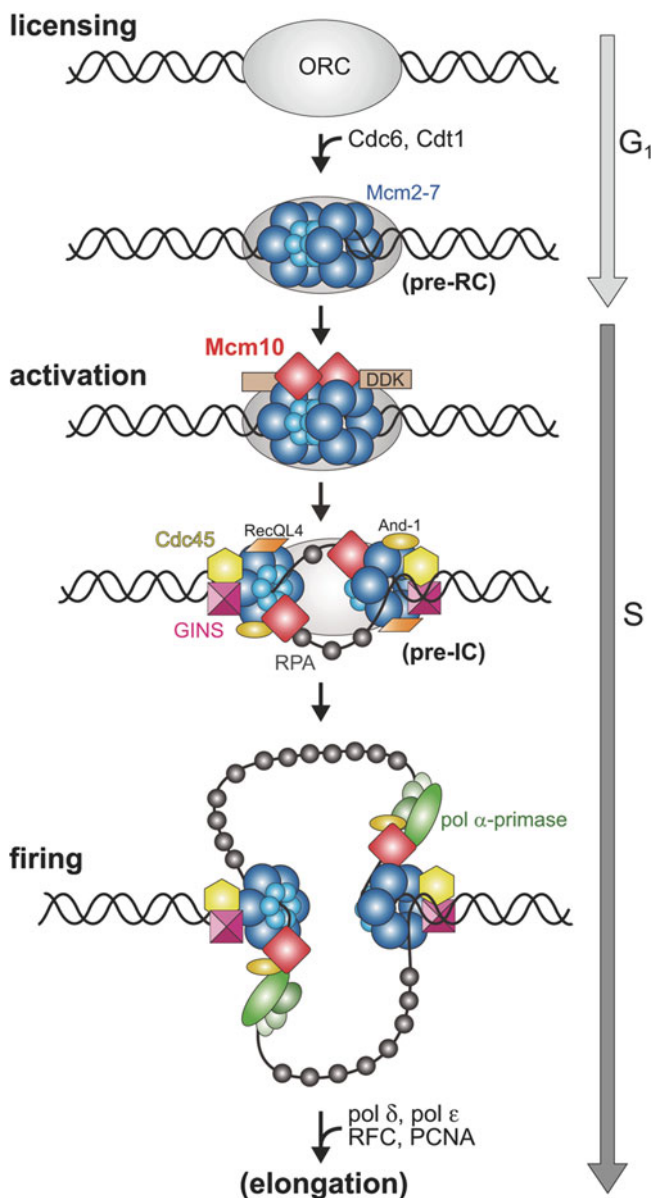


Fig. 11.1 A simplified view of the initiation phase of eukaryotic replication, highlighting key steps involved in replisome assembly. Many replication factors are omitted for clarity. At the end of the G₁ phase of the cell cycle, chromatin is licensed for replication at the origin by formation of a pre-replicative complex (pre-RC), which includes an inactive Mcm2-7 helicase. At the onset of S-phase, the pre-RC is activated by Dbf4-dependent kinase (DDK) phosphorylation. Mcm10 loads in early S-phase and is required for loading of Cdc45 and GINS, which form the CMG helicase complex with Mcm2-7 and help constitute a pre-initiation complex (pre-IC). Denaturation of origin DNA allows for binding of DNA polymerases and the rest of the elongation machinery, stimulating origin firing. Mcm10 and And-1/Ctf4 have been implicated in coupling Pol α to the replisome

11.2 Role of Mcm10 in Replication

The gene encoding Mcm10 was first identified in genetic screens in yeast. Referred to at the time as Cdc23, Mcm10 was shown to be necessary for cell division in *Schizosaccharomyces pombe* (Aves et al. 1998; Nasmyth and Nurse 1981). Bulk DNA synthesis was disrupted in temperature sensitive alleles of *cdc23*, and thus DNA replication and mitosis were blocked. Similar genes, referred to as *DNA43* and *MCM10*, were identified in screens in *Saccharomyces cerevisiae* and shown to encode homologs of Cdc23 (Dumas et al. 1982; Maine et al. 1984). *DNA43* was found to be essential for entering S-phase and maintaining cell viability (Solomon et al. 1992). Ricke and Bielinsky (2004) showed that the recruitment of *S. cerevisiae* Mcm10 (scMcm10) to replication origins is cell cycle regulated and dependent on pre-RC assembly, and that scMcm10 is required to maintain Pol α on chromatin independently of Cdc45. The importance of Mcm10 to replication initiation in yeast is evident from the number of genetic and physical interactions identified between Mcm10 and proteins involved in origin recognition, replisome assembly, and fork progression (Gregan et al. 2003; Hart et al. 2002; Homesley et al. 2000; Kawasaki et al. 2000; Merchant et al. 1997; Sawyer et al. 2004).

Mcm10 homologs have also been identified and characterized in higher eukaryotes, including humans, *Xenopus* and *Drosophila* (Christensen and Tye 2003; Izumi et al. 2000; Wohlschlegel et al. 2002). Human Mcm10 (hMcm10) interacts with chromatin at the G1/S-phase transition and dissociates in G2-phase (Izumi et al. 2000). It is important for activation of pre-RCs and functional assembly of the replisome and is regulated by phosphorylation-dependent proteolysis during late M- and early G1-phase (Izumi et al. 2001). Studies in *Xenopus* extracts showed that Mcm10 (xMcm10) binds to the pre-RC at the onset of S-phase, with roughly one xMcm10 bound per 5,000 bp of DNA (approximately two Mcm10s per active origin) (Wohlschlegel et al. 2002). These studies also showed Mcm10 to be essential for loading downstream proteins Cdc45 and RPA (Wohlschlegel et al. 2002), which are in turn required for chromatin unwinding and the association of Pol α at the origin (Walter and Newport 2000). The *Drosophila* homolog of Mcm10 was able to complement an *mcm10*-null strain of *S. cerevisiae* and was shown to interact with many members of the pre-RC in KC cells, including Mcm2, ORC and Cdc45 (Christensen and Tye 2003). Depletion of Mcm10 from KC cells led to defects in chromosome condensation (Christensen and Tye 2003).

The human, *Xenopus* and *Drosophila* Mcm10 orthologs have high sequence similarity but are distinct from the yeast proteins in several ways. First, the vertebrate proteins have an additional C-terminal domain (Robertson et al. 2008) (Fig. 11.2a). Second, phosphorylated and mono- and diubiquitylated forms of hMcm10 have been identified (Izumi et al. 2001), whereas only diubiquitylated Mcm10 has been shown to be associated with chromatin in yeast (Das-Bradoo et al. 2006). Finally, *S. pombe* Mcm10/Cdc23 (spMcm10) has been reported to contain primase activity (Fien and Hurwitz 2006), a function not observed in other orthologs.

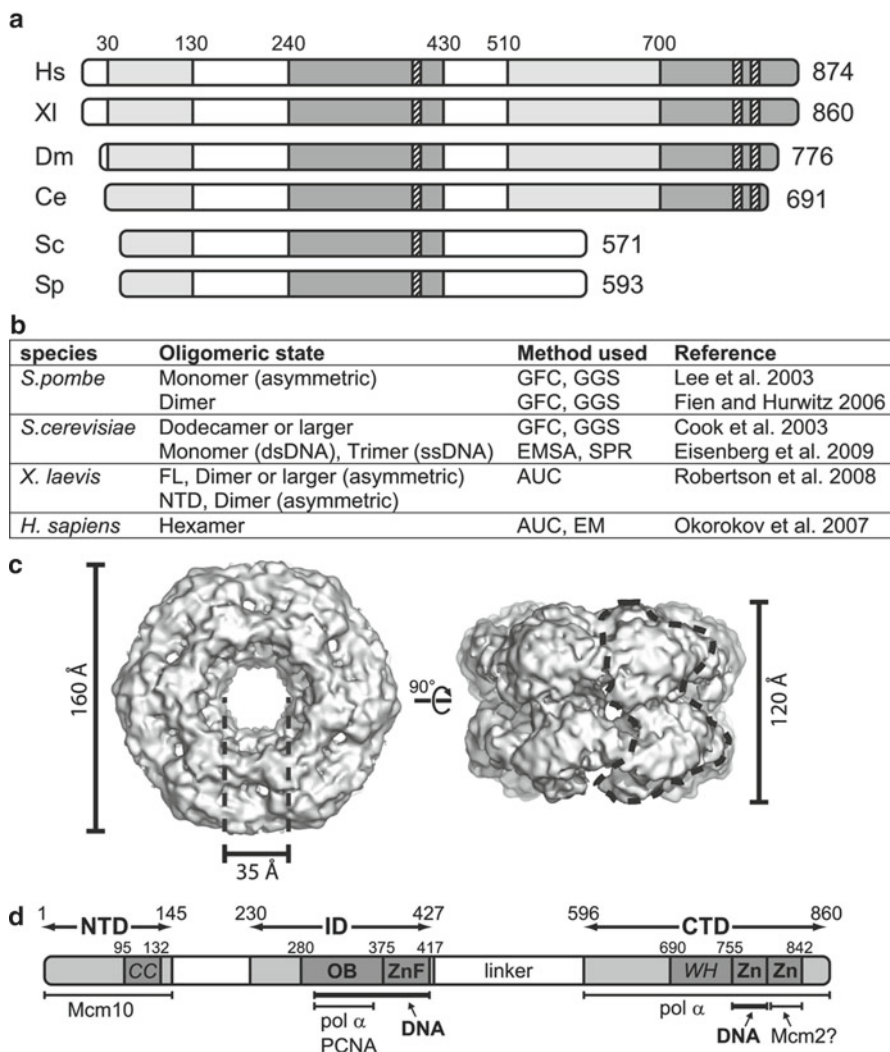


Fig. 11.2 Mcm10 sequence homology, oligomerization, and domain architecture. **(a)** A schematic sequence alignment of Mcm10 from *Homo sapiens* (*Hs*), *Xenopus laevis* (*Xl*), *Drosophila melanogaster* (*Dm*), *Caenorhabditis elegans* (*Ce*), *Saccharomyces cerevisiae* (*Sc*), and *Schizosaccharomyces pombe* (*Sp*). Light and dark grey bars indicate regions of moderate and high sequence conservation, respectively, and hatched boxes represent invariant cysteine/histidine clusters involved in zinc coordination. **(b)** Various oligomeric states of Mcm10 reported in the literature. GFC gel filtration chromatography; GGS glycerol gradient sedimentation; EMSA electrophoretic mobility shift assay, SPR surface plasmon resonance, AUC analytical ultracentrifugation, EM electron microscopy. **(c)** Orthogonal views of an EM reconstruction of human Mcm10 at 16 Å resolution and contoured at 1 σ . The dashed line represents one 95-kDa subunit. **(d)** Domain architecture of *Xenopus laevis* Mcm10. NTD N-terminal domain, ID internal domain, CTD C-terminal domain, CC predicted coiled coil, OB oligonucleotide/oligosaccharide binding fold, ZnF zinc-finger, WH predicted winged helix, Zn zinc ribbon. Interactions with proteins and DNA are shown below the schematic

Physical interactions have been observed between Mcm10 and multiple proteins found in the pre-RC and at the replication fork, including ORC (Hart et al. 2002; Izumi et al. 2000), Mcm2-7 (see below), Pol α (Chattopadhyay and Bielinsky 2007; Ricke and Bielinsky 2004, 2006) and the recently identified sister chromatid cohesion protein And-1 and the RecQ-like helicase RecQ4 (Xu et al. 2009; Zhu et al. 2007). spMcm10 interacts with Mcm4/6/7 and Dfp1, the *S. pombe* homolog of Dbf4 (Lee et al. 2003). Furthermore, spMcm10 has been reported to stimulate DDK phosphorylation of Mcm2-7 (Lee et al. 2003) and is thus believed to play a role in helicase activation. Studies in *S. cerevisiae* have shown that scMcm10 facilitates assembly of the Cdc45/Mcm2-7/GINS helicase complex (Gambus et al. 2009) and physical interactions have been observed with Mcm2, Mcm4, Mcm5, Mcm6, and Mcm7 subunits (Apger et al. 2010; Hart et al. 2002; Homesley et al. 2000; Merchant et al. 1997). Recent work suggested that scMcm10 serves as a functional linker between the MCM helicase and Pol α by coordinating their activities and ensuring their physical stability and integrity at the replication fork (Lee et al. 2010), a role also identified for Ctf4 (Gambus et al. 2009). *Drosophila* Mcm10 interacts with Mcm2, Dup (Cdt1), Orc2, Cdc45 and Hp1 in yeast two-hybrid assays (Christensen and Tye 2003). xMcm10 interacts with And-1/Ctf4 (Zhu et al. 2007) and with the helicase/nuclease Dna2 (Wawrousek et al. 2010). hMcm10 interacts with Orc2, Mcm2, and Mcm6 (Izumi et al. 2000) and assembly of the Cdc45/Mcm2-7/GINS complex in human cells requires Mcm10 as well as the And-1/Ctf4 and RecQL4 proteins (Im et al. 2009). hMcm10 also regulates the helicase activity of RecQ4 by direct binding (Xu et al. 2009).

Biochemical studies of Mcm10 have focused on its interactions with DNA and Pol α . Mcm10 binds to both single-stranded (ss) and double-stranded (ds) DNA, with about a three- to five-fold preference for ssDNA (Eisenberg et al. 2009; Fien et al. 2004; Robertson et al. 2008). Fien et al. (2004) showed that spMcm10 can stimulate DNA polymerase activity by interacting with both ssDNA and Pol α , leading to the idea that Mcm10 may facilitate the binding of polymerase complexes to primed DNA. Indeed, Mcm10 affects the localization and stability of Pol α , further supporting the idea that Mcm10 acts as a molecular chaperone for Pol α *in vivo* (Chattopadhyay and Bielinsky 2007; Ricke and Bielinsky 2004, 2006; Yang et al. 2005). Mcm10 interacts with the p180 subunit of Pol α in both yeast and *Xenopus* (Fien et al. 2004; Lee et al. 2010; Robertson et al. 2008). In fact, ssDNA and the N-terminal region of p180 compete for binding to the conserved internal domain of Mcm10 (Warren et al. 2009). Mcm10 can stabilize Pol α throughout the cell cycle by preventing its degradation by the proteasome (Chattopadhyay and Bielinsky 2007; Ricke and Bielinsky 2004, 2006). Moreover, Mcm10 appears to be a cofactor for Pol α activity by increasing its affinity for DNA (Fien et al. 2004; Zhu et al. 2007).

11.3 Overall Architecture

The Mcm10 protein exists only in eukaryotes; no orthologs have been identified in archaea or bacteria, although loose homology has been observed between regions of Mcm10 and the Mcm2-7 proteins (Robertson et al. 2010). Mcm10 proteins range in

size from 571 amino acids in yeast to 874 in humans, with regions of sequence homology clustered in the central and extreme N- and C-terminal regions (Fig. 11.2a). The spacing of homologous regions suggests the presence of three distinct structured domains tethered by unstructured linkers. Zinc finger motifs, initially identified from sequence alignments (Homesley et al. 2000; Izumi et al. 2000) and later confirmed by structural analysis (Robertson et al. 2010; Warren et al. 2008, 2009), are present in both the central and C-terminal regions. The yeast homologs lack the C-terminal region altogether (Robertson et al. 2008), suggesting that in lower organisms, the essential functions of Mcm10 reside within its N-terminal and central regions.

Biochemical and structural studies using vertebrate and yeast Mcm10 orthologs have been rather controversial with regard to the architecture and oligomeric state of the full-length Mcm10 protein (Fig. 11.2b). scMcm10 has been reported to form large, 800 kDa homocomplexes consisting of ~12 molecules when analyzed by size-exclusion chromatography (Cook et al. 2003), although the shape of the molecule could potentially confound this analysis. Self-association in yeast was shown to be mediated by the central zinc finger-containing domain, and mutations in the zinc-binding residues rendered yeast cells temperature-sensitive, with demonstrable replication defects (Cook et al. 2003; Homesley et al. 2000). A more recent surface plasmon resonance study showed that in the presence of ssDNA, scMcm10 forms complexes with three subunits (Eisenberg et al. 2009). On dsDNA, however, scMcm10 interacted as a monomer with a stoichiometry directly proportional to the length of the DNA (~1 scMcm10 per 21–24 bp). Work from the Hurwitz laboratory has reported highly asymmetric monomeric and dimeric forms of spMcm10 using glycerol gradient centrifugation (Fien and Hurwitz 2006; Lee et al. 2003). Analytical ultracentrifugation of xMcm10 was consistent with self-associated, asymmetric complexes, although the precise oligomeric state could not be determined from the data (Robertson et al. 2008). More recent work using size exclusion chromatography with multi-angle light scattering is indicative of xMcm10 complexes containing two to three subunits in the absence of DNA (W. Du and B.F. Eichman, unpublished). This is consistent with the presence of a coiled-coil domain at the N-terminus of the protein (Robertson et al. 2008) and the calculation of two molecules per replication origin based on the concentration of chromatin-bound Mcm10 in *Xenopus* extracts (Wohlschlegel et al. 2002).

hMcm10 was reported to form a ring-shaped hexameric structure using electron microscopy (EM) and single-particle analysis (Okorokov et al. 2007). The particle has dimensions of 160 Å × 120 Å, a 35 Å central channel (Fig. 11.2c) and a system of smaller lateral channels and inner chambers. The volume of the electron density calculated at the 1σ contour level using Chimera (Pettersen et al. 2004) is consistent with a particle of molecular weight 570 kDa or six 95-kDa subunits (unpublished result). From the side, individual subunits appear to adopt two distinct lobes. Model fitting with the structures available at the time suggested that each subunit within the hexamer had the same orientation, with the zinc molecules positioned toward the upper and lower edges of the ring (Okorokov et al. 2007). Subsequent crystal and NMR structures of individual Mcm10 domains, discussed below, cannot be unambiguously positioned into the EM density. The hexameric structure was reportedly consistent with its sedimentation behavior by analytical ultracentrifugation, although the experimental data were not presented (Okorokov et al. 2007).

The authors of the EM structure provided two explanations for hexamerization of Mcm10. The first was that this architecture may enable a topological link with DNA to allow for processive DNA binding like many other ring-shaped DNA-binding proteins. Another explanation was that Mcm10 inherited the hexameric fold from a DNA helicase ancestor but lost the helicase activity during evolution and instead now serves as a “docking” module to facilitate protein-protein interactions in DNA replication, such as Mcm2-7 helicase and Pol α (Chen et al. 2005; Okorokov et al. 2007; Pape et al. 2003; Patel and Picha 2000). It is enticing to speculate that a hexameric Mcm10 structure would provide an extensive binding interface for the six subunits of Mcm2-7, although there are no data to support such a hexamer-hexamer interaction and Mcm10 does not travel with the helicase that has been uncoupled from the replisome by inhibition of the polymerase with aphidicolin (Pacek et al. 2006). In light of the facts that a hexameric form of Mcm10 has not been reported in non-human orthologs, that other studies identify Mcm10 assemblies composed of two to three subunits and that only two molecules of Mcm10 are likely present at the origin, we offer an additional explanation—that the hexamer is simply one of several states occurring in cellular equilibrium and is needed under specific conditions during the onset of replication. For example, hexamerization may be used for sequestering the molecule at the replication fork or as a compact storage state of the protein during replication inactivity. We note that to be consistent with the available oligomerization data, the Mcm10 hexamer may in fact be a trimer of dimers.

11.4 Mcm10 Domain Structure

Biochemical and structural studies have been performed using xMcm10, which has 84% sequence similarity and 58% identity to the human protein. Limited proteolysis and mass spectrometry revealed that full-length xMcm10 is composed of three structured domains at the N-terminal (NTD; residues 1–145), internal (ID; 230–427) and C-terminal (CTD; 596–860) regions of the protein (Fig. 11.2d) (Robertson et al. 2008). The functional significance of the NTD is currently undefined, while the ID and CTD each bind DNA and Pol α (Robertson et al. 2008). Interdomain linkers are predicted to be largely unstructured by secondary structure and disorder predictions and by virtue of their extreme proteolytic sensitivity in purified preparations (Robertson et al. 2008).

11.4.1 Mcm10-NTD

Circular dichroism indicates that the NTD is predominantly α helical and random coil, consistent with secondary structure predictions. The NTD alone is a dimeric assembly as judged by analytical ultracentrifugation, consistent with the presence of a predicted coiled-coil motif comprising residues 93–132 (Robertson et al. 2008).

A strong yeast one-hybrid interaction from the first 100 residues of *Drosophila* Mcm10 was recently reported (Apger et al. 2010), suggesting that the NTD might function as an oligomerization domain for the full-length protein. Interestingly, self-interaction of ID and CTD regions was observed in yeast 2-hybrid assays when the NTD was deleted, suggesting that the NTD may not be the only point of contact between Mcm10 subunits. Nonetheless, the strong NTD self-interaction supports a proposed model in which Mcm10 forms a dimer with two subunits oriented in the same direction, which provides a plausible explanation for interaction of Mcm10 with both leading and lagging strand polymerases at a replication fork. Unlike the ID and CTD, the NTD does not bind to DNA (Robertson et al. 2008).

11.4.2 *Mcm10-ID*

The ID (residues 230–427) is homologous across all species from vertebrates to yeast and is the most conserved region in the entire protein (Izumi et al. 2000). Mutations in this region were identified in yeast genetic screens to affect minichromosome maintenance and replication *in vivo* (Grallert and Nurse 1997; Liang and Forsburg 2001; Maine et al. 1984; Nasmyth and Nurse 1981). The ID contains a CCCH-type zinc finger and an oligonucleotide/oligosaccharide binding (OB)-fold (Fig. 11.2d) that are now known to facilitate interactions with a number of proteins and DNA (Izumi et al. 2000; Ricke and Bielinsky 2006; Warren et al. 2008). Specifically, the ID has been shown to interact with ssDNA and the N-terminal 323 residues of Pol α (Robertson et al. 2008). In addition, a PCNA interacting peptide (PIP) region was identified in the sequence of scMcm10's ID (Das-Bradoo et al. 2006). Mutations within the PIP box abrogated the interaction between diubiquitylated Mcm10 and PCNA (Das-Bradoo et al. 2006; Warren et al. 2009).

The crystal structure of xMcm10-ID revealed that this region forms a globular domain consisting of an α -helical/random coil region (α A- α B, residues 230–283), an OB-fold (β 1- β 5.2, residues 286–375) and a C-terminal zinc finger motif (β C- α E, residues 378–418) (Fig. 11.3a). The α -helical/coil region packs onto the back of the OB-fold, opposite the canonical DNA-binding cleft of the OB-fold, to form a flat molecular surface (Warren et al. 2008). The zinc finger protrudes sideways relative to the OB-fold cleft and makes extensive electrostatic and van der Waals contacts with both the L23 loop of the OB-fold and the α -helical/coil region. It is interesting to note that the sequential arrangement of the OB-fold and the zinc finger in Mcm10-ID is different from other DNA processing proteins that contain both structural motifs. In the structures of the archaeal MCM helicase (see Chap. 6, this volume; Fletcher et al. 2003), the RPA trimerization core (see Chap. 10, this volume; Bochkareva et al. 2002), T4 gp32 (Shamoo et al. 1995), and NAD⁺-dependent DNA ligase (Lee et al. 2000), a zinc ribbon is inserted into the OB-fold L12 loop, whereas in Mcm10-ID the zinc finger is C-terminal to the entire OB-fold (Warren et al. 2008) (Figs. 11.3a, d). The unique arrangement of the OB-fold/zinc finger suggests that, in Mcm10, this domain assembly may have a unique function.

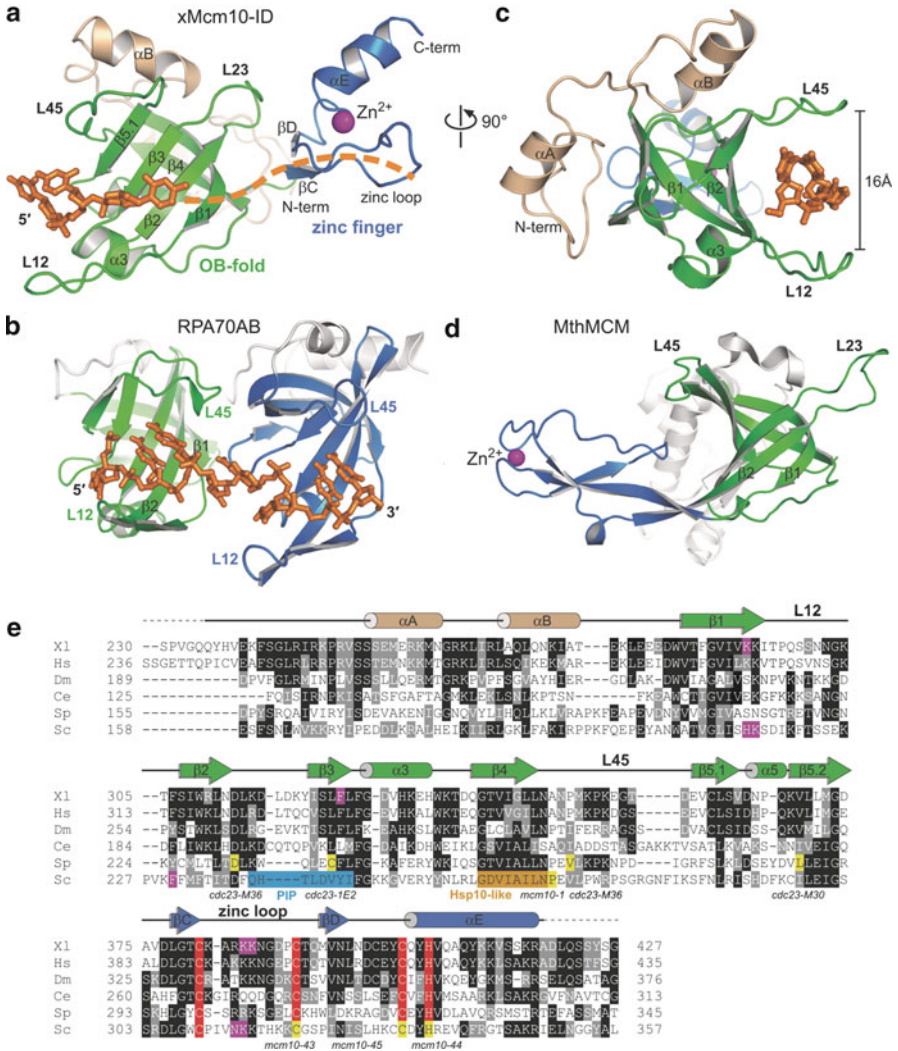


Fig. 11.3 The crystal structure of xMcm10-ID bound to ssDNA. (a) xMcm10-ID (residues 235–419) bound to ssDNA, with the OB-fold in green, the zinc finger in blue, the Zn²⁺ ion in magenta, and the N-terminal α -helical/coil region in tan. The three nucleotides of ssDNA observed in the crystal structure are shown as orange sticks. ssDNA traverses both the OB-fold cleft and the zinc loop. The trajectory of bound DNA determined by NMR is represented by the orange dashed line. (b) Crystal structure of RPA70AB subunit bound to ssDNA (PDB ID 1JMC). The OB-folds are colored green and blue, and the ssDNA orange. (c) Crystal structure of the xMcm10/ssDNA complex viewed 90° with respect to the view shown in panel (a). (d) Crystal structure of *Methanobacterium thermoautotrophicum* MCM (PDB ID 1LTL). (e) Sequence alignment of Mcm10-ID. Zn²⁺-coordinating residues are highlighted in red. Mutations identified in yeast genetic screens to affect cell growth and DNA replication are highlighted in yellow. Residues that affect xMcm10 binding to DNA *in vitro* or that increase the sensitivity of *S. cerevisiae* to HU are highlighted in pink. The PIP-box and Hsp10-like motif are highlighted in blue and gold, respectively

The tandem OB-fold-zinc finger arrangement in xMcm10-ID is reminiscent of the high affinity ssDNA binding surface created by side-by-side OB-folds in the RPA70AB sub-domain (Bochkarev et al. 1997) (Fig. 11.3b). NMR chemical shift perturbation of xMcm10-ID indicated that ssDNA binds to both the OB-fold cleft and to the extended loop of the highly basic zinc finger (Warren et al. 2008). The nature of the ssDNA interaction with the concave cleft of the OB-fold was revealed by the crystal structure of xMcm10-ID in complex with ssDNA (Warren et al. 2009). A tricytidine oligonucleotide was clearly observed within the OB-fold cleft, traversing β strands β 1- β 3 and β 5.1 (Fig. 11.3a). The channel created by loops L12 and L45 was ~ 16 Å in diameter (Fig. 11.3c), allowing the ssDNA a degree of flexibility that precluded observation of atomic-level interactions. However, the polarity of the ssDNA was unmistakably defined, with the 5' end oriented toward β 5.1 and the 3' end toward β 1 and the zinc finger, similar to the polarity reported for the RPA70AB structure (Bochkarev et al. 1997) (Fig. 11.3b). Also similar to RPA70AB, the L12 loop was unobservable in the unliganded structure, presumably due to flexibility (Warren et al. 2008), but upon DNA binding, its electron density was readily visible (Warren et al. 2009).

The Mcm10-ID zinc finger extends the ssDNA binding surface of the OB-fold in a manner analogous to the RPA70B subunit (Bochkarev et al. 1997). A crystal lattice contact occluded DNA binding by the zinc finger in the Mcm10-ID/ssDNA complex structure (Warren et al. 2009). Nonetheless, NMR chemical shift perturbation had unequivocally showed both this region and the cleft between it and the OB-fold to be affected by ssDNA binding, and residues in these regions were shown to affect DNA binding by xMcm10-ID and replication in yeast (Warren et al. 2008). A Lys385Glu/Lys386Glu double mutant on the extended zinc loop reduced ssDNA binding affinity by tenfold, and a Lys293Ala mutant in the cleft reduced it by fivefold (Warren et al. 2008). Transferring these mutations to yeast for assessment of their functional consequences showed that they increased the sensitivity of yeast cells to hydroxyurea (Warren et al. 2008). The Lys293Ala mutation (His215Ala/Lys216Ala in scMcm10) caused a twofold decline in cell survival, while the Lys385Glu/Lys386Glu mutation (Asn313Glu/Lys314Glu in yeast) led to a striking sevenfold decrease. Cell survival was also significantly compromised ($\sim 60\%$) in cells containing the Phe306Ala mutation (Phe230Ala/Phe231Ala in yeast), which resides in the cleft between the OB-fold and zinc finger. Interestingly, the zinc finger domain was also found to be affected by dsDNA binding (Warren et al. 2008). The presence of an extended loop in the zinc finger renders it structurally distinct from the archetypical Zif268 zinc finger that binds dsDNA in a sequence dependent manner, so it remains to be seen how dsDNA binds to this motif.

The Mcm10-ID crystal structures elucidated the yeast mutations originally identified to affect minichromosome maintenance and DNA replication (Fig. 11.3e). The *cdc23-1E2* (Cys239Tyr) (Grallert and Nurse 1997) and *cdc23-M30* (Leu287Pro) (Liang and Forsburg 2001) mutations, which correspond to xMcm10 Leu323 and Leu369, respectively, are located in the interior of the OB-fold's β -barrel, and thus are likely to cause structural perturbations that disrupt protein folding. Other mutations probably affected protein interactions necessary for replisome formation and/or

progression. These include *cdc23-M36* (Asp232Gly), corresponding to the invariant xMcm10 Asp313 that lies on the interior of the L23 loop, and *cdc23-M36* (Val265Ile) and *mcm10-1* (Pro269Leu), which map to solvent exposed positions in the L45 loop (Maine et al. 1984; Nasmyth and Nurse 1981). The human counterpart to the xMcm10-ID has been crystallized (Jung et al. 2008), but the structure was never determined and is expected to be virtually identical to the reported *Xenopus* domain on the basis of high sequence homology (58% identity; 84% similarity).

11.4.3 *Mcm10-CTD*

Vertebrate homologs of Mcm10 contain a CTD that is unique to higher eukaryotes; yeast Mcm10 is not predicted by sequence alignments to have this domain (Fig. 11.2a). Interactions between xMcm10-CTD (residues 596–860) and ssDNA, dsDNA, and Pol α have been mapped to a proteolytically stable subdomain (residues 690–842) that consists of a putative winged helix motif (residues 690–755) followed by tandem CCCH- and CCCC-type zinc motifs (residues 756–842) (Fig. 11.4a) (Robertson et al. 2010; Robertson et al. 2008). Heteronuclear NOE experiments on this region showed that the putative winged helix contains high backbone flexibility while the zinc motif is more rigid (Robertson et al. 2010). The two Zn²⁺ atoms in xMcm10-CTD, originally identified by atomic absorption spectroscopy, likely play a structural role based on the observations that, in the presence of EDTA, the CTD is more proteolytically sensitive and DNA binding affinity decreases (Robertson et al. 2008).

The solution NMR structure of the zinc binding region of xMcm10-CTD revealed a V-shaped globular domain in which the two zinc binding motifs are tethered by a hinge and the two zinc atoms bind at the tips of the V (Fig. 11.4b) (Robertson et al. 2010). The N-terminal CCCH zinc motif (residues 756–795) consists of a three-stranded antiparallel β -sheet capped with a short perpendicular α -helix with a Zn²⁺ ion embedded in between. DNA binding maps to the CCCH zinc motif, the structure of which is unique to Mcm10. The residues involved in DNA binding trace a nearly continuous 35 Å path around the CCCH arm (Robertson et al. 2010). The length of DNA required for maximal binding affinity was between 10 and 15 nucleotides, suggesting that all of the residues along that path are involved to some extent in interactions with DNA.

The CCCC zinc motif (residues 796–830) adopts a twisted antiparallel β -sheet with the zinc coordinated between the loops by the four cysteines. This motif is not involved in DNA binding and, interestingly, is identical in structure to a zinc ribbon motif in the N-terminal domain of *Methanobacterium thermoautotrophicum* MCM helicase (mtMCM) (Fig. 11.4c, see also Chap. 6, this volume). This MCM zinc motif mediates the head-to-head double hexamer assembly observed in mtMCM crystals (Fletcher et al. 2003, 2005) and in scMcm2-7 loaded onto DNA (Remus et al. 2009). The sequence of the CCCC zinc motif in xMcm10-CTD is highly conserved relative to those in the metazoan Mcm2-7 subunits and

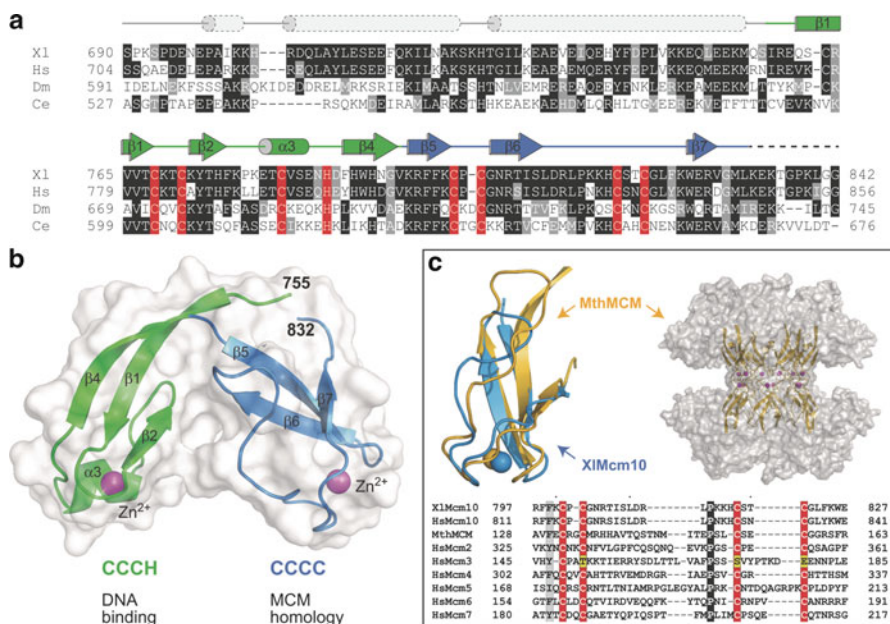


Fig. 11.4 NMR structure of xMcm10-CTD zinc binding region. (a) Sequence alignment of Mcm10-CTD. Secondary structure elements shown above the sequence are either predicted (grey) or determined from the NMR structure (green, blue). CCCH and CCCC zinc coordinating residues are highlighted in red. (b) The NMR structure of xMcm10 (756–842) is shown as a ribbon with a transparent grey molecular surface. (c) Structural and sequence alignment of the CCCC zinc ribbon from xMcm10 (aa 796–830) (blue) and MthMCM (gold, PDB ID 1LTL). In the crystal structure of the MthMCM N-terminal domain (top right of panel), the CCCC zinc ribbon mediates head-to-head double hexamer formation. The sequence alignment of this region (bottom of panel) shows the CCCC motif to be conserved in human Mcm2-7

in Mcm8 and Mcm9 proteins as well (Robertson et al. 2010). Mcm10 has been shown to interact directly with several subunits of Mcm2-7 helicase (Gambus et al. 2006; Izumi et al. 2000; Lee et al. 2003) and the recent finding that *Drosophila* Mcm10's interaction with Mcm2 is localized to the CTD (Aperger et al. 2010) suggests that the CCCC zinc motif in both proteins may be the point of contact. The Mcm2-7 double hexamer that is loaded onto chromatin in the pre-RC (Remus et al. 2009) is able to separate during DNA unwinding (Yardimci et al. 2010), leading us to predict that if Mcm10-Mcm2-7 interactions are indeed facilitated by the zinc motifs, then this interaction would take place only after fork firing. Of course this is highly speculative and additional experiments are needed to define this aspect of Mcm10's function. What is known with certainty is that Mcm10 interacts with both the helicase and Pol α . Most likely, Mcm10 serves as a scaffold to co-localize the essential players within the replisome during the initiation and elongation phases of replication (Lee et al. 2010; Ricke and Bielinsky 2004; Robertson et al. 2010).

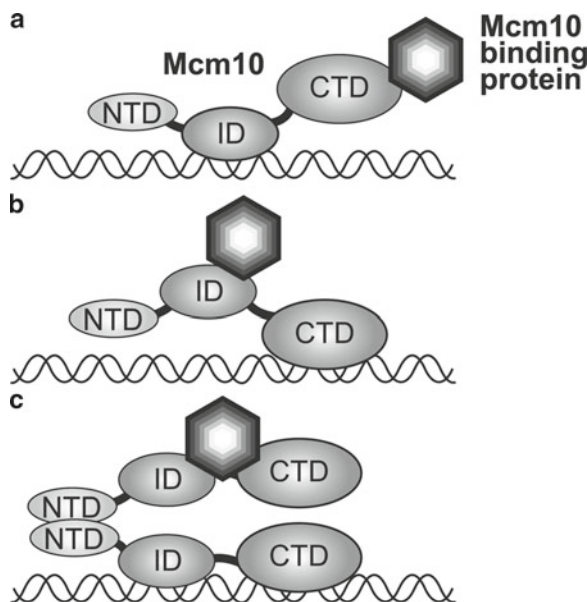
11.5 Implications of Modular Architecture for Function

Modular architecture is a common feature of DNA processing proteins that allows for the coordination of distinct biochemical activities (Stauffer and Chazin 2004). Flexible linkers between structured domains allow those domains to accommodate DNA and protein binding partners simultaneously, by virtue of the fact that the domains retain their structure while distance and angular adjustments are made between them. When tandem domains bind to the same entity, affinity for that entity is often increased relative to the strength of binding by one domain or the other. In addition, many DNA processing proteins contain bifunctional folds, which are known to bind both DNA and other proteins. Thus, when two different entities compete for the same binding site, it promotes molecular hand-off that facilitates the progression of the DNA processing pathway.

The structural organization of Mcm10 exemplifies each of these general features of DNA processing proteins. First, the attachment of the functional domains of Mcm10 by unstructured linkers, coupled with the spatial separation of protein and DNA binding sites, may allow it to bind both DNA and proteins simultaneously. Robertson et al. reported NMR spectra showing that the ID and the CTD of xMcm10 retain their individual structural properties in the context of a larger “ID+CTD” construct containing both domains and that the interdomain linker region is unstructured and flexible (Robertson et al. 2010). Second, the full-length xMcm10 protein, as well as the ID+CTD construct, binds DNA with 10-fold greater affinity than either the ID or CTD alone (Robertson et al. 2008; Warren et al. 2009), and ID+CTD binds Pol α p180 with 15-fold greater affinity than Mcm10-ID alone (Warren et al. 2009). Lastly, ssDNA and the N-terminal region of p180 compete for binding to the OB-fold cleft of Mcm10-ID (Warren et al. 2009). dsDNA also binds to essentially the same site on Mcm10-ID (Warren et al. 2008). Moreover, the PIP box predicted in the scMcm10 sequence (Ricke and Bielinsky 2006) coincides with the OB-fold β 3 strand, suggesting that the OB-fold can bind to PCNA as well. Indeed, xMcm10 Phe324, which corresponds to the residue in scMcm10 that mediates interaction with PCNA (Das-Bradoo et al. 2006), had a modest effect on DNA binding (Warren et al. 2008).

The interaction of multiple binding partners with identical sites on Mcm10 fits with two distinct models of molecular hand-off (Fig. 11.5a, b). In the first, Mcm10-ID binds to ssDNA while the CTD is used to recruit a protein partner (e.g., Pol α p180). In the second model, the CTD binds ssDNA and ID recruits a protein binding partner via its OB-fold. Hand-off would be facilitated in either scenario by competition between the binding partner and Mcm10 for the exposed ssDNA or alternatively, for the OB-fold in Mcm10-ID. Depending on the oligomeric state of Mcm10 *in vivo*, there could be more than one subunit of the ID and the CTD present at the origin, increasing the number of possible interaction points and competition events (Fig. 11.5c). Future studies of the oligomerization state of Mcm10 and of the order of events at the origin will be needed to clarify which model is in play at each stage of replication.

Fig. 11.5 Three possible models for Mcm10 hand-off of other proteins (e.g., Pol α) onto DNA. (a, b) Either Mcm10-ID or CTD interact with DNA, while the non-DNA bound domain is free to bind protein cargo. (c) ID+CTD together create a high-affinity DNA binding platform, and Mcm10 self-association through the NTD would present an additional free binding platform to localize proteins to the DNA



Its modular architecture and lack of enzymatic activity suggest that Mcm10 serves as a scaffold for the coupling of protein and DNA interactions during replication initiation. For example, a head-to-head Mcm10 dimer could couple events on the leading and lagging strands, or physically tether the helicase and Pol α , while retaining the polarity necessary for fork progression. Interestingly, recent studies have shown that yeast Mcm10 displays differential packing on ssDNA versus dsDNA (Eisenberg et al. 2009), suggestive of an Mcm10-DNA scaffold during origin melting or helicase unwinding. The authors speculated that a change in Mcm10 conformation or oligomeric state could facilitate strand separation.

11.6 Summary and Future Perspectives

Despite decades of work, Mcm10 remains an essential yet mysterious player in DNA replication. As one of the first proteins to load after pre-RC formation, Mcm10 is needed for subsequent protein loading and downstream events in DNA replication initiation. Mcm10 interacts with multiple replisome components and DNA. Its interactions with ssDNA and Pol α are mediated by the conserved ID and CTD through OB-fold and zinc finger structural elements. Crystal and NMR structures have elucidated the details of the ID-DNA interactions and have begun to address the binding activity within the CTD. Full-length xMcm10 forms a number of oligomeric species, which may be assembled through coiled coil interactions within the NTD. The functional significance of Mcm10's self-assembly and its interactions

within the replisome, the structure of the NTD, the mechanisms of multi-domain DNA binding activities, and the effects of ubiquitylation and other post-translational modifications on Mcm10 structure and function are all questions that remain unanswered.

The nature of Mcm10 self-assembly is critical for understanding its role at the replication fork, although the structural and functional relationship between the apparent multiple oligomeric states is not at all clear from the literature. A dimerization model best explains the physical and genetic evidence for Mcm10's interaction with both leading and lagging strand polymerases at a replication fork (Fien et al. 2004; Ricke and Bielinsky 2006; Robertson et al. 2008), but this remains to be determined. Detailed structural analyses of the N-terminal domain may help to address this issue. In addition, studies of the configuration(s) of the tandem ID and CTD in complex with DNA, Pol α and other protein binding partners will yield insight into the mechanisms by which Mcm10 acts as a scaffold at the replication origin.

A growing body of research also suggests Mcm10 plays a role in elongation. The first glimpse of a potential role for Mcm10 in fork progression came from the observation in *S. cerevisiae* that Mcm10 mutants delayed completion of DNA synthesis after cells were released from hydroxyurea arrest (Kawasaki et al. 2000). Mcm10 interacted genetically with Pol δ and Pol ϵ (Kawasaki et al. 2000) and a physical interaction with replisome progression complexes, which exist at DNA replication forks, has been observed in yeast (Gambus et al. 2006). In addition, a diubiquitylated form of scMcm10 interacts with PCNA, suggesting that Mcm10 directly participates in DNA elongation (Das-Bradoo et al. 2006). Pacek et al. showed that Mcm10 travels with the replication fork by inducing specific replication fork pausing on biotin-streptavidin-modified plasmids in *Xenopus* egg extracts (Pacek et al. 2006). In these experiments, Mcm10 was found to localize to the vertebrate DNA replication fork by chromatin immunoprecipitation. Finally, recent work in yeast suggested that Mcm10 coordinates the activities of the Mcm2-7 helicase and Pol α and ensures their physical stability at the elongating replication fork (Lee et al. 2010).

Although the majority of work to date has been focused on its role in DNA replication, Mcm10 has also been shown to be important for transcriptional gene silencing (Apger et al. 2010; Douglas et al. 2005; Liachko and Tye 2005, 2009). scMcm10 physically interacts with Sir2 and Sir3, two essential silencing factors in *S. cerevisiae* (Douglas et al. 2005). Moreover, Mcm10 mediates interactions between Sir2 and subunits 3 and 7 of the Mcm2-7 helicase via a ~100-residue segment at its C-terminus. Mutations to this region of Mcm10 caused silencing defects, but had no detrimental effect on replication (Liachko and Tye 2009). The corresponding segment in the *Xenopus* protein resides within an unstructured linker between the ID and CTD (Fig. 11.1b), which suggests that either the yeast protein has an organism-specific function or the vertebrate Mcm10s have an as yet uncharacterized role in gene silencing. The yeast sequence between residues 515 and 523 is predicted to be an amphipathic helix (Liachko and Tye 2009), a finding which warrants further investigation into the analogous segments of its orthologs.

In summary, Mcm10 lies at the heart of the replication initiation pathway. It loads early onto licensed replication origins and is necessary for pre-RC activation and

origin melting as a result of its interactions with DNA and many of the enzymes involved in fork progression. In addition to its essential role in establishing active replication forks at the origin, Mcm10 is involved in other aspects of genome utilization. The structures and interactions between Mcm10 and its binding partners are adding to a growing body of knowledge for how multi-conformation scaffolding proteins are used to maintain the integrity of the genome. To further enhance our understanding of the mechanisms involved in replisome assembly and function, the next step is to utilize the existing structures of the Mcm10 DNA binding domains as a foundation to build up larger sub-complexes, taking advantage of the extensive network of Mcm10 interactions. This higher resolution picture of the replisome will be critical to understand the transactions involved at the replication fork, including DNA synthesis, damage response and repair, and cell cycle regulation.

References

- Apger J, Reubens M, Henderson L, Gouge CA, Ilic N, Zhou HH, Christensen TW (2010) Multiple functions for *Drosophila* Mcm10 suggested through analysis of two *mcm10* mutant alleles. *Genetics* 185:1151–1165
- Aves SJ, Tongue N, Foster AJ, Hart EA (1998) The essential *Schizosaccharomyces pombe cdc23* DNA replication gene shares structural and functional homology with the *Saccharomyces cerevisiae DNA43 (MCM10)* gene. *Curr Genet* 34:164–171
- Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem* 71:333–374
- Bochkarev A, Pfuetzner RA, Edwards AM, Frappier L (1997) Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* 385:176–181
- Bochkareva E, Korolev S, Lees-Miller SP, Bochkarev A (2002) Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J* 21:1855–1863
- Chattopadhyay S, Bielinsky AK (2007) Human Mcm10 regulates the catalytic subunit of DNA polymerase α and prevents DNA damage during replication. *Mol Biol Cell* 18:4085–4095
- Chen YJ, Yu X, Kasiviswanathan R, Shin JH, Kelman Z, Egelman EH (2005) Structural polymorphism of *Methanothermobacter thermautotrophicus* MCM. *J Mol Biol* 346:389–394
- Christensen TW, Tye BK (2003) *Drosophila* MCM10 interacts with members of the prereplication complex and is required for proper chromosome condensation. *Mol Biol Cell* 14:2206–2215
- Cook CR, Kung G, Peterson FC, Volkman BF, Lei M (2003) A novel zinc finger is required for Mcm10 homocomplex assembly. *J Biol Chem* 278:36051–36058
- Costa A, Ilves I, Tamberg N, Petojevic T, Nogales E, Botchan MR, Berger JM (2011) The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol* 18:471–477
- Das-Bradoo S, Ricke RM, Bielinsky AK (2006) Interaction between PCNA and diubiquitinated Mcm10 is essential for cell growth in budding yeast. *Mol Cell Biol* 26:4806–4817
- Douglas NL, Dozier SK, Donato JJ (2005) Dual roles for Mcm10 in DNA replication initiation and silencing at the mating-type loci. *Mol Biol Rep* 32:197–204
- Dumas LB, Lussky JP, McFarland EJ, Shampay J (1982) New temperature-sensitive mutants of *Saccharomyces cerevisiae* affecting DNA replication. *Mol Gen Genet* 187:42–46
- Eisenberg S, Korza G, Carson J, Liachko I, Tye BK (2009) Novel DNA binding properties of the Mcm10 protein from *Saccharomyces cerevisiae*. *J Biol Chem* 284:25412–25420
- Fien K, Hurwitz J (2006) Fission yeast Mcm10p contains primase activity. *J Biol Chem* 281:22248–22260
- Fien K, Cho YS, Lee JK, Raychaudhuri S, Tappin I, Hurwitz J (2004) Primer utilization by DNA polymerase α -primase is influenced by its interaction with Mcm10p. *J Biol Chem* 279:16144–16153

- Fletcher RJ, Bishop BE, Leon RP, Sclafani RA, Ogata CM, Chen XS (2003) The structure and function of MCM from archaeal *M. thermoautotrophicum*. *Nat Struct Biol* 10:160–167
- Fletcher RJ, Shen J, Gomez-Llorente Y, Martin CS, Carazo JM, Chen XS (2005) Double hexamer disruption and biochemical activities of *Methanobacterium thermoautotrophicum* MCM. *J Biol Chem* 280:42405–42410
- Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, Labib K (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol* 8:358–366
- Gambus A, van Deursen F, Polychronopoulos D, Foltman M, Jones RC, Edmondson RD, Calzada A, Labib K (2009) A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase α within the eukaryotic replisome. *EMBO J* 28:2992–3004
- Garg P, Stith CM, Majka J, Burgers PM (2005) Proliferating cell nuclear antigen promotes translesion synthesis by DNA polymerase ζ . *J Biol Chem* 280:23446–23450
- Grallert B, Nurse P (1997) An approach to identify functional homologues and suppressors of genes in fission yeast. *Curr Genet* 32:27–31
- Gregan J, Lindner K, Brimage L, Franklin R, Namdar M, Hart EA, Aves SJ, Kearsey SE (2003) Fission yeast Cdc23/Mcm10 functions after pre-replicative complex formation to promote Cdc45 chromatin binding. *Mol Biol Cell* 14:3876–3887
- Hart EA, Bryant JA, Moore K, Aves SJ (2002) Fission yeast Cdc23 interactions with DNA replication initiation proteins. *Curr Genet* 41:342–348
- Homesley L, Lei M, Kawasaki Y, Sawyer S, Christensen T, Tye BK (2000) Mcm10 and the MCM2-7 complex interact to initiate DNA synthesis and to release replication factors from origins. *Genes Dev* 14:913–926
- Ilves I, Petojevic T, Pesavento JJ, Botchan MR (2010) Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell* 37:247–258
- Im JS, Ki SH, Farina A, Jung DS, Hurwitz J, Lee JK (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 USA* 106:15628–15632
- Izumi M, Yanagi K, Mizuno T, Yokoi M, Kawasaki Y, Moon KY, Hurwitz J, Yatagai F, Hanaoka F (2000) The human homolog of *Saccharomyces cerevisiae* Mcm10 interacts with replication factors and dissociates from nuclease-resistant nuclear structures in G(2) phase. *Nucleic Acids Res* 28:4769–4777
- Izumi M, Yatagai F, Hanaoka F (2001) Cell cycle-dependent proteolysis and phosphorylation of human Mcm10. *J Biol Chem* 276:48526–48531
- Jung NY, Bae WJ, Chang JH, Kim YC, Cho Y (2008) Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of the central zinc-binding domain of the human Mcm10 DNA replication factor. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64:495–497
- Kawasaki Y, Hiraga S, Sugino A (2000) Interactions between Mcm10p and other replication factors are required for proper initiation and elongation of chromosomal DNA replication in *Saccharomyces cerevisiae*. *Genes Cells* 5:975–989
- Lee JY, Chang C, Song HK, Moon J, Yang JK, Kim HK, Kwon ST, Suh SW (2000) Crystal structure of NAD⁺-dependent DNA ligase: modular architecture and functional implications. *EMBO J* 19:1119–1129
- Lee JK, Seo YS, Hurwitz J (2003) The Cdc23 (Mcm10) protein is required for the phosphorylation of minichromosome maintenance complex by the Dfp1-Hsk1 kinase. *Proc Natl Acad Sci USA* 100:2334–2339
- Lee C, Liachko I, Bouten R, Kelman Z, Tye BK (2010) Alternative mechanisms for coordinating polymerase α and MCM helicase. *Mol Cell Biol* 30:423–435
- Lei M, Kawasaki Y, Young MR, Kihara M, Sugino A, Tye BK (1997) Mcm2 is a target of regulation by Cdc7-Dbf4 during the initiation of DNA synthesis. *Genes Dev* 11:3365–3374
- Liachko I, Tye BK (2005) Mcm10 is required for the maintenance of transcriptional silencing in *Saccharomyces cerevisiae*. *Genetics* 171:503–515
- Liachko I, Tye BK (2009) Mcm10 mediates the interaction between DNA replication and silencing machineries. *Genetics* 181:379–391

- Liang DT, Forsburg SL (2001) Characterization of *Schizosaccharomyces pombe* *mcm7⁺* and *cdc23⁺* (*MCM10*) and interactions with replication checkpoints. *Genetics* 159:471–486
- Maine GT, Sinha P, Tye BK (1984) Mutants of *S. cerevisiae* defective in the maintenance of minichromosomes. *Genetics* 106:365–385
- Merchant AM, Kawasaki Y, Chen Y, Lei M, Tye BK (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* 17:3261–3271
- Moyer SE, Lewis PW, Botchan MR (2006) Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci USA* 103:10236–10241
- Nasmyth K, Nurse P (1981) Cell division cycle mutants altered in DNA replication and mitosis in the fission yeast *Schizosaccharomyces pombe*. *Mol Gen Genet* 182:119–124
- Okorokov AL, Waugh A, Hodgkinson J, Murthy A, Hong HK, Leo E, Sherman MB, Stoerber K, Orlova EV, Williams GH (2007) Hexameric ring structure of human MCM10 DNA replication factor. *EMBO Rep* 8:925–930
- Pacek M, Tutter AV, Kubota Y, Takisawa H, Walter JC (2006) Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol Cell* 21:581–587
- Pape T, Meka H, Chen S, Vicentini G, van Heel M, Onesti S (2003) Hexameric ring structure of the full-length archaeal MCM protein complex. *EMBO Rep* 4:1079–1083
- Patel SS, Picha KM (2000) Structure and function of hexameric helicases. *Annu Rev Biochem* 69:651–697
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera – a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612
- Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JF (2009) Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139:719–730
- Ricke RM, Bielinsky AK (2004) Mcm10 regulates the stability and chromatin association of DNA polymerase α . *Mol Cell* 16:173–185
- Ricke RM, Bielinsky AK (2006) A conserved Hsp10-like domain in Mcm10 is required to stabilize the catalytic subunit of DNA polymerase α in budding yeast. *J Biol Chem* 281:18414–18425
- Robertson PD, Warren EM, Zhang H, Friedman DB, Lary JW, Cole JL, Tutter AV, Walter JC, Fanning E, Eichman BF (2008) Domain architecture and biochemical characterization of vertebrate Mcm10. *J Biol Chem* 283:3338–3348
- Robertson PD, Chagot B, Chazin WJ, Eichman BF (2010) Solution NMR structure of the C-terminal DNA binding domain of MCM10 reveals a conserved MCM motif. *J Biol Chem* 285:22942–22949
- Sawyer SL, Cheng IH, Chai W, Tye BK (2004) Mcm10 and Cdc45 cooperate in origin activation in *Saccharomyces cerevisiae*. *J Mol Biol* 340:195–202
- Shamoo Y, Friedman AM, Parsons MR, Konigsberg WH, Steitz TA (1995) Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. *Nature* 376:362–366
- Sheu YJ, Stillman B (2006) Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. *Mol Cell* 24:101–113
- Solomon NA, Wright MB, Chang S, Buckley AM, Dumas LB, Gaber RF (1992) Genetic and molecular analysis of *DNA43* and *DNA52*: two new cell-cycle genes in *Saccharomyces cerevisiae*. *Yeast* 8:273–289
- Stauffer ME, Chazin WJ (2004) Structural mechanisms of DNA replication, repair, and recombination. *J Biol Chem* 279:30915–30918
- Tanaka T, Nasmyth K (1998) Association of RPA with chromosomal replication origins requires an Mcm protein, and is regulated by Rad53, and cyclin- and Dbf4-dependent kinases. *EMBO J* 17:5182–5191
- 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* 445:328–332

- Walter J, Newport J (2000) Initiation of eukaryotic DNA replication: origin unwinding and sequential chromatin association of Cdc45, RPA and DNA polymerase α . *Mol Cell* 5:617–627
- Warren EM, Vaithiyalingam S, Haworth J, Greer B, Bielinsky AK, Chazin WJ, Eichman BF (2008) Structural basis for DNA binding by replication initiator Mcm10. *Structure* 16:1892–1901
- Warren EM, Huang H, Fanning E, Chazin WJ, Eichman BF (2009) Physical interactions between Mcm10, DNA and DNA polymerase α . *J Biol Chem* 284:24662–24672
- Wawrousek KE, Fortini BK, Polaczek P, Chen L, Liu Q, Dunphy WG, Campbell JL (2010) *Xenopus* DNA2 is a helicase/nuclease that is found in complexes with replication proteins And-1/Ctf4 and Mcm10 and DSB response proteins Nbs1 and ATM. *Cell Cycle* 9:1156–1166
- Wohlschlegel JA, Dhar SK, Prokhorova TA, Dutta A, Walter JC (2002) *Xenopus* Mcm10 binds to origins of DNA replication after Mcm2-7 and stimulates origin binding of Cdc45. *Mol Cell* 9:233–240
- Xu X, Rochette PJ, Feyissa EA, Su TV, Liu Y (2009) MCM10 mediates RECQ4 association with MCM2-7 helicase complex during DNA replication. *EMBO J* 28:3005–3014
- Yang X, Gregan J, Lindner K, Young H, Kearsy SE (2005) Nuclear distribution and chromatin association of DNA polymerase α -primase is affected by TEV protease cleavage of Cdc23 (Mcm10) in fission yeast. *BMC Mol Biol* 6:13
- Yardimci H, Loveland AB, Habuchi S, van Oijen AM, Walter JC (2010) Uncoupling of sister replisomes during eukaryotic DNA replication. *Mol Cell* 40:834–840
- Zegerman P, Diffley JF (2007) Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* 445:281–285
- Zhu W, Ukomadu C, Jha S, Senga T, Dhar SK, Wohlschlegel JA, Nutt LK, Kornbluth S, Dutta A (2007) Mcm10 and And-1/CTF4 recruit DNA polymerase α to chromatin for initiation of DNA replication. *Genes Dev* 21:2288–2299
- Zou L, Stillman B (2000) Assembly of a complex containing Cdc45p, replication protein A and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbf4p kinase. *Mol Cell Biol* 20:3086–3096

Chapter 12

Structure and Function of Eukaryotic DNA Polymerase δ

Tahir H. Tahirov

Abstract DNA polymerase δ (Pol δ) is a member of the B-family DNA polymerases and is one of the major replicative DNA polymerases in eukaryotes. In addition to chromosomal DNA replication it is also involved in DNA repair and recombination. Pol δ is a multi-subunit complex comprised of a catalytic subunit and accessory subunits. The latter subunits play a critical role in the regulation of Pol δ functions. Recent progress in the structural characterization of Pol δ , together with a vast number of biochemical and functional studies, provides the basis for understanding the intriguing mechanisms of its regulation during DNA replication, repair and recombination. In this chapter we review the current state of the Pol δ structure-function relationship with an emphasis on the role of its accessory subunits.

Keywords DNA polymerase delta • Crystal structure • Catalytic subunit • Accessory subunits • Iron-sulfur cluster

12.1 Introduction

Eukaryotic DNA polymerase δ (Pol δ) belongs to the B-family of DNA polymerases (Pavlov et al. 2006). In addition to its central role in chromosomal DNA replication (Downey et al. 1990; Garg and Burgers 2005), Pol δ is involved in DNA repair and recombination (Baranovskiy et al. 2008; Gibbs et al. 2005; Huang et al. 2000, 2002; Lawrence 2002; Lydeard et al. 2007). In concert with Pol α -primase, it can synthesize both leading and lagging DNA strands in the SV40 system and in yeast

T.H. Tahirov (✉)

Eppley Institute for Research in Cancer and Allied Diseases,
University of Nebraska Medical Center, Omaha, NE 68198-7696, USA
e-mail: ttahirov@unmc.edu

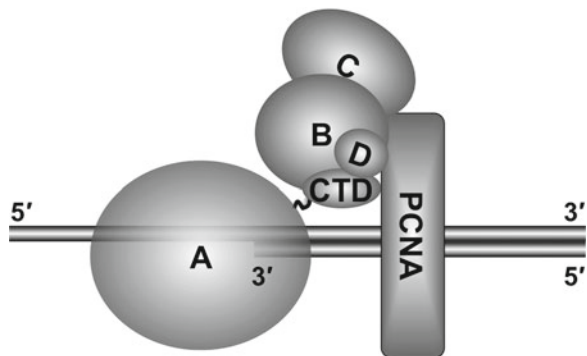
Table 12.1 Pol δ subunit designations

Designation	Human	<i>Sp</i>	<i>Sc</i>	Comments
Catalytic or A-subunit	p125	Pol3	Pol3p	Catalytic subunit; contains the polymerase and 3'→5' exonuclease active sites; interacts with B- and D-subunits and with PCNA
B-subunit	p50	Cdc1	Pol31p	Accessory subunit; interacts with A-, C- and D-subunits and with PCNA
C-subunit	p66	Cdc27	Pol32p	Accessory subunit; interacts with B-subunit and with PCNA
D-subunit	p12	Cdm1	–	Accessory subunit; interacts with A- and B-subunits and with PCNA

strains carrying a catalytically inactive Pol ϵ (Kesti et al. 1999; Waga et al. 1994). The best studied Pol δ s are those from human, the fission yeast *Schizosaccharomyces pombe* (*Sp*) and the budding yeast *Saccharomyces cerevisiae* (*Sc*). The former two Pol δ s are found as four-subunit complexes (Liu et al. 2000; Zuo et al. 1997), while the latter Pol δ is found only as a three-subunit complex (Gerik et al. 1998). A unified nomenclature for the subunits of these three Pol δ s has been proposed (MacNeill et al. 2001) and is provided in Table 12.1 along with updated comments on their roles and intersubunit interactions.

Currently published data on intermolecular interactions supports the overall architecture of Pol δ in which the A-, B- and D-subunits interact with each other and the C-subunit interacts only with the B-subunit, while all four subunits somehow interact with PCNA (Fig. 12.1). Progress on the three-dimensional structural characterization of Pol δ was slow for a long time and limited to the crystal structure of the p66 C-terminal peptide with PCNA (Bruning and Shamoo 2004). Initial significant progress was achieved when the crystal structure was solved for a complex of human p50 subunit with the N-terminal p50-interacting domain of the p66 subunit (Baranovskiy et al. 2008). The next notable progress made was the crystal structure determination of the Pol3p catalytic core in ternary complex with a template primer and an incoming nucleotide (Swan et al. 2009). Based on these recent achievements and a vast amount of biochemical, biophysical and functional studies, in this chapter we provide a review rationalizing the structure-function relationship for Pol δ . The first part of the chapter is devoted to the crystal structure of the A-subunit catalytic core and possible structural consequences of cancer-causing mutations, the organization and role of the A-subunit C-terminal domain (CTD), its metal-binding motifs (MBM) and comparisons with CTDs of other B-family members. The second part is devoted to a complex of the B- and C-subunits, its interaction with CTD and the interaction of the C-subunit with PCNA. The final third part is devoted to the least structurally characterized D-subunit, its interactions and functions, and speculation on the possibility of its evolution from the B-subunit.

Fig. 12.1 Cartoon representation of Pol δ in complex with DNA-loaded PCNA



12.2 Catalytic Subunit (A-subunit)

The purification of eukaryotic Pol δ from the cytoplasm of erythroid hyperplastic bone marrow was first reported in 1976 (Byrnes et al. 1976). Pol δ has been shown to possess both polymerase and 3'–5' exonuclease activities within a purified single 122 kDa polypeptide (Goscin and Byrnes 1982). This polypeptide is now known as a catalytic subunit of Pol δ and designated as an A-subunit (Table 12.1). The proof-reading exonuclease activity significantly enhances the fidelity of the A-subunit (Byrnes 1984; Simon et al. 1991). Cloning and sequencing of *Sp*, *Sc*, bovine and human Pol δ A-subunits revealed that, in general, their structures are similar (Chung et al. 1991; Pignede et al. 1991; Zhang et al. 1991). Primary structure analysis indicates that A-subunits could be considered as comprising of less conserved N-terminal sequences, the highly conserved catalytic core and CTD. Until recently, information for understanding the three-dimensional structure of the Pol δ catalytic core was derived mainly from the crystal structures of bacteriophage RB69 as well as its replicating and editing complexes (Franklin et al. 2001; Shamoo and Steitz 1999; Wang et al. 1997). Recently the 2.0 Å resolution crystal structure of *Sc* Pol3p catalytic core (residues 68–985) in ternary complex with a template primer and an incoming dCTP nucleotide has been reported by Swan and co-authors (Swan et al. 2009). This structure, which is reviewed below, enables precise mapping of functionally important residues and provides the basis for targeted genetic and functional studies.

12.2.1 Crystal Structure of Catalytic Core

The overall structure of the Pol3p catalytic core replicating complex is shown in Fig. 12.2a. The DNA is located in a clasp comprised of the thumb, palm, fingers and exonuclease domains. The thumb forms extensive contacts with the sugar-phosphate backbones of template primer around the minor groove while the palm interacts with

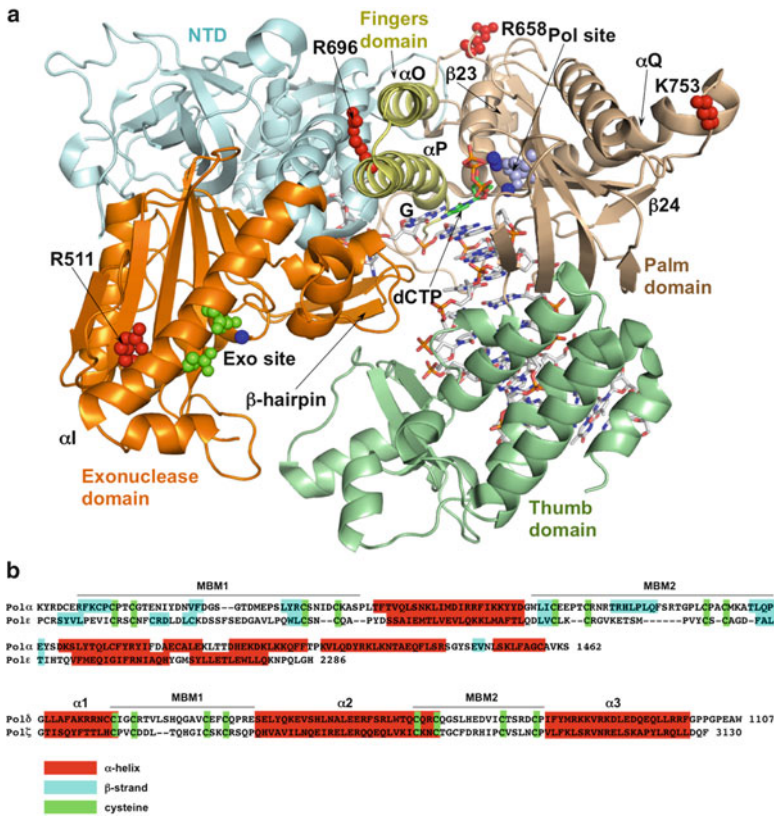


Fig. 12.2 A-subunit structure. (a) Cartoon representation of *Sc* Pol δ catalytic core (PDB code 3iay) (Swan et al. 2009). The NTD, palm, thumb, fingers, and exonuclease domains are color coded. The side chain atoms of exonuclease and polymerase active site residues are shown as light green and light cyan balls, respectively. The calcium ions are shown as blue balls. The side chain atoms of *Sc* Pol δ residues corresponding to cancer-causing mutations are shown as red balls. DNA and dCTP molecules are drawn as sticks. The panel (a) was prepared with PyMol software (Delano Scientific). (b) Comparison of human Pol α CTD amino acid sequences with Pol ϵ , and Pol δ with Pol ζ . The secondary structure elements predicted with Phyre server (Kelley and Sternberg 2009) and metal-coordinating cysteines are marked

the replicative end of the DNA. The nascent G-dCTP base pair packs against the second α -helix (αP) of the fingers domain. A long β -hairpin of exonuclease domain is extended toward the DNA major groove forming a hydrogen bond with a DNA base and a number of van der Waals contacts with the backbone of a template strand. Finally, the N-terminal domain (NTD), which is wedged between the exonuclease and fingers domains and interacting with the palm domain, contacts only the unpaired segment of the template strand. Two active site residues (Asp608 and Asp764) that catalyze the nucleotidyl transfer reaction are in the palm domain. The exonuclease active site residues Asp321, Glu323 and Asp407 are located close to a single Ca^{2+} . The distance between the polymerase and exonuclease active sites is ~ 45 Å.

Pol3p catalytic core exhibits several notable differences compared with the RB69 Pol. Their structures can be superimposed with a root-mean-square deviation of 2.50 Å for 631 matching α -carbons. The most notable differences are in NTD and the fingers domain. The NTD is significantly larger in Pol3p and contains an OB-fold and a RNA-binding motif, raising the possibility of RNA involvement in Pol δ function. Unlike the NTD, the fingers domain in Pol3p is comprised of only two shorter and straighter antiparallel α -helices than in RB69 Pol. Another significant difference is in DNA recognition by the long β -hairpin of exonuclease domain. In RB69 Pol structure this β -hairpin does not associate with DNA. Contrarily, in the Pol3p structure it is tightly bound to the template, which is consistent with its role in strand separation. The strand separation is proposed to be facilitated by holding the template strand in place by a β -hairpin that allows the primer strand to separate and migrate to the exonuclease active site (Hogg et al. 2007; Stocki et al. 1995).

12.2.2 Cancer-Causing Mutations

High-fidelity DNA replication is necessary for cell to avoid disease-causing mutations. In the case of Pol δ , DNA replication fidelity is maintained by accurate nucleotide selectivity and exonucleolytic proofreading activity. Furthermore, the DNA mismatch repair pathway contributes to an additional increase of fidelity. Engineering mutations leading to defects in Pol δ fidelity in mice increased the genome instability and accelerated tumorigenesis, supporting the mutator hypothesis for cancer (Albertson et al. 2009; Goldsby et al. 2001, 2002; Venkatesan et al. 2007). Unlike the mutations in mice Pol δ , which were intentionally modeled to destroy specific polymerase functions, the mechanisms of actions of cancer-causing mutations in the human Pol δ A-subunit are not apparent. However, using the structure of the Pol3p catalytic core it became possible to map the location of cancer-causing mutations and provide the rationale for their structure-function relationship (Swan et al. 2009).

Examination of Pol δ mRNA in six colon cancer cell lines (DLD-1, HCT116, SW48, HT29, SW480 and SW620) and seven sporadic human colorectal cancers revealed Arg506His, Arg689Trp and Ser746Ile substitutions located in the catalytic core (Fig. 12.2a – note that in the figure these are numbered according to their location in *Sc* Pol3p, as described below) (da Costa et al. 1995; Flohr et al. 1999).

Arg506 of p125 is aligned with Arg511 of Pol3p that is located in a long α -helix (α I) opposite the active site of the exonuclease domain. Replacement of this arginine residue by histidine may alter its interactions with the surrounding residues from loop β 13 β 14 and consequently shift the position of the α I helix. Such a shift of the α I helix may cause an allosteric effect on the exonuclease activity of Pol δ . The Arg511His substitution in yeast Pol δ resulted in a small but significant (2.5-fold) increase in the rate of mutagenesis in a MMR-deficient strain (Dae et al. 2010). Because the amino acid sequences and hence local structures are different between the yeast and human Pol δ s, a more profound effect of Arg506His substitution is a possibility for human Pol δ .

Arg689 of p125 is the equivalent of Pol3p Arg696 which is located in the second α -helix (α P) of the fingers domain. The Arg696 side chain is sandwiched between the NTD and the fingers domain. Modeling of Trp696 in place of Arg696 shows that, in any allowable conformation, the bulky tryptophan side chain creates significant clashes with the surrounding residues. In order to release the clashes and sterical hindrance introduced by Arg696Trp substitution, it is necessary to have a significant shift of, and/or conformational changes in, the α P helix. Because the α P helix is one of the most important determinants for the selectivity of incoming nucleotide, any shift in its position will inevitably affect the accuracy of base selection. Indeed, the Arg696Trp substitution is lethal in yeast, with the lethality caused by a catastrophic increase in spontaneous mutagenesis attributed to low-fidelity DNA synthesis (Daee et al. 2010).

The effect of Ser746Ile mutation on the catalytic activities of Pol δ is less clear. The equivalent residue in Pol3p is Lys753 which is located in loop α Q β 24 at the edge of the palm domain. The Lys753 side chain is completely solvent-exposed. However, it is possible that the local structure of the loop α Q β 24 is different in human Pol δ and that Ser746 is less exposed; consequently, its replacement by a bulkier isoleucine may result in conformational changes. Another possibility is that substitution of serine by isoleucine introduces conformational changes in the loop α Q β 24 by packing the isoleucine side chain in the nearby hydrophobic pocket. In both such scenarios an allosteric effect of Ser746Ile mutation to the polymerase active site is possible.

Pol δ from highly malignant Novikoff hepatoma cells was found to contain an Arg648Gln substitution (Popanda et al. 1999). Several biochemical characteristics of this polymerase were found to be altered, including a decrease in copying fidelity. The Arg648 equivalent in Pol3p is Arg658, which is located in loop β 23 α O connecting the fingers and palm domains (Fig. 12.2a). Thus the conformational changes caused by Arg648Gln substitution could allosterically alter the selectivity of incoming nucleotide via the shift of the α O and α P helices.

In summary, an inspection of the structure-function relationship of the reported cancer-causing amino acid replacements in the Pol δ A-subunit catalytic core indicates that all these mutations have a potential to alter its exonuclease or polymerase activities via allosteric effects. That is consistent with the mutator properties of Pol δ in cancer.

12.2.3 C-terminal Domain

Two putative zinc finger motifs (ZnF1 and ZnF2) each with four conserved metal-coordinating cysteines has been identified in the CTD of Pol δ s from a variety of sources (Sanchez Garcia et al. 2004). ZnF2 was found to be both necessary and sufficient for binding to the B-subunit *in vivo* and *in vitro*. The ZnF2 metal-binding site is critical for the activity of *Sc* and *Sp* Pol δ s since substitution of ZnF2 cysteines with alanines abolishes B-subunit binding and their *in vivo* function. Recently Netz

and co-authors have established that the cysteines in ZnF2, but not in ZnF1, coordinate the [4Fe-4S] cluster instead of the previously assumed zinc ion (Netz et al. 2011). They also determined that the ZnF1 is required for PCNA-binding and PCNA-mediated Pol δ processivity. However the nature of the metal in ZnF1 motif has not yet been clarified and in further discussions the ZnF1 and ZnF2 will be referenced as the metal-binding motifs 1 and 2 (MBM1 and MBM2).

At present, the only reported CTD structure is the structure of yeast Pol α CTD in complex with its B-subunit (Klinge et al. 2009) (see Chap. 9, this volume). Pol α CTD forms an elongated bi-lobal shape. The two lobes are connected by a three-helix bundle and each lobe contains an MBM. In contrast to what has been observed with the Pol δ CTD, the conserved cysteines in both MBM1 and MBM2 of Pol α CTD are coordinated with zinc ions. The CTD of Pol α forms a stable complex with the B-subunit. However, in case of Pol δ the CTD interaction with the B-subunit is dependent on the stability of the [4Fe-4S] cluster. The complex becomes unstable and dissociates at ambient conditions due to loss of iron from the Fe-S cluster. In addition to the differences in MBM2 metal content, the primary structure of Pol δ CTD is nearly two times shorter than in Pol α CTD. The secondary structure prediction of Pol δ CTD using the Phyre server (Kelley and Sternberg 2009) revealed three helices alternated with MBM1 and MBM2 in the order: helix-MBM1-helix-MBM2-helix (Fig. 12.2b). Assuming the resemblance of overall architecture of Pol α and Pol δ CTDs, the predicted helices in Pol δ CTD would be packed as a three-helix bundle flanked by the MBM1 and MBM2 lobes. However, unlike Pol α CTD having central three-helix bundle separate of MBMs, MBM2 in particular, the three-helix bundle of Pol δ CTD appears to comprise a part of its MBMs. In Pol δ the first conserved cysteine residue of MBM1 is located at the C-terminus of the first helix while the first and second conserved cysteine residues of MDM2 are protruding from the second helix, pointing to significant structural differences between the MBMs of Pol α and Pol δ . Contrarily, in Pol α MBMs the metal-coordinating cysteines are located in either loops or β -strands. The discrepancy in structure is especially apparent for MBM2 which results in the different nature of their ligands: Zn²⁺ in Pol α and [4Fe-4S] cluster in Pol δ .

12.2.4 Similarities Between C-terminal Domains of Pol δ and Pol ζ

Unlike the bulkier CTDs of Pol α and Pol ϵ , the CTD of Pol ζ appears to have a structure which is highly similar to Pol δ 's CTD (Fig. 12.2b). The alignment of amino acid sequences and secondary structures of the latter two CTDs shows that the length and positions of their helices as well as the positions of metal-coordinating cysteines coincide (Fig. 12.2b). Their similarity raises the possibility of similarity in the structures of their binding partners as well. The CTDs of Pol α , Pol ϵ and Pol δ are known to bind their respective B-subunits. However, a B-subunit was not found in Pol ζ . The three following explanations can be posited for this. First, Pol ζ does

not possess a B-subunit. Second, a B-subunit was not discovered due to low expression levels or other technical complications. Third, Pol ζ shares a B-subunit with Pol δ . The latter case would fit with a plausible model of efficient polymerase switch at replication-blocking lesions during translesion DNA synthesis (TLS). According to such a model, Pol δ and Pol ζ will exchange only their catalytic subunits on a preassembled complex while the shared accessory subunits of Pol δ will be retained by DNA-loaded PCNA. This model is in accordance with previously observed involvement of Pol δ accessory subunits in TLS (Gerik et al. 1998; Gibbs et al. 2005; Huang et al. 2000, 2002; Lydeard et al. 2007).

12.3 B- and C-subunits

The B- and C-subunits of both yeast and mammalian Pol δ s have been cloned, purified and characterized (Gerik et al. 1998; Hughes et al. 1999; Mo et al. 2000; Zhang et al. 1995; Zuo et al. 1997). The sequence homology is high among the B-subunits with amino acid sequence similarity of 23–28%. However, the homology is low among the C-subunits. Using the two-hybrid screening, human p66 has been shown to contain p50- and PCNA-binding domains within the 144 N- and 20 C-terminal amino acids, respectively (Pohler et al. 2005). A similar domain arrangement was found also for Pol32p (Johansson et al. 2004). The B-subunit interacts with the A-, C- and D-subunits, as well as PCNA (Lu et al. 2002). The C-subunit also interacts with PCNA (Bruning and Shamo 2004; Zhong et al. 2006). The parts of the B-subunit responsible for interactions with the D-subunit have not been defined. Interestingly, many essential functions of Pol δ , including the regulation of replication, TLS and BIR, are mediated by the C-subunit (Gerik et al. 1998; Gibbs et al. 2005; Huang et al. 2000, 2002; Lydeard et al. 2007) and thus apparently depend on the interaction between its B- and C-subunits. The crystal structure of human p50 in complex with the N-terminal domain of human p66 is described below.

12.3.1 Crystal Structure of p50•p66_N

The B-subunit of the human Pol δ is the first among B-subunits of the B-family DNA polymerases whose crystal structure has been solved, providing insights into the overall architecture of B-subunits and their structure-function relationship (Baranovskiy et al. 2008). The structure of p50 was determined in complex with the 144-amino-acids N-terminal domain of the C-subunit (p66_N) and the complex hereinafter will be referred as p50•p66_N. The structure revealed an extended cashew-shaped molecule with three domains: phosphodiesterase-like (PDE) and oligonucleotide/oligosaccharide-binding (OB) domains in p50, and a winged helix-turn-helix (wHTH) domain in p66_N (Fig. 12.3a). The PDE domain is located in the center of the molecule and bound to an OB domain on one side, and a wHTH domain on the other side. The fusion of the

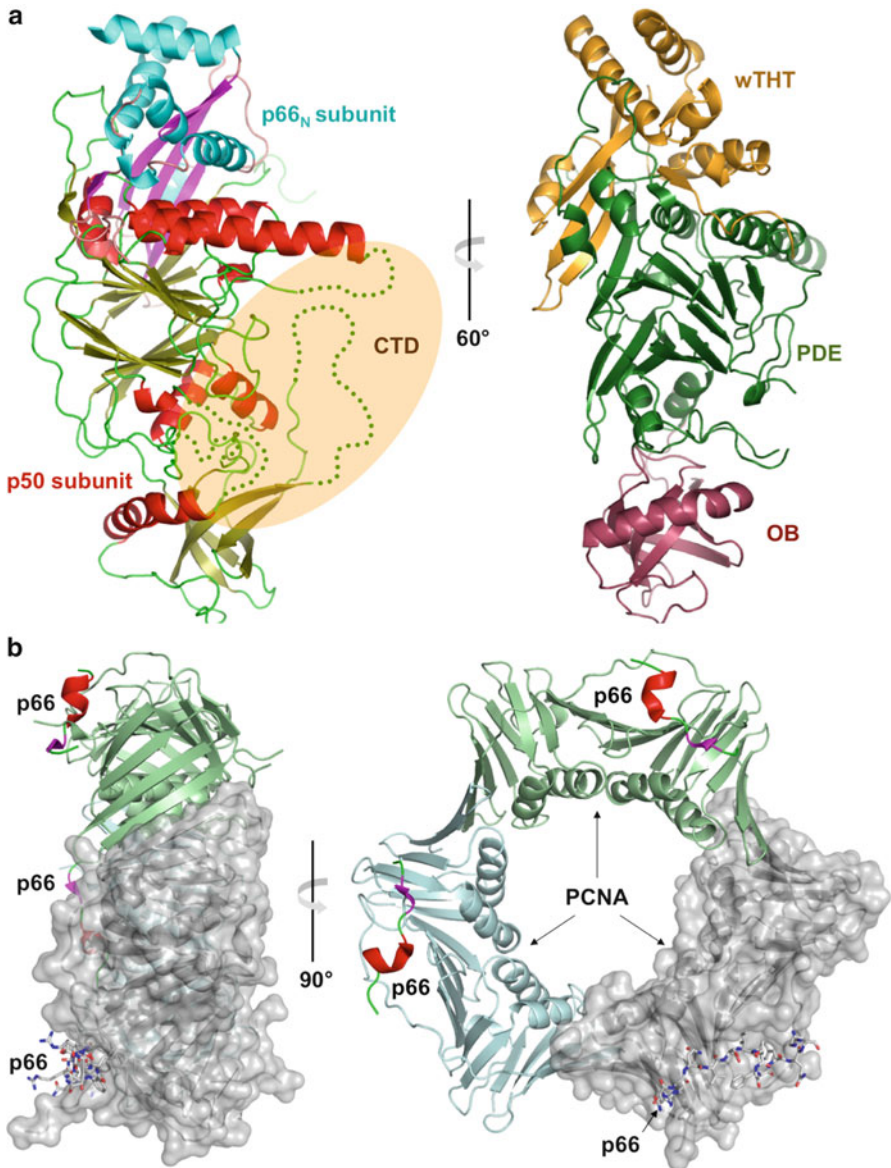


Fig. 12.3 B- and C-subunit structures. **(a)** Cartoon representation of p50•p66_N (PDB code 3e0j) (Baranovskiy et al. 2008). On the *left side* view the secondary structure elements are *color coded* and labeled. α -helices, β -strands and coils are *red, yellow and green* in p50, and *cyan, magenta and light pink* in p66_N. The disordered loops are shown as *dotted lines*. The expected CTD docking site is highlighted in *pink color*. On the *right side* the molecule was rotated to optimize the view of domains. The wTHT, PDE and OB domains are *color coded*. **(b)** Cartoon representation of p66 (residues 453–466) in complex with human PCNA (PDB code 1u76) (Bruning and Shamoo 2004). Two orthogonal views are shown. The surface of one PCNA protomer is shown with bound p66 peptide represented as *sticks*. Both panels were prepared with PyMol software (Delano Scientific)

PDE and OB domains is a distinct feature of the second subunits of replicative B-family DNA polymerases covering all species from archaea to humans (Aravind and Koonin 1998; Makiniemi et al. 1999).

The PDE domain of p50 is comprised of a two-layer β -sheet with two α -helices flanking on one side, and three α -helices flanking on the other side. However, unlike the catalytically active PDE domains, the B-subunits of the eukaryotic B-family DNA polymerases, including p50, do not contain conserved histidine and aspartate residues involved in metal coordination and catalysis (Aravind and Koonin 1998). Instead, the catalytically inactive PDE domains of eukaryotic polymerases play a central role in protein-protein interactions. They have nearly twice the size and surface area compared with catalytically active PDE domains. For example, comparison of p50 with the structure of a monomeric single subunit archaeal phosphodiesterase protein MJ0936 (Chen et al. 2004) (PDB code 1s3l) highlights significant differences in their structures. The MJ0936 has a compact globular shape with a surface area of 6,579 Å². In contrast, the PDE domain of p50 has a surface area of 12,858 Å² and contains many protruding parts and flexible loops. These differences make p50 a perfect docking platform for Pol δ subunits and possibly for other regulatory proteins.

The OB domain is connected with the PDE domain via two linkers, 17 interdomain hydrogen bonds and several scattered interdomain hydrophobic contacts. The OB domain consists of a five-stranded β -barrel (Murzin 1993) wrapped by a bended α -helix on one side. Another side of the β -barrel remains fully accessible and may have a potential for DNA and protein binding as in the case of the OB domains of replication protein A (RPA) (Bochkarev et al. 1997; Bochkareva et al. 2005).

p50-bound p66_N is comprised of a four-stranded, anti-parallel β -sheet surrounded by five α -helices and a C-terminal tail, and has a V-shaped structure. It contains a novel variation of the wHTH motif that is often found in dsDNA-binding domains of transcriptional activators and repressors.

12.3.2 p50•p66 Interactions

p66_N appears as an extension of p50 because one of the layers of a PDE domain two-layer β -sheet is joined with the β -sheet of p66 via parallel β -strands, resulting in an extended 10-stranded β -sheet (Fig. 12.3a). In addition, one more inter-subunit β - β interaction is observed between the parallel β -strands (the first β -strand of p50 and the last β -strand of p66_N). Overall, 37 inter-subunit hydrogen bonds, including 12 main-chain to main-chain hydrogen bonds, inter-subunit hydrophobic interactions covering one larger area and three smaller areas and inter-subunit electrostatic interactions contribute to p50•p66_N dimer formation. The latter interactions are formed by the negatively charged p66-interacting surface of p50 and the complementary, positively charged surface of p66. The p50•p66 complex formation produces significantly large (5,398 Å²) surface area buried at the dimer interface

Because of extended inter-subunit surface and a large number of main-chain to main-chain hydrogen bonds, the impact of one or two interacting side chains

truncations (for example by substitution to alanine) to the stability of p50•p66 complex is expected to be low. However, amino acid substitution with a bulky side chain at the interface, which results in a sterical hindrance, or substitution with a side chain, which results in a same charge repulsion, would be the efficient ways to prevent the complex formation. Three sites in p50 (Leu217, Gly224 and Glu231) were selected to generate single and double tryptophan substitutions, aiming to disrupt the p50•p66 interactions by sterical hindrance and the interactions were examined by two-hybrid assays. The data confirmed that sterical hindrance causing amino acid substitutions leads to disruption of p50•p66 interaction (Baranovskiy et al. 2008).

12.3.3 Functional Studies

The crystal structure of Pol δ p50•p66_N provides a platform for the explanation of earlier accumulated genetic and functional studies as well as to set up novel structure-based rational genetic and functional studies. Surprisingly, the structure also revealed that in spite of the functional importance, the residues from the p50•p66 dimerization interface belong to the least conserved regions. This significantly complicated the mapping of p50•p66 interacting residues by genetic studies. As result, the earlier genetic studies of p50 orthologs in yeast, Pol31p (Vijeh Motlagh et al. 2006) and *Sp* Cdc1 (MacNeill et al. 1996) were performed mainly targeting the highly conserved regions. For example, the temperature-sensitive Pol31p mutants (Vijeh Motlagh et al. 2006) are mapped to the secondary structure elements of p50 that are necessary for its proper folding. After inspection of the structure of p50•p66_N we know that these temperature-sensitive phenotypes are due to deformations in the overall folding of Pol31. In another example the deletion of the C-terminal 20 residues of Cdc1 abolished its interaction with Cdc27 in yeast two-hybrid assays (MacNeill et al. 1996). Indeed, corresponding residues of p50 provide 7 out of the 12 inter-subunit backbone interactions with p66. It has also become apparent why the deletion of up to 100 N-terminal residues of the Pol32p abolishes its interaction with Pol31p (Johansson et al. 2004).

Sc Pol δ was used to verify the functional importance of the interaction between the B- and C-subunits (Baranovskiy et al. 2008). The Pol31p•Pol32p complex was disrupted by a combination of charge repulsion and sterical hindrance. Two structure-guided Pol31p variants were constructed. In the first Pol31p variant, an Arg-Arg-Gly is inserted in the loop located in the center of the Pol31p•Pol32p interface. In the second Pol31p variant, an Arg-Arg-Gly insertion was combined with two point mutations (Gly242Trp, Asp249Trp). These mutations were introduced into the yeast strain allowing for concomitant detection of induced and spontaneous mutations in three genetic loci (Pavlov et al. 2002). Neither of the mutations led to a growth defect or temperature sensitivity, indicating that the mutations in this essential gene do not affect the general properties of the second subunit. However, in comparison to the parent strain, both mutations led to mild cold-sensitivity, increased sensitivity to hydroxyurea, elevated sensitivity to UV light irradiation and reduced UV mutagenesis.

Overall, the mutations behaved as strains with the complete deletion of the third subunit. These experiments confirm that the mutations to affect specifically the interaction between the B- and C-subunits, which is consistent with the structure-predicted functional consequences of the amino acid changes.

In another set of experiments by the MacNeill group, random and directed mutagenesis techniques were used to create a collection of 30 genes encoding mutant Cdc1 proteins (Sanchez Garcia et al. 2009). Each mutant protein was tested for function in fission yeast and for binding to Pol3 and Cdc27 using the two-hybrid system. Mapping of the location of each mutant onto the three-dimensional structure of p50 provided a plausible explanation for *in vivo* function of each Cdc1 mutant and also identified the amino acid residues and regions within the Cdc1 protein that are essential for interaction with Pol3 and Cdc27. Mutants specifically defective in Cdc1•Cdc27 and Cdc1•Pol3 interactions allowed the identification of Cdc27- and Pol3-binding areas on Cdc1. The Cdc27-binding area of Cdc1 coincides with the p66-binding area of p50. The Cdc1 mutations disrupting Pol3 binding were found mainly to affect the four disordered regions which are localized in close proximity to each other and protruding to the same surface area (Fig. 12.3a). Based on these observations, the disordered area of Cdc1 was predicted to bind the CTD of Pol3. Similar conclusions were obtained by Maloisel group for *Sc* Pol δ (Brocas et al. 2010). Later, the crystal structure of the Pol α B-subunit in complex with the C-terminal domain of its catalytic subunit has been reported (Klinge et al. 2009). The structure confirmed that in Pol δ , the B-subunit area that was predicted to interact with the CTD matches well with the corresponding interaction area in the Pol α B-subunit.

12.3.4 Crystal Structure of p66•PCNA

The crystal structure of the C-terminal PCNA-interacting protein motif domain (or PIP-box) of p66 (residues 452–466) has been determined in complex with human PCNA (Fig. 12.3b) (Bruning and Shamoo 2004). The structure shows that the binding mode of p66 and site of interaction is similar to the PIP-box structures of FEN1 and p21^{CIP1} (Bruning and Shamoo 2004; Gulbis et al. 1996). As in these structures, the PIP-box of p66 is comprised of an extended N-terminal region, a central conserved region containing 3₁₀-helix, and a region C-terminal to the PIP-box. A PIP-box exists in proteins involved in cell cycle regulation and DNA processing. Sequential and regulated binding of PIP-box-containing proteins to PCNA has been suggested to contribute to the ordering of events during DNA replication and repair (Warbrick 2000).

12.4 D-subunit

The fourth D-subunit of Pol δ was the latest addition to the list of Pol δ subunits. Hurwitz and coauthors purified and characterized the composition of *Sp* Pol δ (Zuo et al. 1997). Further analysis of the complex revealed four subunits with

molecular masses of 125, 55, 54 and 22 kDa (Zuo et al. 2000). Western blot analysis confirmed that the first three subunits correspond to Pol3, Cdc1, and Cdc27, respectively. The identity of the smallest subunit was determined by sequence analysis, indicating that this is identical to Cdm1, which was previously described as a multi-copy suppressor of the temperature-sensitive *cdc1-P13* mutant (MacNeill et al. 1996). Unlike *Sp* Pol δ , *Sc* Pol δ is comprised of three subunits only and lacks the D-subunit (Gerik et al. 1998).

Shortly after the discovery of D-subunit in yeast Pol δ , Lee and coauthors discovered the D-subunit of mammalian Pol δ exhibiting a 25% identity to Cdm1 (Liu et al. 2000). The D-subunit of human Pol δ , designated as p12, is a 107 amino acid protein with a molecular mass of 12.4 kDa. The four-subunit human Pol δ complex was reconstituted by overexpression of the four subunits in Sf9 insect cells by two groups (Podust et al. 2002; Xie et al. 2002). The purified complex displayed a specific activity comparable with that of the human, bovine, and fission yeast proteins isolated from natural sources. Recently a large-scale production of human Pol δ has been achieved by its overexpression in silkworm larvae, providing a basis for its better structural and functional characterization, especially the role of p12 (Zhou et al. 2011).

12.4.1 D-subunit Structure and Inter-Subunit Interactions

A stable three-subunit *Sp* Pol δ complex containing Pol3, Cdc1, and Cdm1 was reconstituted by simultaneous expression of corresponding polypeptides in baculovirus-infected insect cells (Zuo et al. 2000). The presence of such a complex pointed to an interaction of Cdm1 with either or both Pol3 and Cdc1 (Zuo et al. 2000). Further clarification of D-subunit interactions was reported in the case of human Pol δ . The yeast two-hybrid and pull-down assays revealed that p12 interacts with both the catalytic subunit and the B-subunit of Pol δ (Li et al. 2006).

The three-dimensional structure of the D-subunit from any source has not been determined. Compared with other Pol δ subunits the human p12 is a relatively small polypeptide. Secondary structure predictions using the Phyre server (Kelley and Sternberg 2009) shows that N-terminal residues may fold as two short β -strands followed by a long stretch of disordered region that is rich in acidic residues and three well-defined helices that are located at C-terminal half. The detailed mapping of p12 regions interacting with the catalytic subunit and the B-subunit as well as determination of its docking areas on the surface of these subunits has not been performed. In addition to linking the Pol δ with PCNA, interactions with p50 and p125 subunits (Li et al. 2006) point to several likely scenarios for the architectural role of p12 within the Pol δ complex. One possible scenario is when p12 binds simultaneously to the p50 and the CTD of p125. However, it may also form a bridge between the p50 and the catalytic core of p125, or even interact with both CTD and the catalytic core of p125 as well as p50. The p50-interacting CTD is flexible relative to the catalytic core of p125 (Jain et al. 2009). Consequently, in the latter two scenarios p12 will reduce the overall flexibility of the Pol δ complex.

D-subunits	B-subunits	Hs 356	SSMED-----	-----HLEILEWTL
		Dr 354	SSVDD-----	-----HLEILESTL
		At 337	SEAKS-----	-----KLDVERTL
		Sp 354	HPKSS-----	-----LQCMENLTL
		Sc 367	VIPSNONGESENKVEEGESNDFKDDIEH-----	-----RLDLMECTM
D-subunits	B-subunits	Hs 21	AGHSKGE ^{LAPEL} GEEPPQPRDEEEAELEL-----	-----LRQFDL
		Dr 32	KRSKSPSPKAE ^{PEP} QLSEREKDIOE-----	-----LKNFDL
		At 42	GSDVTQPAALISHGSVDLNEDYDKEEEM-----	-----LRQFDM
		Sp 44	PTPDVTTITTKTL ^{DERIK} EDDELSKEVEENQIMAERISE ^{PIHCENIT} KVEFILHAWHFDT	
D-subunits	B-subunits	Hs	RVRHISPTAPDTL--GCYPFYKTDPFIFPECPHVYFCGNTPSFGSKIIRGPEQDQTVLL	425
		Dr	RLRH ^L APTAPDTL--GCYPFYQKDPFILEECPHVYFSGNAPSYSQSKRV ^T GPEGQDVL	423
		At	R ^W RHLAPTAPNTL--GCYPFTDRDPFLIETCPHVYFVGNQDKYDNR ^L IKGSEGQVLR	406
		Sp	LWNHITPTSPDTL--WCYPFTDKDTFVMEEMPDLYLCGNQPKFGCKT ^V IN-EGNRIQL	422
		Sc	KW ^Q NIAPTAPDTL--WCYPYTDKDFVLDK ^W PHVYIVANQPYFGTRVVEI-GGKNIKI	458
D-subunits	B-subunits	Hs	AWQYGPCTGITRLQR ^W CRAKHMGLEPPPEV ^W QVLKTHPGDPRFQCSL ^W HLYPL	107
		Dr	DWKFGPCTGISRLQR ^W ERAALHGLNPPQEIKDILKEDT ^D PEY ^T QSLWRDYPF	117
		At	NITYGPCLG ^M TRLD ^R WERAVRLGMNPPNEIEKLLK---TGK ^V QQDCLWQGRV	124
		Sp	TARYGPYLG ^M TRMQR ^W KRAKNFNLNPPETV ^G KILMLEEADEENR ^K RESLFYDLQ ^T IPG	160

Fig. 12.4 Amino-acid sequence alignment of B- and D-subunits of Pol δ from different organisms. Hs (human), Dr (*Danio rerio*), At (*Arabidopsis thaliana*), Sp (*S. pombe*) and Sc (*S. cerevisiae*). Note that *S. cerevisiae* does not encode a D-subunit. The aspartate and glutamate residues in Asp/Glu-rich insert of the *S. cerevisiae* B-subunit Pol31p and in corresponding regions of D-subunits are marked in cyan. The tryptophan residues discussed in Sect. 12.4.1 are marked in magenta

As it was mentioned above, *Sc* Pol δ lacks the D-subunit and functions as a three-subunit complex (Gerik et al. 1998) indicating that some of these subunits may compensate, at least in part, for the function of the D-subunit. Analysis of the amino acid sequences of *Sc* Pol δ subunits revealed a highly surprising similarity between the amino acid sequences close to the C-terminal part of Pol31 and the human p12 (Tahirov, unpublished data, see Fig. 12.4). This supports the idea that upon the evolution the D-subunit was derived by duplication of the part of B-subunit. In particular, Pol31 contains an insertion that is rich in acidic residues. Such acidic sequences are absent in B-subunits of four-subunit Pol δ s but may have migrated to their D-subunits. Another interesting finding is an apparent matching of three tryptophan residues in Pol31p and p12. An evolutionary relationship could be observed upon changing from a lower eukaryotic organism to a higher eukaryotic organism: during these changes, fewer of these tryptophan residues remained in B-subunit and more appeared in D-subunit, but their total number in both subunits remained at least three (Fig. 12.4). For example, in the case of *Sp* Pol δ only one tryptophan appears in Cdm1 and two remain in Cdc1, and, finally, in the case of human Pol δ all four tryptophan residues appear in p12 and none remain in p50.

The revealed evolutionary relationship of the B- and D-subunits along with the crystal structure of the B-subunit (Baranovskiy et al. 2008) may help to map the location of Pol31 sequences that “migrated” to form the D-subunit, and hence provide ideas about potential sites where the D-subunit might dock. In particular, the position of the insertion that is rich in acidic residues is mapped to the loop between the α -helices α_7 and α_8 of B-subunit, and the sequences harboring the tryptophan

residues are in an extended loop between the α -helix α_8 and β -strand β_{14} of the B-subunit. Both of these loops protrude towards the B-subunit surface that is proposed to bind the C-terminal domain of the catalytic subunit (Fig. 12.3a) (Sanchez Garcia et al. 2009). In summary, these observations point out that the most likely scenario for D-subunit docking is when the D-subunit binds simultaneously to the B-subunit and the C-terminal domain of the catalytic subunit close to their interaction interface; also, additional interaction with the catalytic core cannot be excluded since p12 alters the catalytic activity of Pol δ (Meng et al. 2009, 2010).

12.4.2 D-subunit Function

Earlier studies with *Sp* Pol δ showed that cells deleted for the D-subunit are viable, indicating that the D-subunit is non-essential for *Sp* mitotic growth (Reynolds et al. 1998). Subsequent characterization of the D-subunit and its functional role were extensively pursued using mainly human Pol δ . In initial studies the four-subunit Pol δ and three-subunit Pol δ lacking the p12 subunit were reconstituted by co-expressing the corresponding subunits in insect cells (Podust et al. 2002; Xie et al. 2002). The four-subunit Pol δ activity was comparable to that isolated from natural sources and it efficiently replicated singly primed M13 DNA in the presence of RPA, PCNA, and replication factor C (RFC) and was active in the SV40 DNA replication system (Podust et al. 2002). The DNA polymerizing activity of three-subunit Pol δ was 15-fold lower than that of the four-subunit Pol δ . However, addition of p12 stimulated the activity of the three-subunit Pol δ fourfold on poly(dA)-oligo(dT) template-primer, confirming that the p12 subunit is required to reconstitute fully active recombinant human DNA polymerase (Podust et al. 2002). In further studies, a PCNA-binding motif KRLITDSY (residues 4–11) at the N terminus of p12 was determined by binding assays (Li et al. 2006). The interaction with PCNA was confirmed by site-directed mutagenesis by replacing the p12 residues Ile7, Ser10, and Tyr11 with alanines. Comparison of DNA polymerizing activities of four-subunit Pol δ containing wild-type p12 with Pol δ containing mutated p12 with lost PCNA binding has been performed in the presence of increasing amounts of PCNA using singly primed M13 DNA as the template. The data revealed that Pol δ with mutant p12 is less active than the wild-type Pol δ in the presence of similar concentrations of PCNA. Thus, based on obtained results authors proposed that p12 not only stabilizes Pol δ but also plays an important role in integration of Pol δ with PCNA (Li et al. 2006).

Lee and coauthors have recently discovered that genotoxic agents such as UV and alkylating chemicals trigger a DNA damage response in which the four-subunit Pol δ is converted to a three-subunit Pol δ by rapid degradation of p12 (Meng et al. 2009, 2010; Zhang et al. 2007). The p12 degradation required an active ubiquitin pathway and could be inhibited by proteasome inhibitors. Degradation was regulated by ATR kinase that controls the damage response in S-phase. The three-subunit

Pol δ exhibits an increased capacity for proofreading 3'–5' exonuclease activity compared with four-subunit Pol δ (Meng et al. 2010). In particular, it cleaves single-stranded DNA two times faster and transfers mismatched DNA from the polymerase site to the exonuclease site nine times faster. However, Pol δ extends mismatched primers three times more slowly in the absence of p12.

In parallel studies, Suzuki and coauthors applied the short hairpin RNA (shRNA) to control the levels of p12 in a variety of human cells (Huang et al. 2010a). Reduction of p12 resulted in a marked decrease in colony formation activity by Calu6, ACC-LC-319, and PC-10 cells but increased population of karyomere-like cells. The latter cells retained an ability to progress through the cell cycle suggesting that p12 reduction induces modest genomic instability (Huang et al. 2010a). Furthermore they also found that reduced expression of p12 plays a role in genomic instability in lung cancer (Huang et al. 2010b).

The p12 subunit has been shown to be a target of posttranslational modifications. Analysis of the p12 sequence for putative phosphorylation sites revealed a single potential casein kinase 2 (CK2) phosphorylation site at Ser24 (Gao et al. 2008). Mutation of this serine to aspartate resulted in resistance to phosphorylation, confirming that Ser24 is indeed a target for CK2 phosphorylation. Dephosphorylation of Pol δ subunits, including p12, is catalyzed by the protein phosphatase-1 (PP1) (Gao et al. 2008). In addition to phosphorylation and dephosphorylation, the p12 subunit was found to be target for ubiquitination (Liu and Warbrick 2006). However the regulation of the levels of p12 by the proteasome was found to be through a mechanism that is not dependent upon p12 ubiquitination (Liu and Warbrick 2006). The relation of these post-translational modifications to the above-described Pol δ functions has not been established.

p12 alone and within Pol δ has been reported to interact specifically *in vitro* and *in vivo* with BLM (the Bloom's syndrome protein) and stimulate its DNA helicase activity, and, reciprocally, BLM stimulates Pol δ strand displacement activity (Selak et al. 2008). A region of p12 comprising amino acids 31–60 was responsible for interaction with BLM that indeed includes the disordered acidic region of p12. Interestingly, BLM was found to co-localize with Pol δ and prevent the degradation of p12 when cells were treated with hydroxyurea (HU). These findings for the first time provide a link between BLM and the replicative machinery in human cells and points to the possible involvement of Pol δ in Bloom's syndrome (BS) (Selak et al. 2008). BS, which is caused by mutation of the *BLM* gene, is associated with excessive chromosomal instability and a high incidence of cancers of all types.

12.5 Conclusions and Prospects

Decades of research has resulted in the identification and characterization of eukaryotic B-family DNA polymerases. Now it is evident that the accessory subunits of these high-molecular weight complexes are playing crucial roles in regulating their activity. However, deciphering the exact mechanisms of their action remains highly

challenging. Among the major hurdles toward achieving this goal is the lack of comprehensive high-resolution three-dimensional structural information, especially for the complexes containing accessory subunits. As we described in this chapter, in the case of Pol δ , partial but significant progress has recently been achieved in this direction. Importantly, the crystal structures were determined for the human Pol δ p50•p60_N complex and the *Sc* Pol δ catalytic core. However, other important pieces of the puzzle related to Pol δ structure and function remain unknown. These include the structure of the D-subunit and its complexes with other subunits of Pol δ , the structure of the CTD with its bound iron-sulfur cluster, the interaction of the CTD with the B-subunit, the structure of the C-subunit C-terminal, the effect of post-translational modifications, the interaction of Pol δ subunits with PCNA, the interaction of Pol δ subunits with other polymerases and regulatory factors, etc. Further collective efforts of structural biologists will be required to find the missing pieces of this puzzle and put them together in order to achieve a more complete understanding of Pol δ .

References

- Albertson TM, Ogawa M, Bugni JM, Hays LE, Chen Y, Wang Y, Treuting PM, Heddle JA, Goldsby RE, Preston BD (2009) DNA polymerase ϵ and δ proofreading suppress discrete mutator and cancer phenotypes in mice. *Proc Natl Acad Sci U S A* 106:17101–17104
- Aravind L, Koonin EV (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res* 26:3746–3752
- Baranovskiy AG, Babayeva ND, Liston VG, Rogozin IB, Koonin EV, Pavlov YI, Vassilyev DG, Tahirov TH (2008) X-ray structure of the complex of regulatory subunits of human DNA polymerase δ . *Cell Cycle* 7:3026–3036
- Bochkarev A, Pfuetzner RA, Edwards AM, Frappier L (1997) Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* 385:176–181
- Bochkareva E, Kaustov L, Ayed A, Yi GS, Lu Y, Pineda-Lucena A, Liao JC, Okorokov AL, Milner J, Arrowsmith CH, Bochkarev A (2005) Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. *Proc Natl Acad Sci U S A* 102:15412–15417
- Brocas C, Charbonnier JB, Dherin C, Gangloff S, Maloisel L (2010) Stable interactions between DNA polymerase δ catalytic and structural subunits are essential for efficient DNA repair. *DNA Repair (Amst)* 9:1098–1111
- Bruning JB, Shamoo Y (2004) Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase δ p66 subunit and flap endonuclease-1. *Structure* 12:2209–2219
- Byrnes JJ (1984) Structural and functional properties of DNA polymerase delta from rabbit bone marrow. *Mol Cell Biochem* 62:13–24
- Byrnes JJ, Downey KM, Black VL, So AG (1976) A new mammalian DNA polymerase with 3' to 5' exonuclease activity: DNA polymerase δ . *Biochemistry* 15:2817–2823
- Chen S, Yakunin AF, Kuznetsova E, Busso D, Pufan R, Proudfoot M, Kim R, Kim SH (2004) Structural and functional characterization of a novel phosphodiesterase from *Methanococcus jannaschii*. *J Biol Chem* 279:31854–31862
- Chung DW, Zhang JA, Tan CK, Davie EW, So AG, Downey KM (1991) Primary structure of the catalytic subunit of human DNA polymerase δ and chromosomal location of the gene. *Proc Natl Acad Sci U S A* 88:11197–11201
- da Costa LT, Liu B, el-Deiry W, Hamilton SR, Kinzler KW, Vogelstein B, Markowitz S, Willson JK, de la Chapelle A, Downey KM et al (1995) Polymerase δ variants in RER colorectal tumours. *Nat Genet* 9:10–11

- Dae DL, Mertz TM, Shcherbakova PV (2010) A cancer-associated DNA polymerase δ variant modeled in yeast causes a catastrophic increase in genomic instability. *Proc Natl Acad Sci U S A* 107:157–162
- Downey KM, Tan CK, So AG (1990) DNA polymerase δ : a second eukaryotic DNA replicase. *Bioessays* 12:231–236
- Flohr T, Dai JC, Buttner J, Popanda O, Hagemuller E, Thielmann HW (1999) Detection of mutations in the DNA polymerase δ gene of human sporadic colorectal cancers and colon cancer cell lines. *Int J Cancer* 80:919–929
- Franklin MC, Wang J, Steitz TA (2001) Structure of the replicating complex of a pol α family DNA polymerase. *Cell* 105:657–667
- Gao Y, Zhou Y, Xie B, Zhang S, Rahmeh A, Huang HS, Lee MY, Lee EY (2008) Protein phosphatase-1 is targeted to DNA polymerase δ via an interaction with the p68 subunit. *Biochemistry* 47:11367–11376
- Garg P, Burgers PM (2005) DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit Rev Biochem Mol Biol* 40:115–128
- Gerik KJ, Li X, Pautz A, Burgers PM (1998) Characterization of the two small subunits of *Saccharomyces cerevisiae* DNA polymerase δ . *J Biol Chem* 273:19747–19755
- Gibbs PE, McDonald J, Woodgate R, Lawrence CW (2005) The relative roles *in vivo* of *Saccharomyces cerevisiae* Pol η , Pol ζ , Rev1 protein and Pol32 in the bypass and mutation induction of an abasic site, T-T (6–4) photoadduct and T-T *cis-syn* cyclobutane dimer. *Genetics* 169:575–582
- Goldsbey RE, Lawrence NA, Hays LE, Olmsted EA, Chen X, Singh M, Preston BD (2001) Defective DNA polymerase δ proofreading causes cancer susceptibility in mice. *Nat Med* 7:638–639
- Goldsbey RE, Hays LE, Chen X, Olmsted EA, Slayton WB, Spangrude GJ, Preston BD (2002) High incidence of epithelial cancers in mice deficient for DNA polymerase δ proofreading. *Proc Natl Acad Sci U S A* 99:15560–15565
- Goscin LP, Byrnes JJ (1982) DNA polymerase δ : one polypeptide, two activities. *Biochemistry* 21:2513–2518
- Gulbis JM, Kelman Z, Hurwitz J, O'Donnell M, Kuriyan J (1996) Structure of the C-terminal region of p21^{WAF1/CIP1} complexed with human PCNA. *Cell* 87:297–306
- Hogg M, Aller P, Konigsberg W, Wallace SS, Doublet S (2007) Structural and biochemical investigation of the role in proofreading of a β hairpin loop found in the exonuclease domain of a replicative DNA polymerase of the B family. *J Biol Chem* 282:1432–1444
- Huang ME, de Calignon A, Nicolas A, Galibert F (2000) Pol32, a subunit of the *Saccharomyces cerevisiae* DNA polymerase δ , defines a link between DNA replication and the mutagenic bypass repair pathway. *Curr Genet* 38:178–187
- Huang ME, Rio AG, Galibert MD, Galibert F (2002) Pol32, a subunit of *Saccharomyces cerevisiae* DNA polymerase δ , suppresses genomic deletions and is involved in the mutagenic bypass pathway. *Genetics* 160:1409–1422
- Huang QM, Akashi T, Masuda Y, Kamiya K, Takahashi T, Suzuki M (2010a) Roles of POLD4, smallest subunit of DNA polymerase δ , in nuclear structures and genomic stability of human cells. *Biochem Biophys Res Commun* 391:542–546
- Huang QM, Tomida S, Masuda Y, Arima C, Cao K, Kasahara TA, Osada H, Yatabe Y, Akashi T, Kamiya K, Takahashi T, Suzuki M (2010b) Regulation of DNA polymerase POLD4 influences genomic instability in lung cancer. *Cancer Res* 70:8407–8416
- Hughes P, Tratner I, Ducoux M, Piard K, Baldacci G (1999) Isolation and identification of the third subunit of mammalian DNA polymerase δ by PCNA-affinity chromatography of mouse FM3A cell extracts. *Nucleic Acids Res* 27:2108–2114
- Jain R, Hammel M, Johnson RE, Prakash L, Prakash S, Aggarwal AK (2009) Structural insights into yeast DNA polymerase δ by small angle X-ray scattering. *J Mol Biol* 394:377–382
- Johansson E, Garg P, Burgers PM (2004) The Pol32 subunit of DNA polymerase δ contains separable domains for processive replication and proliferating cell nuclear antigen (PCNA) binding. *J Biol Chem* 279:1907–1915

- Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4:363–371
- Kesti T, Flick K, Keranen S, Syvaaja JE, Wittenberg C (1999) DNA polymerase ϵ catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. *Mol Cell* 3:679–685
- Klinge S, Nunez-Ramirez R, Llorca O, Pellegrini L (2009) 3D architecture of DNA Pol α reveals the functional core of multi-subunit replicative polymerases. *EMBO J* 28:1978–1987
- Lawrence CW (2002) Cellular roles of DNA polymerase ζ and Rev1 protein. *DNA Repair (Amst)* 1:425–435
- Li H, Xie B, Zhou Y, Rahmeh A, Trusa S, Zhang S, Gao Y, Lee EY, Lee MY (2006) Functional roles of p12, the fourth subunit of human DNA polymerase δ . *J Biol Chem* 281:14748–14755
- Liu G, Warbrick E (2006) The p66 and p12 subunits of DNA polymerase δ are modified by ubiquitin and ubiquitin-like proteins. *Biochem Biophys Res Commun* 349:360–366
- Liu L, Mo J, Rodriguez-Belmonte EM, Lee MY (2000) Identification of a fourth subunit of mammalian DNA polymerase δ . *J Biol Chem* 275:18739–18744
- Lu X, Tan CK, Zhou JQ, You M, Carastro LM, Downey KM, So AG (2002) Direct interaction of proliferating cell nuclear antigen with the small subunit of DNA polymerase δ . *J Biol Chem* 277:24340–24345
- Lydeard JR, Jain S, Yamaguchi M, Haber JE (2007) Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* 448:820–823
- MacNeill SA, Moreno S, Reynolds N, Nurse P, Fantes PA (1996) The fission yeast Cdc1 protein, a homologue of the small subunit of DNA polymerase δ , binds to Pol3 and Cdc27. *EMBO J* 15:4613–4628
- MacNeill SA, Baldacci G, Burgers PM, Hubscher U (2001) A unified nomenclature for the subunits of eukaryotic DNA polymerase δ . *Trends Biochem Sci* 26:16–17
- Makiniemi M, Pospiech H, Kilpelainen S, Jokela M, Vihinen M, Syvaaja JE (1999) A novel family of DNA-polymerase-associated B subunits. *Trends Biochem Sci* 24:14–16
- Meng X, Zhou Y, Zhang S, Lee EY, Frick DN, Lee MY (2009) DNA damage alters DNA polymerase δ to a form that exhibits increased discrimination against modified template bases and mismatched primers. *Nucleic Acids Res* 37:647–657
- Meng X, Zhou Y, Lee EY, Lee MY, Frick DN (2010) The p12 subunit of human polymerase δ modulates the rate and fidelity of DNA synthesis. *Biochemistry* 49:3545–3554
- Mo J, Liu L, Leon A, Mazloun N, Lee MY (2000) Evidence that DNA polymerase δ isolated by immunoaffinity chromatography exhibits high-molecular weight characteristics and is associated with the KIAA0039 protein and RPA. *Biochemistry* 39:7245–7254
- Murzin AG (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J* 12:861–867
- Netz DJ, Stith CM, Stumpfig M, Kopf G, Vogel D, Genau HM, Stodola JL, Lill R, Burgers PM, Pierik AJ (2011) Eukaryotic DNA polymerases require an iron-sulfur cluster for the formation of active complexes. *Nat Chem Biol* 8:125–132
- Pavlov YI, Newlon CS, Kunkel TA (2002) Yeast origins establish a strand bias for replicational mutagenesis. *Mol Cell* 10:207–213
- Pavlov YI, Shcherbakova PV, Rogozin IB (2006) Roles of DNA polymerases in replication, repair, and recombination in eukaryotes. *Int Rev Cytol* 255:41–132
- Pignede G, Bouvier D, de Recondo AM, Baldacci G (1991) Characterization of the *POL3* gene product from *Schizosaccharomyces pombe* indicates inter-species conservation of the catalytic subunit of DNA polymerase delta. *J Mol Biol* 222:209–218
- Podust VN, Chang LS, Ott R, Dianov GL, Fanning E (2002) Reconstitution of human DNA polymerase delta using recombinant baculoviruses: the p12 subunit potentiates DNA polymerizing activity of the four-subunit enzyme. *J Biol Chem* 277:3894–3901
- Pohler JR, Otterlei M, Warbrick E (2005) An *in vivo* analysis of the localisation and interactions of human p66 DNA polymerase δ subunit. *BMC Mol Biol* 6:17
- Popanda O, Flohr T, Fox G, Thielmann HW (1999) A mutation detected in DNA polymerase δ cDNA from Novikoff hepatoma cells correlates with abnormal catalytic properties of the enzyme. *J Cancer Res Clin Oncol* 125:598–608

- Reynolds N, Watt A, Fantes PA, MacNeill SA (1998) Cdm1, the smallest subunit of DNA polymerase δ in the fission yeast *Schizosaccharomyces pombe*, is non-essential for growth and division. *Curr Genet* 34:250–258
- Sanchez Garcia J, Ciuffo LF, Yang X, Kearsey SE, MacNeill SA (2004) The C-terminal zinc finger of the catalytic subunit of DNA polymerase δ is responsible for direct interaction with the B-subunit. *Nucleic Acids Res* 32:3005–3016
- Sanchez Garcia J, Baranovskiy AG, Knatko EV, Gray FC, Tahirov TH, MacNeill SA (2009) Functional mapping of the fission yeast DNA polymerase δ B-subunit Cdc1 by site-directed and random pentapeptide insertion mutagenesis. *BMC Mol Biol* 10:82
- Selak N, Bachrati CZ, Shevelev I, Dietschy T, van Loon B, Jacob A, Hubscher U, Hoheisel JD, Hickson ID, Stagljar I (2008) The Bloom's syndrome helicase (BLM) interacts physically and functionally with p12, the smallest subunit of human DNA polymerase δ . *Nucleic Acids Res* 36:5166–5179
- Shamoo Y, Steitz TA (1999) Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* 99:155–166
- Simon M, Giot L, Faye G (1991) The 3' to 5' exonuclease activity located in the DNA polymerase δ subunit of *Saccharomyces cerevisiae* is required for accurate replication. *EMBO J* 10:2165–2170
- Stocki SA, Nonay RL, Reha-Krantz LJ (1995) Dynamics of bacteriophage T4 DNA polymerase function: identification of amino acid residues that affect switching between polymerase and 3' \rightarrow 5' exonuclease activities. *J Mol Biol* 254:15–28
- Swan MK, Johnson RE, Prakash L, Prakash S, Aggarwal AK (2009) Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase δ . *Nat Struct Mol Biol* 16:979–986
- Venkatesan RN, Treuting PM, Fuller ED, Goldsby RE, Norwood TH, Gooley TA, Ladiges WC, Preston BD, Loeb LA (2007) Mutation at the polymerase active site of mouse DNA polymerase δ increases genomic instability and accelerates tumorigenesis. *Mol Cell Biol* 27:7669–7682
- Vijeh Motlagh ND, Seki M, Branzei D, Enomoto T (2006) Mgs1 and Rad18/Rad5/Mms2 are required for survival of *Saccharomyces cerevisiae* mutants with novel temperature/cold sensitive alleles of the DNA polymerase δ subunit, Pol31. *DNA Repair (Amst)* 5:1459–1474
- Waga S, Hannon GJ, Beach D, Stillman B (1994) The p21 inhibitor of cyclin-dependent kinases controls DNA replication by interaction with PCNA. *Nature* 369:574–578
- Wang J, Sattar AK, Wang CC, Karam JD, Konigsberg WH, Steitz TA (1997) Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. *Cell* 89:1087–1099
- Warbrick E (2000) The puzzle of PCNA's many partners. *Bioessays* 22:997–1006
- Xie B, Mazloum N, Liu L, Rahmeh A, Li H, Lee MY (2002) Reconstitution and characterization of the human DNA polymerase δ four-subunit holoenzyme. *Biochemistry* 41:13133–13142
- Zhang J, Chung DW, Tan CK, Downey KM, Davie EW, So AG (1991) Primary structure of the catalytic subunit of calf thymus DNA polymerase δ : sequence similarities with other DNA polymerases. *Biochemistry* 30:11742–11750
- Zhang J, Tan CK, McMullen B, Downey KM, So AG (1995) Cloning of the cDNAs for the small subunits of bovine and human DNA polymerase delta and chromosomal location of the human gene (*POLD2*). *Genomics* 29:179–186
- Zhang S, Zhou Y, Trusa S, Meng X, Lee EY, Lee MY (2007) A novel DNA damage response: rapid degradation of the p12 subunit of DNA polymerase δ . *J Biol Chem* 282:15330–15340
- Zhong X, Garg P, Stith CM, Nick McElhinny SA, Kissling GE, Burgers PM, Kunkel TA (2006) The fidelity of DNA synthesis by yeast DNA polymerase ζ alone and with accessory proteins. *Nucleic Acids Res* 34:4731–4742
- Zhou Y, Chen H, Li X, Wang Y, Chen K, Zhang S, Meng X, Lee EY, Lee MY (2011) Production of recombinant human DNA polymerase δ in a *Bombyx mori* bioreactor. *PLoS One* 6:e22224
- Zuo S, Gibbs E, Kelman Z, Wang TS, O'Donnell M, MacNeill SA, Hurwitz J (1997) DNA polymerase δ isolated from *Schizosaccharomyces pombe* contains five subunits. *Proc Natl Acad Sci U S A* 94:11244–11249
- Zuo S, Bermudez V, Zhang G, Kelman Z, Hurwitz J (2000) Structure and activity associated with multiple forms of *Schizosaccharomyces pombe* DNA polymerase δ . *J Biol Chem* 275:5153–5162

Chapter 13

DNA Polymerase ϵ

Matthew Hogg and Erik Johansson

Abstract DNA polymerase ϵ (Pol ϵ) is one of three replicative DNA polymerases in eukaryotic cells. Pol ϵ is a multi-subunit DNA polymerase with many functions. For example, recent studies in yeast have suggested that Pol ϵ is essential during the initiation of DNA replication and also participates during leading strand synthesis. In this chapter, we will discuss the structure of Pol ϵ , the individual subunits and their function.

Keywords DNA replication • Leading strand • Pre-loading complex • Processivity

13.1 Introduction

In the decades since the first DNA polymerase was discovered, the number and types of known polymerases has expanded dramatically. In human cells there are at least 15 DNA polymerases that play a part in a wide variety of activities in the replication and maintenance of the genome (Shcherbakova et al. 2003a). DNA polymerase ϵ (Pol ϵ) is a large, multi-subunit enzyme found in all eukaryotic organisms studied to date. The enzyme possesses two catalytic activities: template directed DNA polymerization and the exonucleolytic removal of mispaired primer termini. The role of Pol ϵ , however, is not limited to DNA replication; it has been implicated in such pathways as epigenetic silencing, cell cycle regulation, sister chromatid adhesion and possibly DNA recombination during repair of DNA lesions (Pursell and Kunkel 2008). All Pol ϵ enzymes discovered to date consist of the same basic

M. Hogg • E. Johansson (✉)
Department of Medical Biochemistry and Biophysics,
Umeå University, Umeå SE-90187, Sweden
e-mail: erik.johansson@medchem.umu.se

architecture. The core of the holoenzyme is the large, catalytic subunit that can be divided into two subdomains: the N-terminal portion of the molecule is the catalytic domain and contains the polymerase and exonuclease active sites, while the C-terminal domain is catalytically inactive and appears to play a structural role in the enzyme. All Pol ϵ holoenzymes contain a large B-subunit as well as two small subunits. The accessory subunits in Pol ϵ do not appear to contain any catalytic activities and do not appear to influence the catalytic rates of the polymerase or exonuclease active sites of the N-terminal portion of the catalytic subunit. They do, however, play important roles in the biological pathways in which Pol ϵ is found. The following discussion looks at each of these subunits individually in terms of what is known about their structures and then discusses how they function together as higher order molecular complexes in the various cellular pathways in which Pol ϵ plays a role.

13.2 Structure of Pol ϵ Subunits

13.2.1 *Pol2*

While the crystal structure of Pol ϵ has not yet been reported, many aspects of its structure can be inferred from its classification with other DNA polymerases of known structure. The N-terminal portion of Pol2 is classified as a B-family DNA polymerase based on primary sequence similarities with other B-family DNA polymerases including eukaryotic DNA polymerase α and δ , the DNA polymerases from bacteriophages T4 and RB69 and the DNA polymerases from several archaea such as *Desulfurococcus sp.* (strain Tok), *Thermococcus gorgonarius* and *Pyrococcus furiosus*. X-ray crystal structures are available for several members of the B-family of DNA polymerases and all of these reveal a highly conserved arrangement of protein domains (Fig. 13.1). All of the B-family DNA polymerases whose structures have been solved consist of five domains: a catalytic portion likened to a right hand with a thumb, palm and fingers domain, an exonuclease domain and an N-terminal domain.

The thumb domain associates with the duplex DNA upstream from the polymerase active site and has been shown to play a critical role in establishing the balance between polymerizing and editing modes of the B-family DNA polymerases. Mutations in the thumb domain have been shown to act as mutators or antimutators depending on whether they cause the enzyme to spend more time in the polymerizing mode or the editing mode respectively (Stocki et al. 1995; Wu et al. 1998). The palm domain is the most highly conserved sub-domain among all DNA polymerases, not just those from the B-family, and contains the catalytic residues involved in the nucleotidyl transferase reaction of addition of nucleoside triphosphates to the growing 3' end of the primer terminus. All DNA polymerases in this family possess three conserved sequence motifs called motif A, B and C, with motifs A and C containing

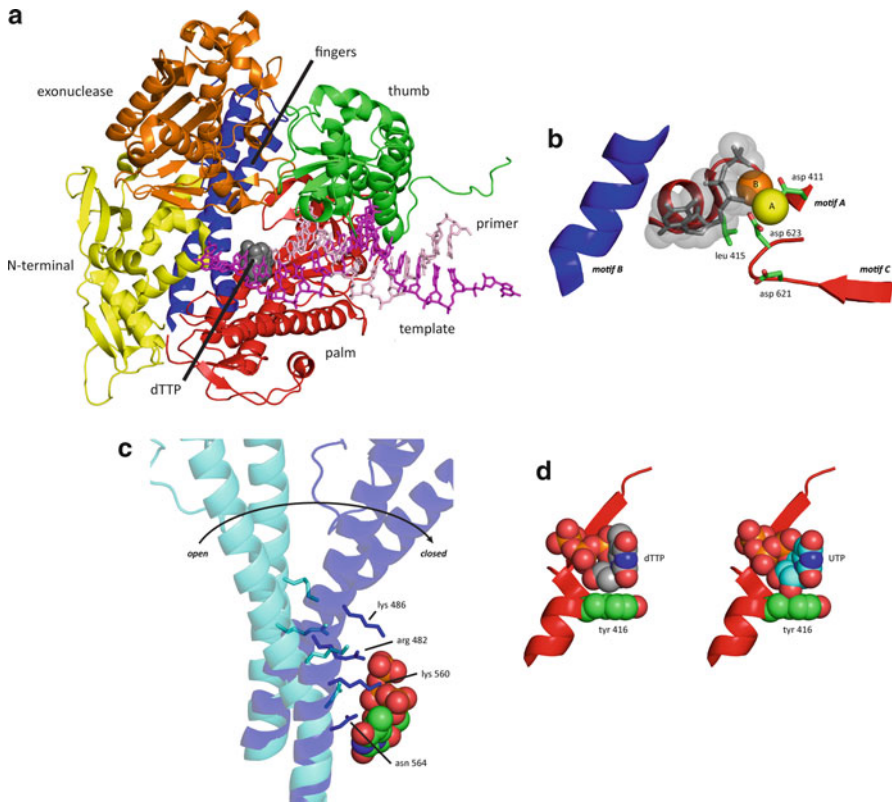


Fig. 13.1 (a) The arrangement of conserved functional domains in the B-family DNA polymerases as exemplified by the crystal structure of the ternary complex of the DNA polymerase from bacteriophage RB69 (Franklin et al. 2001). The N-terminal domain (residues 1–108 and 340–382) is shown in *yellow*, the exonuclease domain (residues 109–339) is shown in *orange*, the palm domain (residues 383–468 and 573–729) is shown in *red*, the fingers domain (residues 469–572) is shown in *blue* and the thumb domain (residues 730–903) is shown in *green*. The duplex DNA co-crystallized with the protein is represented by a *stick* model with the primer strand appearing in *pink* and the templating strand in *magenta*. The incoming dTTP molecule in the ternary complex is shown as *grey spheres*. (b) The polymerase catalytic core of the DNA polymerase from bacteriophage RB69 (Franklin et al. 2001). The three conserved motifs A, B and C (Delarue et al. 1990) are *colored* according to the scheme in part a. The incoming dTTP is shown as *grey sticks* with *transparent spheres*. The three conserved aspartates in the B-family DNA polymerases are shown as *green sticks*. The two catalytic metal ions A and B are shown as *yellow* and *orange spheres* respectively. Leu 415 corresponds to Met 644 in Pol ϵ . (c) Conformational changes in the fingers domain of the DNA polymerase from bacteriophage RB69 (Franklin et al. 2001). The fingers from the open apo complex are shown in *cyan* and those from the closed ternary complex in *blue*. The incoming dTTP is shown as *spheres*. The conserved residues that coordinate the residues of the triphosphate tail during catalysis are shown as *sticks*. (d) Sugar selectivity in the B-family DNA polymerases from bacteriophage RB69 (PDB ID: 1IG9 (Franklin et al. 2001)). The image on the *left* shows the incoming dTTP molecule in *grey spheres* and the steric gate residue tyrosine 416 in *green spheres*. The image on the *right* has a UTP ribonucleotide superimposed onto the dTTP to show the steric clash between the 2'-OH group and the tyrosine ring. Tyr 416 corresponds to Tyr 645 in Pol ϵ

the highly conserved catalytic aspartates (Fig. 13.1b) (Delarue et al. 1990). Motif C consists of a β -hairpin loop with the invariant DTD sequence at its tip. These conserved aspartate residues participate in the two-metal catalytic mechanism (Steitz 1993) that appears to be utilized by all DNA polymerases discovered to date. It has been shown in genetic experiments that substitutions of the two catalytic aspartates for alanine in the catalytic subunit of Pol ϵ (Pol2) rendered the cells inviable, suggesting that DNA synthesis by Pol ϵ is essential in yeast (Dua et al. 1999). Inserted within the palm domain sequence is the fingers domain that contains motif B. This motif consists of residues at the junction of the palm domain and the fingers domain, the movement of which is a critical component in the catalytic activity of B-family DNA polymerases. This domain also contains conserved amino acid residues that bind the triphosphate tail of the incoming nucleotide triphosphates and which undergo large conformational changes during catalysis (Fig. 13.1c) (Franklin et al. 2001; Yang et al. 1999). Another motif that may be located in the palm domain or the vicinity of the thumb domain is a 66 amino acid insertion that is only found in Pol ϵ among all B-family DNA polymerases (Shcherbakova et al. 2003b). The function of this motif is still unknown.

The exonuclease domain in B-family DNA polymerases contains four conserved residues that participate in the catalytic removal of mis-incorporated nucleotides and provide an approximately 100-fold increase in replication fidelity (McCulloch and Kunkel 2008). The exonuclease active site resides about 40 Å away from the polymerase active site in RB69 and so at least three residues must be unwound from the template in order to span this distance. Steric restraints prevent duplex DNA from entering the exonuclease active site. The Pol ϵ mutant allele *pol2-4*, carrying the double mutation D290A and E292A in the conserved motif, has no exonuclease activity in *in vitro* experiments and yeast strains with the *pol2-4* allele also have a significantly increased mutation rate (Morrison et al. 1991; Shcherbakova et al. 2003b)

The N-terminus of B-family polymerases consists of the aptly named N-terminal domain. In archaeal DNA polymerases, this domain contains conserved cysteine residues that form a binding pocket for uracil that can discriminate against the four canonical DNA bases (Fogg et al. 2002). This activity has been proposed to allow these polymerases to pause upon encountering uracil in the template to allow the cell time to repair the uracil residue before replication proceeds across this potentially mutagenic base (Connolly 2009). Interestingly, this read-ahead capability appears to be unique to the archaeal B-family DNA polymerases as no other B-family DNA polymerases, including Pol ϵ from *S. cerevisiae*, are blocked by the presence of uracil in the templating DNA (Wardle et al. 2008). A distinct function of the N-terminal domain of Pol ϵ has not yet been reported.

The catalytic core subunit of Pol ϵ is unique among all of the eukaryotic B-family DNA polymerases in that it appears to consist of two distinct polymerase domains. While all of the domains discussed above exist within the catalytic Pol2 subunit of Pol ϵ , these domains only make up about 140 kD of the estimated 259 kD enzyme in *S. cerevisiae*. Thus there is a large portion of Pol2 that appears to possess no catalytic activity. Secondary structure predictions, fold recognition and sequence similarity searches, however, result in high similarities between the 120 kD C-terminus

of human Pol ϵ and B-family DNA polymerases such as DNA polymerase II in *Escherichia coli* and the archaeal DNA polymerases from *Desulfurococcus sp.* (strain Tok) and *Thermococcus kodakaraensis* (Tahirov et al. 2009). The conserved motifs for polymerase and exonuclease domains in the C-terminus of Pol ϵ in both humans and *S. cerevisiae* align with the catalytic domains of other B-family DNA polymerases but the catalytic residues in both the polymerase and exonuclease domains are disrupted such that catalytic activity has been lost. Such a disruption in catalytic domains appears to be a common feature among a diverse group of archaeal polymerase homologs (Rogozin et al. 2008). Interestingly, the alignments of the C-terminal domain appear to be more similar to bacterial B-family DNA polymerases than the N-terminal domain of Pol ϵ itself (Tahirov et al. 2009). Thus the evolutionary pathway for the creation of Pol ϵ appears to be more complex than a simple gene duplication event of the N-terminal domain.

The very C-terminal end of the protein is a cysteine rich sequence containing two putative zinc fingers of the sequence CX₂CX₁₈CX₂CX₃₀CX₂CX₁₁CXC (Dua et al. 1998). The zinc fingers have been demonstrated to bind zinc ions, though the binding efficiency of the two fingers is not identical (Dua et al. 2002). Sequence analyses of the zinc fingers suggest that they are more similar to the single zinc finger in the archaeal PolD polymerase than to other B-family Zn fingers (Tahirov et al. 2009). The zinc-finger domain is essential and interacts with the B-subunit Dpb2 (presented below). Point mutations that support growth exhibit sensitivity to the alkylating agent methyl methane sulfonate (MMS) (Dua et al. 1999) suggesting that the mutant is deficient in DNA damage repair or avoidance. Interestingly, it is the inter-zinc finger region that is essential for cell viability, not the conserved cysteine residues that constitute the zinc fingers (Dua et al. 1998).

13.2.2 Dpb2

The B-subunit of Pol ϵ (Dpb2) consists of three domains: an N-terminal region that shows structural similarity to AAA+ proteins, a predicted OB fold in the center of the protein and a C-terminal calcineurin-like domain (Nuutinen et al. 2008). The N-terminus of human Dpb2 has been solved by NMR spectroscopy (Nuutinen et al. 2008). So far, there have not been any functions assigned to the three domains. The C-terminal calcineurin-like phosphodiesterase domain is active in archaea and appears to be involved in PCNA binding. In eukaryotic B subunits, this domain is disrupted and appears to no longer possess catalytic activity (Aravind and Koonin 1998). Dpb2 contains a putative PCNA binding domain (Dua et al. 2002). However, mutant Pol ϵ with mutations in the Dpb2 PIP (PCNA interacting protein) motif was not affected in PCNA-dependent holoenzyme assays (unpublished observations).

Dpb2 is essential for cell viability as disruption of the *DPB2* gene in *Saccharomyces cerevisiae* and in *Schizosaccharomyces pombe* is lethal to the cell. Temperature-sensitive mutants of Dpb2 showed partial defects at the restrictive temperature (Araki et al. 1991), suggesting that Dpb2 participates during the establishment

of the replication fork and is required for proper chromosomal replication in yeast. In *S. pombe*, Dpb2 binds to origins in early S-phase supporting a function during the initiation process (Feng et al. 2003). The C-terminal domain of Pol2 also associates with origin DNA at the same time as Dpb2 (Feng et al. 2003).

Dpb2 is phosphorylated by Cdc28 during late G1 phase. Inactivation of cyclin-dependent kinase (CDK) phosphorylation sites in Dpb2 leads to slow cell growth and decreased spore viability. It was suggested that phosphorylation of Dpb2 regulates its interaction with Pol2 or the activity of the Pol ϵ holoenzyme (Kesti et al. 2004), something which remains to be tested *in vitro*. Temperature sensitive *dpb2* alleles were shown to give an increased spontaneous mutation rate (Jaszczur et al. 2008). The level of spontaneous mutation rates were correlated with loss of interaction with Pol2 in two-hybrid assays (Jaszczur et al. 2009). Whether Dpb2 directly influences the fidelity of Pol2 remains to be shown.

13.2.3 *Dpb3/Dpb4 Dimer*

The primary amino acid sequence of Dpb3 and Dpb4 suggests that the N-terminal region of Dpb4 contains a histone fold motif consisting of a helix-strand-helix motif (Ohya et al. 2000). This motif was first identified as being required for dimerization of histone H2A/H2B and H3/H4 pairs (Arents and Moudrianakis 1993) and has been shown to be present in numerous other protein-protein and protein-DNA interactions (Baxevanis et al. 1995). In fact Dpb4 also has another partner, Dls1, in a chromatin remodeling complex. The structure of the Dls1/Dpb4 complex (CHRAC14-CHRAC16) from *D. melanogaster* revealed that Dpb4 indeed has a histone-fold (Hartlepp et al. 2005). The role of Dls1/Dpb4 appears to tether the chromatin remodeling complex to the double-stranded DNA, a function which also has been proposed for the Dpb3/Dpb4 complex in Pol ϵ (Dang et al. 2007; Hartlepp et al. 2005; Tsubota et al. 2003).

Pol ϵ has the ability to bind double-stranded DNA with high affinity, a property not normally associated with DNA polymerases (Tsubota et al. 2003). The double-stranded DNA binding affinity is a property of the C-terminal portion and/or the C-terminal associated subunits of Pol ϵ while the more common single-stranded DNA binding property resides in the N-terminal portion of the enzyme that contains the polymerase and exonuclease motifs. Binding of double-stranded DNA by a heterodimer of Dpb3 and Dpb4, however, was very weak. Subsequent work showed that the Dpb3/Dpb4 heterodimer acts in concert with the Pol2/Dpb2 heterodimer to bind double-stranded DNA with an affinity much higher than the individual heterodimers (Tsubota et al. 2006). When Pol2/Dpb2 and Dpb3/Dpb4 were purified separately, binding to double-stranded DNA as assayed in gel shift assays only occurred at high concentrations of protein but when the two complexes were mixed together, the affinity for double-stranded DNA was similar to that of the four subunit holoenzyme. Homology modeling of the Dpb3/Dpb4 heterodimer onto the crystal structure of the chicken H2A-H2B-dsDNA complex (Harp et al. 2000) suggested regions of the proteins that might interact with double-stranded DNA (Fig. 13.2).

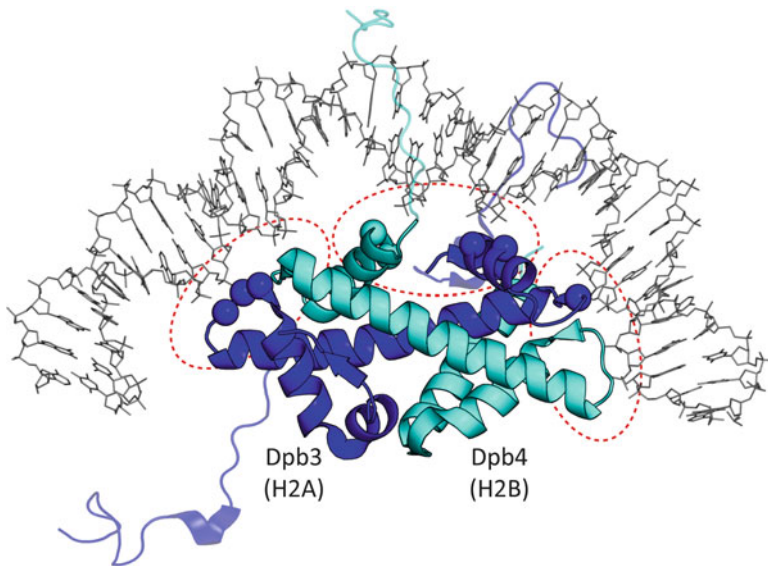


Fig. 13.2 Theoretical DNA binding motifs in Dpb3 and Dpb4 based on the crystal structure of histone fold containing proteins. The structures of H2A (blue) and H2B (cyan) are taken from PDB ID 1EQZ (Harp et al. 2000). The regions of the proteins onto which Dpb3 and Dpb4 were homology modeled are shown in darker colors with black outlines. The portion of the DNA duplex in direct contact with the H2A and H2B proteins is shown as grey sticks. The residues in H2A and H2B corresponding to those that were mutated in Dpb3 and Dpb4 are indicated by spheres and the proposed DNA binding regions in Dpb3 and Dpb4 are indicated by the red dashed lines (Tsubota et al. 2006). This figure is based on Figure 3 in Tsubota et al. (2006)

Indeed, mutations of several lysine residues in Dpb3 reduced double-stranded DNA binding and this defect could be rescued by mutations of a serine and threonine to lysine in Dpb4. Thus the conclusion was made that the heterodimer of Dpb3 and Dpb4 is involved in double-stranded DNA binding (Tsubota et al. 2006).

Dpb4 and Dpb3 are not essential for cell viability in *S. cerevisiae* and appear to offer no enhancement to the catalytic activity of Pol ϵ (Aksenova et al. 2010; Ohya et al. 2000). However, genetic interactions with Dpb11 and Rad53, as well as altered cell cycle progression during S-phase support a model where Dpb3 and Dpb4 are required for normal replication fork progression (Ohya et al. 2000).

13.3 Structure of Pol ϵ Holoenzyme

Based on a combination of sedimentation velocity experiments and gel filtration analysis, the quaternary structure of Pol ϵ appears to be a 1:1:1:1 stoichiometric ratio of all four subunits (Chilkova et al. 2003). The molecular mass of Pol ϵ was determined to be 371 kDa, demonstrating that Pol ϵ purifies as a monomer in all conditions tested (Chilkova et al. 2003).

Structural studies of Pol ϵ have been hampered by poor yields of recombinant protein and by proteolytic cleavage of the Pol2 subunit (Dua et al. 2002). Expression of the Pol ϵ holoenzyme in *S. cerevisiae*, however, yielded sufficient quantities of material to begin attempts at structural studies of the holoenzyme (Chilkova et al. 2003). The structure of Pol ϵ has been solved by cryo-electron microscopy (Asturias et al. 2006). An iterative projection mapping procedure (Penczek et al. 1994) on approximately 19,000 individual Pol ϵ holoenzyme molecules resulted in a reconstruction of the holoenzyme to about 20 Å with a volume of approximately 380 kDa that corresponded well with the previously estimated molecular weight of Pol ϵ (Fig. 13.3a) (Asturias et al. 2006). In this reconstruction, Pol ϵ appears as an elongated protein with a globular head domain and an extended tail domain. In an attempt to locate the subdomains of Pol ϵ , identical reconstructions were made using only the Pol2 subunit (Fig. 13.3b). The 20 Å reconstructions of Pol2 are strikingly similar to the globular domain of the holoenzyme suggesting that the globular head domain consists of the catalytic core. This, then, leaves the extended tail domain to contain the three subunits Dpb2, Dpb3 and Dpb4. A third construct was subjected to the same imaging protocol and consisted of the catalytic core Pol2 and the B-subunit Dpb2. This reconstruction (again to about 20 Å) resembled the structure of Pol2 alone but with extra volume (Fig. 13.3c). This extra volume, however, does not account for all of the expected volume from Dpb2 suggesting that, at least in the absence of Dpb3 and Dpb4, Dpb2 or the C-terminus of Pol2 is highly mobile in the Pol2/Dpb2 heterodimer.

While particles preserved in amorphous ice tend to adopt similar conformations, particles preserved in stain can show variability in conformational changes within a protein structure (Burgess et al. 2004). Images of single Pol ϵ molecules preserved in stain were divided into subcategories based on the relative orientations of the head and tail domains. This classification suggested that the tail domain can move up to 25° and 70° relative to the head domain depending on the direction of motion (see Figure 8 in Asturias et al. 2006). Such domain motions provide a tempting mechanism for how Pol ϵ could bind duplex DNA by way of a flexible linker between the head domain containing the catalytic subunit and the tail domain containing the three accessory subunits. Such a flexible linker domain is evidenced by the intrinsic mobility (and loss of density resolution) for Dpb2 in the Pol2/Dpb2 heterodimer. If duplex DNA were, indeed, to bind to the Pol ϵ holoenzyme in the manner suggested in Fig. 13.3d, this would imply that there would be a minimum length of DNA that would be stably bound by the holoenzyme of about 40 nucleotides. Primer extension assays support this hypothesis (Asturias et al. 2006). When duplexes of varying lengths were provided to Pol ϵ holoenzyme, processivity, as measured by the termination probability at each template position (Kokoska et al. 2003; McCulloch et al. 2004), increased once the duplex region was 40 nucleotides or greater. Such an increase in processivity did not occur with the Pol2 catalytic subunit in the absence of the three accessory subunits Dpb2, Dpb3 and Dpb4. Early experiments with Pol ϵ also showed an increase in processivity when the three accessory subunits were added to the Pol2 subunit (Hamatake et al. 1990).

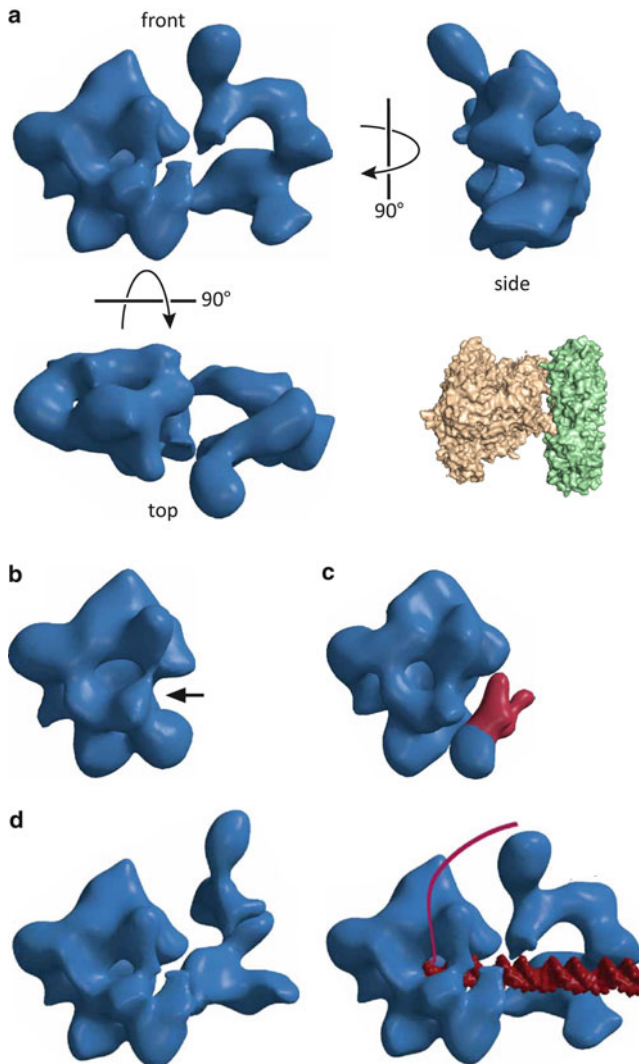


Fig. 13.3 Cryo-electron microscopy (cryo-EM) images of Pol ϵ from *S. cerevisiae*. (a) *Front*, *side* and *top* views of the four subunit holoenzyme. The crystal structure of the DNA polymerase from bacteriophage RB69 (*brown*) (Franklin et al. 2001) docked to the sliding clamp from the same organism (*green*) (Shamoo and Steitz 1999) is shown as a surface representation for scale comparison to the cryo-EM structure of Pol ϵ . (b) The cryo-EM construction of the Pol2 catalytic subunit alone. (c) The cryo-EM construction of the catalytic subunit of Pol2 in complex with Dpb2. The *red* density represents the increase in density over just the Pol2 domain alone. (d) Cryo-EM reconstructions of molecules preserved in stain show flexibility of the tail domain with respect to the globular head domain. This motion suggests a mechanism by which the active site could be made accessible to duplex DNA, which would then be held in place by the closure of the C-terminal domain of the polymerase and presumably, Dpb2, Dpb3 and Dpb4. All cryo-EM reconstructions are from (Asturias et al. 2006)

Interactions between the accessory subunits and duplex DNA as shown in Fig. 13.3d would correlate with the proposed protein folds in these subunits. Dpb3 and Dpb4 both contain histone fold motifs that are presumed to interact with duplex DNA. Indeed, deletion of the two subunits drastically reduces the processivity of the catalytic core along the DNA template (Aksenova et al. 2010). Thus the Dpb3/Dpb4 heterodimer may be functioning as a processivity factor for Pol ϵ in place of PCNA as discussed later.

One question in particular that cannot be answered by the cryo-EM structure is how the accessory subunits are associated with Pol2. It would be tempting to believe that the globular head domain contains all of Pol2 and that the tail domain contains just Dpb2, Dpb3 and Dpb4. It is known, however, that the zinc fingers in *Xenopus* Pol2 are required for binding of the B subunit p60 but not p12-p17. These interact with a motif closer to the N-terminus of Pol2 (Shikata et al. 2006). Yeast two-hybrid assays have also shown that yeast Dpb2 interacts with the zinc-finger domain (Dua et al. 1998; Jaszczur et al. 2009). It is entirely possible, therefore, that the very C-terminal domain of Pol2 extends away from the globular head domain and provides binding sites for the accessory subunits. This C-terminal tail would likely be quite flexible and thus not visible in the cryo-EM reconstructions of Pol2 alone.

13.4 Higher Order Structures

The amino terminal half of Pol ϵ that contains the polymerase and exonuclease domains is not required for cell viability in *S. cerevisiae* (Dua et al. 1999; Kesti et al. 1999) or *S. pombe* (Feng and D'Urso 2001). However, the replication process is impaired and it was suggested that Pol δ can rescue the cells from death (Dua et al. 1999; Ohya et al. 2002). In agreement with these results, the catalytic domain of Pol ϵ appears to be required for replication of the *Xenopus* genome (Shikata et al. 2006). In all cases studied, the C-terminal portion of Pol ϵ is essential for cell viability suggesting that higher order protein interactions between Pol ϵ and other cellular components are required for cellular viability.

13.4.1 Initiation of DNA Replication

The initial step in replication is the formation of a pre-replication complex at an origin of replication (autonomously replicating sequence (ARS) in yeast – see Stinchcomb et al. 1979) during the G1 phase of the cell cycle. This complex consists of the six subunit origin replication complex (ORC), which recruits Cdc6, and together these proteins load Cdt1-Mcm2-7 complexes onto DNA in an ATP hydrolysis-dependent manner (Randell et al. 2006) (see Chaps. 3, 4, 5, 6 and 7, this volume). In budding yeast, upon entry of the cell into S-phase, activated cyclin dependent kinases phosphorylate Cdc6 leading to its degradation to prevent re-initiation of

replication (Drury et al. 1997). The order by which the assembly of replication forks occurs is still not fully understood, however Pol ϵ participates in a very early step. It has recently been reported that Pol ϵ participates in a pre-loading complex in *S. cerevisiae* consisting of Pol ϵ , Dpb11, Sld2, GINS (Muramatsu et al. 2010). This complex forms in a CDK dependent manner and in the absence of association with replication origins. It was proposed that this pre-LC functions as a carrier of GINS to the replication fork, which is required to form the active helicase known as the CMG complex (Cdc45-Mcm2-7-GINS). The Dpb2 subunit has been reported to interact with the Psf1 subunit of GINS (Takayama et al. 2003) as well as with the Orc1 and Orc4 subunits of ORC (Krogan et al. 2006). Thus the C-terminal portion of Pol ϵ may be required for cell viability due to its association with Dpb2 and subsequent formation of protein complexes required for the initiation of genomic replication.

13.4.2 Role at the Replication Fork

While the N-terminal catalytic domain of Pol ϵ was not absolutely required for cell viability in *S. cerevisiae*, several lines of evidence suggested that it plays an important role during normal DNA synthesis. Thermosensitive mutants of Pol2 in *S. cerevisiae* were defective in chromosomal replication, failed to produce chromosome-sized DNA molecules and exhibited a dumbbell shape phenotype, which is indicative of DNA replication problems in budding yeast (Araki et al. 1992; Budd and Campbell 1993). Furthermore, site-directed mutagenesis of the two conserved aspartates in Pol2 that are required for polymerase activity in all other DNA polymerases demonstrated that the polymerase activity of Pol ϵ is essential in *S. cerevisiae* (Dua et al. 1999). Pol ϵ has also been demonstrated to move together with the replication fork in *S. cerevisiae* (Aparicio et al. 1997).

The current view of the role for Pol ϵ at the replication fork emerged in part from a series of studies focused on the fidelity by which Pol ϵ and Pol δ replicate DNA. Both Pol ϵ and Pol δ have proofreading activity (3'-5'exonuclease activity) that significantly increases their fidelity during DNA synthesis (Morrison et al. 1993). Early genetic experiments in *S. cerevisiae* asked if inactivation of the proofreading activity of either Pol δ or Pol ϵ would give a bias in the mutation rates of leading and lagging strand (Shcherbakova and Pavlov 1996). A nucleotide analog, 6-N-hydroxylamine (HAP), base pairs with both T and C and can lead to G-C to A-T and A-T to G-C transitions depending on whether the HAP is in the templating strand or is the incoming nucleotide. Mutagenesis by HAP is only affected by the exonuclease proofreading activity of polymerases in correcting mis-incorporations with this analog. Using the reporter gene *URA3* downstream from the origin of replication *ARS306*, it was shown that mutagenesis by proofreading deficient Pol ϵ and Pol δ changed when the orientation of the *URA3* reporter gene was changed relative to the origin of replication. Interestingly, the frequencies of reversion events were opposite when comparing exonuclease deficient Pol ϵ and Pol δ (Shcherbakova and

Pavlov 1996). Similar studies looking at mutational spectra in the *SUP4* gene with both Pol ϵ and Pol δ exonuclease deficient mutants confirmed the results (Karthikeyan et al. 2000). While these experiments do not specifically demonstrate leading or lagging strand synthesis for either enzyme, they do suggest that Pol δ and Pol ϵ proofread opposite strands.

In vitro studies of Pol ϵ revealed that Pol ϵ is a highly accurate DNA polymerase (Shcherbakova et al. 2003b). In fact, estimates suggest that Pol ϵ has the highest fidelity among all DNA polymerases in *S. cerevisiae* (Fortune et al. 2005). Still, Pol ϵ makes errors and it was found that Pol ϵ has a unique propensity for pyrimidine mismatches, in particular T-dTTP (Shcherbakova et al. 2003b). This characteristic was later enhanced in genetic experiments suggesting that Pol ϵ participates in leading strand synthesis (Pursell et al. 2007). These experiments utilized the concept of asymmetric mutators – polymerases that had a propensity for certain mismatch combinations over others. The original idea came from the E710A mutation in the large Klenow fragment of *E. coli* pol I that showed reduced fidelity (Minnick et al. 1999) but, importantly for the experiments in question, showed a strong preference for forming A-dCTP mismatches over T-dGTP mismatches (Minnick et al. 2002). These two mismatches would both lead to A-T to G-C transitions depending on whether the A or the T were in the templating strand, thus giving a mutational marker as to which strand was being copied. The glutamate at position 710 in the Klenow fragment is likely located in the same position as the invariant tyrosine at position 645 in Pol ϵ and at position 416 in RB69 (Fig. 13.1d). Replacement of the adjacent methionine in Pol ϵ with glycine (*pol2-M644G*) results in a polymerase that still retains high levels of activity but exhibits an *in vitro* error rate for forming T-dTTP mismatches that is approximately 39-fold greater than forming A-dATP mismatches (Pursell et al. 2007). When the *URA3* reporter gene was placed in both orientations both to the right and the left of the *ARS306* origin of replication, it was observed that mutational hot spots occurred in which A-T to T-A transitions were the result of T-dTTP misincorporations by *pol2-M644G*. The data lead to the conclusion that Pol ϵ participates during leading strand synthesis at the replication fork (Pursell et al. 2007). Subsequent studies, using the same approach on Pol δ with the corresponding *pol3-L612M* mutant, showed that Pol δ primarily synthesizes DNA on the lagging strand where the reporter gene was located (Nick McElhinny et al. 2008). Recently a whole genome analysis of errors made in a *pol3-L612M* strain confirmed that Pol δ is the major lagging strand polymerase (Larrea et al. 2010). Together these experiments suggests that during normal DNA replication Pol ϵ is the leading strand polymerase and Pol δ is the lagging strand polymerase.

Other observations that support this model are that Pol δ can proofread errors made by Pol α , while Pol ϵ does not (Pavlov et al. 2006). During lagging strand DNA synthesis, Pol α is continually laying down RNA primers followed by about 20 nucleotides of DNA before the primer terminus moves from Pol α to the lagging strand DNA polymerase (see Chap. 9, this volume). Because Pol α lacks 3'–5' exonuclease activity, it is more error prone than Pol ϵ and Pol δ (Kunkel et al. 1989, 1991). The mutator activity of Pol α has been shown to act synergistically with proofreading-deficient Pol δ (Pavlov et al. 2006). In these experiments, the *pol1-L868M* active site

mutant had a limited mutator effect *in vivo* and the *pol3-exo⁻* exonuclease deficient mutant of Pol δ had an approximately seven-fold increase in mutation rates. The double mutant, however, showed an approximately 70-fold increase in mutation rate. Such synergism in mutagenesis was not observed with a proofreading deficient mutant of Pol ϵ suggesting that Pol δ is able to correct errors by Pol α on the lagging strand but that Pol ϵ is not able to correct such errors and is thus not likely to be active on the lagging strand during DNA replication.

During Okazaki fragment maturation, strand displacement is carried out by the lagging strand polymerase in tight regulation with the flap endonuclease FEN1 which removes the 5' flap of RNA/DNA (see Chap. 16). Pol δ , in contrast to Pol ϵ , functionally interacts with FEN1 and DNA ligase I during the processing of primers in the Okazaki fragments (Jin et al. 2001; Garg et al. 2004). This supports the hypothesis that Pol ϵ is not a lagging strand DNA polymerase.

13.4.3 PCNA

PCNA (proliferating cell nuclear antigen, see Chap. 15, this volume) is a large, trimeric, ring-shaped molecule that wraps around duplex DNA. PCNA is loaded onto the DNA by the activity of the clamp loading protein RFC (see Chap. 14). Many molecules interact with PCNA such that it acts as a structural platform, recruiting and maintaining other molecules on the DNA. The association of DNA polymerases with PCNA at the replication fork serves to prevent dissociation of the polymerase from the DNA. Pol ϵ and Pol δ have both been shown to be stimulated by PCNA (Burgers 1991; Lee et al. 1991). Under single hit conditions in which a polymerase that dissociates from the DNA molecule will not re-associate with another previously extended DNA molecule (Bambara et al. 1995), the processivity of Pol ϵ and Pol δ were very similar, with a less than two-fold longer processivity for Pol δ compared to Pol ϵ (Chilkova et al. 2007). Interestingly, however, under these experimental conditions Pol ϵ processivity was increased only about sixfold in the presence of PCNA and clamp loader while Pol δ processivity was increased at least 100-fold. Thus, while both polymerases are stimulated by PCNA, the effect appears to be much greater for Pol δ .

Pol ϵ appears to have a canonical PCNA-interacting protein (PIP) box motif. That is a putative PCNA binding sequence of QTSLTKFF, which fits the consensus sequence Qxxhxxaa in which 'h' is a hydrophobic residue and 'a' is an aromatic residue (Warbrick et al. 1998). Unlike other proteins known to interact with PCNA, however, the PIP box is not located at the N- or C-terminal end of the protein but is instead located in the middle of the protein. The PIP box is likely where one would expect to find it at the C-terminus of the catalytic domain, but upon the gene duplication event (as discussed earlier) the PIP box became buried in front of the second set of polymerase motifs that make up the C-terminal tail of Pol ϵ . The location of the PIP box suggests that it may no longer interact directly with PCNA. Surface plasmon resonance experiments using immobilized PCNA showed no

interaction with PCNA in solution at similar concentrations at which Pol δ showed strong interactions with PCNA (Chilkova et al. 2007) suggesting that stimulation of Pol ϵ and Pol δ by PCNA occurs by two distinct mechanisms.

It is clear that PCNA stimulates Pol ϵ *in vitro* but genetic experiments demonstrated that the putative PIP-box in Pol ϵ is not necessary for cell viability. Site directed mutagenesis of the conserved residues as well as deletion of the entire PIP box had no effect on cell growth at both 23°C and 37°C (Dua et al. 2002). Instead deletion of the entire PIP box as well as mutations of the conserved hydrophobic (Leu 1196) and aromatic (Phe 1199 and Phe 1200) residues increased the sensitivity of the resulting *S. cerevisiae* cells to the DNA damaging agent methyl methanesulfate (MMS).

Pol δ has at least two different types of interactions with PCNA. It is possible that the interaction between the PIP box and PCNA is important for the loading onto the primer termini and the second interaction supports the distance by which Pol δ replicates DNA before it dissociates from the template. In contrast, Pol ϵ only interacts with PCNA when already loaded on the primer-template. Several enzymes predicted to operate on the lagging strand of the replication fork contain PIP boxes known to interact with PCNA. Pol δ interacts with PCNA through a C-terminal PIP box in its C-subunit p66, Pol32 or Cdc27 (Bermudez et al. 2002; Ducoux et al. 2001; Gerik et al. 1998; Johansson et al. 2004), the flap endonuclease FEN1 interacts with PCNA via a C-terminal PIP box (Warbrick et al. 1997) and DNA ligase I interacts with PCNA via a PIP box in its N-terminus (Levin et al. 1997). Thus there may be distinct mechanisms of PCNA utilization between the two strands at the replication fork. Leading strand replication, as carried out by Pol ϵ , was hypothesized to be PIP box-independent while lagging strand synthesis, as carried out by Pol δ , FEN1 and DNA ligase I, was proposed to be dependent on the conserved PIP box motif for interactions with PCNA (Chilkova et al. 2007). Physical interactions between Pol ϵ and the CMG complex might explain why Pol ϵ primarily replicates the leading strand: Pol ϵ is known to interact with both GINS and Cdc45 but whether this occurs when these proteins are part of the CMG complex is not yet known.

13.4.4 Checkpoint Activation in S Phase

Inhibition of replication fork progression can lead to genomic instability and subsequent chromosomal rearrangements and translocations, which play an important role in cancer development (Lengauer et al. 1998). Replication fork inhibition or blockage can be caused by natural impediments such as DNA binding proteins, collisions with the transcription machinery and aberrant DNA structures such as those caused by nucleotide repeat sequences (reviewed in Mirkin and Mirkin 2007). Replication forks are also stalled by the presence of DNA damage such as bulky adducts (Shiotani et al. 2006), abasic sites (Boiteux and Guillet 2004), DNA-protein crosslinks (Payne et al. 2006) and interstrand crosslinks (Niedernhofer et al. 2005). Sensing of replicative stress leads to a signaling cascade culminating in the phosphorylation of Rad53 in budding yeast.

Decoupling of the helicase from the replicative polymerases is suggested to generate large amounts of single-stranded DNA that recruits RPA and initiates signaling pathways (Navadgi-Patil and Burgers 2009). During chromosome replication, the lagging strand always has a certain amount of single-stranded DNA due to the synthesis of the Okazaki fragments. In contrast, it is less likely to find large amounts of single-stranded DNA on the leading strand during normal replication. Thus the leading strand polymerase is ideally positioned to participate in the sensory mechanism for the generation of checkpoint signals. Indeed, Pol ϵ has been implicated to play a role in this function. The *pol2-12* allele in *S. cerevisiae*, with Gln 2195 converted to a stop codon at the C-terminus of Pol2, leads to a loss of S phase checkpoint function (Navas et al. 1995). Pol ϵ acts as a sensor of DNA replication progression because the *pol2-12* mutants fail to activate both the Dun1 protein kinase and transcription of *RNR3* in response to DNA damaging agents, are inviable in the presence of hydroxyurea (an inhibitor of ribonucleotide reductase (Krakoff et al. 1968)), and enter into mitosis before completion of DNA replication (Navas et al. 1995). Pol ϵ was found to act as a transducer of the DNA damage checkpoint signal only during S phase and operates in a parallel pathway to *RAD9* that transduces the signal during G1 and G2 phases (Navas et al. 1996). It is still unclear how Pol ϵ participates in the checkpoint activation.

Mrc1 is a replication fork associated protein that has been implicated to be important both during DNA replication and mediating an S phase checkpoint. Interestingly, Mrc1 interacts with the helicase component Mcm6 (Komata et al. 2009) and Pol ϵ (Lou et al. 2008) and it was suggested that Mrc1 stabilizes a hypothetical interaction between the Mcm helicase and Pol ϵ (Lou et al. 2008). Pol ϵ associates with Mrc1 via both its N-terminus and C-terminus and the association with the Pol2 C-terminus appears to be modulated by Dpb2. Phosphorylation of Mrc1 during S phase checkpoint eliminates N-terminal association but not C-terminal association (Lou et al. 2008).

Another model is proposed in which stalling of leading strand synthesis by Pol ϵ signals the Dpb11/Sld2-Mec1-Rad53 signaling cascade leading finally to cell cycle arrest. This activity appears to be dependent on Dpb4 and suggests that leading and lagging strands sense DNA damage and signal this to the cell via different pathways (Puddu et al. 2011). This is an interesting model since Pol ϵ is inhibited by single-stranded DNA, while Pol δ is less sensitive to the presence of single-stranded DNA (Chilkova et al. 2007).

13.5 Ribose vs Deoxyribose Discrimination

Most DNA polymerases have mechanisms by which they can discriminate between deoxyribonucleotide and ribonucleotide triphosphates (Brown and Suo 2011; Joyce 1997). For the B-family polymerases such as Pol ϵ , this consists of a conserved tyrosine residue (position 645 in Pol ϵ) that acts as a steric gate against the 2'OH group of the sugar in ribonucleotides (Bonnin et al. 1999; Gardner and Jack 1999; Yang et al. 2002) (Fig. 13.1d). Discrimination between nucleotide sugars, however,

is not absolute and the problem is confounded by the relative abundance of the sugar moieties for which *in vivo* NTP concentrations can be orders of magnitude higher than dNTP concentrations in yeast (Nick McElhinny et al. 2010b) and mammalian cells (Traut 1994). Primer extension assays with Pol ϵ suggest that it may incorporate approximately one NMP for every 1,250 dNMPs incorporated and that DNA synthesis by Pol ϵ is inhibited by the presence of NMPs in the template (Nick McElhinny et al. 2010b; Watt et al. 2011). Pol ϵ can incorporate ribonucleotides *in vivo*, as demonstrated by the increased presence of NMP residues detected by alkali hydrolysis in the M644G mutant and the double mutant of M644G with ribonuclease H2 knockout mutation in *S. cerevisiae* (Nick McElhinny et al. 2010a). These mutations also resulted in mutator phenotypes. These results, along with the fact that ribonuclease H2 contains a PIP box and associates with PCNA (Bubeck et al. 2011), suggests that incorporation of NMP by Pol ϵ may be a significant source of genomic instability and may be repaired by ribonuclease H2 traveling with the replication fork.

13.6 Concluding Remarks

We have in this chapter only discussed some of the properties of Pol ϵ . In addition to what has been discussed, pol ϵ mutants have been reported to affect telomere length and silencing near the telomeres. There are reports suggesting that Pol ϵ interacts with proteins that participate in sister chromatid establishment and in DNA repair pathways, and which may influence chromatin structure. Its position as an important protein during the assembly of the replisome and during leading strand synthesis suggests that Pol ϵ will interact with many processes involved in the maintenance and duplication of the genome. The low-resolution cryo-EM structure has in part offered an explanation to some of the unique properties of Pol ϵ . However, it cannot give us the molecular details on how the high fidelity during DNA synthesis is achieved or why Pol ϵ incorporates ribonucleotides relatively frequently. Both these properties imply that the structure of the active site in Pol ϵ will differ from Pol δ or other B-family polymerases. A high-resolution structure would be invaluable for our understanding of how the active site functions and also for future studies of the interactions with DNA and other proteins which have been discussed in this chapter.

Acknowledgements This work is supported by Kempe Stiftelsen (M.H. and E.J.), the Swedish Research Council, the Swedish Cancer Society, and Smärtafonden (E.J.).

References

- Aksenova A, Volkov K, Maceluch J, Pursell ZF, Rogozin IB, Kunkel TA, Pavlov YI, Johansson E (2010) Mismatch repair-independent increase in spontaneous mutagenesis in yeast lacking non-essential subunits of DNA polymerase ϵ . *PLoS Genet* 6:e1001209

- Aparicio OM, Weinstein DM, Bell SP (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 RK, Johnston LH, Sugino A (1991) *DPB2*, the gene encoding DNA polymerase II subunit B, is required for chromosome replication in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 88:4601–4605
- Araki H, Ropp PA, Johnson AL, Johnston LH, Morrison A, Sugino A (1992) DNA polymerase II, the probable homolog of mammalian DNA polymerase ϵ , replicates chromosomal DNA in the yeast *Saccharomyces cerevisiae*. *EMBO J* 11:733–740
- Aravind L, Koonin EV (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res* 26:3746–3752
- Arents G, Moudrianakis EN (1993) Topography of the histone octamer surface: repeating structural motifs utilized in the docking of nucleosomal DNA. *Proc Natl Acad Sci USA* 90:10489–10493
- Asturias FJ, Cheung IK, 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
- Bambara RA, Fay PJ, Mallaber LM (1995) Methods of analyzing processivity. *Methods Enzymol* 262:270–280
- Baxevas AD, Arents G, Moudrianakis EN, Landsman D (1995) A variety of DNA-binding and multimeric proteins contain the histone fold motif. *Nucleic Acids Res* 23:2685–2691
- Bermudez VP, MacNeill SA, Tappin I, Hurwitz J (2002) The influence of the Cdc27 subunit on the properties of the *Schizosaccharomyces pombe* DNA polymerase δ . *J Biol Chem* 277:36853–36862
- Boiteux S, Guillet M (2004) Abasic sites in DNA: repair and biological consequences in *Saccharomyces cerevisiae*. *DNA Repair (Amst)* 3:1–12
- Bonnin A, Lazaro JM, Blanco L, Salas M (1999) A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type phi29 DNA polymerase. *J Mol Biol* 290:241–251
- Brown JA, Suo Z (2011) Unlocking the sugar “steric gate” of DNA polymerases. *Biochemistry* 50:1135–1142
- Bubeck D, Reijns MA, Graham SC, Astell KR, Jones EY, Jackson AP (2011) PCNA directs type 2 RNase H activity on DNA replication and repair substrates. *Nucleic Acids Res* 39:3652–3666
- Budd ME, Campbell JL (1993) DNA polymerases delta and epsilon are required for chromosomal replication in *Saccharomyces cerevisiae*. *Mol Cell Biol* 13:496–505
- Burgers PM (1991) *Saccharomyces cerevisiae* replication factor C. II. Formation and activity of complexes with the proliferating cell nuclear antigen and with DNA polymerases δ and epsilon. *J Biol Chem* 266:22698–22706
- Burgess SA, Walker ML, Thirumurugan K, Trinick J, Knight PJ (2004) Use of negative stain and single-particle image processing to explore dynamic properties of flexible macromolecules. *J Struct Biol* 147:247–258
- Chilkova O, Jonsson BH, Johansson E (2003) The quaternary structure of DNA polymerase ϵ from *Saccharomyces cerevisiae*. *J Biol Chem* 278:14082–14086
- Chilkova O, Stenlund P, Isoz I, Stiith CM, Grabowski P, Lundstrom EB, Burgers PM, 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* 35:6588–6597
- Connolly BA (2009) Recognition of deaminated bases by archaeal family-B DNA polymerases. *Biochem Soc Trans* 37:65–68
- Dang W, Kagalwala MN, Bartholomew B (2007) The Dpb4 subunit of ISW2 is anchored to extra-nucleosomal DNA. *J Biol Chem* 282:19418–19425
- Delarue M, Poch O, Tordo N, Moras D, Argos P (1990) An attempt to unify the structure of polymerases. *Protein Eng* 3:461–467
- Drury LS, Perkins G, Diffley JF (1997) The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. *EMBO J* 16:5966–5976

- Dua R, Levy DL, Campbell JL (1998) Role of the putative zinc finger domain of *Saccharomyces cerevisiae* DNA polymerase ϵ in DNA replication and the S/M checkpoint pathway. *J Biol Chem* 273:30046–30055
- Dua R, Levy DL, Campbell JL (1999) Analysis of the essential functions of the C-terminal protein/protein interaction domain of *Saccharomyces cerevisiae* pol ϵ and its unexpected ability to support growth in the absence of the DNA polymerase domain. *J Biol Chem* 274:22283–22288
- Dua R, Levy DL, Li CM, Snow PM, Campbell JL (2002) *In vivo* reconstitution of *Saccharomyces cerevisiae* DNA polymerase ϵ in insect cells. Purification and characterization. *J Biol Chem* 277:7889–7896
- Ducoux M, Urbach S, Baldacci G, Hubscher U, Koundrioukoff S, Christensen J, Hughes P (2001) Mediation of proliferating cell nuclear antigen (PCNA)-dependent DNA replication through a conserved p21^{Cip1}-like PCNA-binding motif present in the third subunit of human DNA polymerase δ . *J Biol Chem* 276:49258–49266
- Feng W, D'Urso G (2001) *Schizosaccharomyces pombe* cells lacking the amino-terminal catalytic domains of DNA polymerase ϵ are viable but require the DNA damage checkpoint control. *Mol Cell Biol* 21:4495–4504
- Feng W, Rodriguez-Menocal L, Tolun G, D'Urso G (2003) *Schizosaccharomyces pombe* Dpb2 binds to origin DNA early in S phase and is required for chromosomal DNA replication. *Mol Biol Cell* 14:3427–3436
- Fogg MJ, Pearl LH, Connolly BA (2002) Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat Struct Biol* 9:922–927
- Fortune JM, Pavlov YI, Welch CM, Johansson E, Burgers PM, Kunkel TA (2005) *Saccharomyces cerevisiae* DNA polymerase δ : high fidelity for base substitutions but lower fidelity for single- and multi-base deletions. *J Biol Chem* 280:29980–29987
- Franklin MC, Wang J, Steitz TA (2001) Structure of the replicating complex of a pol α family DNA polymerase. *Cell* 105:657–667
- Gardner AF, Jack WE (1999) Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Res* 27:2545–2553
- Garg P, Stith CM, Sabouri N, Johansson E, Burgers PM (2004) Idling by DNA polymerase δ maintains a ligatable nick during lagging-strand DNA replication. *Genes Dev* 18:2764–2773
- Gerik KJ, Li X, Pautz A, Burgers PM (1998) Characterization of the two small subunits of *Saccharomyces cerevisiae* DNA polymerase δ . *J Biol Chem* 273:19747–19755
- Hamatake RK, Hasegawa H, Clark AB, Bebenek K, Kunkel TA, Sugino A (1990) Purification and characterization of DNA polymerase II from the yeast *Saccharomyces cerevisiae*. Identification of the catalytic core and a possible holoenzyme form of the enzyme. *J Biol Chem* 265:4072–4083
- Harp JM, Hanson BL, Timm DE, Bunick GJ (2000) Asymmetries in the nucleosome core particle at 2.5 Å resolution. *Acta Crystallogr D Biol Crystallogr* 56:1513–1534
- Hartlepp KF, Fernandez-Tornero C, Eberharter A, Grune T, Muller CW, Becker PB (2005) The histone fold subunits of *Drosophila* CHRAC facilitate nucleosome sliding through dynamic DNA interactions. *Mol Cell Biol* 25:9886–9896
- Jaszczur M, Flis K, Rudzka J, Kraszewska J, Budd ME, Polaczek P, Campbell JL, Jonczyk P, Fijalkowska IJ (2008) Dpb2p, a noncatalytic subunit of DNA polymerase ϵ , contributes to the fidelity of DNA replication in *Saccharomyces cerevisiae*. *Genetics* 178:633–647
- Jaszczur M, Rudzka J, Kraszewska J, Flis K, Polaczek P, Campbell JL, Fijalkowska IJ, Jonczyk P (2009) Defective interaction between Pol2p and Dpb2p, subunits of DNA polymerase ϵ , contributes to a mutator phenotype in *Saccharomyces cerevisiae*. *Mutat Res* 669:27–35
- Jin YH, Obert R, Burgers PM, Kunkel TA, Resnick MA, Gordenin DA (2001) The 3'→5' exonuclease of DNA polymerase δ can substitute for the 5' flap endonuclease Rad27/Fen1 in processing Okazaki fragments and preventing genome instability. *Proc Natl Acad Sci USA* 98:5122–5127
- Johansson E, Garg P, Burgers PM (2004) The Pol32 subunit of DNA polymerase δ contains separable domains for processive replication and proliferating cell nuclear antigen (PCNA) binding. *J Biol Chem* 279:1907–1915
- Joyce CM (1997) Choosing the right sugar: how polymerases select a nucleotide substrate. *Proc Natl Acad Sci U S A* 94:1619–1622

- Karthikeyan R, Vonarx EJ, Straffon AF, Simon M, Faye G, Kunz BA (2000) Evidence from mutational specificity studies that yeast DNA polymerases δ and ϵ replicate different DNA strands at an intracellular replication fork. *J Mol Biol* 299:405–419
- Kesti T, Flick K, Keranen S, Syvaoja JE, Wittenberg C (1999) DNA polymerase ϵ catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. *Mol Cell* 3:679–685
- Kesti T, McDonald WH, Yates JR 3rd, Wittenberg C (2004) Cell cycle-dependent phosphorylation of the DNA polymerase ϵ subunit, Dpb2, by the Cdc28 cyclin-dependent protein kinase. *J Biol Chem* 279:14245–14255
- Kokoska RJ, McCulloch SD, Kunkel TA (2003) The efficiency and specificity of apurinic/apryrimidinic site bypass by human DNA polymerase η and *Sulfolobus solfataricus* Dpo4. *J Biol Chem* 278:50537–50545
- 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* 29:5008–5019
- Krakoff IH, Brown NC, Reichard P (1968) Inhibition of ribonucleoside diphosphate reductase by hydroxyurea. *Cancer Res* 28:1559–1565
- Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MH, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'Shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440:637–643
- Kunkel TA, Hamatake RK, Motto-Fox J, Fitzgerald MP, Sugino A (1989) Fidelity of DNA polymerase I and the DNA polymerase I-DNA primase complex from *Saccharomyces cerevisiae*. *Mol Cell Biol* 9:4447–4458
- Kunkel TA, Roberts JD, Sugino A (1991) The fidelity of DNA synthesis by the catalytic subunit of yeast DNA polymerase α alone and with accessory proteins. *Mutat Res* 250:175–182
- Larrea AA, Lujan SA, Nick McElhinny SA, Mieczkowski PA, Resnick MA, Gordenin DA, Kunkel TA (2010) Genome-wide model for the normal eukaryotic DNA replication fork. *Proc Natl Acad Sci U S A* 107:17674–17679
- Lee SH, Pan ZQ, Kwong AD, Burgers PM, Hurwitz J (1991) Synthesis of DNA by DNA polymerase ϵ *in vitro*. *J Biol Chem* 266:22707–22717
- Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. *Nature* 396:643–649
- Levin DS, Bai W, Yao N, O'Donnell M, Tomkinson AE (1997) An interaction between DNA ligase I and proliferating cell nuclear antigen: implications for Okazaki fragment synthesis and joining. *Proc Natl Acad Sci U S A* 94:12863–12868
- Lou H, Komata M, Katou Y, Guan Z, Reis CC, Budd M, Shirahige K, Campbell JL (2008) Mrc1 and DNA polymerase ϵ function together in linking DNA replication and the S phase checkpoint. *Mol Cell* 32:106–117
- McCulloch SD, Kunkel TA (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* 18:148–161
- McCulloch SD, Kokoska RJ, Chilkova O, Welch CM, Johansson E, Burgers PM, Kunkel TA (2004) Enzymatic switching for efficient and accurate translesion DNA replication. *Nucleic Acids Res* 32:4665–4675
- Minnick DT, Bebenek K, Osheroff WP, Turner RM Jr, Astatke M, Liu L, Kunkel TA, Joyce CM (1999) Side chains that influence fidelity at the polymerase active site of *Escherichia coli* DNA polymerase I (Klenow fragment). *J Biol Chem* 274:3067–3075
- Minnick DT, Liu L, Grindley ND, Kunkel TA, Joyce CM (2002) Discrimination against purine-pyrimidine mispairs in the polymerase active site of DNA polymerase I: a structural explanation. *Proc Natl Acad Sci USA* 99:1194–1199
- Mirkin EV, Mirkin SM (2007) Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev* 71:13–35

- Morrison A, Bell JB, Kunkel TA, Sugino A (1991) Eukaryotic DNA polymerase amino acid sequence required for 3'→5' exonuclease activity. *Proc Natl Acad Sci U S A* 88:9473–9477
- Morrison A, Johnson AL, Johnston LH, Sugino A (1993) Pathway correcting DNA replication errors in *Saccharomyces cerevisiae*. *EMBO J* 12:1467–1473
- Muramatsu S, Hirai K, Tak YS, Kamimura Y, Araki H (2010) CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol ϵ , and GINS in budding yeast. *Genes Dev* 24:602–612
- Navadgi-Patil VM, Burgers PM (2009) A tale of two tails: activation of DNA damage checkpoint kinase Mec1/ATR by the 9-1-1 clamp and by Dpb11/TopBP1. *DNA Repair (Amst)* 8:996–1003
- Navas TA, Zhou Z, Elledge SJ (1995) DNA polymerase ϵ links the DNA replication machinery to the S phase checkpoint. *Cell* 80:29–39
- Navas TA, Sanchez Y, Elledge SJ (1996) RAD9 and DNA polymerase ϵ form parallel sensory branches for transducing the DNA damage checkpoint signal in *Saccharomyces cerevisiae*. *Genes Dev* 10:2632–2643
- Nick McElhinny SA, Gordenin DA, Stith CM, Burgers PM, Kunkel TA (2008) Division of labor at the eukaryotic replication fork. *Mol Cell* 30:137–144
- Nick McElhinny SA, Kumar D, Clark AB, Watt DL, Watts BE, Lundstrom EB, Johansson E, Chabes A, Kunkel TA (2010a) Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol* 6:774–781
- Nick McElhinny SA, Watts BE, Kumar D, Watt DL, Lundstrom EB, Burgers PM, Johansson E, Chabes A, Kunkel TA (2010b) Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Natl Acad Sci U S A* 107:4949–4954
- Niedernhofer LJ, Lalai AS, Hoeijmakers JH (2005) Fanconi anemia (cross)linked to DNA repair. *Cell* 123:1191–1198
- Nuutinen T, Tossavainen H, Fredriksson K, Pirila P, Permi P, Pospiech H, Syvaoja JE (2008) The solution structure of the amino-terminal domain of human DNA polymerase ϵ subunit B is homologous to C-domains of AAA+ proteins. *Nucleic Acids Res* 36:5102–5110
- Ohya T, Maki S, Kawasaki Y, Sugino A (2000) Structure and function of the fourth subunit (Dpb4p) of DNA polymerase ϵ in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 28:3846–3852
- Ohya T, Kawasaki Y, Hiraga S, Kanbara S, Nakajo K, Nakashima N, Suzuki A, Sugino A (2002) The DNA polymerase domain of pol(ϵ) is required for rapid, efficient, and highly accurate chromosomal DNA replication, telomere length maintenance, and normal cell senescence in *Saccharomyces cerevisiae*. *J Biol Chem* 277:28099–28108
- Pavlov YI, Frahm C, Nick McElhinny SA, Niimi A, Suzuki M, Kunkel TA (2006) Evidence that errors made by DNA polymerase α are corrected by DNA polymerase δ . *Curr Biol* 16:202–207
- Payne BT, van Knippenberg IC, Bell H, Filipe SR, Sherratt DJ, McGlynn P (2006) Replication fork blockage by transcription factor-DNA complexes in *Escherichia coli*. *Nucleic Acids Res* 34:5194–5202
- Penczek PA, Grassucci RA, Frank J (1994) The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* 53:251–270
- Puddu F, Piergiorganni G, Plevani P, Muzi-Falconi M (2011) Sensing of replication stress and Mec1 activation act through two independent pathways involving the 9-1-1 complex and DNA polymerase ϵ . *PLoS Genet* 7:e1002022
- Pursell ZF, Kunkel TA (2008) DNA polymerase epsilon: a polymerase of unusual size (and complexity). *Prog Nucleic Acid Res Mol Biol* 82:101–145
- Pursell ZF, Isov I, Lundstrom EB, Johansson E, Kunkel TA (2007) Yeast DNA polymerase ϵ participates in leading-strand DNA replication. *Science* 317:127–130
- Randell JC, Bowers JL, Rodriguez HK, Bell SP (2006) Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol Cell* 21:29–39
- Rogozin IB, Makarova KS, Pavlov YI, Koonin EV (2008) A highly conserved family of inactivated archaeal B family DNA polymerases. *Biol Direct* 3:32

- Shamoo Y, Steitz TA (1999) Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* 99:155–166
- Shcherbakova PV, Pavlov YI (1996) 3'→5' Exonucleases of DNA polymerases ϵ and δ correct base analog induced DNA replication errors on opposite DNA strands in *Saccharomyces cerevisiae*. *Genetics* 142:717–726
- Shcherbakova PV, Bebenek K, Kunkel TA (2003a) Functions of eukaryotic DNA polymerases. *Sci Aging Knowledge Environ* 2003:RE3
- Shcherbakova PV, Pavlov YI, Chilkova O, Rogozin IB, Johansson E, Kunkel TA (2003b) Unique error signature of the four-subunit yeast DNA polymerase ϵ . *J Biol Chem* 278:43770–43780
- Shikata K, Sasa-Masuda T, Okuno Y, Waga S, Sugino A (2006) The DNA polymerase activity of Pol ϵ holoenzyme is required for rapid and efficient chromosomal DNA replication in *Xenopus* egg extracts. *BMC Biochem* 7:21
- Shiotani B, Kobayashi M, Watanabe M, Yamamoto K, Sugimura T, Wakabayashi K (2006) Involvement of the ATR- and ATM-dependent checkpoint responses in cell cycle arrest evoked by p18INK4. *Mol Cancer Res* 4:125–133
- Steitz TA (1993) DNA- and RNA-dependent DNA polymerases. *Curr Opin Struct Biol* 3:31–38
- Stinchcomb DT, Struhl K, Davis RW (1979) Isolation and characterisation of a yeast chromosomal replicator. *Nature* 282:39–43
- Stocki SA, Nonay RL, Reha-Krantz LJ (1995) Dynamics of bacteriophage T4 DNA polymerase function: identification of amino acid residues that affect switching between polymerase and 3'→5' exonuclease activities. *J Mol Biol* 254:15–28
- Tahirov TH, Makarova KS, Rogozin IB, Pavlov YI, Koonin EV (2009) Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol ϵ and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. *Biol Direct* 4:11
- 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* 17:1153–1165
- Traut TW (1994) Physiological concentrations of purines and pyrimidines. *Mol Cell Biochem* 140:1–22
- Tsubota T, Maki S, Kubota H, Sugino A, Maki H (2003) Double-stranded DNA binding properties of *Saccharomyces cerevisiae* DNA polymerase ϵ and of the Dpb3p-Dpb4p subassembly. *Genes Cells* 8:873–888
- 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 ϵ , promotes epigenetic silencing in *Saccharomyces cerevisiae*. *J Biol Chem* 281:32898–32908
- Warbrick E, Lane DP, Glover DM, Cox LS (1997) Homologous regions of Fen1 and p21Cip1 compete for binding to the same site on PCNA: a potential mechanism to co-ordinate DNA replication and repair. *Oncogene* 14:2313–2321
- Warbrick E, Heatherington W, Lane DP, Glover DM (1998) PCNA binding proteins in *Drosophila melanogaster*: the analysis of a conserved PCNA binding domain. *Nucleic Acids Res* 26:3925–3932
- Wardle J, Burgers PM, Cann IK, Darley K, Heslop P, Johansson E, Lin LJ, McGlynn P, Sanvoisin J, Stith CM, Connolly BA (2008) Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. *Nucleic Acids Res* 36:705–711
- Watt DL, Johansson E, Burgers PM, Kunkel TA (2011) Replication of ribonucleotide-containing DNA templates by yeast replicative polymerases. *DNA Repair (Amst)* 10:897–902
- Wu P, Nossal N, Benkovic SJ (1998) Kinetic characterization of a bacteriophage T4 antimutator DNA polymerase. *Biochemistry* 37:14748–14755
- Yang G, Lin T, Karam J, Konigsberg WH (1999) Steady-state kinetic characterization of RB69 DNA polymerase mutants that affect dNTP incorporation. *Biochemistry* 38:8094–8101
- Yang G, Franklin M, Li J, Lin TC, Konigsberg W (2002) A conserved Tyr residue is required for sugar selectivity in a Pol α DNA polymerase. *Biochemistry* 41:10256–10261

Chapter 14

The RFC Clamp Loader: Structure and Function

Nina Y. Yao and Mike O'Donnell

Abstract The eukaryotic RFC clamp loader couples the energy of ATP hydrolysis to open and close the circular PCNA sliding clamp onto primed sites for use by DNA polymerases and repair factors. Structural studies reveal clamp loaders to be heteropentamers. Each subunit contains a region of homology to AAA+ proteins that defines two domains. The AAA+ domains form a right-handed spiral upon binding ATP. This spiral arrangement generates a DNA binding site within the center of RFC. DNA enters the central chamber through a gap between the AAA+ domains of two subunits. Specificity for a primed template junction is achieved by a third domain that blocks DNA, forcing it to bend sharply. Thus only DNA with a flexible joint can bind the central chamber. DNA entry also requires a slot in the PCNA clamp, which is opened upon binding the AAA+ domains of the clamp loader. ATP hydrolysis enables clamp closing and ejection of RFC, completing the clamp loading reaction.

Keywords Clamp loader • Sliding clamp • DNA polymerase • Replisome • RFC • PCNA • AAA+ machine

14.1 Overview of Clamp Loaders and Sliding Clamps

Clamp loaders are so called for their action in loading ring-shaped sliding clamps onto DNA (see Fig. 14.1a). Sliding clamps encircle DNA and slide along the duplex while binding DNA polymerases, tethering them to DNA for high processivity

N.Y. Yao (✉) • M. O'Donnell
Howard Hughes Medical Institute, The Rockefeller University,
1230 York Avenue, New York, NY 10065, USA
e-mail: yaon@rockefeller.edu; odonnel@rockefeller.edu

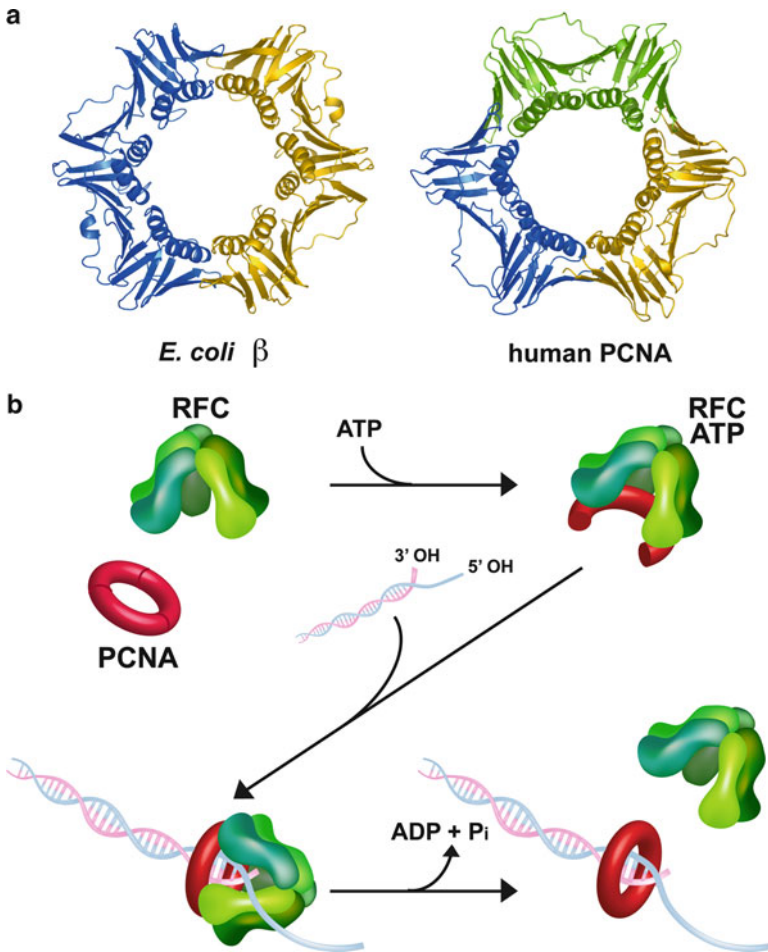


Fig. 14.1 Scheme of clamp loader function. (a) The structures of the *E. coli* β and human PCNA sliding clamps (PDB: 2POL and PDB: AXC, respectively). (b) Schematic of clamp loader function using eukaryotic RFC and PCNA as the example. ATP binding to RFC enables RFC to bind and open PCNA. In the presence of a primed template, RFC places PCNA onto DNA and then hydrolyzes ATP to eject from the PCNA-DNA complex (Reproduced with permission from Figure 1a of Bowman et al. (2004))

during chain extension (reviewed in O'Donnell and Kuriyan 2006). Clamps and clamp loaders are ubiquitous in all life forms and thus must have evolved in the progenitor ancestor cell from which all different cell types arose.

Sliding clamps were the first type of protein known to function by encircling DNA (Kong et al. 1992; Stukenberg et al. 1991). Today, many DNA metabolic proteins are known to encircle DNA for performance of their function. The structures of the *E. coli* β clamp and the eukaryotic PCNA (proliferating cell nuclear antigen) clamp are shown in Fig. 14.1a (Gulbis et al. 1996; Krishna et al. 1994). The eukaryotic PCNA clamp is a homotrimer, while the bacterial clamp is a homodimer. Despite

Table 14.1 Three component subunit structures of replicases from diverse organisms

Organism	Polymerase	Clamp loader pentamer	Sliding clamp
<i>E. coli</i>	Pol III	$(\gamma/\tau)_3\delta\delta'$	β dimer
T4 phage	gp43	gp44/62 pentamer	gp45 trimer
Archaea	Pol B	RFC pentamer	PCNA trimer
Eukaryotes	Pol δ , Pol ϵ	RFC pentamer	PCNA trimer

this difference in oligomeric state, the eukaryotic and bacterial clamps have essentially the same structure. In both cases, the clamps are six domain rings, and each domain has the same chain folding pattern. The domains have fused together during evolution in various ways giving rise to the different oligomerization states. In PCNA, the individual protomers of the trimer are composed of two domains each. In bacteria, the clamps are homodimers, and each protomer consists of three domains. The structure of sliding clamps from archaeal cells and bacteriophage T4 have also been solved; they share the same general chain fold and trimeric oligomerization state as PCNA (Dore et al. 2006; Matsumiya et al. 2001; Moarefi et al. 2000; Shamoo and Steitz 1999). The detailed structure and function of PCNA is the subject of the Chap. 15.

The crystal structures showed the sliding clamp rings were closed, implying that another factor was needed to crack these ring-shaped clamps open and then reclose them onto DNA. In fact, a second protein was known to be required for the clamp to function with DNA polymerase. In all systems, this “clamp loader” protein is a pentameric ATPase. The eukaryotic clamp loader was first identified as a necessary protein for SV40 replication *in vitro* (Fairman et al. 1988). The exact function was not known at the time and it was named replication factor C (RFC) (Waga and Stillman 1994) or activator-1 (Lee et al. 1989). The name RFC gained widespread use and is the term used today for eukaryotic and archaeal clamp loaders (Grabowski and Kelman 2003). The names of the clamp loader, clamp and replicative DNA polymerase in different cell types are given in Table 14.1.

Clamp loaders are heteropentamers and they hydrolyze ATP to assemble their respective clamps onto a primed DNA site (see Fig. 14.1b) (O'Donnell and Kuriyan 2006). However, the clamp loader binds the same surface of the clamp as the DNA polymerase and prevents the interaction of DNA polymerase with the clamp (Jonsson et al. 1998; Naktinis et al. 1996). Thus, upon placing the clamp on DNA, the clamp loader must eject from the clamp to enable the polymerase access to the clamp. The hydrolysis of ATP accomplishes clamp loader ejection from the clamp as illustrated in the second step of Fig. 14.1b (Ason et al. 2003). The polymerase then binds to the clamp for highly processive function.

14.2 Clamp Loader Structure

The RFC clamp loader is composed of five essential “clamp loading” subunits referred to as RFC1 through 5 (Cullmann et al. 1995). With the exception of the large RFC1 subunit (approximately 128 kDa in human), the RFC2, 3, 4 and 5 subunits

are approximately 38–41 kDa each. The five subunits contain a region of homology with one another (O'Donnell et al. 1993). This region of homology defines a large family of proteins referred to as AAA+ proteins (ATPases associated with a variety of cellular functions) (Erzberger and Berger 2006). The AAA+ region encodes two domains, the structure of which will be discussed shortly. AAA+ proteins generally perform protein remodeling reactions in a wide variety of cellular pathways (Neuwald et al. 1999). Many AAA+ proteins are circular hexamers, although other oligomerization states exist and in fact, the heteropentameric clamp loader is one of these exceptions. Numerous examples of AAA+ proteins participate in the replication process, including origin binding proteins (i.e. several subunits of the ORC complex, Cdc6/Cdc18, the Mcm2-7 helicase and of course, the RFC clamp loader) (Davey et al. 2002).

The AAA+ homologous region folds into two domains, and these domains bind ATP (Guenther et al. 1997). The P-loop and DEAD box ATP site motifs are located on the larger, N-terminal domain, and the smaller domain contains several residues that are important to binding and/or hydrolysis. In all clamp loader subunits there is at least one additional domain that is C-terminal to the AAA+ domains. The C-terminal domain that is outside of the region of AAA+ homology is mostly composed of α -helix and it mediates the strongest intersubunit interactions that hold the pentamer together. RFC1, the largest subunit of RFC, contains both N- and C-terminal extensions in addition to these three domains (Bunz et al. 1993). Although the N-terminal extension of RFC1 contains a region of homology to DNA ligases (i.e. the BRCT domain), it does not have ligase activity. The solution structure of the human RFC1 N-terminal BRCT domain has recently been solved and a model for BRCT DNA binding presented (Kobayashi et al. 2010). The N-terminal extension of RFC1 is not essential for cell viability (Gomes et al. 2000), nor is it required for *in vitro* clamp loading activity (Uhlmann et al. 1997), but removal of this region results in sensitivity to DNA damage *in vivo* (Gomes et al. 2000).

The subunits of RFC, like clamp loaders of all cell types, are arranged in a circular shape (Bowman et al. 2004; Jeruzalmi et al. 2001a). The crystal structures of the bacterial and eukaryotic clamp loaders, in Fig. 14.2a and b, respectively, show that the AAA+ domains of the five subunits are arranged in a spiral, while the C-terminal domains define a nearly planar circle, referred to as a “collar”. One subunit of the *E. coli* clamp loader is shown at the right of Fig. 14.2a, to illustrate the three domains structure of clamp loader subunits (Jeruzalmi et al. 2001a). By convention, clamp loaders are viewed from the “side” with the C-terminal domain at the top, and the N-terminal AAA+ domains at the bottom (Jeruzalmi et al. 2001a). Proceeding counterclockwise around the circle from the subunit at the far right, the subunit positions are referred to as the A, B, C, D and E subunits (see Fig. 14.2b and c). The C-terminal domains of the collar form a tightly closed circle with no gap and are the main connections that hold the complex together. In all clamp loaders, there is a gap between the AAA+ domains of subunits A and E (i.e. see Fig. 14.2a). In RFC this gap is present between RFC1 and RFC5 (Fig. 14.2b and c). In RFC, this gap is somewhat narrower than in the *E. coli* clamp loader because RFC1 has a C-terminal region that

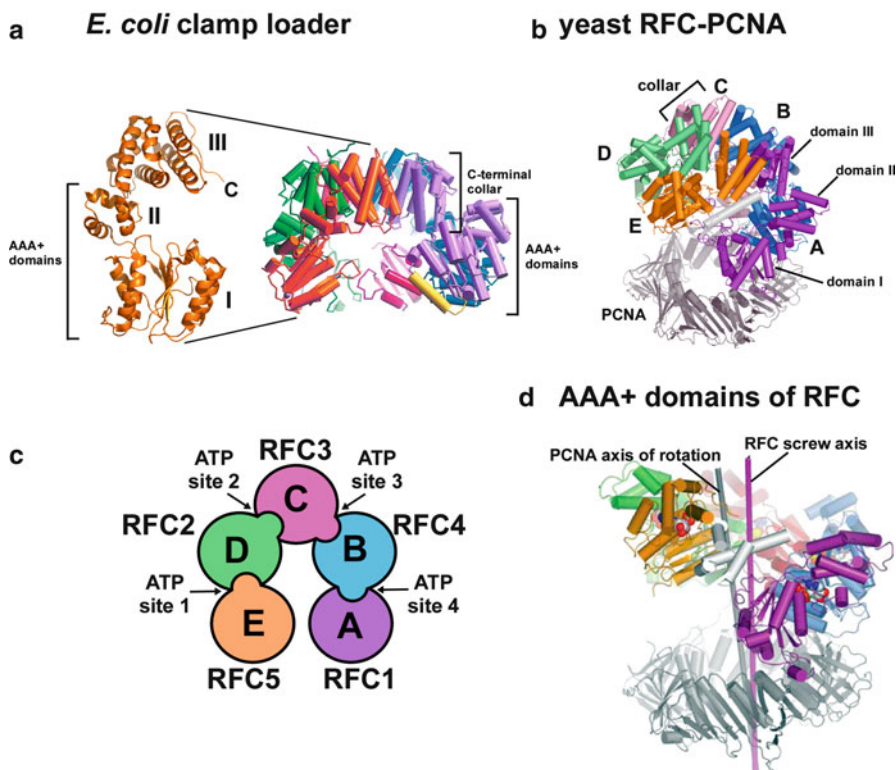


Fig. 14.2 Architecture of the clamp loader. (a) *E. coli* minimal clamp loader heteropentamer of $\gamma_3\delta\delta'$ (PDB: 1JR3). The *left* part of the figure shows the three-domain architecture of each subunit. The *right* part of the figure shows the gap between the AAA+ domains of subunits δ (purple) and δ' (orange). The yellow α helix of δ binds the β clamp at one end (see text for details) (Adapted with permission from Figure 1a of Jeruzalmi et al. (2001a)). (b) The yeast RFC-PCNA-ATP γ S complex (PDB: 1SXJ). RFC is in color and PCNA is in grey. The three domains of the subunits are as indicated, and subunits are noted according to their positions (a–e), which for yeast RFC are RFC1, RFC4, RFC3, RFC2 and RFC5, respectively (Adapted with permission from Figure 1b of Bowman et al. (2004)). (c) Cartoon of the arrangement of yeast RFC subunits. The location of ATP sites at subunit interfaces is indicated. ATP binding sites are within subunits RFC2 (ATP site 1), RFC3 (ATP site 2), RFC4 (ATP site 3) and RFC5 (ATP site 4). Subunits with arginine fingers that are needed for catalysis are in RFC5 (ATP site 1), RFC2 (ATP site 2), RFC3 (ATP site 3) and RFC4 (ATP site 4) (Adapted with permission from Figure 3b in O'Donnell and Kuriyan (2006)). (d) RFC–PCNA structure in which only the AAA+ domains of RFC are shown (color) and the collar is removed. The figure illustrates the *spiral* shape of the AAA+ domains. RFC1 (purple) forms the most extensive contact with PCNA (grey) (Reproduced with permission from Figure 2 of Bowman et al. (2004))

extends across the gap and protrudes down toward the N-terminal face of the clamp loader (colored white in Fig. 14.2b).

Although clamp loader subunits are homologous to AAA+ proteins, they do not all have a functional ATP binding site (e.g. lack the P-loop). Therefore, certain

clamp loader subunits do not hydrolyze ATP. In the eukaryotic RFC clamp loader the RFC5 subunit lacks a P-loop. Thus the RFC pentamer has only four functional ATP sites. In fact, mutational studies show that only three of these sites are needed for PCNA clamp loading; the ATP site of RFC1 is not essential for clamp loading (Gomes and Burgers 2001). A common feature of AAA+ proteins is that the ATP site is located at the interface of two subunits, and catalysis requires residues from both subunits. Specifically, one subunit binds ATP, but residue(s) needed for catalysis of ATP are donated by the adjacent subunit. In RFC, one of these inter-subunit catalytic residues is an “arginine finger” set within a SRC motif that is conserved in clamp loader subunits of all cell types (Bowman et al. 2004; Jeruzalmi et al. 2001a; O'Donnell and Kuriyan 2006). Mutational analysis has shown that the arginine of the SRC motif is required for ATP catalysis (Johnson and O'Donnell 2003; Johnson et al. 2006; Snyder et al. 2004; Williams et al. 2004). The fact that the ATP sites of AAA+ oligomers are formed by residues of two neighboring subunits suggests that the AAA+ architecture may reflect an underlying necessity for intersubunit communication, consistent with the idea that cooperation of subunits within a large complex is important to performance of a complicated protein remodeling task.

14.3 RFC Clamp Loader Interaction with DNA

The structure of RFC in complex with PCNA and ATP γ S revealed an unanticipated mode by which clamp loaders bind to DNA, even though DNA was not present in the structure (Bowman et al. 2004, 2005). Modeling of DNA through the PCNA ring, located below the complex, indicated that DNA may reside inside the center of the clamp loader and be surrounded by all five subunits (see Fig. 14.3a). Surprisingly there is sufficient space inside the center of RFC to accommodate duplex DNA (Fig. 14.3b), but even more telling is the disposition of the AAA+ domains of all five subunits. The AAA+ domains are arranged in a right-handed spiral that closely match the pitch of the DNA duplex (Fig. 14.2d and illustration in Fig. 14.3c). This highly suggested that DNA binds inside the clamp loader. Furthermore, many positive charged residues present in each subunit are within hydrogen bonding distance to DNA modeled into the complex. These potential DNA interactive residues are conserved from bacteria to human, further suggesting their importance in bind to DNA (Bowman et al. 2005). These conserved residues are located on two α helices in each subunit that have positive dipoles pointed toward the direction of DNA modeled into the central chamber (illustrated in Fig. 14.3c). Although the AAA+ domains of the unliganded (no ATP) *E. coli* $\gamma_3\delta\delta'$ complex are also arranged in a spiral, the pitch is not as steep as that seen in the yeast RFC-PCNA-ATP γ S structure. This difference between unliganded *E. coli* clamp loader (no ATP) and ATP γ S bound RFC-PCNA suggests that ATP binding results in a conformational change that brings the AAA+ domains into the correct spiral shape to form a central DNA binding site. Subsequent mutational studies demonstrated that these positive charged residues are indeed required to bind DNA, both for the *E. coli* clamp loader and yeast RFC (Goedken et al. 2005; Yao et al. 2006).

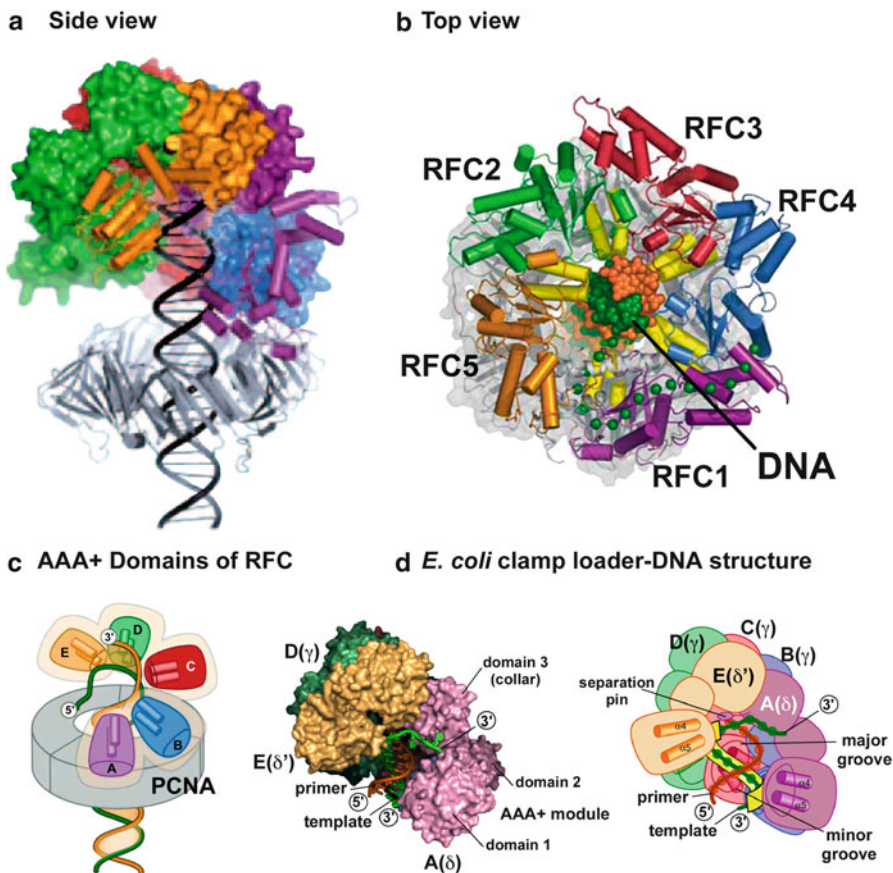


Fig. 14.3 Model of the RFC-PCNA-DNA complex. (a) Duplex DNA is modeled through PCNA and into the center of RFC (Reproduced with permission from Figure 3c in O'Donnell and Kuriyan (2006)). (b) View of the RFC-PCNA-DNA model looking from the “top” of RFC with the C-terminal “collar” removed. There exists space for DNA in the center of RFC, and each subunit contains residues on two α helices (yellow) that point toward the DNA (Adapted with permission from Figure 4b of Bowman et al. (2004)). (c) Cartoon of the spiral disposition of the AAA+ domains of RFC relative to DNA and PCNA (Reproduced with permission from Figure 4c of Bowman et al. (2004)). (d) Structure of the *E. coli* clamp loader bound to a primed DNA (left) and cartoon of the structure (right diagram) (PDB: 3GLF) (Simonetta et al. 2009). The template strand of DNA is colored green (left) or yellow (right) (Reproduced with permission from Figure 1c in Simonetta et al. (2009))

The proposal that DNA binds to the central chamber of the clamp loader provided an immediate answer to the question of why clamp loaders have a gap between two subunits. Since clamps are loaded at primed sites on template strands that do not have DNA “ends”, the DNA cannot simply enter the clamp loader through the “bottom”. Hence, the observed gap between two clamp loading subunits provide an entry port for an “endless” DNA strand to enter into the central cavity of the clamp loader. Entry of a long DNA strand would also require an open interface of the

sliding clamp located under the clamp loader in which the clamp loader gap and the open interface of the clamp are aligned with one another for DNA entry.

Structural support for DNA binding to the central cavity of RFC was obtained from an electron microscopy study of an archaeal RFC-PCNA bound to DNA (Miyata et al. 2005). Although high resolution architectural details are not visible, the DNA appeared to reside inside the clamp loader. The electron microscopy study is described further below in the context of PCNA ring opening. A high resolution structure of DNA bound to the inside of a clamp loader has recently been obtained for the *E. coli* clamp loader in the presence of a primed template DNA (Simonetta et al. 2009). The structure, shown in Fig. 14.3d, confirms that duplex DNA resides within the central cavity of the $\gamma_3\delta\delta'$ clamp loader, and that each subunit interacts with the DNA backbone. Interestingly, the clamp loader only interacts with the template strand, not the primer strand. This finding was unanticipated and may reflect an important biological function. Specifically, the *E. coli* clamp loader must be capable of assembling the clamp onto an RNA primer made by primase, but it must also be capable of assembling the clamp onto a DNA primed site during various types of repair reactions outside the context of chromosome replication. RNA-DNA and DNA-DNA duplexes have different structures. RNA-DNA duplexes prefer the A-form, which has a much larger diameter than B-form duplex DNA. Interestingly, the fit of RNA-DNA modeled into the structure indicates that the template DNA strand interactions can be maintained with very little or no change. Furthermore, the central cavity has sufficient diameter to accommodate the greater diameter of A-form RNA-DNA relative to B-form DNA-DNA. Hence, the fact that the major interactions between the clamp loader and DNA occur through the template strand may facilitate binding to both A-form and B-form structures.

Primed sites are synthesized at nearly random positions during lagging strand synthesis (Kornberg and Baker 1992), and thus clamp loaders must be capable of recognizing the structure of a primed site, not a specific sequence for clamp loading. Structure specific binding to a primed site is made possible by the tight packing of the C-terminal domains in the collar (Bowman et al. 2004, 2005). As discussed above, DNA will enter the central chamber of the clamp loader through the gap between the AAA+ domains of two subunits that are aligned with the open interface of the clamp, but the tightly packed C-terminal domains provide a “cap” that prevents DNA from going straight through the structure. This imposes the requirement that DNA must make a sharp bend in order to bind into the central chamber of the clamp loader. Duplex DNA is too rigid for this sharp bend, but the flexibility of single-strand DNA should allow this bend to occur. This is nicely apparent from the structure of the *E. coli* clamp loader bound to a primed template (Fig. 14.3d), in which the template strand bends out from the side of the clamp loader at the top of the central cavity, just below the “cap” formed by the C-terminal domains. Thus the clamp loader structure can be compared to a “screw cap” in which the screw binds duplex DNA and the cap imposes the sharp bend that provides specificity for a primed template junction that has a flexible template single strand.

14.4 ATP Binding and Opening of the Clamp

Studies in both *E. coli* and eukaryotes have shown that the clamp loader requires ATP binding to interact with the sliding clamp (Gomes and Burgers 2001; Naktinis et al. 1996). However, ATP does not need to be hydrolyzed for this function. Indeed, the *E. coli* clamp loader has been shown to open one interface of the β dimer upon binding ATP (or ATP γ S) (Turner et al. 1999). Therefore hydrolysis is not required for clamp binding or clamp opening, but is needed for completion of the reaction (clamp loader ejection/clamp closing).

Most proteins that bind the PCNA clamp contain a conserved motif, called the PIP (PCNA interacting peptide) motif (Warbrick 2000). The way that the PIP sequence motif binds to PCNA was determined from the structure of a peptide of this motif bound to the human PCNA clamp (Gulbis et al. 1996). The PCNA clamp has a hydrophobic pocket located between the two domains in each subunit. The PIP motif binds into this hydrophobic pocket. The PIP motif is required for most known protein-PCNA interactions. Subsequently, the structure of an *E. coli* clamp loader subunit (δ subunit) bound to the β clamp, and a peptide bound to the T4 clamp, revealed a very similar type of interaction in which a peptide sequence binds into a hydrophobic pocket located between the globular domains (Jeruzalmi et al. 2001b; Shamoo and Steitz 1999). A consensus sequence for bacterial clamp binding peptides has been described, in analogy to the PIP sequence (Dalrymple et al. 2001; Wijffels et al. 2004).

It is important to note that there appear to be additional binding sites to which proteins can interact with the clamp, in addition to the interaction with the hydrophobic pocket of the clamp. For example, one such secondary site is observed in the crystal structure of Pol IV bound to the bacterial clamp (Bunting et al. 2003). Secondary sites of interaction of protein binding to PCNA are suggested by mutational and genetic studies (Ayyagari et al. 1995; Eissenberg et al. 1997; Gomes and Burgers 2000; Johansson et al. 2004). Despite these additional interactions with the clamp, interaction with the hydrophobic pocket in the clamp is thought to be the major source of binding energy between the clamp and the proteins that it binds.

PCNA is a homotrimer and therefore has three identical protein binding sites, one in each subunit. Thus multiple proteins may attach to the PCNA clamp at the same time through binding to these identical sites. This aspect of clamp biology, in which multiple proteins bind the clamp at the same time, is referred to as the “tool belt” hypothesis. This very interesting subject will be briefly mentioned again later, and is expanded upon in Chap. 15.

The N-terminal, or bottom surface, of RFC binds to PCNA (Fig. 14.2b). As described above, this surface is composed of the N-terminal AAA+ domains that are arranged in a right-handed spiral. The spiral shape of the N-terminal bottom surface of RFC can only accommodate binding of two to three RFC subunits, as a spiral geometry is inconsistent with complete interaction with a spiral (i.e. all five subunits of RFC) with a flat closed circular PCNA ring. However, if an open PCNA clamp adopts a right-handed spiral to match the right-handed spiral surface of the clamp

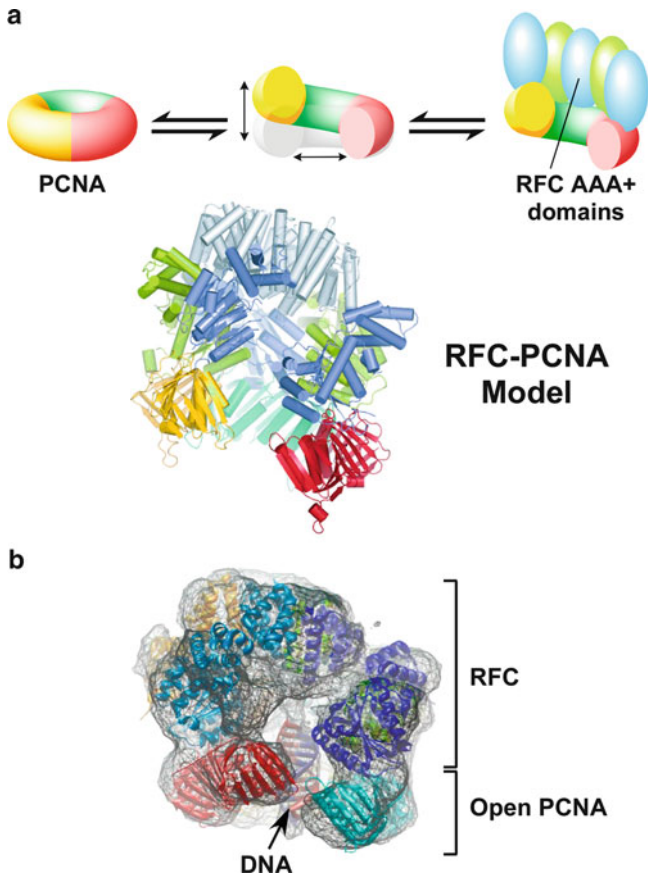


Fig. 14.4 The PCNA clamp opens in a right-handed spiral. **(a)** Molecular simulations indicate that PCNA opens in a right-handed spiral that matches the shape of the RFC AAA+ domains (*top diagram*); the model of open PCNA with the structure of RFC is shown below (Kazmirski et al. 2005) (Reproduced with permission from Figure 5 of Kazmirski et al. (2005)). **(b)** Electron micrograph particle reconstruction of an archaeal RFC-PCNA in the presence of DNA and ATP (Miyata et al. 2005) (Reproduced with permission from Figure 3a in Miyata et al. (2005))

loader, all the RFC subunits could bind the open PCNA clamp. Molecular simulations of PCNA in fact indicate that once an interface of PCNA is broken, the clamp spontaneously opens into a right-handed spiral (Kazmirski et al. 2005). Furthermore, the open PCNA spiral closely matches the N-terminal surface of RFC (illustrated in Fig. 14.4a). The electron micrographic reconstruction of an archaeal RFC-PCNA-DNA also reveals that the clamp is open in a right-handed spiral (Miyata et al. 2005). One problem with both of these studies is that the opening in the clamp ($<15 \text{ \AA}$) is less than the distance needed for passage of a DNA duplex (Fig. 14.4b) (Kazmirski et al. 2005; Miyata et al. 2005). The clamp must open at least 20 \AA to accommodate passage of duplex DNA. Thus the mechanism of clamp opening is still an active issue for future research.

The PCNA clamp is closed in the yeast RFC-PCNA-ATP γ S structure, yet ATP binding should result in an open PCNA clamp when bound to RFC. Indeed, a fluorescent assay for clamp opening indicates that the PCNA clamp is open upon ATP binding to RFC (Zhuang et al. 2006). A likely reason for the closed clamp in the crystal structure is that RFC subunits were mutated in the “arginine finger” active site residue of the SRC motif to prevent any possible hydrolysis of ATP γ S over the long time frames needed for crystal growth. These active site mutations probably prevented the conformation change required to open the clamp. The fact that ATP binding to wild type (i.e. non-mutated) RFC yields an open PCNA clamp is further supported by electron microscopy studies of a archaeal RFC-PCNA-DNA-ATP complex showing the clamp in an open form (Fig. 14.4b) (Miyata et al. 2005).

Substantial insight into the mechanism of clamp opening has been obtained from crystal structure and biochemical analysis of the *E. coli* β clamp in complex with the δ subunit of the clamp loader (i.e. δ is located in position A) (Jeruzalmi et al. 2001b). Biochemical studies demonstrated that the δ subunit of the clamp loader, by itself, opens the clamp (Turner et al. 1999). The δ subunit cannot load clamps on DNA, as the other subunits of the clamp loader are needed to perform the organized tasks of binding DNA, placing the clamp on a primed site, ejection of the clamp loader and clamp closing.

Only one δ subunit binds to a β dimer and surprisingly, the δ subunit binds much more tightly to a monomeric β clamp, mutated at the interface (a half clamp) (Stewart et al. 2001). This observation suggests that the binding energy of δ subunit to β is used to perform “work” on the circular β clamp, presumably to open the clamp on its own. This “work” of clamp opening is manifested in a lower affinity of δ to a β dimer compared to the β monomer mutant. In other words, the β monomer “half clamp” is already open, so δ does not need to expend binding energy to open it and thus binds the β monomer tighter than the β dimer. Biochemical studies have demonstrated that the clamp is not dismantled into monomers during clamp loading, and that only one interface opens during clamp loading onto DNA (Turner et al. 1999).

The crystal structure of δ subunit bound to the monomeric mutant β clamp provided much deeper insight into the clamp opening mechanism (Jeruzalmi et al. 2001b). Specifically, the structure of δ bound to a β monomer revealed that the β monomer adopts a much shallower crescent shape than the β protomers within a β ring. This shallower crescent shape was due to rigid body motions of the three domains of the β protomer. Interestingly, the major rigid body change that contributed to the shallower crescent shape of the β monomer was distant from the δ - β binding site. The implication of these observations for clamp loading is that the protomers within the clamp are under spring tension (i.e. the protomers bend inward because the force of the dimer interfaces are strong), but when one interface of the ring is broken (i.e. by δ), then each protomer can “relax” to a shallower crescent shape, thereby providing a gap for DNA strand passage.

The δ - β structure also indicates a specific mechanism by which the δ subunit forces one β interface open. The connection between δ and the hydrophobic pocket in β is mediated by residues near one end of a long α helix in δ (colored yellow in

Fig. 14.2a). The opposite end of the same α helix extends to the β dimer interface and causes a distortion at the interface (Jeruzalmi et al. 2001a). This distortion may destabilize the β dimer interface and allow it to open.

14.5 ATP Hydrolysis and Closing of the Clamp

Clamp closure around DNA is the step at which ATP is hydrolyzed. In this regard, it is important to note that each subunit of the RFC heteropentamer is encoded by a different gene, and therefore each ATP site is structurally distinct (Cullmann et al. 1995). There are two extremes in which one may view the role of the different ATP sites. At one extreme, all the ATPs are hydrolyzed at once, and the purpose is to simply close the clamp and eject the clamp loader. In this view the clamp loader is a simple switch, either opening or closing the clamp, and DNA is the trigger for the switch in which primed template binding brings all catalytic residues into register at once and hydrolysis is essentially simultaneous at all sites regardless of the structural differences between the sites. At the other extreme, the ATP sites have individual functions enabled by their different molecular structures. In this view, the clamp loader is still a switch, but a more complicated one, where each ATP drives a different step along the path of clamp closure and clamp loader ejection.

Study of ATP binding and ATP site mutations in RFC indicates that different ATP sites have different functions and thus favors a more complicated switch (Gomes and Burgers 2001; Gomes et al. 2001; Johnson et al. 2006; Schmidt et al. 2001). One report demonstrates that the four ATP sites of RFC fill in a specific fashion (Gomes and Burgers 2001). Binding of the first two ATP enable RFC to bind PCNA, followed by a third ATP when it binds PCNA, while the fourth ATP binds upon association with the primed template (Gomes et al. 2001). Mutational studies of ATP site P-loops in different RFC subunits demonstrate that mutation of the P-loop of any single subunit has an effect on activity, although defects due to mutation of the RFC1 P-loop can be overcome by increasing the concentration of ATP (Cai et al. 1998; Podust et al. 1998; Schmidt et al. 2001). Interestingly, P-loop mutants of RFC still bind PCNA but are defective in binding DNA (Cai et al. 1998). This may be explained by inability of the RFC mutants to open PCNA, precluding DNA from the central chamber of RFC.

Studies that mutate the arginine finger in different RFC subunits suggest an order to ATP hydrolysis and further support the proposal that different ATPase sites have distinct functions in the clamp loading mechanism (Johnson et al. 2006). These experiments suggested that ATP binding was sensed by RFC3 to promote DNA association. ATP hydrolysis in RFC2 was specifically stimulated by PCNA, and thus may be the first ATP to be hydrolyzed and coupled to PCNA ring closure around DNA. Remaining ATP appears to be hydrolyzed in an ordered fashion around the clamp loader ring, starting from subunit D (RFC2), then subunits in positions C (RFC3), B (RFC4) and A (RFC1). The RFC-PCNA-ATP γ S structure indicates that upon closure of PCNA, contacts between PCNA to RFC2

and RFC5 will be severed (i.e. these subunit do not bind the closed ring in the structure). This would also disable ATP induced conformations in RFC needed for RFC binding to DNA. Hence, these differential effects conspire to achieve the same goal – specifically to disconnect RFC from PCNA and DNA, allowing ring closure and clamp loader ejection.

Studies of the *E. coli* clamp loader also suggest that different ATP sites may have distinct functions. The *E. coli* clamp loader pentamer ($\gamma_3\delta\delta'$) has three ATP binding sites; only the γ subunits bind ATP while δ and δ' do not. Even though the three ATP sites are each in a γ subunit, they are not identical due to their formation by the union of two subunits. Thus only two sites are structurally similar (those formed by γ - γ junctions at positions B-C and C-D), while the third site (position D-E) involves an arginine finger from an SRC motif within δ' . A mutation in the arginine finger of δ' results in deficient clamp binding, while mutations in the arginine fingers of γ disrupt DNA binding (Johnson and O'Donnell 2003; Snyder et al. 2004). These results further support the idea that different ATP sites have distinct functions in clamp loader action. Interestingly, after ATP hydrolysis and clamp loader ejection from β , the *E. coli* clamp loader remains inactive for a short time, possibly due to slow ADP release (Ason et al. 2003; Bertram et al. 2000). This may serve to prevent the clamp loader from unloading clamps, an observation for both RFC and bacterial clamp loaders (discussed later in this chapter).

A recent structure of the *E. coli* β clamp bound to a primed DNA site has implications for the clamp loading mechanism (Georgescu et al. 2008) and the recent structure of PCNA-DNA indicates these findings may generalize to RFC (McNally et al. 2010). Both structures show that DNA is highly tilted as it passes through the ring. The electron density of DNA is too sparse in the PCNA-DNA structure to locate the template strand. But the β clamp-DNA structure revealed that the clamp binds the single-strand DNA template strand at the same hydrophobic pocket that is used to bind to proteins (Georgescu et al. 2008). Considering the substantial similarities in clamps and clamp loaders of bacteria and eukaryotes, this structural feature may generalize to PCNA. Both the PCNA and *E. coli* β clamp have several contacts to the duplex DNA. Most notably are large loops that extend from the ring and bind each of the two strands of duplex DNA through conserved residues in the loops. Upon mutation of the duplex DNA interactive residues, replication activity is significantly reduced, but the reduction specifically resides in the clamp loading reaction and not the function of the clamp with DNA polymerase (Georgescu et al. 2008).

Clamp-DNA interactions are proposed to function at the step that draws the clamp closed around DNA during the clamp loading reaction (Georgescu et al. 2008). The residues on the loops that extend from the β clamp and bind DNA result in a pronounced tilt of the clamp on DNA, and this tilt may help disconnect the clamp from some of the subunits of the clamp loader (the first step in Fig. 14.5a). One may presume that the last connection between the clamp loader and clamp to be disrupted is the tight interaction between δ and the β clamp. The δ subunit is located in position A, analogous to the RFC1 subunit which, like δ , binds the clamp tighter than any of the other clamp loading subunits (Yao et al. 2003). The δ subunit

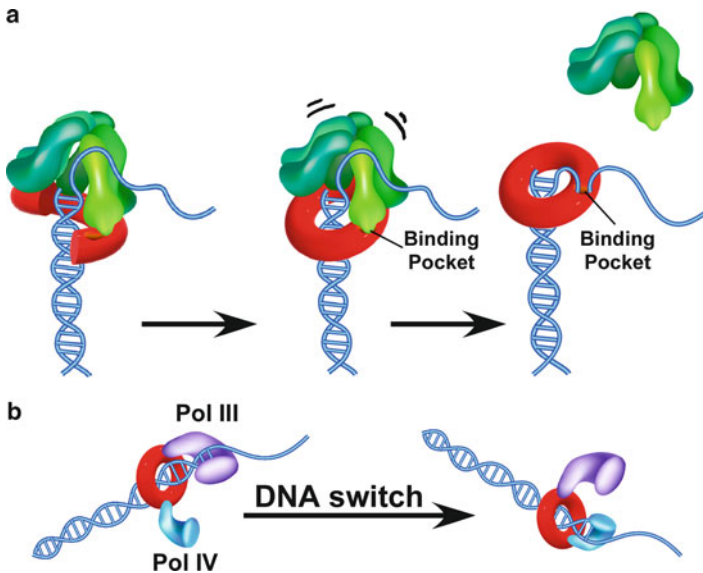


Fig. 14.5 Function of clamp-DNA interactions. (a) Proposed role of DNA-clamp interactions in the clamp loading mechanism. PCNA and β (shown) clamps bind duplex DNA with a high degree of tilt. The first arrow suggests that when the clamp closes, the tilt of DNA through the clamp may sever connections to some clamp loader subunits. The *second* arrow suggests that the template single-strand DNA competes with δ (analogous position to RFC1) to eject the clamp loader from the clamp-DNA complex. (b) The tilt of DNA through PCNA and β may help DNA switch among different proteins bound to the same clamp. Illustrated here is DNA switching among two different polymerases attached to one clamp (Adapted with permission from Figure 7 of Georgescu et al. (2008))

binds to the same hydrophobic pocket to which template single-strand DNA binds, and therefore single-strand DNA may compete δ from β as illustrated in the second step of Fig. 14.5b, thereby facilitating ejection of the clamp loader from the clamp-DNA complex.

Another proposed function of clamp-DNA interaction is to hold the clamp at a primed template junction after clamp loading. In the absence of these interactions, the clamp could conceivably slide on duplex DNA and be lost from the primed template junction. Hence, the clamp-DNA interactions may hold the clamp at the primed template junction after it is loaded but before the polymerase has associated with it, keeping the clamp where it is needed for function with the polymerase. This hypothesis has support in single-molecule studies of clamp sliding on primed DNA (Laurence et al. 2008).

The two protomers of the β clamp are identical, and therefore any particular site is duplicated. DNA may be presumed to rapidly isomerize between the two identical sites, tilting one way and then tilting the other. When two different DNA polymerases bind one clamp, the DNA tilt likely favors binding one polymerase, but when DNA isomerizes and tilts in the opposite direction, it may favor binding to the other DNA

polymerase (see Fig. 14.5b). Ability of the primed terminus to isomerize between different polymerases becomes quite important upon encounter with a template lesion that requires a translesion DNA polymerase to bypass the lesion. Indeed, it has been shown that both DNA polymerase III and translesion DNA polymerase IV can bind to the β dimer at the same time (Benkovic et al. 2001). Studies have shown that the high fidelity DNA polymerase III gains control of the primed DNA in preference to the translesion polymerase IV, but upon DNA polymerase III stalling (e.g. at a template lesion), the translesion polymerase IV gains control of the DNA (Benkovic et al. 2001). Thus DNA isomerization between different DNA polymerases bound to one clamp could enable replication forks that stall at a lesion to rapidly switch to the translesion polymerase for lesion bypass, then switch back to a high fidelity DNA polymerase to continue replication, thereby preventing fork collapse.

14.6 Clamp Loaders also Unload Clamps After Replication

Bacterial systems have demonstrated that the leading and lagging strand DNA polymerases are associated with the helicase, clamp loader and sliding clamps to form a “replisome” machine (reviewed in Benkovic et al. 2001; Johnson and O’Donnell 2005). Eukaryotic replisomes are anticipated to contain both leading and lagging polymerases as well. The presence of both polymerases in one replisome machine implies formation of DNA loops on the lagging strand, due to the antiparallel structure of duplex DNA (Sinha et al. 1980). Specifically, the leading strand proceeds in the direction of helicase unwinding, but the lagging strand must be copied in the opposite direction of fork movement. These opposed directions are made possible by formation of a DNA loop for each Okazaki fragment (Sinha et al. 1980). Upon discovery that both leading and lagging strand polymerases are held to DNA by a sliding clamp for high processivity, it became important to understand how a processive polymerase-clamp complex could dissociate from DNA upon completing each short Okazaki fragment.

The interesting question of how the lagging strand polymerase recycles from the end of one Okazaki fragment to begin extension of the next fragment is made possible by regulated attachment of the polymerase to the clamp. For example, yeast polymerase δ , the lagging strand polymerase (see Chap. 12, this volume), binds tightly to PCNA during processive synthesis but upon completing a substrate rapidly releases from PCNA and dissociates from the DNA (Langston and O’Donnell 2008). The polymerase then transfers to a new PCNA clamp at another primed site (assembled there by RFC) for extension of the next Okazaki fragment. This lagging strand polymerase recycling mechanism, whereby the polymerase hops from one clamp to the next, results in stoichiometric use of clamps, one per Okazaki fragment. This mechanism has been demonstrated to occur in bacterial systems as well (Johnson and O’Donnell 2005; Sinha et al. 1980).

The lagging strand mechanism whereby PCNA clamps are left behind on each completed Okazaki fragment is useful for the process by which RNA is removed at

the 5' terminus of Okazaki fragments. For example, PCNA functions with FEN1, the nuclease that removes RNA primers, and stimulates its action (Ayyagari et al. 2003; Kao and Bambara 2003). PCNA also interacts with ligase which seals Okazaki fragments together (Song et al. 2009). However, there are 10- to 100-fold more Okazaki fragments than there are clamps in both bacterial and eukaryotic systems and therefore clamps must be recycled on and off DNA many times during genome replication. However, PCNA is highly stable on DNA, requiring over half an hour to dissociate from DNA at 37°C (Yao et al. 1996). Therefore clamps must be actively removed from DNA to enable their reuse during duplication of a genome.

Studies have demonstrated that the RFC clamp loader can remove clamps from DNA (Yao et al. 1996). ATP binding is needed, but hydrolysis is not required. However, subassemblies of RFC can also open PCNA and remove it from DNA (Yao et al. 2006), similar to the δ subunit of the *E. coli* clamp loader opening β and unloading it from DNA. Specifically, a complex of RFC2/RFC5 is fully capable of removing PCNA, as is a complex of RFC2/3/4/5. Intracellular concentrations of individual yeast RFC subunits suggest that RFC2/RFC5 complex is in excess over other subunits and thus may be present inside cells, much as *E. coli* contains a fivefold excess of δ for β clamp recycling (Leu et al. 2000). Subassemblies of RFC cannot load PCNA because they are lacking one or more subunits required for clamp loading. Hence, they are capable of unidirectional action in unloading clamps, rather than both loading and unloading as is the case with RFC. It is also interesting to note that certain alternative RFCs (yeast Ctf18-RFC and yeast Rad17-RFC) can also unload PCNA from DNA (Bylund and Burgers 2005; Yao et al. 2006). Alternative RFC's are the subject of the next section.

14.7 Alternative RFCs

The subunit composition of RFC can be altered by replacement of the RFC1 subunit with another protein. Examples include Elg1 (Bellaoui et al. 2003; Ben-Aroya et al. 2003; Kanellis et al. 2003) and Ctf18 (Bellaoui et al. 2003; Mayer et al. 2001). These alternative RFC complexes are thought to load (or unload) PCNA onto DNA for specific processes involved in genome stability (Elg1) and in cohesion (Ctf18). The specific role of these alternative clamp loaders is not yet elucidated. Perhaps the best understood alternative clamp loader is the Rad17-RFC, in which RFC1 is replaced by Rad17 (Rad24 in *S. cerevisiae*) (Green et al. 2000; Lindsey-Boltz et al. 2001). The Rad17-RFC (i.e. subunits RFC 2, 3, 4 and 5, along with Rad17) is involved in the DNA damage checkpoint response and it loads a novel clamp onto DNA. The novel clamp that Rad17-RFC functions with is a heterotrimer of Rad9/Rad1/Hus1 (reviewed in Sancar et al. 2004). This clamp is often referred to as the "911" clamp. The 911 clamp appears to activate a kinase instead of function with a DNA polymerase (Majka et al. 2006b). Phosphorylation of other proteins by the kinase then signals the cell that DNA damage has occurred thereby communicating that an S phase checkpoint is required. Interestingly, the Rad17-RFC appears to load

the 911 clamp onto the 5' terminus of a primed template junction, the opposite polarity of RFC loading PCNA (Ellison and Stillman 2003; Majka et al. 2006a).

14.8 Conclusions

Clamp loaders are ubiquitous in all cellular life forms. The eukaryotic clamp loader is called RFC. Like other clamp loaders, RFC is a heteropentamer and harnesses the energy of ATP hydrolysis to assemble the ring-shaped PCNA processivity factor onto a primed site. The function of RFC has been aided by biochemical studies and structure determination, and also by the large body of research on the clamp loader of *E. coli*, which is also a heteropentamer like RFC and functions in a similar way. The studies to date answer many important questions, including the subunit organization that specifies how DNA is recognized, where clamps are bound to the clamp loader, and how clamp loading is targeted to a primed template. There still exist many important questions about the details of the clamp loading reaction. Multiple ATP molecules are involved for optimum activity, and whether the individual ATP sites perform individual functions is still unresolved. Also uncertain is the way that clamps are opened. It is not certain that RFC binds and then opens the PCNA clamp, or whether it waits for a PCNA clamp to open and then binds and stabilizes the open form of the clamp. The order of ATP hydrolysis in the individual subunits is also understudied and needs clarification. Despite these deficiencies in our knowledge, the RFC clamp loader remains one of the most understood of the AAA+ machines, which are involved in numerous cellular processes. Future research will enable a better grasp of the mechanism and help fill in our gaps in how AAA+ proteins, and clamp loaders in particular, carry out their function.

References

- Ason B, Handayani R, Williams CR, Bertram JG, Hingorani MM, O'Donnell M, Goodman MF, Bloom LB (2003) Mechanism of loading the *Escherichia coli* DNA polymerase III β sliding clamp on DNA. Bona fide primer/templates preferentially trigger the γ complex to hydrolyze ATP and load the clamp. *J Biol Chem* 278:10033–10040
- Ayyagari R, Impellizzeri KJ, Yoder BL, Gary SL, Burgers PM (1995) A mutational analysis of the yeast proliferating cell nuclear antigen indicates distinct roles in DNA replication and DNA repair. *Mol Cell Biol* 15:4420–4429
- Ayyagari R, Gomes XV, Gordenin DA, Burgers PM (2003) Okazaki fragment maturation in yeast. I. Distribution of functions between FEN1 AND DNA2. *J Biol Chem* 278:1618–1625
- Bellaoui M, Chang M, Ou J, Xu H, Boone C, Brown GW (2003) Elg1 forms an alternative RFC complex important for DNA replication and genome integrity. *EMBO J* 22:4304–4313
- Ben-Aroya S, Koren A, Liefshitz B, Steinlauf R, Kupiec M (2003) *ELG1*, a yeast gene required for genome stability, forms a complex related to replication factor C. *Proc Natl Acad Sci U S A* 100:9906–9911
- Benkovic SJ, Valentine AM, Salinas F (2001) Replisome-mediated DNA replication. *Annu Rev Biochem* 70:181–208

- Bertram JG, Bloom LB, Hingorani MM, Beechem JM, O'Donnell M, Goodman MF (2000) Molecular mechanism and energetics of clamp assembly in *Escherichia coli*. The role of ATP hydrolysis when γ complex loads β on DNA. *J Biol Chem* 275:28413–28420
- Bowman GD, O'Donnell M, Kuriyan J (2004) Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature* 429:724–730
- Bowman GD, Goedken ER, Kazmirski SL, O'Donnell M, Kuriyan J (2005) DNA polymerase clamp loaders and DNA recognition. *FEBS Lett* 579:863–867
- Bunting KA, Roe SM, Pearl LH (2003) Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the β -clamp. *EMBO J* 22:5883–5892
- Bunz F, Kobayashi R, Stillman B (1993) cDNAs encoding the large subunit of human replication factor C. *Proc Natl Acad Sci U S A* 90:11014–11018
- Bylund GO, Burgers PM (2005) Replication protein A-directed unloading of PCNA by the Ctf18 cohesion establishment complex. *Mol Cell Biol* 25:5445–5455
- Cai J, Yao N, Gibbs E, Finkelstein J, Phillips B, O'Donnell M, Hurwitz J (1998) ATP hydrolysis catalyzed by human replication factor C requires participation of multiple subunits. *Proc Natl Acad Sci U S A* 95:11607–11612
- Cullmann G, Fien K, Kobayashi R, Stillman B (1995) Characterization of the five replication factor C genes of *Saccharomyces cerevisiae*. *Mol Cell Biol* 15:4661–4671
- Dalrymple BP, Kongsuwan K, Wijffels G, Dixon NE, Jennings PA (2001) A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proc Natl Acad Sci U S A* 98:11627–11632
- Davey MJ, Jeruzalmi D, Kuriyan J, O'Donnell M (2002) Motors and switches: AAA+ machines within the replisome. *Nat Rev Mol Cell Biol* 3:826–835
- Dore AS, Kilkenny ML, Jones SA, Oliver AW, Roe SM, Bell SD, Pearl LH (2006) Structure of an archaeal PCNA1-PCNA2-FEN1 complex: elucidating PCNA subunit and client enzyme specificity. *Nucleic Acids Res* 34:4515–4526
- Eissenberg JC, Ayyagari R, Gomes XV, Burgers PM (1997) Mutations in yeast proliferating cell nuclear antigen define distinct sites for interaction with DNA polymerase δ and DNA polymerase ϵ . *Mol Cell Biol* 17:6367–6378
- Ellison V, Stillman B (2003) Biochemical characterization of DNA damage checkpoint complexes: clamp loader and clamp complexes with specificity for 5' recessed DNA. *PLoS Biol* 1:E33
- Erzberger JP, Berger JM (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct* 35:93–114
- Fairman M, Prelich G, Tsurimoto T, Stillman B (1988) Identification of cellular components required for SV40 DNA replication *in vitro*. *Biochim Biophys Acta* 951:382–387
- Georgescu RE, Kim SS, Yurieva O, Kuriyan J, Kong XP, O'Donnell M (2008) Structure of a sliding clamp on DNA. *Cell* 132:43–54
- Goedken ER, Kazmirski SL, Bowman GD, O'Donnell M, Kuriyan J (2005) Mapping the interaction of DNA with the *Escherichia coli* DNA polymerase clamp loader complex. *Nat Struct Mol Biol* 12:183–190
- Gomes XV, Burgers PM (2000) Two modes of FEN1 binding to PCNA regulated by DNA. *EMBO J* 19:3811–3821
- Gomes XV, Burgers PM (2001) ATP utilization by yeast replication factor C. I. ATP-mediated interaction with DNA and with proliferating cell nuclear antigen. *J Biol Chem* 276:34768–34775
- Gomes XV, Gary SL, Burgers PM (2000) Overproduction in *Escherichia coli* and characterization of yeast replication factor C lacking the ligase homology domain. *J Biol Chem* 275:14541–14549
- Gomes XV, Schmidt SL, Burgers PM (2001) ATP utilization by yeast replication factor C. II. Multiple stepwise ATP binding events are required to load proliferating cell nuclear antigen onto primed DNA. *J Biol Chem* 276:34776–34783
- Grabowski B, Kelman Z (2003) Archaeal DNA replication: eukaryal proteins in a bacterial context. *Annu Rev Microbiol* 57:487–516

- Green CM, Erdjument-Bromage H, Tempst P, Lowndes NF (2000) A novel Rad24 checkpoint protein complex closely related to replication factor C. *Curr Biol* 10:39–42
- Guenther B, Onrust R, Sali A, O'Donnell M, Kuriyan J (1997) Crystal structure of the δ' subunit of the clamp-loader complex of *E. coli* DNA polymerase III. *Cell* 91:335–345
- Gulbis JM, Kelman Z, Hurwitz J, O'Donnell M, Kuriyan J (1996) Structure of the C-terminal region of p21^{WAF1/CIP1} complexed with human PCNA. *Cell* 87:297–306
- Jeruzalmski D, O'Donnell M, Kuriyan J (2001a) Crystal structure of the processivity clamp loader γ complex of *E. coli* DNA polymerase III. *Cell* 106:429–441
- Jeruzalmski D, Yurieva O, Zhao Y, Young M, Stewart J, Hingorani M, O'Donnell M, Kuriyan J (2001b) Mechanism of processivity clamp opening by the δ subunit wrench of the clamp loader complex of *E. coli* DNA polymerase III. *Cell* 106:417–428
- Johansson E, Garg P, Burgers PM (2004) The Pol32 subunit of DNA polymerase δ contains separable domains for processive replication and proliferating cell nuclear antigen (PCNA) binding. *J Biol Chem* 279:1907–1915
- Johnson A, O'Donnell M (2003) Ordered ATP hydrolysis in the γ complex clamp loader AAA+ machine. *J Biol Chem* 278:14406–14413
- Johnson A, O'Donnell M (2005) Cellular DNA replicases: components and dynamics at the replication fork. *Annu Rev Biochem* 74:283–315
- Johnson A, Yao NY, Bowman GD, Kuriyan J, O'Donnell M (2006) The replication factor C clamp loader requires arginine finger sensors to drive DNA binding and proliferating cell nuclear antigen loading. *J Biol Chem* 281:35531–35543
- Jonsson ZO, Hindges R, Hubscher U (1998) Regulation of DNA replication and repair proteins through interaction with the front side of proliferating cell nuclear antigen. *EMBO J* 17:2412–2425
- Kanellis P, Agyei R, Durocher D (2003) Elg1 forms an alternative PCNA-interacting RFC complex required to maintain genome stability. *Curr Biol* 13:1583–1595
- Kao HI, Bambara RA (2003) The protein components and mechanism of eukaryotic Okazaki fragment maturation. *Crit Rev Biochem Mol Biol* 38:433–452
- Kazmirski SL, Zhao Y, Bowman GD, O'Donnell M, Kuriyan J (2005) Out-of-plane motions in open sliding clamps: molecular dynamics simulations of eukaryotic and archaeal proliferating cell nuclear antigen. *Proc Natl Acad Sci U S A* 102:13801–13806
- Kobayashi M, Ab E, Bonvin AM, Siegal G (2010) Structure of the DNA-bound BRCA1 C-terminal region from human replication factor C p140 and model of the protein-DNA complex. *J Biol Chem* 285:10087–10097
- Kong XP, Onrust R, O'Donnell M, Kuriyan J (1992) Three-dimensional structure of the β subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell* 69:425–437
- Kornberg A, Baker TA (1992) DNA replication. W. H. Freeman, New York
- Krishna TS, Kong XP, Gary S, Burgers PM, Kuriyan J (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* 79:1233–1243
- Langston LD, O'Donnell M (2008) DNA polymerase δ is highly processive with proliferating cell nuclear antigen and undergoes collision release upon completing DNA. *J Biol Chem* 283:29522–29531
- Laurence TA, Kwon Y, Johnson A, Hollars CW, O'Donnell M, Camarero JA, Barsky D (2008) Motion of a DNA sliding clamp observed by single molecule fluorescence spectroscopy. *J Biol Chem* 283:22895–22906
- Lee SH, Eki T, Hurwitz J (1989) Synthesis of DNA containing the simian virus 40 origin of replication by the combined action of DNA polymerases α and δ . *Proc Natl Acad Sci U S A* 86:7361–7365
- Leu FP, Hingorani MM, Turner J, O'Donnell M (2000) The δ subunit of DNA polymerase III holoenzyme serves as a sliding clamp unloader in *Escherichia coli*. *J Biol Chem* 275:34609–34618
- Lindsey-Boltz LA, Bermudez VP, Hurwitz J, Sancar A (2001) Purification and characterization of human DNA damage checkpoint Rad complexes. *Proc Natl Acad Sci U S A* 98:11236–11241

- Majka J, Binz SK, Wold MS, Burgers PM (2006a) Replication protein A directs loading of the DNA damage checkpoint clamp to 5'-DNA junctions. *J Biol Chem* 281:27855–27861
- Majka J, Niedziela-Majka A, Burgers PM (2006b) The checkpoint clamp activates Mec1 kinase during initiation of the DNA damage checkpoint. *Mol Cell* 24:891–901
- Matsumiya S, Ishino Y, Morikawa K (2001) Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from *Pyrococcus furiosus*. *Protein Sci* 10:17–23
- Mayer ML, Gygi SP, Aebersold R, Hieter P (2001) Identification of RFC^{Ctf18p, Ctf8p, Dcc1p}: an alternative RFC complex required for sister chromatid cohesion in *S. cerevisiae*. *Mol Cell* 7:959–970
- McNally R, Bowman GD, Goedken ER, O'Donnell M, Kuriyan J (2010) Analysis of the role of PCNA-DNA contacts during clamp loading. *BMC Struct Biol* 10:3
- Miyata T, Suzuki H, Oyama T, Mayanagi K, Ishino Y, Morikawa K (2005) Open clamp structure in the clamp-loading complex visualized by electron microscopic image analysis. *Proc Natl Acad Sci U S A* 102:13795–13800
- Moarefi I, Jeruzalmi D, Turner J, O'Donnell M, Kuriyan J (2000) Crystal structure of the DNA polymerase processivity factor of T4 bacteriophage. *J Mol Biol* 296:1215–1223
- Naktinis V, Turner J, O'Donnell M (1996) A molecular switch in a replication machine defined by an internal competition for protein rings. *Cell* 84:137–145
- Neuwald AF, Aravind L, Spouge JL, Koonin EV (1999) AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* 9:27–43
- O'Donnell M, Kuriyan J (2006) Clamp loaders and replication initiation. *Curr Opin Struct Biol* 16:35–41
- O'Donnell M, Onrust R, Dean FB, Chen M, Hurwitz J (1993) Homology in accessory proteins of replicative polymerases: *E. coli* to humans. *Nucleic Acids Res* 21:1–3
- Podust VN, Tiwari N, Ott R, Fanning E (1998) Functional interactions among the subunits of replication factor C potentiate and modulate its ATPase activity. *J Biol Chem* 273:12935–12942
- Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem* 73:39–85
- Schmidt SL, Gomes XV, Burgers PM (2001) ATP utilization by yeast replication factor C. III. The ATP-binding domains of Rfc2, Rfc3, and Rfc4 are essential for DNA recognition and clamp loading. *J Biol Chem* 276:34784–34791
- Shamoo Y, Steitz TA (1999) Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* 99:155–166
- Simonetta KR, Kazmirski SL, Goedken ER, Cantor AJ, Kelch BA, McNally R, Seyedin SN, Makino DL, O'Donnell M, Kuriyan J (2009) The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. *Cell* 137:659–671
- Sinha NK, Morris CF, Alberts BM (1980) Efficient *in vitro* replication of double-stranded DNA templates by a purified T4 bacteriophage replication system. *J Biol Chem* 255:4290–4293
- Snyder AK, Williams CR, Johnson A, O'Donnell M, Bloom LB (2004) Mechanism of loading the *Escherichia coli* DNA polymerase III sliding clamp: II. Uncoupling the beta and DNA binding activities of the γ complex. *J Biol Chem* 279:4386–4393
- Song W, Pascal JM, Ellenberger T, Tomkinson AE (2009) The DNA binding domain of human DNA ligase I interacts with both nicked DNA and the DNA sliding clamps, PCNA and hRad9-hRad1-hHus1. *DNA Repair* 8:912–919
- Stewart J, Hingorani MM, Kelman Z, O'Donnell M (2001) Mechanism of β clamp opening by the δ subunit of *Escherichia coli* DNA polymerase III holoenzyme. *J Biol Chem* 276:19182–19189
- Stukenberg PT, Studwell-Vaughan PS, O'Donnell M (1991) Mechanism of the sliding β -clamp of DNA polymerase III holoenzyme. *J Biol Chem* 266:11328–11334
- Turner J, Hingorani MM, Kelman Z, O'Donnell M (1999) The internal workings of a DNA polymerase clamp-loading machine. *EMBO J* 18:771–783

- Uhlmann F, Cai J, Gibbs E, O'Donnell M, Hurwitz J (1997) Deletion analysis of the large subunit p140 in human replication factor C reveals regions required for complex formation and replication activities. *J Biol Chem* 272:10058–10064
- Waga S, Stillman B (1994) Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication *in vitro*. *Nature* 369:207–212
- Warbrick E (2000) The puzzle of PCNA's many partners. *Bioessays* 22:997–1006
- Wijffels G, Dalrymple BP, Prosselkov P, Kongsuwan K, Epa VC, Lilley PE, Jergic S, Buchardt J, Brown SE, Alewood PF, Jennings PA, Dixon NE (2004) Inhibition of protein interactions with the β_2 sliding clamp of *Escherichia coli* DNA polymerase III by peptides from β_2 -binding proteins. *Biochemistry* 43:5661–5671
- Williams CR, Snyder AK, Kuzmic P, O'Donnell M, Bloom LB (2004) Mechanism of loading the *Escherichia coli* DNA polymerase III sliding clamp: I. Two distinct activities for individual ATP sites in the γ complex. *J Biol Chem* 279:4376–4385
- Yao N, Turner J, Kelman Z, Stukenberg PT, Dean F, Shechter D, Pan ZQ, Hurwitz J, O'Donnell M (1996) Clamp loading, unloading and intrinsic stability of the PCNA, β and gp45 sliding clamps of human, *E. coli* and T4 replicases. *Genes Cells* 1:101–113
- Yao N, Coryell L, Zhang D, Georgescu RE, Finkelstein J, Coman MM, Hingorani MM, O'Donnell M (2003) Replication factor C clamp loader subunit arrangement within the circular pentamer and its attachment points to proliferating cell nuclear antigen. *J Biol Chem* 278:50744–50753
- Yao NY, Johnson A, Bowman GD, Kuriyan J, O'Donnell M (2006) Mechanism of proliferating cell nuclear antigen clamp opening by replication factor C. *J Biol Chem* 281:17528–17539
- Zhuang Z, Yoder BL, Burgers PM, Benkovic SJ (2006) The structure of a ring-opened proliferating cell nuclear antigen-replication factor C complex revealed by fluorescence energy transfer. *Proc Natl Acad Sci U S A* 103:2546–2551

Chapter 15

PCNA Structure and Function: Insights from Structures of PCNA Complexes and Post-translationally Modified PCNA

Lynne M. Dieckman, Bret D. Freudenthal, and M. Todd Washington

Abstract Proliferating cell nuclear antigen (PCNA), the eukaryotic DNA sliding clamp, forms a ring-shaped homo-trimer that encircles double-stranded DNA. This protein is best known for its ability to confer high processivity to replicative DNA polymerases. However, it does far more than this, because it forms a mobile platform on the DNA that recruits many of the proteins involved in DNA replication, repair, and recombination to replication forks. X-ray crystal structures of PCNA bound to PCNA-binding proteins have provided insights into how PCNA recognizes its binding partners and recruits them to replication forks. More recently, X-ray crystal structures of ubiquitin-modified and SUMO-modified PCNA have provided insights into how these post-translational modifications alter the specificity of PCNA for some of its binding partners. This article focuses on the insights gained from structural studies of PCNA complexes and post-translationally modified PCNA.

Keywords DNA replication • DNA repair • DNA polymerase • Processivity factor • Protein-protein interactions • Sliding clamp

L.M. Dieckman • B.D. Freudenthal
Department of Biochemistry, University of Iowa College of Medicine,
Iowa City, IA 52242-1109, USA

M.T. Washington (✉)
Department of Biochemistry, University of Iowa College of Medicine,
4-403 Bowen Science Building, Iowa City, IA 52242-1109, USA
e-mail: todd-washington@uiowa.edu

15.1 Introduction

DNA sliding clamp proteins are found in all three domains of life. Despite little sequence homology among the sliding clamps from bacteria, archaea, and eukaryotes, these proteins have similar overall structures. They all form ring-shaped proteins that encircle double-stranded DNA. These sliding clamps are most widely known for their ability to confer high processivity to classical DNA polymerases – those involved in normal DNA replication and repair. Sliding clamps, however, do far more than this; they form mobile platforms on the DNA that recruit many of the enzymes involved in DNA replication, repair, and recombination.

In eukaryotes, the sliding clamp protein is proliferating cell nuclear antigen (PCNA). In a recent review article on PCNA, this protein was called the “maestro of the replication fork” (Moldovan et al. 2007). This is indeed an apt metaphor because PCNA coordinates the recruitment of many proteins to sites of DNA replication and in many cases regulates their activities. In this capacity, PCNA plays a critical role in a wide range of nuclear processes including DNA replication, translesion DNA synthesis, base excision repair, nucleotide excision repair, mismatch repair, recombination, chromatin assembly and remodeling, sister chromatid cohesion, and cell cycle control (Maga and Hubscher 2003; Moldovan et al. 2007; Naryzhny 2008; Tsurimoto 1999; Zhuang and Ai 2010).

The X-ray crystal structure of eukaryotic PCNA was first determined in 1994 (Krishna et al. 1994). Over the last 18 years, various X-ray crystal structures of PCNA bound to peptides derived from PCNA-binding proteins have been determined. More recently, X-ray structures and lower-resolution structures of PCNA bound to full-length protein partners have been determined. These structures have provided valuable insights into how PCNA recognizes PCNA binding proteins and recruits them to replication forks.

A paradigm that has emerged over the last decade is that the specificity of PCNA for some of its binding partners is regulated by post-translational modifications of PCNA (Bergink and Jentsch 2009; Shaheen et al. 2010; Ulrich 2009; Ulrich and Walden 2010; Watts 2006). For example, mono-ubiquitylation of PCNA facilitates translesion synthesis by recruiting non-classical DNA polymerases to stalled replication forks. SUMOylation of PCNA inhibits unwanted recombination by recruiting anti-recombinogenic helicases to replication forks. Recently, X-ray crystal structures of ubiquitin-modified and SUMO-modified PCNA have been determined, and these have provided insights into how these modifications alter the specificity of PCNA for some of its binding partners (Freudenthal et al. 2010, 2011).

This review article focuses on the insights gained in recent years from structural studies of PCNA, PCNA complexes and post-translationally modified PCNA. In particular, we will discuss how PCNA recognizes many PCNA-interacting proteins. We will discuss how these structures begin to help us understand how PCNA regulates the activity of some of these proteins and how PCNA facilitates multi-step enzymatic processes on DNA. Finally, we will discuss how ubiquitin and SUMO modifications impact the function of PCNA.

15.2 Structure of PCNA

Sliding clamps from bacteria, archaea and eukaryotes all form ring-shaped proteins with pseudo-sixfold symmetry. However, there are differences in the number of domains that comprise each subunit and the way that the subunits assemble to form the ring. For example, the bacterial sliding clamp, called the β clamp, is a component of the DNA polymerase III holoenzyme. It is a ring-shaped homo-dimer with each subunit containing three domains (Kong et al. 1992). By contrast, PCNA, the eukaryotic sliding clamp, is a ring-shaped homo-trimer with each subunit containing two domains (Krishna et al. 1994). The archaeal sliding clamp is also called PCNA. Like eukaryotic PCNA, it is a ring-shaped trimer with each subunit containing two domains. However, in some archaeal species, PCNA is a homo-trimer and in others it is a hetero-trimer.

The X-ray crystal structure of eukaryotic PCNA shows that each subunit consists of two independent and similarly folded domains (Fig. 15.1) (Krishna et al. 1994). The N-terminal domain (residues 1–117) is referred to as domain A, and the C-terminal domain (residues 135–258) is referred to as domain B. These independent domains are held together by an extended β sheet across the interdomain boundary on each subunit. Furthermore, the two domains are connected through a long, flexible linker (residues 118–134) called the interdomain connector loop (IDCL). The three subunits assemble in a head-to-tail manner with domain A of one subunit interacting with domain B on an adjacent subunit. This interaction is stabilized through an extended β sheet comprised of β strands from domain A of one subunit and β strands from domain B of an adjacent subunit at each subunit interface.

The PCNA ring has a diameter of approximately 80 Å. The central hole in the ring has a diameter of approximately 35 Å. The outer surface of the PCNA ring is a circular collar of the aforementioned six β sheets (three interdomain β sheets and three intersubunit β sheets). The inner surface of the PCNA ring is a set of 12 α helices, two from each domain. While the overall electrostatic potential of PCNA is negative, the inner surface is positively charged due to the presence of lysine and arginine residues on these α helices. These localized positive charges facilitate the passage of the negatively charged DNA through the central hole.

The PCNA ring is approximately 30 Å wide and contains distinct front and back faces. The front face of PCNA contains the IDCL and is involved in many protein-protein interactions (see Sect. 15.3 below). This is notable as many replication proteins, such as DNA polymerases and DNA ligases, carry out their operations on the DNA at the front face of the PCNA ring. The role of the back face of PCNA is currently less clear. The back face is emerging as a site of PCNA post-translational modification and is likely involved in recruiting protein factors to replication forks and holding them in reserve until they are needed on the front face of PCNA (see Sect. 15.5 below).

X-ray crystal structures of the bacterial β clamp and of eukaryotic PCNA bound to DNA show that as the DNA passes through the central hole of the ring, it is tilted significantly away from the axis of symmetry (Georgescu et al. 2008). The angle of

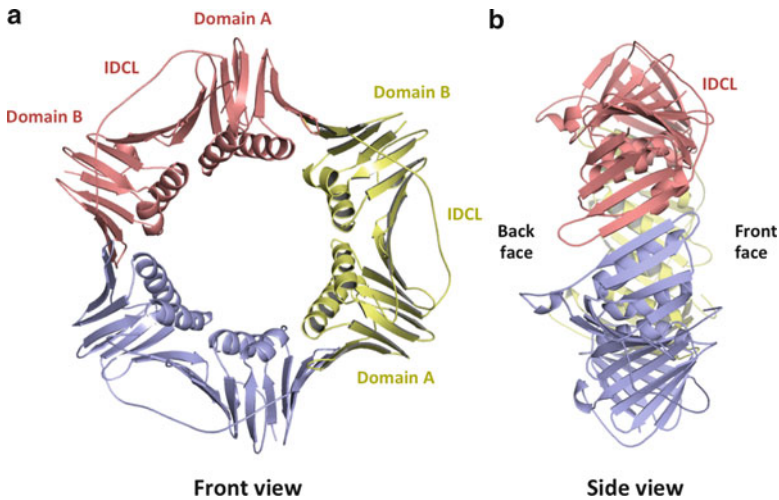


Fig. 15.1 The structure of PCNA. Ribbon diagram of the PCNA trimer (PDB ID: 1PLQ) shown from the *front* (a) and the *side* (b) with the individual PCNA subunits colored *red*, *yellow* and *blue*. The inter-domain connector loop (IDCL) is indicated

the DNA is 22° in the case of the β clamp and 40° in the case of the PCNA. In addition, single-particle electron microscopy analysis of other PCNA-containing complexes also shows that the DNA is tilted as it passes through the central hole of the PCNA ring (see 15.3.3 below) (Mayanagi et al. 2009, 2011). Moreover, single-molecule studies have shown that PCNA can diffuse along the DNA in two distinct modes (Kochaniak et al. 2009). The first mode involves rotation and translation as it tracks the helical pitch of the DNA duplex. The second mode, which is less common, involves faster translation that does not involve tracking the helical pitch. This angular, rotational, and translational flexibility of PCNA on DNA may allow it to accommodate the many diverse proteins with which it must interact.

15.3 Structures of PCNA Complexes

PCNA provides a structural platform for many cellular processes including DNA replication and repair. To do this, PCNA must interact with many of the enzymes involved in these processes. Structural studies of PCNA bound to several of its binding partners have been carried out and these have provided valuable insights into how PCNA interacts with these proteins. We will first discuss X-ray crystal structures of PCNA bound to peptides derived from a variety of PCNA-interacting proteins. We will then discuss X-ray crystal structures of PCNA bound to full-length PCNA-interacting proteins. Finally, we will discuss the architecture of other PCNA-containing complexes determined by single-particle electron microscopy and small angle X-ray scattering (SAXS).

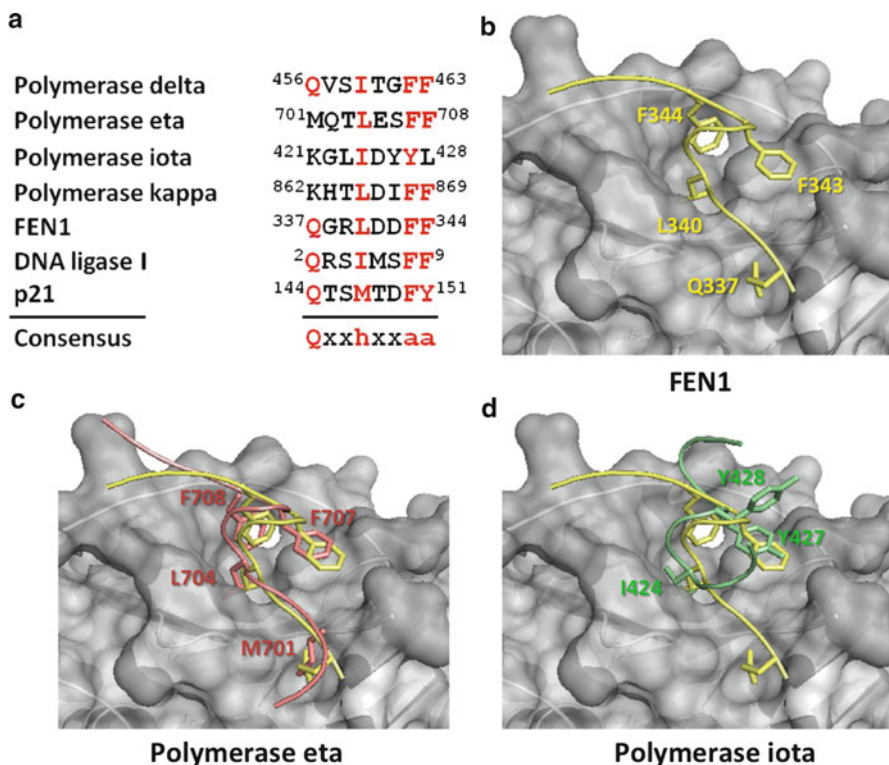


Fig. 15.2 Structures of PCNA bound to PIP peptides. (a) Sequence alignment of PIP peptides from several human PCNA-binding proteins. In the PIP consensus sequence, the ‘h’ can be isoleucine, leucine or methionine, and the ‘a’ can be phenylalanine or tyrosine. (b) The structure of the canonical PIP motif from FEN1 binding to PCNA (PDB ID: 1U7B) is shown in *yellow*. (c) The structure of the PIP motif from DNA polymerase η bound to PCNA (PDB ID: 2ZVK) shown in *red* overlaid with the structure of the PIP motif from FEN1 shown in *yellow*. (d) The structure of the PIP motif from DNA polymerase ι bound to PCNA (PDB ID: 2ZVM) shown in *green* overlaid with the structure of the PIP motif from FEN1 shown in *yellow*

15.3.1 Structures of PCNA Bound to PIP Peptides

Many proteins that bind PCNA do so through a conserved PCNA-interacting protein (PIP) motif (Hingorani and O’Donnell 2000; Maga and Hubscher 2003; Tsurimoto 1999). The PIP motifs of several proteins are shown in Fig. 15.2a. These motifs usually interact with PCNA on a single subunit in a region between the two domains near the IDCL. The canonical PIP motif contains eight amino acid residues. The conserved glutamine of the PIP motif normally inserts into a small pocket in PCNA (Fig. 15.2b). The last five residues of the PIP motif, which include the conserved hydrophobic residue (methionine, leucine, or isoleucine) and the two conserved phenylalanine or tyrosine residues, form a 3_{10} helix that binds in a large hydrophobic pocket between the two domains and also contacts the IDCL. Generally,

the structure of PCNA is not changed upon the binding of PIP peptides; only small alterations in the structure of the IDCL are observed.

PIP motifs are often thought to be a flexible tether that anchors the PCNA-binding protein to PCNA. PIP motifs are often found at the C-termini of PCNA binding proteins, such as classical DNA polymerase δ (the p66 subunit), non-classical DNA polymerase η , and the cyclin-dependent kinase inhibitor p21. PIP motifs, however, can occur elsewhere in the primary structure of the PCNA-binding proteins, including the N-termini (such as DNA ligase I) and the interiors of the proteins (such as non-classical DNA polymerase ι). Deletion of the PIP motif or mutations in its conserved residues can significantly weaken or abolish PCNA interactions *in vivo* and *in vitro*. Thus, even though the PCNA-PIP interactions involve rather small regions of these proteins, these interactions are often necessary to recruit many enzymes to replication forks.

Classical DNA polymerases are responsible for synthesizing DNA during DNA replication and DNA repair. They achieve high processivity by interacting with PCNA, and this interaction is dependent on their PIP motifs. DNA polymerase δ is the classical polymerase that is responsible for lagging strand synthesis in eukaryotes (see Chap. 13, this volume). In humans, DNA polymerase δ is composed of four subunits (p125, p66, p50, and p12). The catalytic activity resides in the p125 subunit. DNA polymerase δ interacts with PCNA via the PIP motif on the p66 subunit. The X-ray crystal structure of PCNA bound to the PIP peptide of p66 shows that the PIP motif forms the normal 3_{10} helix that fits into the large hydrophobic pocket of PCNA (Bruning and Shamoo 2004).

Upon encountering DNA damage in the template strand, the replication fork stalls. This is because classical DNA polymerases are unable to incorporate nucleotides across from damaged DNA templates. Non-classical DNA polymerases, such as DNA polymerases η , κ , and ι , are recruited to stalled replication forks to carry out translesion synthesis (Prakash et al. 2005; Prakash and Prakash 2002; Washington et al. 2009). The recruitment of these non-classical DNA polymerases is governed in part by the mono-ubiquitylation of PCNA; this aspect of non-classical polymerase recruitment will be described later (see Sect. 15.5 below). Nevertheless, the PIP motifs of these non-classical polymerases are necessary for their recruitment to stalled replication forks.

The X-ray crystal structures of PCNA bound to the PIP motifs of DNA polymerases η , κ , and ι have been determined (Hishiki et al. 2009). The structures of the PIP motifs of DNA polymerases η and κ are similar to that of the classical DNA polymerase δ in that they form the normal 3_{10} helix (Fig. 15.2c). There are, however, some minor differences in the specific contacts made by these PIP motifs, because the sequences of the PIP motifs of these non-classical polymerases differ slightly from the PIP consensus sequence. For example, neither of these PIP motifs have the conserved glutamine residue. DNA polymerase η , for instance, has a methionine residue that inserts into the small pocket where the glutamine normally fits. The structure of the PIP motif of DNA polymerase ι , however, differs significantly from that of any other PIP motif structure. It does not form the normal 3_{10} helix, but instead forms a β -bend-like structure (Fig. 15.2d). Taken together, it is likely that

the divergence of the non-classical polymerase PIP motifs from the consensus PIP sequence reduces their affinities for PCNA relative to other PIP motifs (Hishiki et al. 2009). This could be important for preventing the recruitment of non-classical polymerases to replication forks until the PCNA is mono-ubiquitylated and their activities are needed.

In the X-ray crystal structures of PCNA bound to some PIP peptides, secondary contacts (i.e., those that occur outside of the PIP motif) are observed between PCNA and the portions of the peptide flanking the PIP motif. For example, DNA ligases catalyze the linkage of 5' phosphates and a 3'OH groups during DNA repair and Okazaki fragment processing (see Chap. 17, this volume). The yeast Cdc9 DNA ligase has a PIP motif that forms the conventional 3_{10} helix. However, the residues flanking the N-terminal sides of the PIP motif form an anti-parallel β -sheet with the C-terminus of PCNA (Vijayakumar et al. 2007).

The presence of DNA damage triggers an increase in expression of the tumor suppressor protein p21 leading to DNA replication arrest. The inhibition of DNA replication by p21 requires that it bind directly to PCNA (Flores-Rozas et al. 1994; Gibbs et al. 1997; Waga et al. 1994). The X-ray crystal structure of the p21 PIP motif bound to PCNA reveals that this PIP motif binds in the normal manner. However, secondary contacts between PCNA and the peptide in the regions immediately flanking the PIP motif are observed. The N-terminal and the C-terminal flanking regions form anti-parallel β -sheets with the C-terminus and the IDCL of PCNA, respectively (Gulbis et al. 1996). It has been suggested that these extensive interactions are responsible for the higher affinity PIP motif-PCNA interaction observed with the p21 PIP motif relative to other PIP motifs. This tighter binding may allow the p21 PIP to inhibit DNA replication by effectively competing with DNA polymerases for binding PCNA.

15.3.2 Structures of PCNA Bound to Full-Length Proteins

While most structures of PCNA have been of complexes of PCNA with PIP motif peptides, a few structures have been determined of complexes of PCNA with full-length proteins. These have provided insights into the secondary contacts between PCNA and PCNA-binding proteins that occur in addition to and alongside the contacts mediated by PIP motifs. For example, the X-ray crystal structure of PCNA bound to full-length flap endonuclease 1 (FEN1), which catalyzes the removal of 5' single-stranded DNA overhangs that occur during DNA repair and during the processing of the ends of Okazaki fragments (see Chap. 16, this volume), has been determined (Fig. 15.3a) (Sakurai et al. 2005). FEN1 consists of a nuclease core domain (residues 1–332) and a C-terminal tail region (333–380). The main PCNA-interacting interface of FEN1 is the N-terminal half of the C-terminal tail region, which contains a PIP motif.

Although the primary contact made between FEN1 and PCNA is mediated by the PIP motif, there are secondary contacts between PCNA and the regions flanking

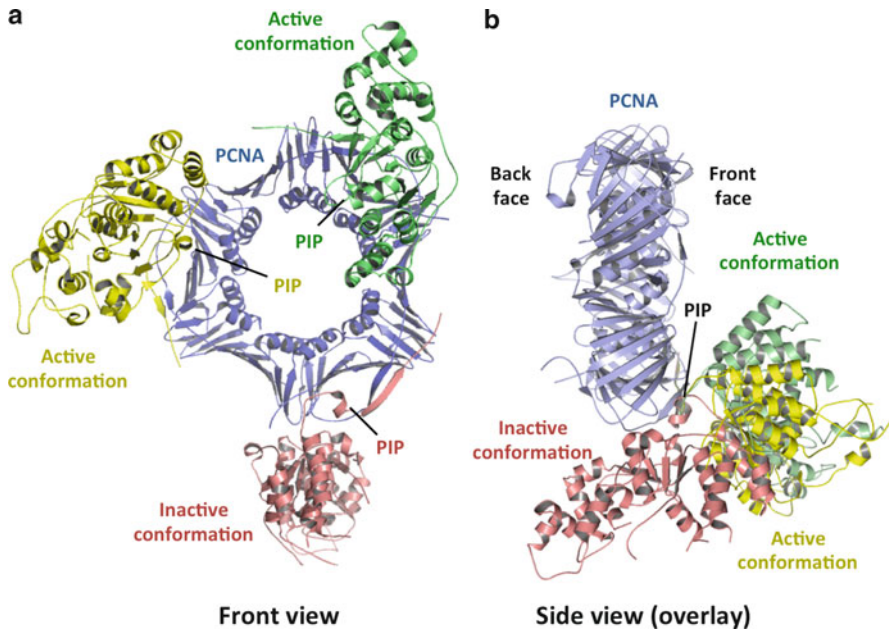


Fig. 15.3 Structure of PCNA bound to FEN1. (a) Ribbon diagram of the PCNA trimer shown in *blue* bound to three molecules of FEN1 shown in *red*, *yellow* and *green* (PDB ID: 1UL1). (b) Overlay showing the three positions of FEN1 relative to the PCNA subunit to which they are bound. The PCNA is shown in *blue*, the inactive conformation is shown in *red*, and the active conformations are shown in *yellow* and *green*

the PIP motif and between PCNA and the core domain of FEN1. Residues of the core domain make several intramolecular contacts with the PIP motif as well as several intermolecular interactions with both the IDCL and C-terminus of PCNA. Moreover, the core domain of FEN1 is connected to its C-terminal tail through a 4-residue linker. It has been suggested that this linker acts as a hinge to allow the core domain of FEN1 to be positioned near its DNA substrate.

The structure of the FEN1-PCNA complex had three FEN1 molecules bound to PCNA, and each FEN1 molecule was in a different position relative to the PCNA subunit to which it was bound (Fig. 15.3b). One of the observed FEN1 positions had the active site of the core domain swung away from the front face of PCNA, and this may represent an inactive conformation of FEN1. In the other two positions, the core domain is located closer to the PCNA central cavity near the expected position of the DNA. These latter positions may reflect active conformations in which FEN1 can bind the DNA flap and bring itself into a position to cleave it.

Replication factor C (RFC) is the ATP-dependent clamp loading protein that binds to PCNA, opens the ring, and deposits it on the DNA. The structure and mechanism of RFC is described in more detail in a companion chapter (Chap. 14, this volume). Here, however, we will briefly mention the key features observed in the X-ray crystal structure of the RFC-PCNA complex (Bowman et al. 2004). RFC

sits on the front face of the closed PCNA ring. The five subunits of RFC form a right-handed spiral that is tilted by approximately 9° relative to the threefold axis of PCNA. Only three of the five subunits of RFC (RFC-A, RFC-B, and RFC-C) make contacts with the PCNA. In the case of RFC-A and RFC-C, these are contacts mediated by PIP motifs. RFC-B, by contrast, makes several secondary contacts with PCNA at the intersubunit regions.

15.3.3 Low Resolution Structures of PCNA Complexes

Lower resolution approaches, such as small angle X-ray scattering (SAXS) and single particle electron microscopy (EM), have been used to examine the architecture of other PCNA complexes. Although much of this work has been done using archaeal PCNA from either *Sulfolobus solfataricus* or *Pyrococcus furiosus*, these studies have uncovered principles about PCNA complexes that are likely applicable to eukaryotic systems. Like eukaryotic PCNA, *P. furiosus* PCNA is a homotrimer. Each subunit has a similar overall fold to eukaryotic PCNA including a protein binding pocket near the IDCL (Matsumiya et al. 2001). Thus *P. furiosus* PCNA trimers have three identical protein binding sites. Unlike eukaryotic PCNA, *S. solfataricus* PCNA is a heterotrimer comprised of three subunits: PCNA1, PCNA2, and PCNA3. These three subunits share the same overall fold with one another and with eukaryotic PCNA (Williams et al. 2006). Consequently, *S. solfataricus* PCNA trimers have three distinct protein binding sites.

The architecture of the *S. solfataricus* PCNA bound to DNA ligase in the absence of DNA was examined using SAXS (Pascal et al. 2006). The ligase and PCNA trimer form a 1:1 complex with the ligase binding to the PCNA3 subunit. *Ab initio* shape predictions suggested that DNA ligase has a preferred orientation with respect to the PCNA ring and is extended out from the side of the ring. Structures of PCNA and DNA ligase obtained from X-ray crystallography were docked into the SAXS molecular envelope showing that the DNA ligase was in the open conformation. It is suggested that the interface between the DNA ligase and the PCNA is malleable enough to accommodate the conformational change in the DNA ligase from the open state to the closed state that is needed for catalysis when DNA is present.

Insight into the architecture of PCNA-DNA ligase complex in the presence of DNA came from single particle EM studies of *P. furiosus* PCNA and DNA ligase (Mayanagia et al. 2009). The 3D map, with a resolution estimated to be 15 Å, revealed a two-tier structure. The lower tier was a hexagonal ring into which the structure of PCNA nicely fits. The upper tier was crescent-shaped and corresponded well to the structures of the domains of the DNA ligase. The DNA was visible in the 3D map as a rod-shaped component that went through the center of the PCNA ring. The DNA ligase wrapped half way around the DNA. In this complex, the DNA was tilted about 16° from the threefold axis of the PCNA ring.

Single particle EM studies were also used to examine the structure of the complex of *P. furiosus* PCNA and DNA polymerase B bound to DNA (Mayanagi et al. 2011).

Again, the 3D map, with a resolution estimated to be 19 Å, was a two-tier structure. The lower layer corresponded to PCNA, and the upper layer corresponded to DNA polymerase B. The DNA was visible through the central channel of PCNA and was tilted about 13° from the threefold axis of PCNA. Interestingly, the DNA polymerase directly contacted the PCNA trimer at two sites. One contact site was the normal interaction mediated through the PIP motif of the polymerase. The other contact site was with a different PCNA subunit than the one contacted by the PIP motif. It has been suggested that this secondary contact helps to properly orient the polymerase, which is difficult to do with only a PIP-mediated contact as this latter contact is rather flexible. This secondary contact may also preclude other proteins from binding at this other subunit.

The Msh2-Msh6 protein recognizes DNA mismatches and initiates mismatch repair. The architecture of the complex of eukaryotic PCNA and the Msh2-Msh6 mismatch repair protein was analyzed by SAXS (Shell et al. 2007). First, the N-terminal region of Msh6, which contains a PIP motif and binds tightly to PCNA, was shown by SAXS to be intrinsically disordered. Upon binding to PCNA, the N-terminal region does not acquire structure suggesting that this region functions as a disordered tether. SAXS analysis was also performed on the full Msh2-Msh6 protein bound to PCNA, and these results did not favor a model in which the folded regions of the Msh2-Msh6 protein directly contacted the PCNA ring. Instead they suggested that the interaction is solely mediated through the long, unstructured tether. This tether likely allows the structured regions of Msh2-Msh6 to remain associated with PCNA, but also reach around other protein factors at the replication fork in search of mismatches. It is likely that this type of PCNA interaction is common, because other proteins containing PIP motifs also have adjacent regions predicted to be intrinsically unstructured (Shell et al. 2007).

15.3.4 *Unresolved Issues*

PCNA interacts with a variety of proteins. How does PCNA discriminate between these different partners? How does PCNA regulate when a protein should be recruited to a replication fork or released from a replication fork? Because most PIP motifs make very similar contacts with PCNA, they are unlikely to contribute much toward this specificity. Notable exceptions include the non-classical DNA polymerases and p21. The PIP motifs of non-classical polymerases are thought to bind PCNA with lower affinity than those of classical DNA polymerases (Hishiki et al. 2009). This could be important for preventing non-classical polymerases from binding PCNA until PCNA is mono-ubiquitylated. By contrast, the PIP motif of p21 binds PIP with higher affinity than those of classical polymerases (Bruning and Shamo 2004). This could be important for arresting DNA replication when there is DNA damage.

In most cases, the specificity of PCNA for its binding partners probably comes from contacts outside the PIP motif. Secondary contacts between PCNA and PCNA-binding proteins involving regions that immediately flank the PIP motif or elsewhere on the PCNA-binding protein likely play a major role in specificity. This emphasizes the

need for additional X-ray crystal structures of PCNA bound to full-length proteins. Moreover, contact with the DNA may also play an important role in specificity in cells. For example, specificity for FEN1 likely arises in part due to the secondary contacts observed in the X-ray crystal structure and in part due to the presence of a DNA substrate containing a 5' flap. Moreover, post-translational modifications of PCNA such as mono-ubiquitylation and SUMOylation clearly control the specificity of PCNA interactions (see Sect. 15.5 below). Similarly, post-translational modifications of PCNA-binding proteins such as phosphorylation have been observed with p21 and FEN1, and these modifications inhibit PCNA binding (Henneke et al. 2003; Scott et al. 2000). Further studies will be needed to flesh out some of these mechanisms and uncover yet others.

Many of the processes in which PCNA participates are multi-step processes that involve the handing off of the DNA from one enzyme to another. For example, in Okazaki fragment processing, FEN1 must cleave off the 5' flap on the DNA before handing it off to DNA ligase that seals the nick (see Chap. 16, this volume). How does such a DNA handoff occur? One possibility is that PCNA forms toolbelts by simultaneously binding several different enzymes and that these toolbelts facilitate the handoff. For example, the simultaneous binding of DNA polymerase, FEN1, and DNA ligase to a single PCNA trimer has been observed in *S. solfataricus* (Dionne et al. 2003). Currently, there is no clear evidence for eukaryotic PCNA functioning as a toolbelt but this seems to be a very likely scenario.

15.4 Structures of Mutant PCNA Proteins

A variety of PCNA mutant proteins have been identified that increase the sensitivity of cells to DNA damaging agents (Ayyagari et al. 1995). Here we will focus on two mutant proteins that block translesion synthesis. The first of these mutant proteins, which was identified in a yeast genetic screen, has a glycine to serine substitution at residue 178 (Zhang et al. 2006). Yeast cells producing this G178S mutant form of PCNA have an increased sensitivity to DNA-damaging agents and are completely defective in translesion synthesis. Interestingly, this mutant form of PCNA functions normally in all other respects, such as DNA replication and repair. The second of these mutant proteins, which was identified in another yeast genetic screen, has a glutamate to glycine substitution at residue 113 (Amin and Holm 1996). Yeast cells producing this E113G mutant protein have a very similar phenotype to those with the aforementioned G178S substitution.

Gly178 is located in domain B at the subunit interface of the PCNA trimer. Glu-113 is located in domain A at the subunit interface directly across from Gly-178 on the adjacent subunit (Fig. 15.4a). Steady state kinetic studies show that while wild-type PCNA stimulates incorporation by the non-classical DNA polymerase η opposite an abasic site, the G178S PCNA protein actually inhibits incorporation opposite this DNA lesion (Freudenthal et al. 2008). Similarly, the E113G PCNA mutant protein is unable to stimulate incorporation by DNA polymerase η opposite

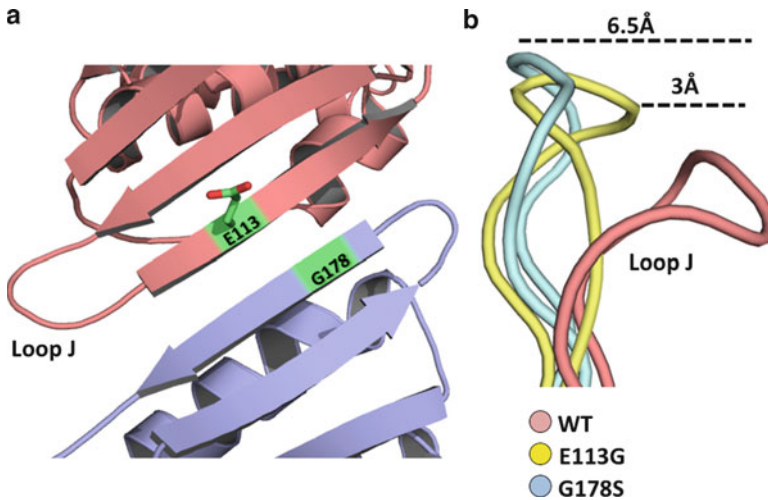


Fig. 15.4 Structures of the G178S and E113G PCNA mutant proteins. (a) A ribbon diagram of the subunit interface of PCNA with domain A of one subunit shown in *red* and domain B of the adjacent subunit shown in *blue*. The position of Glu113, the position of Gly178 and loop J are indicated. (b) The structure of backbone of loop J in wild-type PCNA protein shown in *red* (PDB ID: 1PLQ) is superimposed on the structures of the backbones of loop J in the E113G PCNA mutant protein shown in *yellow* (PDB ID: 3GPM) and the G178S PCNA mutant protein shown in *blue* (PDB ID: 3F1W)

this lesion (Freudenthal et al. 2008). Furthermore, the E113G PCNA mutant protein is unable to stimulate the activity of non-classical DNA polymerase ζ involved in translesion synthesis (Northam et al. 2006). The E113G PCNA mutant protein, however, is capable of being mono-ubiquitylated on Lys164 (Northam et al. 2006), which is required for translesion synthesis *in vivo* (see Sect. 15.5 below). This suggests that the inability of these PCNA mutant proteins to support translesion synthesis is independent of their mono-ubiquitylation.

X-ray crystal structures of the G178S and E113G PCNA mutant proteins have provided insight into how these substitutions disrupt translesion synthesis by non-classical DNA polymerases (Freudenthal et al. 2008, 2009). The G178S PCNA mutant protein has little effect on the structure of domain B, which is the domain in which the amino acid substitution occurs. Instead, a significant, local structural change occurs in domain A of the adjacent subunit. This difference between the G178S PCNA mutant protein and the wild type PCNA structures is limited to a single, extended loop (residues 105–110), which is called loop J. In the mutant protein structure, loop J adopts a very different conformation in which the protein backbone has moved by as much as 6.5 Å from its position in the wild type structure (Fig. 15.4b) (Freudenthal et al. 2008). The E113G mutant protein structure has a similar, but somewhat smaller (only about 3 Å) shift in loop J (Freudenthal et al. 2008). These structures suggest a key role for loop J in facilitating translesion synthesis by non-classical polymerases, perhaps as a novel site of a secondary contact between the polymerases and PCNA.

15.5 Structures of Post-translationally Modified PCNA

The recruitment of proteins to sites of replication via interactions with PCNA is regulated in some cases by post-translational modifications of PCNA. For example, PCNA is mono-ubiquitylated on Lys164 by Rad6 (an ubiquitin-conjugating enzyme) and Rad18 (an ubiquitin ligase) in a DNA damage-dependent manner (Hoegge et al. 2002; Stelter and Ulrich 2003). The mono-ubiquitylation of PCNA is required for translesion synthesis by non-classical DNA polymerases. Several of these non-classical polymerases contain ubiquitin-binding motifs (Bienko et al. 2005) and the switch between the classical and non-classical DNA polymerases only occurs when PCNA is mono-ubiquitylated (Zhuang et al. 2008).

The mono-ubiquitin on Lys164 can be converted to Lys63-linked poly-ubiquitin chains by the Mms2-Ubc13 complex (an ubiquitin conjugating enzyme) and Rad5 (an ubiquitin ligase) (Hoegge et al. 2002). The poly-ubiquitylation of PCNA is required for an error-free damage bypass pathway that is currently poorly understood. It has been suggested that this pathway involves switching of the stalled replicative polymerase from the damaged template to the newly synthesized sister strand.

In addition to mono-ubiquitylation and poly-ubiquitylation, PCNA is also subject to SUMOylation on Lys164 by Ubc9 (a SUMO conjugating enzyme) and Siz1 (a SUMO ligase) (Hoegge et al. 2002). PCNA SUMOylation inhibits unwanted recombination by recruiting the anti-recombinogenic Srs2 helicase (Papouli et al. 2005; Pfander et al. 2005) which contains a SUMO-binding motif. The Srs2 helicase then disrupts the Rad51 nucleoprotein filaments needed to carry out the strand exchange reaction (Krejci et al. 2004; Veaute et al. 2003). SUMOylation has also been observed to a lesser extent on Lys127 on the IDCL but the biological implications of this SUMOylation are unclear.

Our understanding of the structural and mechanistic basis of the recruitment of these factors to post-translationally modified PCNA has come in part from recently determined X-ray crystal structures of ubiquitin-modified and SUMO-modified PCNA (Freudenthal et al. 2010, 2011). In the sections that follow we discuss these structures and their implications.

15.5.1 Structure of Ubiquitin-Modified PCNA

Obtaining sufficient quantities of ubiquitin-modified PCNA for X-ray crystallography had been an obstacle for years. A breakthrough came when it was shown that large amounts of ubiquitin-modified PCNA could be created by splitting the PCNA protein into two polypeptides at residue 164 (the site of ubiquitylation) (Freudenthal et al. 2010). These two polypeptides self-assemble *in vivo*. This allowed the ubiquitin to be fused in-frame to the C-terminal portion of the split PCNA generating a split ubiquitylated PCNA analog that supported UV resistance *in vivo* and translesion synthesis *in vitro* (Freudenthal et al. 2010). This analog allowed for the determination of the X-ray crystal structure of ubiquitin-modified PCNA.

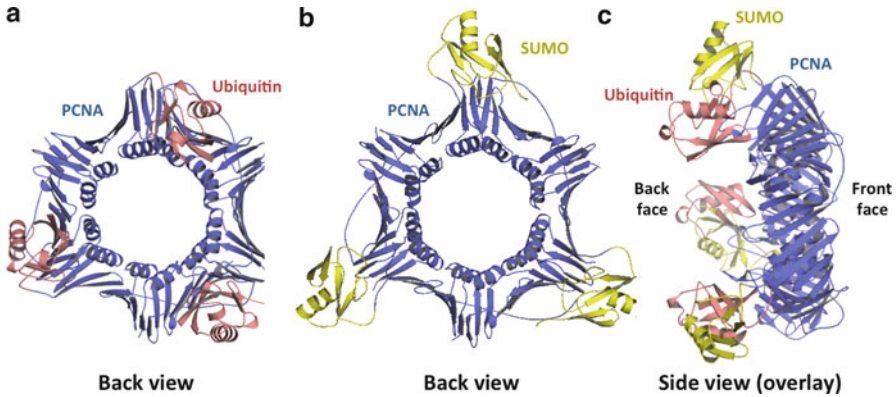


Fig. 15.5 Structures of ubiquitin-modified and SUMO-modified PCNA. **(a)** Ribbon diagram of the ubiquitin-modified PCNA trimer (PDB ID: 3L10) shown from the back with the PCNA ring shown in *blue* and the ubiquitin moieties shown in *red*. **(b)** Ribbon diagram of the SUMO-modified PCNA trimer (PDB ID: 3PGE) shown from the back with the PCNA ring shown in *blue* and the SUMO moieties shown in *yellow*. **(c)** Overlay of the structures of ubiquitin-modified PCNA and SUMO-modified PCNA shown from the side

The structure of ubiquitin-modified PCNA shows that the ubiquitin moiety occupies a position on the back face of the PCNA ring (Fig. 15.5a) (Freudenthal et al. 2010). It interacts primarily with a long loop on the back of PCNA called loop P (residues 184–195). Moreover, the attachment of ubiquitin to PCNA does not alter the conformation of the PCNA in any significant way. This suggests that the ubiquitin moiety does not act as an allosteric modifier to increase the affinity of PCNA for the non-classical polymerase. Instead, it argues for a simpler model in which the ubiquitin moiety provides an additional binding surface to which the non-classical polymerases can attach.

The position of the ubiquitin moiety on the back face of PCNA is consistent with a variation of the typical toolbelt model. In the typical toolbelt model, different PCNA-binding partners interact with different subunits on the front face of the PCNA ring. In the case of PCNA ubiquitylation, PCNA should be able to interact with other protein factors such as the classical DNA polymerase on its front face, while at the same time binding non-classical polymerases on its back face. Here the non-classical polymerase can be held in reserve until needed without interfering with on-going activity on the front face of the PCNA ring. While there is not yet experimental evidence for such a toolbelt in eukaryotes, there is convincing experimental evidence for the analogous toolbelt in prokaryotes. The classical DNA polymerase III and the non-classical DNA polymerase IV have been shown to simultaneously bind the β sliding clamp (Indiani et al. 2005).

The residues of ubiquitin that interact with the non-classical polymerases have been mapped by NMR spectroscopy (Bomar et al. 2007). In the case of DNA polymerase η , this interaction is mediated by the same hydrophobic residues (Leu7, Ile44, and Val70) that ubiquitin uses to interact with a wide range of other proteins. It turns out that these residues of ubiquitin are buried at the ubiquitin-PCNA interface (Freudenthal et al. 2010). This means that the conformation of

ubiquitin-modified PCNA observed in this structure is not the conformation to which the non-classical polymerases bind. This implies that the ubiquitin moiety must either be capable of re-orienting itself on the back face of the PCNA ring or be capable of moving around to occupy other positions on the PCNA ring including possibly the side of the ring. Such alternative conformations would be necessary to recruit the non-classical polymerase.

15.5.2 Structure of SUMO-Modified PCNA

Once it was shown that one could obtain sufficient quantities of ubiquitin-modified PCNA for structural studies using the split/fusion strategy, large quantities of SUMO-modified PCNA were produced using the same approach (Freudenthal et al. 2011). The X-ray crystal structure of SUMO-modified PCNA was then determined. In this structure, the SUMO was also found to be on the back face of the PCNA ring interacting predominantly with loop P of PCNA (Fig. 15.5b) (Freudenthal et al. 2011). Interestingly, the SUMO occupied a different, more radial, position on the PCNA relative to the position occupied by ubiquitin in the ubiquitin-modified PCNA structure (Fig. 15.5c).

The attachment of SUMO does not change the structure of PCNA suggesting that allosteric models for the recruitment of the anti-recombinogenic Srs2 helicase are unlikely. Instead, the SUMO moiety likely provides an additional binding surface to which Srs2 can bind. Moreover, the finding of the SUMO moiety on the back face of the PCNA ring also argues for a toolbelt model. In such a model, Srs2 could be recruited to the back face of PCNA where it can be held in reserve until needed.

15.5.3 Unresolved Issues

Understanding precisely how ubiquitin and SUMO modifications regulate the recruitment of non-classical polymerases and anti-recombinogenic helicases to replication forks requires that we learn more about the dynamics of these modified PCNA proteins. What other conformations do ubiquitin-modified PCNA or SUMO-modified PCNA adopt? Which of these conformations recruit the polymerase or helicase to the replication fork? Which of these conformations supports the enzymatic activity of the polymerase or helicase? Answering these questions will require further structural studies of the modified PCNA proteins as well as structures of the modified PCNA proteins bound to target proteins. It is likely that a combination of high resolution approaches such as X-ray crystallography and low resolution approaches such as single-particle EM and SAXS will be required.

Despite an intense effort, it remains unclear whether one subunit or all three subunits of PCNA are modified in cells. This raises the possibility that PCNA trimers may have some subunits mono-ubiquitylated and other subunits SUMOylated at the same time. Do the ubiquitin and SUMO modifications work together? Does the

SUMO modification inhibit recombination and therefore allow the ubiquitin modification time to promote translesion synthesis? The ability to produce constitutively ubiquitylated or constitutively SUMOylated PCNA in cells using the split/fusion strategy may help answer these questions.

Finally, the details of the error-free damage bypass pathway promoted by Lys63-linked poly-ubiquitylation of PCNA are still unclear. What protein factors are recruited to replication forks when PCNA is poly-ubiquitylated? How do these proteins allow replication to proceed past the DNA damage? These questions are important because, unlike translesion synthesis, this damage bypass pathway does not contribute to genome instability.

15.6 Concluding Remarks

All of the structural studies discussed here have contributed to our understanding of the complex and dynamic processes in which PCNA participates. They have revealed how PCNA binds PIP motifs from various PCNA-binding proteins, and they have suggested mechanisms by which these interactions may be regulated. Nevertheless, many unanswered questions remain regarding the regulation of these processes. One important issue deals with how PCNA recognizes specific binding partners in certain circumstances. Does this involve secondary contacts between PCNA and its binding partner outside the PIP motif? Does this involve post-translational modifications of either PCNA or its binding partner? Some of the structural studies discussed here represent the first steps toward addressing these issues.

Another important issue deals with how the DNA substrate is handed off from one PCNA binding partner to another. Does this involve the sequential binding of these enzymes or does PCNA form specific toolbelts among the three protein-binding sites on the front face of the PCNA ring to facilitate this handoff? Do ubiquitin and SUMO modifications also allow PCNA to function as a toolbelt by opening up new binding sites on the back face of the PCNA ring? The answers to these questions await further biochemical and structural studies of complexes of unmodified and modified PCNA.

Acknowledgements This article was supported by Award Number GM081433 from the National Institute of General Medical Sciences to M.T.W. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. We thank Christine Kondratik, John Pryor and Marc Wold for valuable discussions.

References

- Amin NS, Holm C (1996) *In vivo* analysis reveals that the interdomain region of the yeast proliferating cell nuclear antigen is important for DNA replication and DNA repair. *Genetics* 144:479–493

- Ayyagari R, Impellizzeri KJ, Yoder BL, Gary SL, Burgers PMJ (1995) A mutational analysis of the yeast proliferating cell nuclear antigen indicates distinct roles in DNA replication and DNA repair. *Mol Cell Biol* 15:4420–4429
- Bergink S, Jentsch S (2009) Principles of ubiquitin and SUMO modifications in DNA repair. *Nature* 458:461–467
- Bienko M, Green CM, Crosetto N, Rudolf F, Zapart G, Coull B, Kannouche P, Wider G, Peter M, Lehmann AR, Hofmann K, Dikic I (2005) Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science* 310:1821–1824
- Bomar MG, Pai MT, Tzeng SR, Li SSC, Zhou P (2007) Structure of the ubiquitin-binding zinc finger domain of human DNA γ -polymerase η . *EMBO Rep* 8:247–251
- Bowman GD, O'Donnell M, Kuriyan J (2004) Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature* 429:724–730
- Bruning JB, Shamoo Y (2004) Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase δ p66 subunit and flap endonuclease-1. *Structure* 12:2209–2219
- Dionne I, Nookala RK, Jackson SP, Doherty AJ, Bell SD (2003) A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Cell* 11:275–282
- Flores-Rozas H, Kelman Z, Dean FB, Pan ZQ, Harper PW, Elledge SJ, O'Donnell M, Hurwitz J (1994) Cdk-interacting protein 1 directly binds with proliferating cell nuclear antigen and inhibits DNA replication catalyzed by the DNA polymerase δ holoenzyme. *Proc Natl Acad Sci USA* 91:8655–8659
- Freudenthal BD, Ramaswamy S, Hingorani MM, Washington MT (2008) Structure of a mutant form of proliferating cell nuclear antigen that blocks translesion DNA synthesis. *Biochemistry* 47:13354–13361
- Freudenthal BD, Gakhar L, Ramaswamy S, Washington MT (2009) A charged residue at the subunit interface of PCNA promotes trimer formation by destabilizing alternate subunit interactions. *Acta Crystallogr D Biol Crystallogr* 65:560–566
- Freudenthal BD, Gakhar L, Ramaswamy S, Washington MT (2010) Structure of monoubiquitinated PCNA and implications for translesion synthesis and DNA polymerase exchange. *Nat Struct Mol Biol* 17:479–484
- Freudenthal BD, Brogie JE, Gakhar L, Kondratick CM, Washington MT (2011) Crystal structure of SUMO-modified proliferating cell nuclear antigen. *J Mol Biol* 406:9–17
- Georgescu RE, Kim SS, Yurieva O, Kuriyan J, Kong XP, O'Donnell M (2008) Structure of a sliding clamp on DNA. *Cell* 132:43–54
- Gibbs E, Kelman Z, Gulbis JM, O'Donnell M, Kuriyan J, Burgers PMJ, Hurwitz J (1997) The influence of the proliferating cell nuclear antigen-interacting domain of p21^{CIP1} on DNA synthesis catalyzed by the human and *Saccharomyces cerevisiae* polymerase δ holoenzymes. *J Biol Chem* 272:2373–2381
- Gulbis JM, Kelman Z, Hurwitz J, O'Donnell M, Kuriyan J (1996) Structure of the C-terminal region of p21^{WAF1/CIP1} complexed with human PCNA. *Cell* 87:297–306
- Henneke G, Koundrioukoff S, Hubscher U (2003) Phosphorylation of human Fen1 by cyclin-dependent kinase modulates its role in replication fork regulation. *Oncogene* 22:4301–4313
- Hingorani MM, O'Donnell M (2000) Sliding clamps: a (tail)ored fit. *Curr Biol* 10:R25–R29
- Hishiki A, Hashimoto H, Hanafusa T, Kamei K, Ohashi E, Shimizu T, Ohmori H, Sato M (2009) Structural basis for novel interactions between human translesion synthesis polymerases and proliferating cell nuclear antigen. *J Biol Chem* 284:10552–10560
- Hoeg C, Pfander B, Moldovan GL, Pyrowolakis G, Jentsch S (2002) *RAD6*-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* 419:135–141
- Indiani C, McInerney P, Georgescu R, Goodman MF, O'Donnell M (2005) A sliding-clamp tool-belt binds high- and low-fidelity DNA polymerases simultaneously. *Mol Cell* 19:805–815
- Kochaniak AB, Habuchi S, Loparo JJ, Chang DJ, Cimprich KA, Walter JC, van Oijen AM (2009) Proliferating cell nuclear antigen uses two distinct modes to move along DNA. *J Biol Chem* 284:17700–17710
- Kong XP, Onrust R, O'Donnell M, Kuriyan J (1992) Three-dimensional structure of the β -subunit of *Escherichia coli* DNA polymerase III holoenzyme – a sliding DNA clamp. *Cell* 69:425–437

- Krejci L, Macris M, Li Y, Van Komen S, Villemain J, Ellenberger T, Klein H, Sung P (2004) Role of ATP hydrolysis in the antirecombinase function of *Saccharomyces cerevisiae* Srs2 protein. *J Biol Chem* 279:23193–23199
- Krishna TSR, Kong XP, Gary S, Burgers PM, Kuriyan J (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* 79:1233–1243
- Maga G, Hubscher U (2003) Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci* 116:3051–3060
- Matsumiya S, Ishino Y, Morikawa K (2001) Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from *Pyrococcus furiosus*. *Protein Sci* 10:17–23
- Mayanagia K, Kiyonari S, Saito M, Shirai T, Ishino Y, Morikawa K (2009) Mechanism of replication machinery assembly as revealed by the DNA ligase-PCNA-DNA complex architecture. *Proc Natl Acad Sci USA* 106:4647–4652
- Mayanagi K, Kiyonari S, Nishida H, Saito M, Kohda D, Ishino Y, Shirai T, Morikawa K (2011) Architecture of the DNA polymerase B-proliferating cell nuclear antigen (PCNA)-DNA ternary complex. *Proc Natl Acad Sci USA* 108:1845–1849
- Moldovan GL, Pfander B, Jentsch S (2007) PCNA, the maestro of the replication fork. *Cell* 129:665–679
- Naryzhny SN (2008) Proliferating cell nuclear antigen: a proteomics view. *Cell Mol Life Sci* 65:3789–3808
- Northam MR, Garg P, Baitin DM, Burgers PMJ, Shcherbakova PV (2006) A novel function of DNA polymerase ζ regulated by PCNA. *EMBO J* 25:4316–4325
- Papouli E, Chen SH, Davies AA, Huttner D, Krejci L, Sung P, Ulrich HD (2005) Crosstalk between SUMO and ubiquitin on PCNA is mediated by recruitment of the helicase Srs2p. *Mol Cell* 19:123–133
- Pascal JM, Tsodikov OV, Hura GL, Song W, Cotner EA, Classen S, Tomkinson AE, Tainer JA, Ellenberger T (2006) A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. *Mol Cell* 24:279–291
- Pfander B, Moldovan GL, Sacher M, Hoegge C, Jentsch S (2005) SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature* 436:428–433
- Prakash S, Prakash L (2002) Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. *Genes Dev* 16:1872–1883
- Prakash S, Johnson RE, Prakash L (2005) Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu Rev Biochem* 74:317–353
- Sakurai S, Kitano K, Yamaguchi H, Hamada K, Okada K, Fukuda K, Uchida M, Ohtsuka E, Morioka H, Hakoshima T (2005) Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J* 24:683–693
- Scott MT, Morrice N, Ball KL (2000) Reversible phosphorylation at the C-terminal regulatory domain of p21^{Waf1/Cip1} modulates proliferating cell nuclear antigen binding. *J Biol Chem* 275:11529–11537
- Shaheen M, Shanmugam I, Hromas R (2010) The role of PCNA post-translational modifications in translesion synthesis. *J Nucl Acids* 2010:761217
- Shell SS, Putnam CD, Kolodner RD (2007) The N-terminus of *Saccharomyces cerevisiae* Msh6 is an unstructured tether to PCNA. *Mol Cell* 26:565–578
- Stelter P, Ulrich HD (2003) Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. *Nature* 425:188–191
- Tsurimoto T (1999) PCNA binding proteins. *Front Biosci* 4:d849–d858
- Ulrich HD (2009) Regulating post-translational modifications of the eukaryotic replication clamp PCNA. *DNA Repair* 8:461–469
- Ulrich HD, Walden H (2010) Ubiquitin signalling in DNA replication and repair. *Nat Rev Mol Cell Biol* 11:479–489
- Veaute X, Jeusset J, Soustelle C, Kowalczykowski SC, Le Cam E, Fabre F (2003) The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* 423:309–312
- Vijayakumar S, Chapados BR, Schmidt KH, Kolodner RD, Tainer JA, Tomkinson AE (2007) The C-terminal domain of yeast PCNA is required for physical and functional interactions with Cdc9 DNA ligase. *Nucleic Acids Res* 35:1624–1637

- Waga S, Hannon GJ, Beach D, Stillman B (1994) The p21 inhibitor of cyclin-dependent kinases controls DNA replication by interaction with PCNA. *Nature* 369:574–578
- Washington MT, Carlson KD, Freudenthal BD, Pryor JM (2009) Variations on a theme: eukaryotic Y-family DNA polymerases. *Biochim Biophys Acta* 1804:1113–1123
- Watts FZ (2006) Sumoylation of PCNA: wrestling with recombination at stalled replication forks. *DNA Repair* 5:399–403
- Williams GJ, Johnson K, Rudolf J, McMahon SA, Carter L, Oke M, Liu HT, Taylor GL, White MF, Naismith JH (2006) Structure of the heterotrimeric PCNA from *Sulfolobus solfataricus*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 62:944–948
- Zhang HS, Gibbs PEM, Lawrence CW (2006) The *Saccharomyces cerevisiae rev6-1* mutation, which inhibits both the lesion bypass and the recombination mode of DNA damage tolerance, is an allele of *POL30*, encoding proliferating cell nuclear antigen. *Genetics* 173:1983–1989
- Zhuang ZH, Ai YX (2010) Processivity factor of DNA polymerase and its expanding role in normal and translesion DNA synthesis. *Biochim Biophys Acta* 1804:1081–1093
- Zhuang ZH, Johnson RE, Haracska L, Prakash L, Prakash S, Benkovic SJ (2008) Regulation of polymerase exchange between Pol η and Pol δ by monoubiquitination of PCNA and the movement of DNA polymerase holoenzyme. *Proc Natl Acad Sci USA* 105:5361–5366

Chapter 16

The Wonders of Flap Endonucleases: Structure, Function, Mechanism and Regulation

L. David Finger, John M. Atack, Susan Tsutakawa, Scott Classen,
John Tainer, Jane Grasby, and Binghui Shen

Abstract Processing of Okazaki fragments to complete lagging-strand DNA synthesis requires coordination among several proteins. RNA primers and DNA synthesised by DNA polymerase α are displaced by DNA polymerase δ to create bifurcated nucleic acid structures known as 5'-flaps. These 5'-flaps are removed by Flap

L.D. Finger (✉) • J.M. Atack • J. Grasby
Department of Chemistry, Centre for Chemical Biology, Krebs Institute,
University of Sheffield, Sheffield S3 7HF, UK
e-mail: d.finger@sheffield.ac.uk; j.m.atack@sheffield.ac.uk; j.a.grasby@sheffield.ac.uk

S. Tsutakawa
Life Sciences Division, Lawrence Berkeley National
Laboratory, Berkeley, CA 94720, USA
e-mail: setsutakawa@lbl.gov

S. Classen
Physical Biosciences Division, The Scripps Research
Institute, La Jolla, CA 92037, USA
e-mail: sclassen@lbl.gov

J. Tainer
Life Sciences Division, Lawrence Berkeley
National Laboratory, Berkeley, CA 94720, USA
Department of Molecular Biology, The Scripps Research
Institute, La Jolla, CA 92037, USA

Skaggs Institute for Chemical Biology,
La Jolla, CA 92037, USA
e-mail: jat@scripps.edu

B. Shen
Division of Radiation Biology, City of Hope National Medical Center
and Beckman Research Institute, Duarte, CA 91010, USA
College of Life Sciences, Zhejiang University,
Hangzhou 310058, China
e-mail: bshen@coh.org

Endonuclease 1 (FEN), a structure-specific nuclease whose divalent metal-ion-dependent phosphodiesterase activity cleaves 5'-flaps with exquisite specificity. FENs are paradigms for the 5' nuclease superfamily, whose members perform a wide variety of roles in nucleic acid metabolism using a similar nuclease core domain that displays common biochemical properties and structural features. A detailed review of FEN structure is undertaken to show how DNA substrate recognition occurs and how FEN achieves cleavage at a single phosphate diester. A proposed double nucleotide unpairing trap (DoNUT) is discussed with regards to FEN and has relevance to the wider 5'-nuclease superfamily. The homotrimeric proliferating cell nuclear antigen protein (PCNA) coordinates the actions of DNA polymerase, FEN and DNA ligase by facilitating the hand-off intermediates between each protein during Okazaki fragment maturation to maximise through-put and minimise consequences of intermediates being released into the wider cellular environment. FEN has numerous partner proteins that modulate and control its action during DNA replication and is also controlled by several post-translational modification events, all acting in concert to maintain precise and appropriate cleavage of Okazaki fragment intermediates during DNA replication.

Keywords Okazaki fragment maturation • Lagging-strand DNA replication • Double nucleotide unpairing • Structure-specific nuclease • Disorder-order transition

16.1 Introduction

Unlike leading-strand DNA replication, lagging-strand DNA is synthesised discontinuously as the replication fork moves in the opposite direction to the polymerase. As the replication fork progresses, newly exposed DNA on the lagging strand is continuously primed by primase/pol α then extended by pol δ with the assistance of PCNA. These segments of DNA on the lagging-strand are known as Okazaki fragments, and it is estimated that human DNA replication generates ~50 million per cell cycle (Burgers 2009). To form a continuous piece of DNA, the RNA primer, and possibly the DNA laid down by pol α , must be removed, with DNA being subsequently ligated to complete Okazaki fragment maturation (Fig. 16.1a). Initial experiments aimed at reconstituting the DNA replication machinery *in vitro* using fractionated nuclear extracts identified a maturation factor (MF1) necessary for the completion of lagging-strand DNA replication (Waga et al. 1994). This maturation factor was later renamed Flap Endonuclease due to the enzyme's preference to cleave bifurcated DNA structures with displaced 5'-single-stranded DNA flaps (Fig. 16.1b–d) (Harrington and Lieber 1994). Flap endonucleases (FENs), which are present across all domains of life, are divalent metal ion-dependent nucleases, whose phosphodiesterase activity enhances the hydrolysis rate of targeted phosphodiester bonds at least $\sim 10^{17}$ fold (Tomlinson et al. 2010). FENs possess a single active site that can perform both endo- and exo-nucleolytic cleavages.

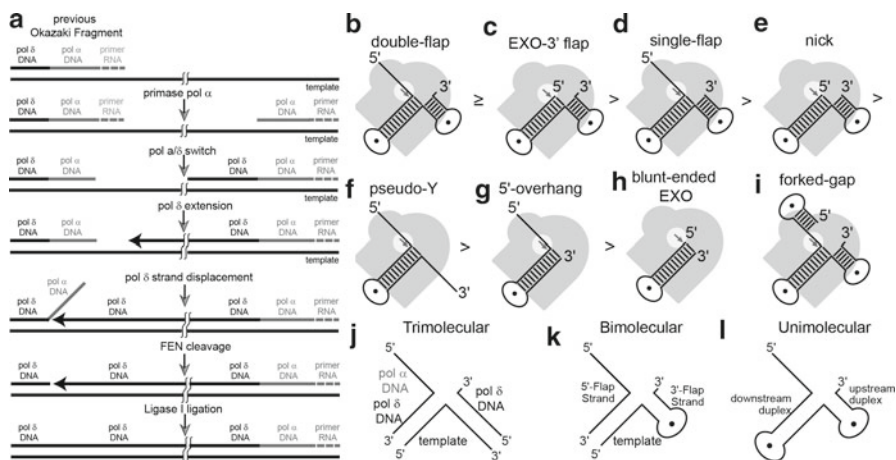


Fig. 16.1 The role of FENs in lagging-strand DNA replication and various FEN substrates *in vitro*. (a) Simplified diagram of the Okazaki fragment maturation (b–i) Various activities that can be observed with FENs and model substrates *in vitro*. The grey image represents the protein with the active site highlighted by the white circle to show how these activities are achieved. (b–h) The strength of each activity is indicated by the order they are placed as illustrated by the symbols between each. (i) Example of a forked-gap substrate. (j–l) Examples of double flap constructs used in biochemical studies. Illustrations are labelled to (j) show how *in vivo* Okazaki fragments correspond to *in vitro* substrates and (k) each region or (l) component of the substrates are referred to as herein

Furthermore, FENs are considered prototypical members of the 5'-nuclease superfamily, which includes enzymes with diverse DNA processing activities such as EXO1, XPG and GEN1 (Finger and Shen 2010; Tomlinson et al. 2010; Grasby et al. 2011), as well as the 5'-exoribonucleases Xrn1 and Rat1 (Solinger et al. 1999; Yang 2010).

Current paradigms of eukaryotic DNA replication are based mainly on studies in yeast (Burgers 2009). FENs were once thought to be solely responsible for cleavage of flaps of any length *in vivo*, but studies identified two additional proteins involved in yeast Okazaki fragment maturation – Pif1 and Dna2 (Budd and Campbell 1997; Budd et al. 2006). Pif1, a member of the IB helicase superfamily (Bochman et al. 2010), has been shown to increase the efficiency of pol δ strand displacement synthesis, thereby resulting in long (>30 nucleotides) 5' ssDNA flaps. Exposure of 30 or more nucleotides of ssDNA recruits Replication Protein A (RPA; ssDNA binding protein), resulting in a 5'-flap-ssDNA-RPA complex that is refractory to FEN1 cleavage. To remove these RPA coated flaps, the RPA-activated Dna2 nuclease/helicase is recruited to imprecisely cleave the long flap, thereby creating a short flap that is then processed by FEN1 (Kang et al. 2010). The relevance of this yeast “two-step” Okazaki fragment maturation model to mammalian systems remains unclear. Nonetheless, both short and long flap pathways require

the ability of FEN to cleave with exquisite precision to create a product that is a substrate for DNA ligase. Primers that are incorrectly processed or not removed by FEN would create gaps or overlaps, respectively, resulting in genomic instability, as seen in studies of budding and fission yeast lacking Rad27/Rad2 (FEN1 homologues in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, respectively) (Johnson et al. 1995; Reagan et al. 1995), which display severe mutator phenotypes (Liu et al. 2004; Navarro et al. 2007).

The importance of FEN is further highlighted in higher eukaryotes, where homozygous deletion of the *fen1* gene (*fen1*^{-/-}) is embryonically lethal in mice (Larsen et al. 2003), indicating that FEN1 is absolutely essential in mammals. FEN1 is expressed in all proliferative tissues in humans, including cancers (Kim et al. 2000; Warbrick et al. 1998). FEN1 expression levels in normal tissue are correlated with proliferative capacity. In addition, FEN1 over-expression in human cancers has been linked to tumour aggressiveness (Finger and Shen 2010); for this reason, the chemotherapeutic potential of FEN1 inhibitors has been investigated (Tumey et al. 2004, 2005). Furthermore, mutations that decrease expression levels or alter FEN1 biochemical properties predispose humans and mice to cancers (Kucherlapati et al. 2002; Larsen et al. 2008; Zheng et al. 2007b). Thus, a paradox of FEN activity emerges: optimal FEN1 activity is essential to prevent cancer, but overabundance or impaired function of FEN1 can promote cancer by increasing the efficiency of DNA replication and repair or reducing fidelity of DNA replication respectively.

An area of controversy regarding FENs is how the enzyme is able to precisely cleave at a single phosphate diester to create a ligatable nick based solely upon the structure instead of the sequence of the DNA. Early studies established that prokaryotic and eukaryotic FENs recognize the structure of bifurcated 5'-flaps rather than sequence (Harrington and Lieber 1995; Lyamichev et al. 1993), but controversy as to how this was accomplished quickly emerged. Dahlberg and co-workers suggested that specificity was achieved by threading the 5'-flap through a hole in the FEN protein. Subsequent biochemical studies on mammalian FENs supported this proposal and suggested that these enzymes thread the 5' flap through a hole in the protein until it encounters the dsDNA, whereupon cleavage occurs (Murante et al. 1995). Structural studies with bacteriophage T5 FEN (T5FEN) revealed a helical archway above the active site whose dimensions could only accommodate ssDNA (Ceska et al. 1996), lending support to threading models. Further biochemical and structural work subsequently suggested instead that the helical arch was actually used to clamp onto 5'-flap ssDNA at the 5'-terminus and to then track along the flap until dsDNA was encountered and subsequently cleaved (Bornarth et al. 1999; Chapados et al. 2004). Alternatively, studies using the *E. coli* FEN homologue led Joyce and colleagues to suggest that FENs initially recognize the dsDNA portion of 5'-flap substrates and then, thread the 5'-flap DNA (Xu et al. 2001). More recent evidence from several groups suggests that the latter model is a better mechanistic description of eukaryotic FENs (Finger et al. 2009; Gloor et al. 2010; Hohl et al. 2007; Stewart et al. 2009). Furthermore, X-ray crystallographic studies of enzyme-substrate and enzyme-product complexes of human FEN1 (hFEN1) have shed light on how FENs identify their substrate and select the scissile phosphate (Tsutakawa et al. 2011). Here, we review

both biochemical and structural aspects of FEN1 that give rise to a structure-specific nuclease with exquisite scissile phosphate diester selectivity, and then, discuss how this protein is assisted and regulated *in vivo* by sub-cellular localization, protein interaction partners, and post-translational modification.

16.2 Biochemical Activity

FENs cleave a large range of substrates *in vitro* with a 5' to 3' polarity both endo- and exo-nucleolytically (Fig. 16.1b–i) (Nazarkina et al. 2008; Shen et al. 2005). The divalent metal ion-dependent phosphodiesterase activity of FENs exclusively generates 5'-phosphate monoester and 3'-hydroxyl products (Pickering et al. 1999). However, the catalytic efficiency on each type of substrate structure varies greatly. In addition to DNA replication, FENs have also been implicated in other DNA metabolic pathways due to the ability to observe activities with certain substrates (Zheng et al. 2011b).

Eukaryotic and archaeal FENs prefer substrates with two dsDNA regions (Fig. 16.1b–e), but substrates having only a single dsDNA region of at least 12 base pairs can be cleaved as well, albeit weakly (Fig. 16.1f–i). Single 5'-flap substrates (Fig. 16.1d) were initially thought to be the preferred substrate of FENs (Harrington and Lieber 1995; Lyamichev et al. 1999), but 5'-flap and EXO substrates also having a single nucleotide 3'-flap (i.e., a double-flap substrate) were later shown to be the preferred substrates for eukaryotic FENs (Fig. 16.1b,c) for three reasons. Substrates bearing 3'-flaps have lower apparent K_D s (Friedrich-Heineken and Hubscher 2004; Kao et al. 2002), are cleaved with greater efficiency than their single-flap cognates, and are cleaved exclusively at the phosphate diester between the first and second nucleotide of the downstream duplex. Importantly, this increased precision results in all dsDNA product being ligatable (Fig. 16.2a) (Finger et al. 2009; Kao et al. 2002). With single-flap substrates, cleavage is less efficient and predominantly occurs at the dsDNA-ssDNA flap junction, but to a lesser extent also one nucleotide into the downstream duplex (Fig. 16.2b) (Finger et al. 2009; Kao et al. 2002). The minor cleavage product results in a 1-nt gap that would require post-replication repair mechanisms to fill in and close the gap (Chapados et al. 2004; Finger and Shen 2010). Lower organism FENs, such as those from the T4 and T5 bacteriophages, do not possess a 3'-flap binding pocket (Friedrich-Heineken and Hubscher 2004; Shen et al. 1998). Thus, FENs from higher organisms have evolved such a feature to ensure that cleavage results in immediately ligatable nicks, avoiding the need for initiation of DNA repair mechanisms.

The substrates used *in vitro* are commonly designed to exclusively form a 5'-flap of a known length, with or without a single nt 3'-flap. These types of flap substrates are referred to as static double-flap and single-flap substrates, respectively (Fig. 16.2c,d), and are prepared either as tri-, bi-, or uni-molecular constructs (Fig. 16.1j). The 'template strand' oligonucleotide corresponds to the lagging-strand template *in vivo*. The strand that base pairs with the template to form the upstream

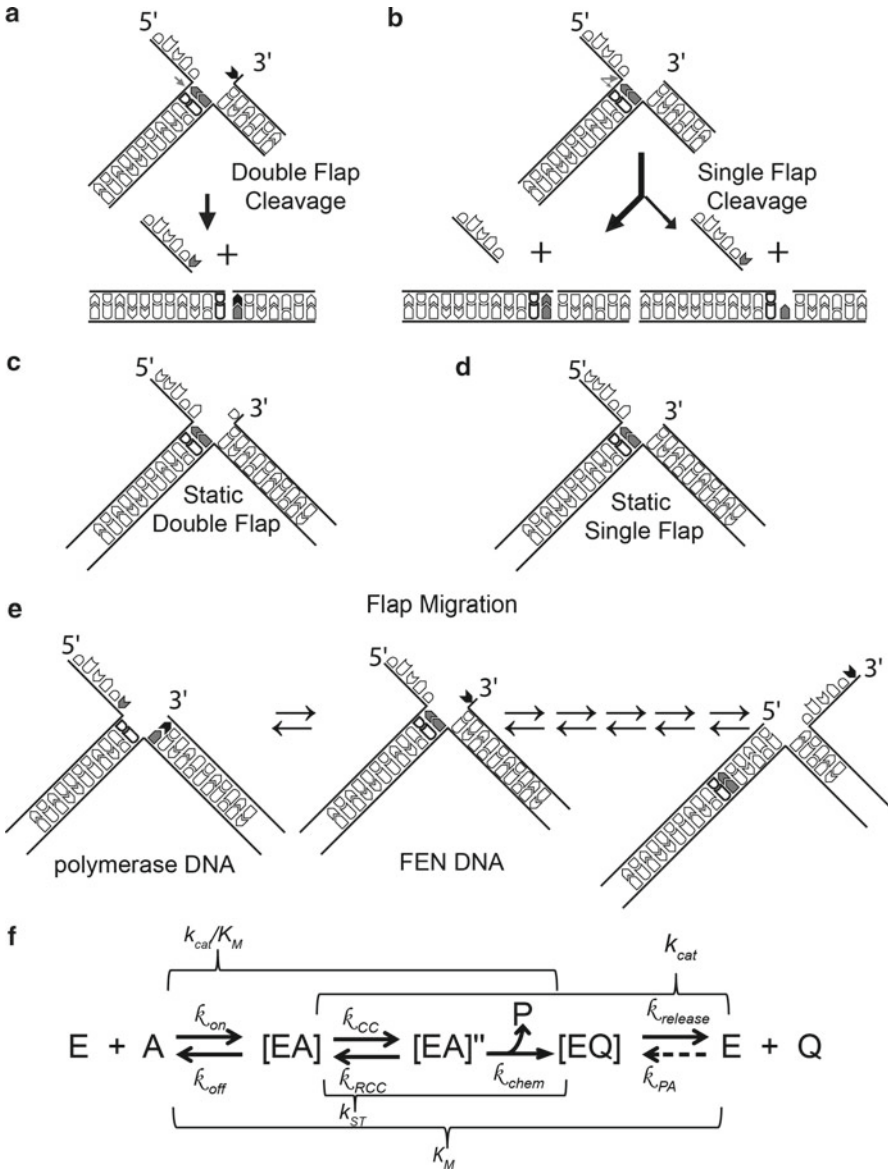


Fig. 16.2 The importance of the 3'-flap *in vitro* and *in vivo*. Reactions on (a) double flaps result in all dsDNA product being a ligatable nick, whereas (b) single flaps produce a minor product that is a 1 nt gap. Examples of static (i.e., single conformation) double- (c) and single- (d) flaps that are commonly used *in vitro*. (e) Flaps generated in replication are potentially migrating flaps (equilibrating flaps) and can theoretically form multiple structures. Those labelled here represent the two conformations important for polymerase and FENs. (f) Diagram of the FEN reaction pathway on double-flap substrates deduced from studies of FENs to date. Note this is only a model, and there may be more steps in the reaction pathway than illustrated. The FEN enzyme (E) binds its substrate (S) to form an enzyme-substrate complex ([ES]). To be able to cleave the substrate, the protein,

duplex region is equivalent to the nascently synthesized pol δ DNA, whereas the strand that anneals to form the downstream dsDNA corresponds to pol α and/or pol δ synthesized DNA of the previous Okazaki fragment depending on the extent of strand displacement synthesis (Fig. 16.1j–l) and assuming that RNaseH has already removed all but the last nucleotide of the RNA primer (Chon et al. 2009; Mesiet-Cladiere et al. 2007; Qiu et al. 1999). *In vivo* pol α and pol δ use the same template, so the sequences they synthesise should be identical. Thus, the 5'-flap structures generated *in vivo* are 'equilibrating' double-flap substrates (Liu et al. 2004); as such, overlapping sequences can form multiple structures of varying 5' and 3' flap lengths by a mechanism analogous to Holliday junction migration (Fig. 16.2e). Model 'equilibrating' double flap structures are cleaved at a single phosphate diester in a manner analogous to the cleavage of a static double-flap substrate (i.e., one nucleotide into the downstream duplex) (Kao et al. 2002). The 3'-flap in the static and equilibrating double-flap substrates corresponds to the last nucleotide added by pol δ during strand displacement synthesis (Figs. 16.1a and 16.2e). Thus, FENs have evolved to recognize the last nucleotide added by pol δ and to identify the scissile phosphate accordingly. In addition to increased specificity, the 3'-flap augments "enzyme commitment" to the forward reaction by increasing first-order rates of reaction after initial enzyme substrate complex formation (Fig. 16.2f) (i.e., $k_{CC} > k_{off}$) (Finger et al. 2009). In fact, the catalytic efficiency of FEN1 on a static double flap substrate approaches enzyme:substrate association rates in solution. Therefore, FEN reactions with static double flap substrates may be diffusion controlled under conditions whereby substrate is limiting ($[E] < [S] < K_M$; k_{cat}/K_M conditions), implying that the enzyme has reached catalytic perfection (Sengerova et al. 2010).

Under saturating multiple turnover (MT) conditions ($[S] >> K_M >> [E]$), the enzyme is rate limited by enzyme product release as single turnover (ST) rates of hydrolysis are faster than MT rates (Fig. 16.2f) (Finger et al. 2009; Williams et al. 2007).

← **Fig. 16.2** (continued) DNA, or both have to change conformation to create a cleavage competent complex ($[ES]''$). Upon cleavage, ssDNA (P) and dsDNA (Q) products are created. The ssDNA (P) likely dissociates from the complex immediately upon cleavage resulting in the enzyme-dsDNA product complex ($[EQ]$). The dissociation of $[EQ]$ results in nicked dsDNA and enzyme turnover. The rates (k) associated with each step are listed above or below the corresponding arrow: k_{on} – bimolecular association (i.e., diffusion), k_{off} – dissociation, k_{CC} – conformational change, k_{RCC} – reverse conformational change, k_{chem} – chemical catalysis, $k_{release}$ – product dissociation, k_{PA} – product association. Note, k_{PA} can be ignored when measuring initial rates of reaction. The macroscopic rate constants commonly measured kinetically are above or below a bracket that encompasses the rates that can influence the measured parameter. Because reactions with FEN have intermediates after initial $[ES]$ formation, the K_M is an overall dissociation constant for all enzyme-bound species ($[ES] + [ES]'' + [EQ]$). The turnover number (k_{cat}) in WT FENs is mainly a reflection of the slowest step (enzyme product release), but can be affected by other first order rates in the reaction pathway. The second order rate constant (k_{cat}/K_M) for WT is mainly a measure of diffusion (k_{on}), but mutants of FENs can sometimes change and represent anyone or some combination of steps within the bracket. The rate measured under single turnover conditions can measure any rates after initial $[ES]$ complex formation and before $[EQ]$ release and is a measure of some physical limitation such as conformational change in WT FENs

On double-flap substrates, FENs produce two products (denoted P and Q), which with a double-flap substrate would be nicked dsDNA product (Q) and a small ssDNA fragment (P). Product inhibition studies have deduced that the ssDNA product (P) is instantaneously released after cleavage or released much faster than the dsDNA product, whereas the dsDNA product (Q) is retained (Finger et al. 2009). Thus, release of the dsDNA product is rate limiting *in vitro* under MT conditions (Fig. 16.2f). The fact that FENs hold onto the dsDNA product is similar to the observation that many DNA metabolic enzymes chaperone their potentially toxic repair intermediates (Parikh et al. 1999). Furthermore, in FEN1-product DNA crystals, only electron density for the hFEN1-dsDNA product (Q) and not for the ssDNA flap (P) was observed, consistent with FEN retaining only the dsDNA product after cleavage (Tsutakawa et al. 2011). Thus, FENs release the 5'-flaps ssDNA product, but retain the dsDNA product, which is the substrate for the next step of Okazaki fragment maturation (i.e., ligation).

Because k_{cat}/K_M conditions are likely diffusion controlled and saturating MT conditions are rate-limited by enzyme product release (Fig. 16.2f), MT measurements of reaction of FENs are probably not physiologically relevant, as these physical limitations likely do not exist in the cell (Berg and von Hippel 1985) due to sequestration of FENs to replication forks (Beattie and Bell 2011) and potential PCNA-mediated handoff (see Chap. 15, this volume) of replication intermediates between processing proteins (i.e., pol δ and Ligase I, respectively) (Chapados et al. 2004; Tsutakawa et al. 2011). A more physiologically relevant measurement of FEN activity is single-turnover experiments where $[S] < K_M < [E]$. These experiments measure the rates of reaction after enzyme-substrate complex formation and before enzyme product release (Fig. 16.2c). Despite knowing the rate of reaction of the rate-limiting step under ST conditions, it is difficult to know exactly what is being measured (e.g., k_{chem} or a physical step in the reaction cycle like k_{CC}/k_{RCC}). So far, studies of T5FEN have shown that a physical limitation (e.g., protein, DNA or concerted protein/DNA conformational change (k_{CC})) is rate limiting under ST conditions (Fig. 16.2f) (Sengerova et al. 2010). This is likely to be the case for eukaryotic FENs *in vitro* as well.

16.3 FEN Structure and Substrate Recognition

16.3.1 Free Protein

Crystal structures of FENs lacking substrate have been solved from a wide variety of organisms (Ceska et al. 1996; Chapados et al. 2004; Hosfield et al. 1998b; Hwang et al. 1998; Mase et al. 2009; Sakurai et al. 2005, 2008). These have revealed a common architecture among FENs. The nuclease core domains of FENs fold into an α/β structure known as a SAM or PIN fold (Fig. 16.3a). Moreover, a mixed twisted β -sheet of (usually) seven-strands is sandwiched between two α -helical

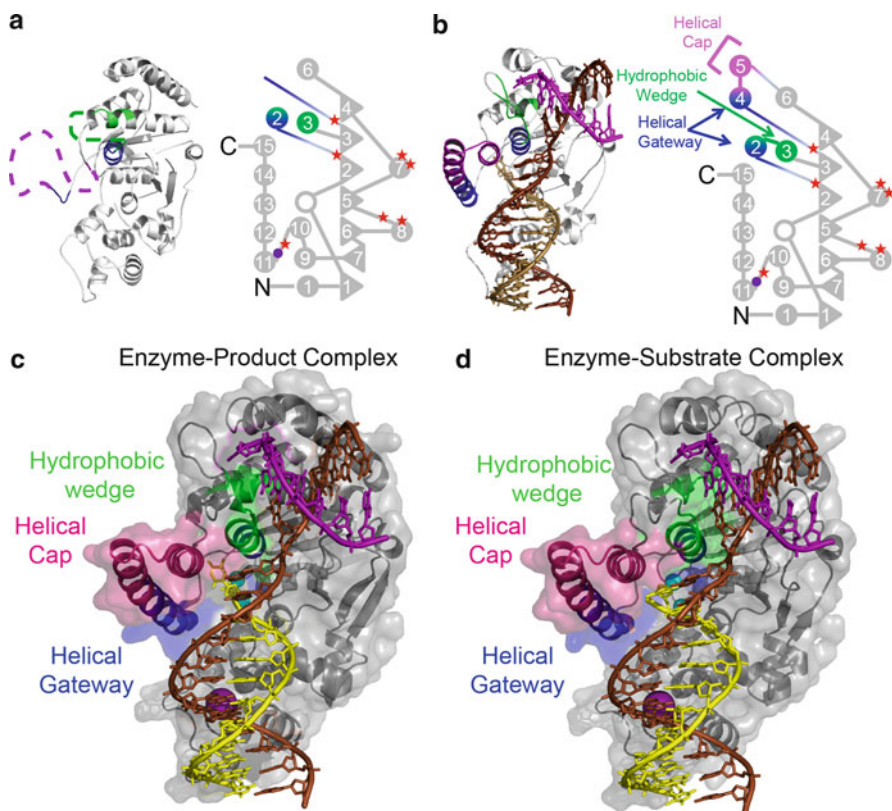


Fig. 16.3 Comparison of DNA-free hFEN1 and DNA-bound hFEN1. Structures of (a) hFEN1 without DNA and (b) hFEN1-product dsDNA with domain maps to highlight the ordering of the helical gateway and cap. *Filled circles and triangles* represent α -helices and β -sheets, respectively, and are numbered accordingly. The *open circle* represents a single helical turn of 4 amino acids. The approximate locations of the seven acidic residues for divalent metal sequestration and of the three amino acid residues responsible for coordination of the K^+ ion are represented by filled *red stars* and *purple circle*, respectively. Structures of the enzyme-product (c) and enzyme-substrate (d) complexes with the protein illustrated as ribbon diagrams with translucent surface representations. The DNA template, 5'-flap and 3'-flap strand are shown in *brown, yellow* and *purple*, respectively

regions and forms a saddle-like structure. Two sites for binding dsDNA are on either side of the β -sheet. Protruding from the saddle-like structure is the region called the helical arch or helical clamp that in some crystal structures is observed as a disordered, flexible loop, and in others is structured as two α helices poised above the active site. The active site of FENs are highly conserved and in eukaryotic and archaeal FENs, and is characterized by the presence of seven acidic residues (Lieber 1997; Shen et al. 1998) (Fig. 16.3a) that sequester the divalent metal ions requisite for catalysis. Consistent with a two metal ion mechanism for phosphate diester hydrolysis

(Yang et al. 2006), these acidic residues in the hFEN1 substrate-free structure hold the divalent metal ion <4 Å apart; however, the occupancy in the crystal structures varies for the metals (Sakurai et al. 2005). An unusual feature of archaeal and eukaryotic FENs is that their N-termini are structured and lie close to the active site. If electron density for the N-terminus can be seen, the sequence starts with Gly2, and thus, N-terminal aminopeptidases have removed the initiator methionine during translation (Bradshaw et al. 1998), at least when expressed as a recombinant protein in *E. coli*. The observation of structured N-termini near the active sites of FENs explains why cloning FENs with N-terminal affinity tags results in proteins that are insoluble (i.e., GST tagged protein) (Zheng and Shen, unpublished data) or have negligible activity (i.e., His-tags that are cleaved off leaving additional N-terminal residues) (Grasby and Tainer, unpublished data) (Mase et al. 2009). This is the reason for the exclusive use of C-terminal affinity tags for recombinant expression.

16.3.2 Protein-Product and Protein-Substrate Complexes

Comparison of the crystal structures of human FEN1 with product and substrate DNA revealed an extraordinarily sophisticated dsDNA binding and ssDNA incision mechanism (Tsutakawa et al. 2011). The proposed model allows FEN to distinguish 5'-flap structures from 3'-flap structures and ssDNA and is biologically elegant in eliminating non-FEN substrates from inadvertent incision. First, we will discuss structural elements common to both the enzyme-substrate and enzyme-product complexes with respect to the template and 3'-flap strand interactions (Fig. 16.11). Then, we will discuss the differences that exist between the two complexes at the extreme 5'-end of the 5'-flap strand.

The protein-substrate and protein-product complexes show that most of the binding surface is to the two dsDNA regions. The upstream dsDNA is bent at approximately a 100° angle at a single phosphate diester relative to the downstream dsDNA. Helix 2 of FEN is wedged between the basepairs formed at either side of the two-way dsDNA junction (Fig. 16.3c, d). Thus, an initial recognition mechanism is binding dsDNA containing a junction that can easily bend at a single phosphate diester. Recognition of the substrate is not mediated through the 5' flap DNA strand as would be expected according to the original threading/clamping and tracking models, but via interactions with the template DNA strand. Moreover, approximately half of the protein:DNA interface is directed to the template strand (Fig. 16.4a–d), which corresponds to the parental strand of DNA in replication (Fig. 16.1j). Part of this interaction is mediated by a potassium ion that is coordinated to amino acid peptide carbonyls from a helix-two-turn-helix motif (H2tH), a side chain hydroxyl of a serine residue, and a phosphate diester from the template DNA (Fig. 16.3b–d, 16.4a–f). In addition, there are contacts to the 5'-flap strand near its 3'-terminus (Fig. 16.4c–d). Another facet to the dsDNA binding is that the protein-dsDNA interaction surface is not contiguous. The downstream portion of the template strand departs from the surface of the protein after the H2tH motif and then, returns to the

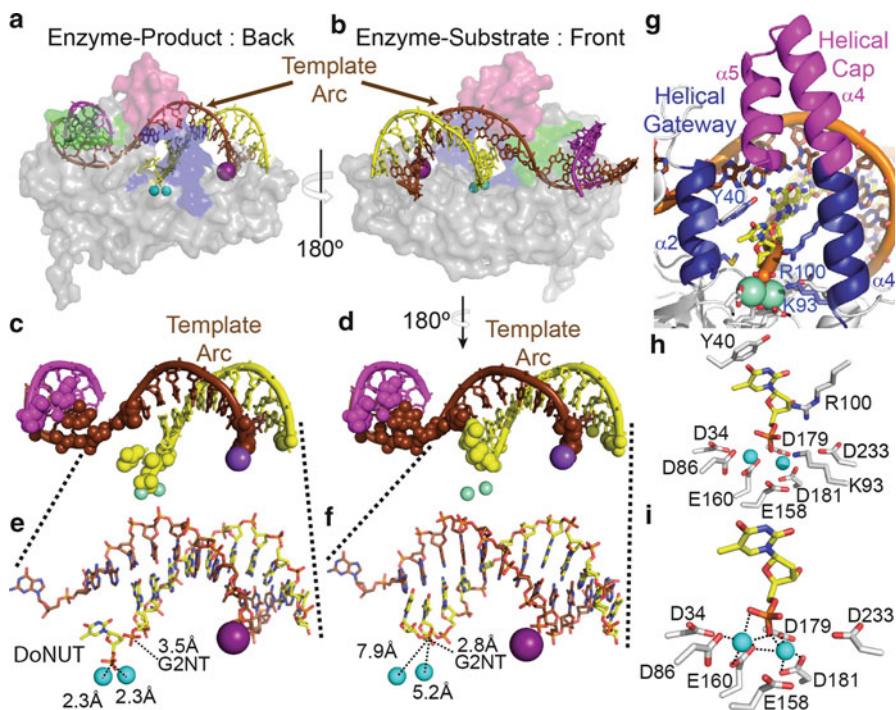


Fig. 16.4 FENs effectively utilize the helical properties of DNA to deliver the scissile phosphate to the active site. Due to the spacing of the dsDNA binding regions and the bend induced on the template strand of the two-way DNA junction, the template arc directs the 5'-flap strand towards the active in both the enzyme product (a) and enzyme substrate (b) complexes. The DNA strands and translucent surface representation of the protein are coloured as in Fig. 16.3. The two trivalent samarium ions (Sm^{3+}) and the K^+ ions are represented as cyan and purple spheres, respectively. Looking at the DNA alone and representing sites of protein contact to DNA by spheres, the product (c) and substrate (d) DNA are almost the same except for important changes near the scissile phosphate. Note the lack of direct contacts in the template arc region. Focussing on the downstream duplex region, comparison of the product (e) and substrate (f) DNA near the active site shows that the two nucleotides have unpaired for the scissile phosphate to interact with the catalytically important metal ions. In addition, the Gly2 N-terminus (G2NT) interacts initially with the scissile phosphate diester in the enzyme substrate complex (f), but interacts with the phosphate diester 3' to the scissile phosphate in the enzyme-product complex (e). (g) View of the active site (coloured as above) from the back of the helical gateway and helical cap to highlight some of the catalytically important residues. (h) View of the complete hFEN1 active site that includes the seven highly conserved carboxylates, two basic residues from helix four (Lys93 and Arg100), and an aromatic stacking partner (Tyr40) from helix two. (i) View of the four active site acidic residues (D86, E160, D179, and D181) and product complex unpaired base phosphate monoester directly coordinating the Sm^{3+} metals (grey dashed lines)

surface of the protein just before the bend in the DNA. The template strand enters a groove where the DNA bends sharply, giving rise to the 100° bend at the two-way dsDNA junction. The arc that is formed by the template strand (template arc) as it leaves and then, returns to the protein surface is used to deliver the 5'-flap strand to

the proximity of the active site (Fig. 16.4a, b). Moreover, as the downstream primer strand follows the template strand, it is the 5'-terminus of the downstream flap strand that is directed towards the active site. When delivered to the proximity of the active site, helix 2 and helix 4 residues contact the 5'-terminal portion of the 5'-flap strand. However, the way in which these residues interact with the 5'-terminal residues of the downstream primer differs between the enzyme substrate and enzyme product complexes, and will be dealt with in the next two sections (Tsutakawa et al. 2011).

Despite being base paired to the upstream nucleotides of the template, the only interactions seen with the 3'-flap strand are localized to its last three nucleotides. The rest of the upstream primer strand does not contribute to binding and passively exits from hFEN1 by following the template DNA (Fig. 16.4c, d). Most of the interactions are to the one nucleotide 3'-flap. The 3' flap binding pocket is constructed by ten amino acid residues, with most of the protein interaction with the DNA to the sugar-phosphate backbone and the 3'-hydroxyl group. None of the interactions are base-specific, as would be expected for a structure rather than sequence specific nuclease. The 3'-flap is bound unpaired, as would be needed to obtain a ligatable product from an equilibrating flap (Fig. 16.2a). The cleft formed by the 3'-flap binding site physically blocks anything larger than a one nucleotide 3'-flap being accommodated, consistent with biochemical experiments with substrates where the 3'-flap sugar moiety was modified (Kao et al. 2002). In conjunction with the 3'-flap binding pocket, a conserved group of acidic residues (denoted as the acid block) in the proximity of the 3'-flap binding pocket and presumably is present to prevent DNA moving beyond this point. Both the 3'-flap binding pocket and the acid block are features that are present only in FENs of the 5'-nuclease superfamily, suggesting this feature has evolved to enhance FEN function. To a lesser extent, there are also interactions between residues of helix 3 and the penultimate nucleotide of the 3'-flap strand. These residues are part of the hydrophobic wedge that stabilizes the bent conformation of the two-way dsDNA junction by interacting with the face of the last base pair created by the 3'-flap strand and template. In the free protein structures (Sakurai et al. 2005), the residues involved in 3'-flap binding are disordered (Fig. 16.3a), suggesting the 3'-flap pocket orders upon recognition and binding of this nucleotide. The only structural feature of the 3'-flap that distinguishes it from all other nucleotides in the upstream primer strand is the presence of a 3'-hydroxyl. Thus, it is not surprising that the 3'-hydroxyl is used a key recognition feature by FEN.

16.3.3 Protein-Product Complex 5'-Strand Interactions

Crystallization of hFEN1 in complex with a four nucleotide 'quasi-equilibrating' double flap substrate resulted in an enzyme product complex where the dsDNA was still in complex with the protein, consistent with the dsDNA product (Q) being a competitive inhibitor (Fig. 16.2f) (Finger et al. 2009). Unlike structures of hFEN1 in the absence of DNA, the residues of the helical arch and the top of helix two are

ordered in a gate-like conformation (Fig. 16.4g). Helix two and most of helix four form the posts of the gateway, whereas the upper portion of helix 4 and all of helix 5 sits atop the two posts, and has been referred to as the helical 'cap' (Fig. 16.3c,d). This nomenclature diverges from the previous arch or archway, which does not include helix two. The helical gateway is 13 Å at its most narrow and is located over the active site at the base of the arch. This region of the protein would not be able to order if dsDNA is in the gateway. Although FENs can cleave flaps containing dsDNA (Fig. 16.1i), a sufficient number of single-stranded nucleotides must be present between the dsDNA in the flap and the two way-dsDNA junction to allow the helical gateway to order around the ssDNA portion of the flap and to bring critical catalytic residues into the active site. The helical cap sterically limits passage through the gateway to substrates with free 5' terminus (i.e., no bubbles, bulges, etc.) (Tsutakawa et al. 2011). Interestingly, the helical gateway is strongly conserved in the 5'-nuclease superfamily. The cap, on the other hand, is conserved only in FEN1 and EXO1. GEN1 and XPG, which cleave Holliday junctions and DNA bubbles, respectively, do not appear to have a cap (Orans et al. 2011; Tsutakawa et al. 2011).

In the product complex, the nucleotide just to the 3'-side of the cleaved phosphate diester is unpaired, with the 5'-phosphate monoester product interacting with the two divalent metal ions in the active site. Comparison of the product and substrate complexes (Fig. 16.4c-f) shows that two downstream dsDNA nucleotides 5' and 3' of the scissile phosphate must unpair to enter the active site. Tyr40 of helix two stacks against the 3'-face of the base of the unpaired nucleotide, thereby stabilizing the unpaired state (Fig. 16.4g). Two residues of helix four, Lys93 and Arg100, interact with the terminal phosphate suggesting that these residues are also important in stabilizing the unpaired state. These conserved residues were also identified in a screen of toxic yeast *rad27* mutants that had a dominant negative effect on cell viability and growth (Storici et al. 2002). In DNA-free FEN structures where the helical cap and gateway are disordered, Lys93 and Arg100 (or their equivalent homologues) are not near the active site (Hosfield et al. 1998b; Hwang et al. 1998; Sakurai et al. 2005). Furthermore, evidence from studies of T5FEN suggests that Lys83, the equivalent of hFEN1 Lys93, acts as an electrophilic catalyst (Sengerova et al. 2010). Once flipped, the N-terminus of Gly2 contacts the phosphate diester of the next nucleotide in the 5'-flap strand. The trajectory of the phosphate backbone through the active site of the enzyme product complex suggests that the substrate 5'-flap would traverse the helical gateway under the helical cap (Tsutakawa et al. 2011) (Fig. 16.4e). This strongly suggests that a bind-then-thread mechanism of action. However, this is still being debated (Patel et al. 2012).

16.3.4 Protein-Substrate Complex 5'-Flap Strand Interactions

Overall, the protein-substrate and protein-product complexes share similar features in binding, except for differences in the location of some downstream/flap

strand primer nucleotides and amino acid residues involved in the interaction with these two residues (Fig. 16.4d). Unlike the enzyme-product complex, the downstream 5'-flap strand of the enzyme substrate complex is base paired to the template strand, and the scissile phosphate is in the proximity of the active site, but not in contact with the critical active site metals or amino acid residues. Despite not being in the active site, the protein-substrate complexes show that binding of the dsDNA alone in this bent, non-contiguous manner results in the scissile phosphate being placed within 5–8 Å of the active site in enzyme substrate complexes. Moreover, the binding of a two-way dsDNA junction in this orientation is important to place the scissile phosphate in the proximity of the active site. By allowing the structural features of the placement of the scissile phosphate diester to be mainly with the template strand and 3'-flap rather than the 5'-flap strand itself may prevent inadvertent incision of ssDNA in the 5'-flap, creating products that would require DNA repair processes to correct the imprecise cleavage. The lack of interaction with the 5'-flap strand allows cleavage of substrates containing RNA or DNA in the 5'-flap or near the site of incision, which is consistent with the fact that deletion of RNase H in yeast does not result in a significant mutator phenotype (Qiu et al. 1999). The spacing between dsDNA binding regions of the protein formed by the potassium ion/H2tH and at the DNA bend, which are approximately one helical turn apart, is also critical to exclude incision of 3'-flaps. Using models, the spacing between dsDNA binding sites is too wide for positioning a 5'-flap near the active site. FENs, therefore, elegantly take advantage of the helical properties of two-way DNA junctions to position the scissile phosphate diester close to the active site and to distinguish 5'-flap and 3'-flap structures (Tsutakawa et al. 2011).

As noted above, one-fourth of the protein-DNA interactions are made by the helical gateway of the protein and the nucleotides at the base of the 5'-flap around the scissile phosphate. In the product complex, there is little direct interaction between the protein and the downstream primer strand between the gateway and the potassium ion binding regions (Fig. 16.4a–d). Tyr40 in helix 2 stacks with the 5'-face of the base of the terminal 5'-flap strand nucleotide, whereas another helix 2 residue, Ile44, contacts the 5'-face of the complimentary nucleotide of the template strand. In addition, the scissile phosphate in the enzyme substrate complex is in contact with the N-terminus (Gly2) (Fig. 16.4d), instead of the phosphate diester 3' to the scissile phosphate in the enzyme-product complex (Fig. 16.4c). In the paired substrate complex, the terminal base pair of the downstream dsDNA makes only one of two possible H-bonds due to a pronounced base pair opening and stagger towards the major groove. Although the DNA helical parameters of the upstream duplex conform to B-DNA, the helical parameters of the downstream duplex differ. The dsDNA in contact with the K⁺ ion and the H2tH motif conforms to B-DNA (Fig. 16.3c,d), but the six base pairs of the downstream dsDNA nearest to the active site deviate from B-DNA parameters and become more like A-DNA (Fig. 16.4d). In this region, the nucleotides of the 5'-flap strand DNA are less stacked (Tsutakawa et al. 2011).

16.3.5 Bind-Then-Thread or Bind-Then-Clamp

Despite this advance in our understanding of FEN substrate recognition, it is still unknown whether the 5'-flap portion of the substrates is threaded through the arch consistent with the bind-then-thread model (Xu et al. 2001), or if the flaps are instead clamped by the helical arch (Chapados et al. 2004; Orans et al. 2011). Evidence for the bind-then-thread model comes from the structure of T4 FEN in complex with a pseudo-Y substrate, which suggests that the 5'-flap traverses the archway (Devos et al. 2007). However, in this structure, a portion of the helical arch is not observed and the 5'-flap is not in the active site due to the absence of the requisite divalent metals. The bind-then-thread model would be consistent with the detrimental effects on incision activity with substrates where the 5'-flap has been modified with streptavidin as passage through the helical arch would be blocked by the addition of a molecule as large as tetrameric streptavidin (Gloor et al. 2010; Murante et al. 1995). However, because the structured helical arch is not large enough to accommodate dsDNA, the ability to cleave a fold-back flap substrate is said to be more consistent with the bind-then-clamp model (Finger et al. 2009). To counter this, the helical arch region in several DNA-free FEN structures from several archaea and eukaryotes has been observed to be unstructured (Fig. 16.3a) (Hosfield et al. 1998a; Hwang et al. 1998; Mase et al. 2009; Sakurai et al. 2005). An unstructured arch is theoretically large enough to accommodate double stranded DNA, and could then reorder around the ssDNA portion of the fold-back flap structure (Tsutakawa et al. 2011). Thus, a bind-then-thread model, whereby substrates containing dsDNA in the flap are threaded through an unstructured arch with subsequent ordering after threading, is plausible (Patel et al. 2012).

16.3.6 Scissile Phosphate Placement: The Double Nucleotide Unpairing Trap (DoNUT)

A disorder-to-order transition of the helical arch (gateway and cap) induced by DNA binding is suggested by comparing DNA-bound hFEN1 structures with DNA-free hFEN1 structures that had previously been solved in complex with PCNA, where three hFEN1 molecules were bound to one PCNA homotrimer with each in a unique crystallographic environment (Sakurai et al. 2005) (Fig. 16.5a,b). In all three DNA-free structures, the helical gateway and cap, as well as the hydrophobic wedge of helix two were disordered (Fig. 16.3a). Furthermore, the catalytically critical residues in helix four (Lys93 and Arg100) and helix two (Tyr40) were not correctly positioned. Thus, we proposed that the binding of the 3'-flap likely orders the hydrophobic wedge, which in turn leads to the ordering of the helical gateway and cap. When ordered, Lys93 and Arg100 are poised in the active site (Fig. 16.4g,h). Because ordering has occurred without unpairing in the fully base-paired substrate complexes, we propose that unpairing occurs after the ordering of the helical gateway. Although

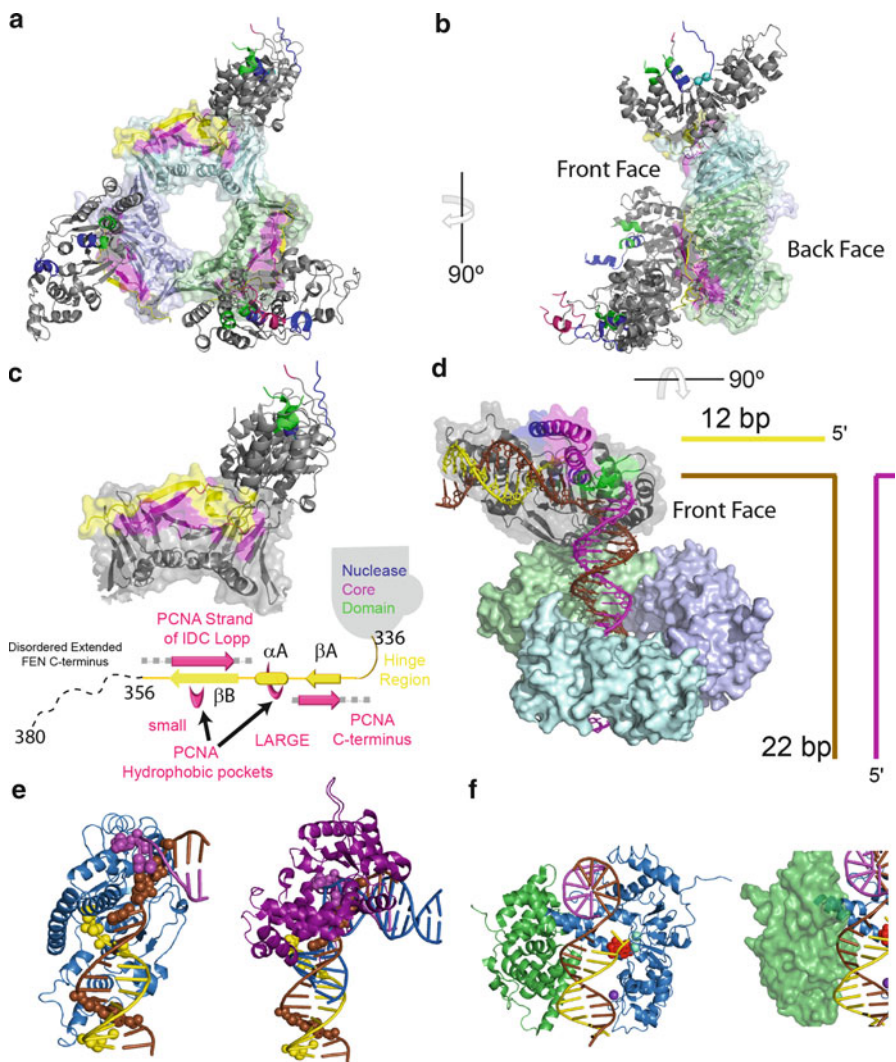


Fig. 16.5 How FENs work with other proteins. (a) Structure of homotrimeric hPCNA in complex with three subunits of hFEN1. The nuclease core domain is coloured as in Fig. 16.3. PCNA subunits are shown as combined ribbon and transparent surface representation. The portion of the extended C-terminus of hFEN1 that was observed in the crystal (residues 336–356) and the regions important for interaction in PCNA are shown in *yellow* and *magenta*, respectively. (b) Turning the structure 90° shows that the protein is positioned on one face of the protein. (c) Closer view of the βA - αA - βB motif with schematic illustration below it coloured as above. Alpha helices are represented as *cylinders* and *arrows* indicate the position of the β -sheets. (d) Model of hFEN1 interacting with PCNA and DNA simultaneously and coloured as above. (e) Comparison of the hFEN1 (*blue*) DNA (coloured as above) structure with the pol β -DNA (*Purple*) structures shows that the only regions available for FENs to initially contact in a handoff model is the downstream dsDNA binding region. (f) Comparison of the DBD domain of ligase (*green*) and its known interaction site (minor groove) in comparison to hFEN1 (*blue*) DNA product complex (coloured as above). The groove necessary for DBD interaction is accessible. Below is a translucent surface representation of the ligase DBD showing a steric clash between the DNA and the helical cap of hFEN1. This steric interference may be the initial manner in which ligase facilitates hFEN1 release of its product

base flipping can occur via movement through the major or minor groove (Bouvier and Grubmuller 2007), the dinucleotide unpairing inferred by comparing enzyme-substrate and enzyme-product complexes would likely occur via the major groove, because as the minor groove pathway is blocked by the hydrophobic wedge of helix two. This is consistent with the pronounced base pair opening and stagger towards the major groove in the enzyme-substrate complex (Tsutakawa et al. 2011).

But what drives the unpairing of the DNA? It has been shown by von Hippel and co-workers that nucleotides at ssDNA-dsDNA junctions are much more prone to spontaneous base flipping than nucleotides farther away from the ssDNA-dsDNA junction (Jose et al. 2009). Therefore, FENs may take advantage of this and simply capture the unpaired state. Comparison of the helical parameters of the enzyme-substrate and enzyme-product complexes shows that some of the distortions from B-DNA parameters in the downstream dsDNA of the enzyme-substrate complex are alleviated in the unpaired state of the enzyme-product complex (i.e., returns to more B-DNA like parameters), thereby suggesting that binding energy may also be used to promote the dinucleotide unpairing (Fersht 1999). Regardless of whether the unpairing process is passive or is actively promoted by the enzyme, the ability of FEN to capture the unpaired state is clearly demonstrated in the product complex (Fig. 16.4c–h).

16.3.7 Cleavage of the Scissile Phosphate Diester: Active Site Structure

In the enzyme-product crystal structure, the active site of hFEN1 contains two samarium (Sm^{3+}) cations, which directly interact with the terminal phosphate of the product complex (Fig. 16.4h,i). The Sm^{3+} ions are directly coordinated to FEN by four of the seven carboxylate residues that are invariantly conserved in the 5'-nuclease superfamily active site and occupy the same sites that Mg^{2+} is known to bind in the DNA-free hFEN1 protein structures (Sakurai et al. 2005). Asp86, Glu160, Asp179, and Asp181 make inner sphere contacts with the metal ions, with Glu160 acting as the crucial bridging residue between the two metals. The other three conserved acidic residues in addition to a phenolic hydroxide from a conserved tyrosine (Tyr234) make outer sphere contacts via water molecules. Using a model of a 'cleavage-competent' enzyme-substrate complex based upon the enzyme-product complex, we have proposed that the attacking water activated by the two divalent metals hydrolyze the scissile phosphate diester bond in a manner similar to a mechanism proposed for EndoIV (Garcin et al. 2008; Ivanov et al. 2007). However, others have proposed a more traditional two divalent ion mechanism for cleavage, where the nucleophilic water is activated by one of the two metal ions (Orans et al. 2011). More work is necessary to distinguish which mechanism accurately describes hFEN1 catalysis and to identify the mechanism responsible for activating the attacking nucleophile (Yang et al. 2006). Nevertheless, there is consensus that phosphate diester hydrolysis requires two ions, in accordance with functional data (Syson et al. 2008).

16.4 Regulation of FEN1 Activity

The activity of FEN must be tightly controlled and coordinated with other components of the DNA replication machinery. FEN has several protein-protein interaction partners that work with FENs to achieve efficient and faithful copying of the DNA. FEN is also the target of several post-translational modifications, adding a further level of control (Zheng et al. 2011b). FEN interacts with its partner proteins mainly through its C-terminal extension (Guo et al. 2008a). This C-terminus has been implicated in DNA binding *in vitro* (Friedrich-Heineken et al. 2003; Stucki et al. 2001), but may in fact be an artefact *in vitro* due to its role of this region in mediating protein-protein interactions *in vivo*, which would interfere with the ability to bind DNA (see Sect. 16.4.1). The C-terminus is likely disordered in the absence of partner proteins; a disorder to order transition upon partner presence may allow for multiple, simultaneous protein-protein interactions. Furthermore, FENs are sequestered where necessary (e.g., nucleus and mitochondria), and evidence also suggests that even subcellular as well as suborganellar location are regulated (Zheng et al. 2011b).

16.4.1 Protein-Protein Interactions

16.4.1.1 PCNA

Proliferating cell nuclear antigen (PCNA) is the major protein involved in coordination of FEN recruitment and activity in the processing of Okazaki fragments. PCNA is a trimeric ‘sliding clamp’ protein, localised at the sites of Okazaki fragment maturation (see Chap. 15, this volume). PCNA binds and coordinates the action of pol δ , FEN and DNA ligase I (Beattie and Bell 2011; Burgers 2009). Each subunit of the homotrimeric PCNA likely interacts with a particular protein. In support of this notion, archaeal PCNA is a heterotrimer, with each member of the trimer having specificity for either pol δ , FEN or ligase, suggesting a precise architecture in the maturation of Okazaki fragments (Dionne et al. 2003). How a eukaryotic cell could ensure that each subunit of the homotrimeric PCNA is loaded with a single pol δ , FEN1, and ligase is unknown.

PCNA has been shown to interact with FEN1 in the absence of DNA (Wu et al. 1996) through a PCNA binding motif QXX(I/L/M)XXF(F/Y) (Frank et al. 2001). Structural data also shows that FENs utilize amino acid residues Thr336 to Leu356 in their C-terminus to form two β -zipper (β A and β B) separated by a small α helix (α A) to interact with PCNA (Fig. 16.5a–c) (Chapados et al. 2004; Sakurai et al. 2005). Disruption of FEN and PCNA interactions by mutating Phe343 and Phe344 of FEN to alanine (FFAA) results in loss of binding as determined using non-equilibrium binding assays, but does not affect FEN nuclease activity *in vitro* or display gross phenotypes in yeast (Frank et al. 2001). However, a homozygous FFAA mutation in

mice results in newborn lethality and other phenotypes indicative of poor replicative capacity (Zheng et al. 2007a). Furthermore, heterozygous FFAA mice (i.e., FFAA/WT *fen1*) also show defects in Okazaki fragment maturation and an increased rate of aneuploidy-associated cancers (Zheng et al. 2011a).

In the hFEN1-PCNA co-crystal structure, the various conformations of three hFEN1s were a result of different torsion angles among four amino acid residues of hFEN1 (aa residues 333-336) (Sakurai et al. 2005) (Fig. 16.5a,b). This 'hinge region' (Fig. 16.5c) is conserved among eukaryotic FENs. Using the recent hFEN1-product structure in combination with one of the conformations in the FEN-PCNA structure, a model of how FEN and PCNA would work together was generated (Fig. 16.5d). Although upstream dsDNA need only be as short as 6 nts when studying FENs *in vitro*, the upstream dsDNA for a PCNA/FEN complex would need to be 6 nts plus ~15 base pairs to allow PCNA-DNA binding as well.

PCNA has been shown to stimulate *in vitro* FEN activity on static single-flap substrates (Frank et al. 2001; Li et al. 1995; Sakurai et al. 2005). Furthermore, kinetic analyses have shown that this PCNA stimulation occurs by facilitating FEN-DNA complex formation (i.e., decreases K_M rather than increasing k_{cat} (Hutton et al. 2009; Tom et al. 2000)). Because k_{cat} for FEN1 is a measure of enzyme-product release under MT conditions (Finger et al. 2009; Williams et al. 2007), it is interesting that the addition of PCNA increases the affinity for substrate but does not affect the rate of enzyme-product release. Obviously more work on this is necessary to determine how PCNA can selectively enhance substrate binding without slowing enzyme-product release.

16.4.1.2 RecQ Helicase Family Interactions

The RecQ helicase family members, WRN (Werner syndrome ATP-dependent helicase) and BLM (Bloom syndrome protein), are also key modulators of FEN activity. WRN co-localises with FEN at stalled replication forks (Sharma et al. 2004), unwinding Holliday junctions and stimulating FEN cleavage in a structure-dependent manner. WRN has also been shown to stimulate both FEN cleavage on single-flap, EXO, and forked-gapped substrates (Sharma et al. 2004; Zheng et al. 2005). FEN-WRN interactions are mediated by a 144 residue domain of WRN that has homology to RecQ helicase, interacting with the terminal 18 residues of FEN's C-terminal domain (Brosh et al. 2001; Guo et al. 2008a; Sharma et al. 2005). These residues are disordered in the (hFEN1)₃-PCNA complex (Fig. 16.5c). WRN, unlike PCNA, purportedly stimulates FEN by increasing its turnover of DNA substrates directly (Brosh et al. 2002). The WRN/BLM binding site is adjacent to the PCNA binding site and should allow direct co-ordination of activities by these partner proteins (Sharma et al. 2005). Both WRN and BLM over-expression can rescue mutants lacking Dna2, indicating their ability to stimulate FEN (Imamura and Campbell 2003; Sharma et al. 2004). WRN is also purportedly required for nucleolar localisation of FEN (Guo et al. 2008b).

16.4.2 Post-translational Modifications

Phosphorylation, methylation and acetylation also play a key role in modulating FEN activity and localisation (Zheng et al. 2011b). FEN is phosphorylated at Ser187 in late S-phase by the cyclin-dependent kinases Cdk1 and Cdk2, in partnership with cyclin A (Cdk1) or cyclin E (Cdk2). Phosphorylation *in vitro* results in a decrease in nuclease activity, but not DNA binding (Henneke et al. 2003). Phosphorylated FEN is unable to associate with PCNA and possibly also abrogates interactions with other proteins like WRN (Zheng et al. 2011b). Interestingly, Ser187 is buried in both DNA-free and DNA-bound structures of FEN (Sakurai et al. 2005; Tsutakawa et al. 2011). Thus, for kinases to phosphorylate FENs, a conformational change will likely be necessary. PRMT5 methylates Arg192 (Guo et al. 2010), a residue whose side chain contacts DNA in the enzyme-product complex. Methylation prevents phosphorylation, and has the opposite effect; addition of a methyl group facilitates interaction with PCNA (Zheng et al. 2011b). Acetylation occurs on multiple lysine residues of FEN both *in vitro* and *in vivo* (Choudhary et al. 2009; Hasan et al. 2001). Acetylation *in vitro* of four lysines in the extreme C-terminus of the protein results in a decrease in PCNA independent nuclease activity (Hasan et al. 2001). It is speculated that acetylation could lead to blocking of the short flap pathway by retarding FEN activity, leading to Okazaki fragment maturation by the long flap method involving Dna2 (Balakrishnan et al. 2009). The roles of Lys80 and Lys267 acetylation, which were discovered in a genome-wide screen for acetylated proteins (Choudhary et al. 2009), remain unclear. Lys80 is solvent exposed and interacts with PCNA residues in the hFEN1-PCNA co-crystal structure (Sakurai et al. 2005); therefore, its modification could influence PCNA interaction as well as other protein interaction partners. Lys267 interacts with the 3'-terminus of the 5'-flap strand (Tsutakawa et al. 2011); therefore, the addition of an acetyl group could regulate downstream dsDNA binding (Zheng et al. 2011b).

16.5 Handoff of DNA Intermediates

Although physical limitations such as diffusion and enzyme-product release sometimes limit reaction rates *in vitro*, these limitations may not occur *in vivo*. To increase the efficiency of metabolic processes, cells have overcome physical limitations by sequestering the requisite proteins to their sites of action and channelling or handing-off intermediates from protein to protein in the metabolic pathway, best exemplified by the multifunctional enzyme tryptophan synthase (Yanofsky 1989). This channelling is important to cellular function as it prevents the release of potentially reactive intermediates into solution and prevents the establishment of enzyme-substrate equilibrium (Ovádi 1991). An analogy to DNA metabolic pathways like Okazaki fragment maturation can be hypothesized, but rather than channelling a small metabolite from one active site to another, the metabolic intermediate (i.e., DNA) is passed from protein to protein. Although DNA nicks or flaps may not necessarily be

reactive intermediates, release of these species *in vivo* during Okazaki fragment maturation could result in competition between ligation versus initiation of DNA repair pathways. Although DNA repair pathways can usually faithfully replicate the DNA, these processes are not perfect (e.g., unequal sister chromatid exchange in trinucleotide repeat regions resulting in contractions and expansions). Thus, an attractive mechanism for the processing of Okazaki fragment maturation is for intermediates to be efficiently passed between each protein in the pathway with PCNA coordinating the overall process.

Although there is no direct biochemical evidence yet for this handoff model, support for it does come from structural studies conducted on proteins involved in the pathway. Currently, the best available polymerase structure in complex with downstream and upstream dsDNA is a structure of pol β (Krahn et al. 2004). Although pol β is not involved in DNA replication, it is involved in an analogous process of long-patch base excision repair (LP-BER), (Robertson et al. 2009). Inspection of the pol β -DNA complex shows that the upstream and downstream dsDNA regions are bent $\sim 100^\circ$ and that the base of the 5'-flap and 3'-flap and upstream dsDNA duplex are in direct contact with the polymerase (Krahn et al. 2004), and thus, would sterically clash with FENs (Fig. 16.5e). The only region exposed for FENs to initially bind is the downstream dsDNA region. To access the 5'- and 3'-flaps, FENs would first interact with the downstream dsDNA and would then displace the polymerase (Tsutakawa et al. 2011). Because flaps *in vivo* can potentially migrate (Fig. 16.2e), the handoff of the substrate in this manner would prevent flaps from forming structures that would need to be extensively remodelled before cleavage, thereby ensuring efficient Okazaki fragment maturation. Still, at some point in this hypothetical transition between the polymerase and FEN, the one nucleotide 3'-flap must bind the single nucleotide binding pocket of FEN (Fig. 16.2e).

The structure of DNA ligase I in complex with DNA shows that the enzyme encircles the nicked DNA using three domains: the DNA binding domain (DBD), adenylation domain (AdD) and OB domain (OB) (see Chap. 17, this volume). It is known that the DBD is essential for interaction with DNA substrates (Tomkinson et al. 2006). The DBD interacts with the minor grooves on either side of the nick, whereas the OB fold interacts with the major groove (Pascal et al. 2004). Inspection of the hFEN1-product complex shows that the two minor grooves for DBD interaction are exposed (Fig. 16.5f). However, the product DNA is bent in this structure unlike in the ligase I-DNA structure, where DNA is not bent in order to align the termini for ligation. Despite this, the fact that the grooves necessary for DBD interaction are exposed in the hFEN1-product complex provides further compelling evidence for the handoff model, and suggests that hFEN1 and the DBD of ligase I could bind DNA simultaneously.

In summary, consideration of the buried interfaces of the pol β , hFEN1 and ligase I in complex with DNA shows that the choreography of handoff would be dictated by the exposed regions of dsDNA. As such, it is not surprising that FENs have a tendency to retain the product that is the substrate of the next step of the DNA metabolic pathway and simply release the other upon cleavage. Such a handoff or 'passing of the baton' concept (Parikh et al. 1999; Wilson and Kunkel 2000) is a simple but efficient way to envisage how the efficiency of Okazaki fragment

maturation can be increased beyond the physical limitations of *in vitro* studies. The proteins necessary for the process are sequestered to the site of action by PCNA, and the intermediates could be passed between one another (Beattie and Bell 2011; Burgers 2009). Although FENs show a remarkable non-sequence specific ability to identify their substrate while preventing inadvertent activity on other DNA structures, further work to determine how FEN interacts with its partner proteins is likely to be as elegant as the structural and biochemical work to date that has dissected what we already know about this remarkable enzyme.

Acknowledgements This article was supported by the European Union 7th Research Framework Programme (FP7)-Marie Curie International Incoming Fellowship Project No. 254386 (LDF), National Cancer Institute grants RO1CA081967, P01 CA092584 (JAT), BBSRC grant BBF0147321 (JAG), and RO1CA073764 (BS).

References

- Balakrishnan L, Stewart J, Polaczek P, Campbell JL, Bambara RA (2009) Acetylation of Dna2 endonuclease/helicase and flap endonuclease 1 by p300 promotes DNA stability by creating long flap intermediates. *J Biol Chem* 285:4398–4404
- Beattie TR, Bell SD (2011) The role of the DNA sliding clamp in Okazaki fragment maturation in archaea and eukaryotes. *Biochem Soc Trans* 39:70–76
- Berg OG, von Hippel PH (1985) Diffusion-controlled macromolecular interactions. *Annu Rev Biophys Chem* 14:131–160
- Bochman ML, Sabouri N, Zakian VA (2010) Unwinding the functions of the Pif1 family helicases. *DNA Repair* 9:237–249
- Bornarth CJ, Ranalli TA, Henricksen LA, Wahl AF, Bambara RA (1999) Effect of flap modifications on human FEN1 cleavage. *Biochemistry* 38:13347–13354
- Bouvier B, Grubmüller H (2007) A molecular dynamics study of slow base flipping in DNA using conformational flooding. *Biophys J* 93:770–786
- Bradshaw RA, Brickey WW, Walker KW (1998) N-terminal processing: the methionine aminopeptidase and N alpha-acetyl transferase families. *Trends Biochem Sci* 23:263–267
- Brosh RM Jr, von Kobbe C, Sommers JA, Karmakar P, Opreško PL, Piotrowski J, Dianova I, Dianov GL, Bohr VA (2001) Werner syndrome protein interacts with human flap endonuclease 1 and stimulates its cleavage activity. *EMBO J* 20:5791–5801
- Brosh RM Jr, Driscoll HC, Dianov GL, Sommers JA (2002) Biochemical characterization of the WRN-FEN-1 functional interaction. *Biochemistry* 41:12204–12216
- Budd ME, Campbell JL (1997) A yeast replicative helicase, Dna2 helicase, interacts with yeast FEN-1 nuclease in carrying out its essential function. *Mol Cell Biol* 17:2136–2142
- Budd ME, Reis CC, Smith S, Myung K, Campbell JL (2006) Evidence suggesting that Pif1 helicase functions in DNA replication with the Dna2 helicase/nuclease and DNA polymerase delta. *Mol Cell Biol* 26:2490–2500
- Burgers PM (2009) Polymerase dynamics at the eukaryotic DNA replication fork. *J Biol Chem* 284:4041–4045
- Ceska TA, Sayers JR, Stier G, Suck D (1996) A helical arch allowing single-stranded DNA to thread through T5 5'-exonuclease. *Nature* 382:90–93
- Chapados BR, Hosfield DJ, Han S, Qiu J, Yelent B, Shen B, Tainer JA (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* 116:39–50
- Chon HG, Vassilev A, DePamphilis ML, Zhao YM, Zhang JM, Burgers PM, Crouch RJ, Cerritelli SM (2009) Contributions of the two accessory subunits, RNASEH2B and RNASEH2C, to the activity and properties of the human RNase H2 complex. *Nucleic Acids Res* 37:96–110

- Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, Mann M (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325:834–840
- Devos JM, Tomanicek SJ, Jones CE, Nossal NG, Mueser TC (2007) Crystal structure of bacteriophage T4 5' nuclease in complex with a branched DNA reveals how flap endonuclease-1 family nucleases bind their substrates. *J Biol Chem* 282:31713–31724
- Dionne I, Robinson NP, McGeoch AT, Marsh VL, Reddish A, Bell SD (2003) DNA replication in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Biochem Soc Trans* 31:674–676
- Fersht A (1999) Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. W.H. Freeman and Company, New York
- Finger LD, Shen B (2010) FEN1 (flap endonuclease 1). In Atlas Genet Cytogenet Oncol Haematol (URL:<http://www.atlasgeneticsoncology.org/Genes/FEN1ID40543ch11q12.html>)
- Finger L, Blanchard M, Theimer C, Sengerova B, Singh P, Chavez V, Liu F, Grasby J, Shen B (2009) The 3'-flap pocket of human flap endonuclease 1 is critical for substrate binding and catalysis. *J Biol Chem* 284:22184–22194
- Frank G, Qiu J, Zheng L, Shen B (2001) Stimulation of eukaryotic flap endonuclease-1 activities by proliferating cell nuclear antigen (PCNA) is independent of its *in vitro* interaction via a consensus PCNA binding region. *J Biol Chem* 276:36295–36302
- Friedrich-Heineken E, Henneke G, Ferrari E, Hubscher U (2003) The acetylatable lysines of human Fen1 are important for endo- and exonuclease activities. *J Mol Biol* 328:73–84
- Friedrich-Heineken E, Hubscher U (2004) The Fen1 extrahelical 3'-flap pocket is conserved from archaea to human and regulates DNA substrate specificity. *Nucleic Acids Res* 32:2520–2528
- Garcin ED, Hosfield DJ, Desai SA, Haas BJ, Bjoras M, Cunningham RP, Tainer JA (2008) DNA apurinic-apyrimidinic site binding and excision by endonuclease IV. *Nat Struct Mol Biol* 15:515–522
- Gloor JW, Balakrishnan L, Bambara RA (2010) Flap endonuclease 1 mechanism analysis indicates flap base binding prior to threading. *J Biol Chem* 285:34922–34931
- Grasby JA, Finger LD, Tsutakawa SE, Attack JM, Tainer JA (2011) Unpairing and gating: sequence-independent substrate recognition by FEN superfamily nucleases. *Trends Biochem Sci* 37:74–84
- Guo Z, Chavez V, Singh P, Finger LD, Hang H, Hegde ML, Shen B (2008a) Comprehensive mapping of the C-terminus of flap endonuclease-1 reveals distinct interaction sites for five proteins that represent different DNA replication and repair pathways. *J Mol Biol* 377:679–690
- Guo Z, Qian L, Liu R, Dai H, Zhou M, Zheng L, Shen B (2008b) Nucleolar localization and dynamic roles of flap endonuclease 1 in ribosomal DNA replication and damage repair. *Mol Cell Biol* 28:4310–4319
- Guo Z, Zheng L, Xu H, Dai H, Zhou M, Pascua MR, Chen QM, Shen B (2010) Methylation of FEN1 suppresses nearby phosphorylation and facilitates PCNA binding. *Nat Chem Biol* 6:766–773
- Harrington JJ, Lieber MR (1994) Functional domains within FEN-1 and RAD2 define a family of structure-specific endonucleases: implications for nucleotide excision repair. *Genes Dev* 8:1344–1355
- Harrington JJ, Lieber MR (1995) DNA structural elements required for FEN-1 binding. *J Biol Chem* 270:4503–4508
- Hasan S, Stucki M, Hassa PO, Imhof R, Gehrig P, Hunziker P, Hubscher U, Hottiger MO (2001) Regulation of human flap endonuclease-1 activity by acetylation through the transcriptional coactivator p300. *Mol Cell* 7:1221–1231
- Henneke G, Koundrioukoff S, Hubscher U (2003) Phosphorylation of human Fen1 by cyclin-dependent kinase modulates its role in replication fork regulation. *Oncogene* 22:4301–4313
- Hohl M, Dunand-Sauthier I, Staresinic L, Jaquier-Gubler P, Thorel F, Modesti M, Clarkson SG, Scharer OD (2007) Domain swapping between FEN-1 and XPG defines regions in XPG that mediate nucleotide excision repair activity and substrate specificity. *Nucleic Acids Res* 35:3053–3063
- Hosfield DJ, Frank G, Weng Y, Tainer JA, Shen B (1998a) Newly discovered archaeobacterial flap endonucleases show a structure-specific mechanism for DNA substrate binding and catalysis resembling human flap endonuclease-1. *J Biol Chem* 273:27154–27161

- Hosfield DJ, Mol CD, Shen B, Tainer JA (1998b) Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell* 95:135–146
- Hutton RD, Craggs TD, White MF, Penedo JC (2009) PCNA and XPF cooperate to distort DNA substrates. *Nucleic Acids Res* 38:1664–1675
- Hwang KY, Baek K, Kim HY, Cho Y (1998) The crystal structure of flap endonuclease-1 from *Methanococcus jannaschii*. *Nat Struct Biol* 5:707–713
- Imamura O, Campbell JL (2003) The human Bloom syndrome gene suppresses the DNA replication and repair defects of yeast *dna2* mutants. *Proc Natl Acad Sci USA* 100:8193–8198
- Ivanov I, Tainer JA, McCammon JA (2007) Unraveling the three-metal-ion catalytic mechanism of the DNA repair enzyme endonuclease IV. *Proc Natl Acad Sci USA* 104:1465–1470
- Johnson RE, Kovvali GK, Prakash L, Prakash S (1995) Requirement of the yeast RTH1 5' to 3' exonuclease for the stability of simple repetitive DNA. *Science* 269:238–240
- Jose D, Datta K, Johnson NP, von Hippel PH (2009) Spectroscopic studies of position-specific DNA “breathing” fluctuations at replication forks and primer-template junctions. *Proc Natl Acad Sci USA* 106:4231–4236
- Kang YH, Lee CH, Seo YS (2010) Dna2 on the road to Okazaki fragment processing and genome stability in eukaryotes. *Crit Rev Biochem Mol Biol* 45:71–96
- Kao HI, Henriksen LA, Liu Y, Bambara RA (2002) Cleavage specificity of *Saccharomyces cerevisiae* flap endonuclease 1 suggests a double flap structure as the cellular substrate. *J Biol Chem* 277:14379–14389
- Kim IS, Lee MY, Lee IH, Shin SL, Lee SY (2000) Gene expression of flap endonuclease-1 during cell proliferation and differentiation. *Biochim Biophys Acta* 1496:333–340
- Krahn JM, Beard WA, Wilson SH (2004) Structural insights into DNA polymerase beta deterrents for misincorporation support an induced-fit mechanism for fidelity. *Structure* 12:1823–1832
- Kucherlapati M, Yang K, Kuraguchi M, Zhao J, Lia M, Heyer J, Kane MF, Fan K, Russell R, Brown AM et al (2002) Haploinsufficiency of Flap endonuclease (Fen1) leads to rapid tumor progression. *Proc Natl Acad Sci USA* 99:9924–9929
- Larsen E, Gran C, Saether BE, Seeberg E, Klungland A (2003) Proliferation failure and gamma radiation sensitivity of Fen1 null mutant mice at the blastocyst stage. *Mol Cell Biol* 23:5346–5353
- Larsen E, Kleppa L, Meza TJ, Meza-Zepeda LA, Rada C, Castellanos CG, Lien GF, Nesse GJ, Neuberger MS, Laerdahl JK et al (2008) Early-onset lymphoma and extensive embryonic apoptosis in two domain-specific Fen1 mice mutants. *Cancer Res* 68:4571–4579
- Li X, Li J, Harrington J, Lieber MR, Burgers PM (1995) Lagging strand DNA synthesis at the eukaryotic replication fork involves binding and stimulation of FEN-1 by proliferating cell nuclear antigen. *J Biol Chem* 270:22109–22112
- Lieber MR (1997) The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. *Bioessays* 19:233–240
- Liu Y, Kao HI, Bambara RA (2004) Flap endonuclease 1: a central component of DNA metabolism. *Annu Rev Biochem* 73:589–615
- Lyamichev V, Brow MA, Dahlberg JE (1993) Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* 260:778–783
- Lyamichev V, Brow MA, Varvel VE, Dahlberg JE (1999) Comparison of the 5' nuclease activities of taq DNA polymerase and its isolated nuclease domain. *Proc Natl Acad Sci USA* 96:6143–6148
- Mase T, Kubota K, Miyazono K, Kawarabayasi Y, Tanokura M (2009) Crystallization and preliminary X-ray analysis of flap endonuclease 1 (FEN1) from *Desulfurococcus amylolyticus*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 65:923–925
- Mesiet-Cladiere L, Norais C, Kuhn J, Briffotiaux J, Sloostra JW, Ferrari E, Hubscher U, Flament D, Myllykallio H (2007) A novel proteomic approach identifies new interaction partners for proliferating cell nuclear antigen. *J Mol Biol* 372:1137–1148
- Murante RS, Rust L, Bambara RA (1995) Calf 5' to 3' exo/endonuclease must slide from a 5' end of the substrate to perform structure-specific cleavage. *J Biol Chem* 270:30377–30383

- Navarro MS, Bi L, Bailis AM (2007) A mutant allele of the transcription factor IIIH helicase gene, RAD3, promotes loss of heterozygosity in response to a DNA replication defect in *Saccharomyces cerevisiae*. *Genetics* 176:1391–1402
- Nazarkina JK, Lavrik OI, Khodyreva SN (2008) Flap endonuclease-1 and its role in the processes of DNA metabolism in eucaryotic cells. *Mol Biol (Mosk)* 42:405–421
- Orans J, McSweeney EA, Iyer RR, Hast MA, Hellinga HW, Modrich P, Beese LS (2011) Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. *Cell* 145:212–223
- Óvádi J (1991) Physiological significance of metabolic channelling. *J Theor Biol* 152:1–22
- Parikh SS, Mol CD, Hosfield DJ, Tainer JA (1999) Envisioning the molecular choreography of DNA base excision repair. *Curr Opin Struct Biol* 9:37–47
- Pascal JM, O'Brien PJ, Tomkinson AE, Ellenberger T (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* 432:473–478
- Patel N, Attack JM, Finger LD, Exell J, Thompson P, Tsutakawa SE, Tainer JA, Williams DM, Grasby JA (2012) Flap endonucleases pass 5'-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5'-ends. *Nucl Acids Res* doi: 10.1093/nar/gks051
- Pickering TJ, Garforth SJ, Thorpe SJ, Sayers JR, Grasby JA (1999) A single cleavage assay for T5 5'→3' exonuclease: determination of the catalytic parameters for wild-type and mutant proteins. *Nucleic Acids Res* 27:730–735
- Qiu J, Qian Y, Frank P, Wintersberger U, Shen B (1999) *Saccharomyces cerevisiae* RNase H(35) functions in RNA primer removal during lagging-strand DNA synthesis, most efficiently in cooperation with Rad27 nuclease. *Mol Cell Biol* 19:8361–8371
- Reagan MS, Pittenger C, Siede W, Friedberg EC (1995) Characterization of a mutant strain of *Saccharomyces cerevisiae* with a deletion of the *RAD27* gene, a structural homolog of the *RAD2* nucleotide excision repair gene. *J Bacteriol* 177:364–371
- Robertson AB, Klungland A, Rognes T, Leiros I (2009) DNA repair in mammalian cells: base excision repair: the long and short of it. *Cell Mol Life Sci* 66:981–993
- Sakurai S, Kitano K, Yamaguchi H, Hamada K, Okada K, Fukuda K, Uchida M, Ohtsuka E, Morioka H, Hakoshima T (2005) Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J* 24:683–693
- Sakurai S, Kitano K, Morioka H, Hakoshima T (2008) Crystallization and preliminary crystallographic analysis of the catalytic domain of human flap endonuclease 1 in complex with a nicked DNA product: use of a DPCS kit for efficient protein-DNA complex crystallization. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64:39–43
- Sengerova B, Tomlinson C, Attack JM, Williams R, Sayers JR, Williams NH, Grasby JA (2010) Bronsted analysis and rate-limiting steps for the T5 flap endonuclease catalyzed hydrolysis of exonucleolytic substrates. *Biochemistry* 49:8085–8093
- Sharma S, Otterlei M, Sommers JA, Driscoll HC, Dianov GL, Kao HI, Bambara RA, Brosh RM Jr (2004) WRN helicase and FEN-1 form a complex upon replication arrest and together process branchmigrating DNA structures associated with the replication fork. *Mol Biol Cell* 15:734–750
- Sharma S, Sommers JA, Gary RK, Friedrich-Heineken E, Hubscher U, Brosh RM Jr (2005) The interaction site of Flap Endonuclease-1 with WRN helicase suggests a coordination of WRN and PCNA. *Nucleic Acids Res* 33:6769–6781
- Shen B, Qiu J, Hosfield D, Tainer JA (1998) Flap endonuclease homologs in archaeobacteria exist as independent proteins. *Trends Biochem Sci* 23:171–173
- Shen B, Singh P, Liu R, Qiu J, Zheng L, Finger LD, Alas S (2005) Multiple but dissectible functions of FEN-1 nucleases in nucleic acid processing, genome stability and diseases. *Bioessays* 27:717–729
- Solinger JA, Pascolini D, Heyer WD (1999) Active-site mutations in the Xrn1p exoribonuclease of *Saccharomyces cerevisiae* reveal a specific role in meiosis. *Mol Cell Biol* 19:5930–5942
- Stewart JA, Campbell JL, Bambara RA (2009) Significance of the dissociation of Dna2 by flap endonuclease 1 to Okazaki fragment processing in *Saccharomyces cerevisiae*. *J Biol Chem* 284:8283–8291
- Storici F, Henneke G, Ferrari E, Gordenin DA, Hubscher U, Resnick MA (2002) The flexible loop of human FEN1 endonuclease is required for flap cleavage during DNA replication and repair. *EMBO J* 21:5930–5942

- Stucki M, Jonsson ZO, Hubscher U (2001) In eukaryotic flap endonuclease 1, the C terminus is essential for substrate binding. *J Biol Chem* 276:7843–7849
- Syson K, Tomlinson C, Chapados BR, Sayers JR, Tainer JA, Williams NH, Grasby JA (2008) Three metal ions participate in the reaction catalyzed by T5 flap endonuclease. *J Biol Chem* 283:28741–28746
- Tom S, Henricksen LA, Bambara RA (2000) Mechanism whereby proliferating cell nuclear antigen stimulates flap endonuclease 1. *J Biol Chem* 275:10498–10505
- Tomkinson AE, Vijayakumar S, Pascal JM, Ellenberger T (2006) DNA ligases: structure, reaction mechanism, and function. *Chem Rev* 106:687–699
- Tomlinson CG, Attack JM, Chapados B, Tainer JA, Grasby JA (2010) Substrate recognition and catalysis by flap endonucleases and related enzymes. *Biochem Soc Trans* 38:433–437
- Tsutakawa SE, Classen S, Chapados BR, Arvai AS, Finger LD, Guenther G, Tomlinson CG, Thompson P, Sarker AH, Shen B et al (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell* 145:198–211
- Tumey LN, Huck B, Gleason E, Wang J, Silver D, Brunden K, Boozer S, Rundlett S, Sherf B, Murphy S et al (2004) The identification and optimization of 2,4-diketobutyric acids as flap endonuclease 1 inhibitors. *Bioorg Med Chem Lett* 14:4915–4918
- Tumey LN, Bom D, Huck B, Gleason E, Wang J, Silver D, Brunden K, Boozer S, Rundlett S, Sherf B et al (2005) The identification and optimization of a N-hydroxy urea series of flap endonuclease 1 inhibitors. *Bioorg Med Chem Lett* 15:277–281
- Waga S, Bauer G, Stillman B (1994) Reconstitution of complete SV40 DNA replication with purified replication factors. *J Biol Chem* 269:10923–10934
- Warbrick E, Coates PJ, Hall PA (1998) Fen1 expression: a novel marker for cell proliferation. *J Pathol* 186:319–324
- Williams R, Sengerova B, Osborne S, Syson K, Ault S, Kilgour A, Chapados BR, Tainer JA, Sayers JR, Grasby JA (2007) Comparison of the catalytic parameters and reaction specificities of a phage and an archaeal flap endonuclease. *J Mol Biol* 371:34–48
- Wilson SH, Kunkel TA (2000) Passing the baton in base excision repair. *Nat Struct Biol* 7:176–178
- Wu X, Li J, Li X, Hsieh CL, Burgers PM, Lieber MR (1996) Processing of branched DNA intermediates by a complex of human FEN-1 and PCNA. *Nucleic Acids Res* 24:2036–2043
- Xu Y, Potapova O, Leschziner AE, Grindley ND, Joyce CM (2001) Contacts between the 5' nuclease of DNA polymerase I and its DNA substrate. *J Biol Chem* 276:30167–30177
- Yang W (2010) Nucleases: diversity of structure, function and mechanism. *Q Rev Biophys* 44:1–93
- Yang W, Lee JY, Nowotny M (2006) Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol Cell* 22:5–13
- Yanofsky C (1989) A second reaction catalyzed by the tryptophan synthetase of *Escherichia coli*. 1959. *Biochim Biophys Acta* 1000:137–145
- Zheng L, Zhou M, Chai Q, Parrish J, Xue D, Patrick SM, Turchi JJ, Yannone SM, Chen D, Shen B (2005) Novel function of the flap endonuclease 1 complex in processing stalled DNA replication forks. *EMBO Rep* 6:83–89
- Zheng L, Dai H, Qiu J, Huang Q, Shen B (2007a) Disruption of the FEN-1/PCNA interaction results in DNA replication defects, pulmonary hypoplasia, pancytopenia, and newborn lethality in mice. *Mol Cell Biol* 27:3176–3186
- Zheng L, Dai H, Zhou M, Li M, Singh P, Qiu J, Tsark W, Huang Q, Kernstine K, Zhang X et al (2007b) Fen1 mutations result in autoimmunity, chronic inflammation and cancers. *Nat Med* 13:812–819
- Zheng L, Dai H, Hegde ML, Zhou M, Guo Z, Wu X, Wu J, Su L, Zhong X, Mitra S et al (2011a) Fen1 mutations that specifically disrupt its interaction with PCNA cause aneuploidy-associated cancer. *Cell Res* 21:1052–1067
- Zheng L, Jia J, Finger LD, Guo Z, Zer C, Shen B (2011b) Functional regulation of FEN1 nuclease and its link to cancer. *Nucleic Acids Res* 39:781–794

Chapter 17

DNA Ligase I, the Replicative DNA Ligase

Timothy R.L. Howes and Alan E. Tomkinson

Abstract Multiple DNA ligation events are required to join the Okazaki fragments generated during lagging strand DNA synthesis. In eukaryotes, this is primarily carried out by members of the DNA ligase I family. The C-terminal catalytic region of these enzymes is composed of three domains: a DNA binding domain, an adenylation domain and an OB-fold domain. In the absence of DNA, these domains adopt an extended structure but transition into a compact ring structure when they engage a DNA nick, with each of the domains contacting the DNA. The non-catalytic N-terminal region of eukaryotic DNA ligase I is responsible for the specific participation of these enzymes in DNA replication. This proline-rich unstructured region contains the nuclear localization signal and a PCNA interaction motif that is critical for localization to replication foci and efficient joining of Okazaki fragments. DNA ligase I initially engages the PCNA trimer via this interaction motif which is located at the extreme N-terminus of this flexible region. It is likely that this facilitates an additional interaction between the DNA binding domain and the PCNA ring. The similar size and shape of the rings formed by the PCNA trimer and the DNA ligase I catalytic region when it engages a DNA nick suggest that these proteins interact to form a double-ring structure during the joining of Okazaki fragments. DNA ligase I also interacts with replication factor C, the factor that loads the PCNA trimeric ring onto DNA. This interaction, which is regulated by phosphorylation of the non-catalytic N-terminus of DNA ligase I, also appears to be critical for DNA replication.

T.R.L. Howes

Biomedical Sciences Graduate Program, University of New Mexico,
Cancer Research Facility MSC08 4640, 1 University of New Mexico,
Albuquerque, NM 87131-0001, USA, Phone: 505-272-5405 FAX 505-925-4459
e-mail: howes.timothy@gmail.com

A.E. Tomkinson (✉)

Department of Internal Medicine and University of New Mexico Cancer Center,
Cancer Research Facility MSC08 4640, 1 University of New Mexico, Albuquerque,
NM 87131-0001, USA, Phone: 505-272-5405 FAX 505-925-4459
e-mail: atomkinson@salud.unm.edu

Keywords Lagging strand DNA synthesis • Okazaki fragments • Genome instability • Cancer predisposition • Phosphodiester bond formation

17.1 Introduction

Since DNA polymerases only synthesize DNA in the 5'–3' direction, one of the two antiparallel strands of duplex DNA is synthesized discontinuously as a series of short Okazaki fragments that are then joined by a DNA ligase to generate an intact strand. In 1967, several laboratories identified DNA ligase activity in extracts from uninfected *E. coli* cells and also from *E. coli* cells infected with bacteriophage T4 (Lehman 1974). The following year DNA ligase activity was described in extracts from mammalian cells (Soderhall and Lindahl 1976). Notably, the *Escherichia coli* DNA ligase is NAD⁺-dependent whereas the bacteriophage and mammalian DNA ligases are ATP-dependent (Lehman 1974; Soderhall and Lindahl 1976). Subsequent studies have revealed the existence of both NAD⁺- and ATP-dependent DNA ligases in prokaryotes. In contrast, eukaryotic and viral DNA ligases are almost exclusively ATP-dependent (Ellenberger and Tomkinson 2008; Tomkinson et al. 2006).

Apart from utilizing a different nucleotide co-factor, the reaction mechanisms of NAD⁺- and ATP-dependent are identical (Fig. 17.1). In the first step, the DNA ligases react with the nucleotide co-factor to form a covalent DNA ligase-adenylate complex in which the AMP moiety is linked to a specific lysine residue via a phosphoramidite bond. When the DNA ligase-adenylate engages a DNA nick with 3' HO and 5' phosphate termini, it transfers the AMP group to the 5' phosphate, forming a covalent DNA adenylate intermediate. Finally, the non-adenylated DNA ligase interacts with the DNA-adenylate, catalyzing phosphodiester bond formation and release of AMP as a result of nucleophilic attack on the 5' DNA adenylate by the 3' HO group.

The first eukaryotic DNA ligase genes were identified in screens for cell division cycle mutants in the yeasts, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (Johnston and Nasmyth 1978; Nasmyth 1977). The DNA ligases encoded by the *CDC9* gene in *Saccharomyces cerevisiae* and the *cdc17⁺* gene in *Schizosaccharomyces pombe* are required for cell viability because of their essential role in DNA replication. Biochemical and immunological characterization of DNA ligase activity in mammalian cell extracts provided the first evidence that eukaryotes contain more than one species of DNA ligase (Soderhall and Lindahl 1976). The presence of multiple DNA ligase enzymes suggested that these may have distinct cellular functions. In the remainder of this chapter, we will briefly describe the eukaryotic DNA ligases and then focus on DNA ligase I, the replicative DNA ligase, and the mechanisms that underlie the specific participation of this enzyme in DNA replication.

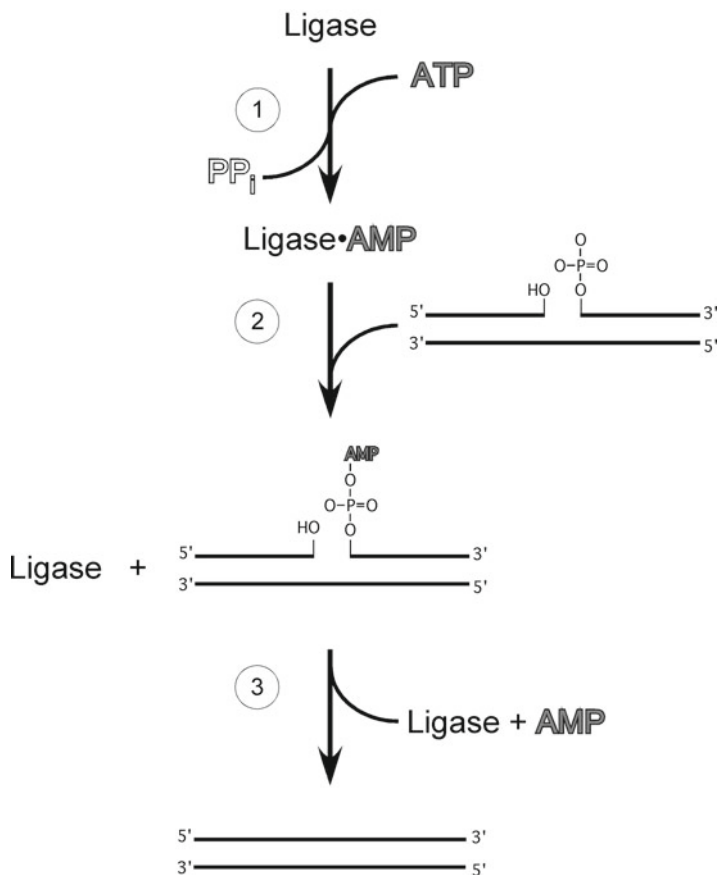


Fig. 17.1 Three-step mechanism of DNA ligation. (1) DNA ligase I binds and hydrolyses adenosine triphosphate (ATP), releasing pyrophosphate (PP_i) and a covalent ligase-adenosine monophosphate (AMP) intermediate. (2) The AMP group is subsequently transferred from the ligase polypeptide to the 5' phosphate termini of a nick in duplex DNA. (3) The non-adenylated ligase catalyzes phosphodiester bond formation in a reaction involving nucleophilic attack by the 3'HO group and release of AMP

17.2 Eukaryotic DNA Ligase Genes

As mentioned above, the DNA ligases encoded by the *CDC9* and *cdc17⁺* genes of *S. cerevisiae* and *Schiz. pombe*, respectively, were the first eukaryotic DNA ligases to be identified (Johnston and Nasmyth 1978; Nasmyth 1977). Human cDNAs that complemented the temperature-sensitive phenotype of a yeast *cdc9* strain were isolated from a human cDNA library (Barnes et al. 1990). Subsequent DNA sequencing revealed that these cDNAs encoded a polypeptide that was highly homologous

with the yeast DNA ligases and contained sequences that were identical to those of peptides from purified mammalian DNA ligase I (Barnes et al. 1990). Thus, human DNA ligase I and the yeast DNA ligases are functional homologs that belong to the DNA ligase I family of eukaryotes.

Two other mammalian genes that encode DNA ligases, *LIG3* and *LIG4*, have been identified (Chen, et al. 1995; Wei et al. 1995). Homologs of the *LIG4* gene have been found in all eukaryotes, whereas the *LIG3* gene appears to be restricted to vertebrates (Ellenberger and Tomkinson 2008). The DNA ligases encoded by the three *LIG* genes share a conserved catalytic region that is flanked by unrelated amino- and/or carboxyl-terminal regions (Fig. 17.2). There is compelling evidence that interactions with specific protein partners mediated by these unique regions flanking the catalytic domain direct the participation of the DNA ligases in different DNA transactions (Ellenberger and Tomkinson 2008).

17.3 DNA Ligase I: Molecular Genetics and Cell Biology

The increased expression of the *LIG1* gene when quiescent cells are induced to proliferate and increased levels of DNA ligase I protein and activity in proliferating cells and tissues, implicated DNA ligase I in DNA replication (Petrini et al. 1991; Soderhall and Lindahl 1976). This linkage was strengthened by studies showing that DNA ligase I co-localized with replication foci in S-phase cells (Lasko et al. 1990). The location of distinct amino acid sequences within the non-catalytic N-terminal region of DNA ligase I that function as nuclear localization and replication foci targeting sequences (Cardoso et al. 1997; Montecucco et al. 1995) are shown in Fig. 17.2. The mechanism underlying the recruitment of DNA ligase I to replication foci is described below.

A single case of human DNA ligase I-deficiency has been described. This individual, whose symptoms included retarded growth, delayed development, recurrent ear and chest infections and lymphoma, died at age 19 as result of complications following a chest infection (Webster et al. 1992). Sequencing of genomic DNA revealed the presence of two different mutant *lig1* alleles. The maternally inherited *lig1* allele encodes a DNA ligase polypeptide with reduced catalytic activity whereas the other mutant allele, whose origin is not known, encodes a DNA ligase polypeptide with essentially no catalytic activity. In both mutant *lig1* alleles, the DNA sequence change results in a single amino acid substitution within the conserved catalytic region of DNA ligase I (Barnes et al. 1992). The locations of the amino acid changes are described in the section below.

Primary (46BR) and SV40-immortalized (46BR.1G1) fibroblasts established from the DNA ligase I-deficient individual exhibit defective joining of Okazaki fragments and sensitivity to a wide range of DNA damaging agents, particularly DNA alkylating agents (Teo et al. 1983). Since both mutant *lig1* alleles are present in the primary fibroblasts whereas only the maternally inherited allele is present in the SV40-immortalized (46BR.1G1) fibroblasts, it appears that the maternal *lig1*

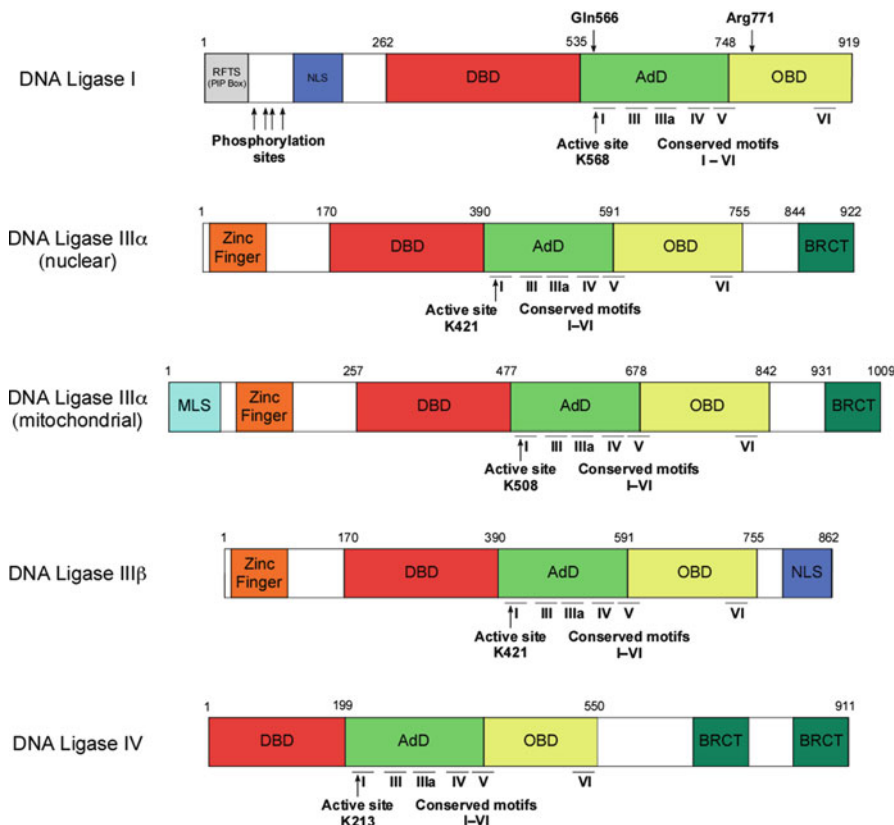


Fig. 17.2 Alignment of the polypeptides encoded by the three human *LIG* genes. The conserved catalytic regions of the human DNA ligases each contain a DNA binding (DBD, red), adenylation (AdD, green) and OB-fold (OBD, yellow) domain. The AdD and OBD, which make up the catalytic core of the nucleotidyl transferase family that also includes RNA ligases and mRNA capping enzymes, contain six highly conserved motifs (I, III, IIIa, IV, V and VI). The position of the active site lysine residue within motif I that forms the covalent bond with AMP is indicated for each DNA ligase. The non-catalytic regions that flank the DNA ligase catalytic region determine the subcellular distribution and cellular functions of the DNA ligases. The positions of the nuclear localization signals (NLS, blue) of DNA ligase I and DNA ligase III β , and the mitochondrial leader sequence (MLS, cyan) of DNA ligase III α are shown. The replication factory targeting sequence (RFTS), which is also a PCNA interacting peptide (PIP) box that targets DNA ligase I replication and interacts with PCNA, is indicated in grey. Sites of phosphorylation on serine residues within the non-catalytic N-terminal region of DNA ligase I are indicated. Amino acids, glutamine 566 and arginine 771, are replaced with lysine and tryptophan, respectively in the polypeptides encoded by the mutant *ligI* alleles of the only known DNA ligase I deficient human identified to date. All the DNA ligases encoded by the *LIG3* gene have an N-terminal zinc finger (orange) that is involved in DNA binding. Both DNA ligase III α and DNA ligase IV contain breast and ovarian cancer susceptibility protein-1 C-terminal (BRCT, dark green) domain that are involved in protein-protein interactions

allele is responsible for the individual's symptoms and the phenotype of the cell lines (Barnes et al. 1992). As expected, both the DNA replication and repair defects of the 46BR.1G1 fibroblasts are complemented by expression of wild type DNA ligase I (Levin et al. 2000).

Although the levels of DNA ligase I protein and activity are reduced by about 50% and 90%, respectively in the 46BR.1G1 fibroblasts compared with SV40-immortalized fibroblasts from a normal individual, there are no significant differences in cell cycle progression despite the defect in converting Okazaki fragments into high molecular weight DNA (Barnes et al. 1992). In fact, the results of pulse-labeling studies indicate that the majority of Okazaki fragments are degraded rather than ligated (Henderson et al. 1985; Levin et al. 2004). Thus, it appears that when DNA ligase I is not available to ligate the nick between adjacent Okazaki fragments, the downstream fragment is displaced by DNA synthesis and then degraded. This model predicts that the lagging strand in 46BR.1G1 fibroblasts is synthesized as a series of longer fragments that are joined by either the defective DNA ligase I polypeptide or one of the other DNA ligases.

Based on the results of genetic studies in the yeasts *S. cerevisiae* and *Schiz. pombe* (Johnston and Nasmyth 1978; Nasmyth 1977), it was expected that the mammalian *LIG1* gene would be essential. In accordance with this prediction, *lig1* null mouse embryonic stem cells could only be obtained when wild-type DNA ligase I cDNA was ectopically expressed (Petrini et al. 1995). Surprisingly, *lig1* null embryos generated by crossing heterozygous mice were detectable until day 16 (Bentley et al. 1996, 2002). Furthermore, it was possible to establish *lig1* null embryonic fibroblasts (MEFs) from these embryos, demonstrating that *LIG1* is not an essential gene in mouse somatic cells. Similar to the human 46BR.1G1 fibroblasts, the *lig1* null MEFs had a defect in converting Okazaki fragments into high molecular weight DNA but no defect in proliferation (Bentley et al. 1996, 2002). In contrast to the 46BR.1G1 fibroblasts, the *lig1* null MEFs have no apparent DNA repair defect (Bentley et al. 1996, 2002; Teo et al. 1983). Thus, it appears that one of the other mammalian DNA ligases can substitute for DNA ligase I in DNA replication. The presence of additional DNA ligases in mammals encoded by the *LIG3* gene provides a possible explanation as to why the *LIG1* gene homolog is essential for cell viability in yeast but not in mammals. For example, DNA ligase III α and its partner protein XRCC1 are recruited to participate in the repair of DNA single strand breaks by an interaction with the poly(ADP-ribosylated) version of poly(ADP-ribose) polymerase 1 (PARP-1), an abundant nuclear protein that binds to and is activated by DNA single strand breaks (Okano et al. 2003, 2005). Since the defect in Okazaki fragment processing caused by DNA ligase I deficiency is likely to result in relatively long-lived, single-strand interruptions on the lagging strand, it is possible that these breaks are recognized and joined by the PARP-1/DNA ligase III α single-strand break repair pathway. The sensitivity of human 46BR.1G1 fibroblasts and *lig1* null MEFs to a PARP-1 inhibitor (Lehmann et al. 1988) is consistent with the single-strand break repair pathway joining single-strand breaks between Okazaki fragments that remain after lagging strand DNA synthesis.

It is likely that cases of DNA ligase I-deficiency in humans will be extremely rare. A mouse model that reiterates the mutation in human 46BR.1G1 cells has been generated.

These animals are small and have hematopoietic defects (Harrison et al. 2002). Other notable features of these animals include increased genomic instability and an increased incidence of epithelial tumors (Harrison et al. 2002). As with the *lig1* null MEFs, the MEFs harboring the equivalent mutation to that in the human 46BR.1G1 fibroblasts have a defect in replication but not repair, suggesting that the increased genome instability and cancer incidence is due to accumulation of abnormal replication intermediates (Harrison et al. 2002). Together, these studies indicate that DNA ligase I plays an important role in DNA repair in human but not mouse cells (Harrison et al. 2002; Teo et al. 1983). It is conceivable that, while there is functional redundancy between DNA ligases I and III α in DNA repair as well as in DNA replication, the relative contribution of DNA ligase III α -dependent repair may be greater in mouse cells compared with human cells. However, recent studies with mouse *lig3* null MEFs do not support this model (Gao et al. 2011; Simsek et al. 2011).

17.4 DNA Ligase I Protein: Structure and Function

The 919 amino acid polypeptide encoded by human DNA ligase I cDNA has a highly asymmetric shape, which causes anomalous behavior during density gradient sedimentation and gel filtration experiments (Tomkinson et al. 1990). Using limited proteolysis, it was found that catalytic activity resides within a relatively protease-resistant C-terminal fragment of about 78 kDa whereas the N-terminal fragment is extremely protease sensitive, indicative of an unstructured region (Tomkinson et al. 1990). Notably, this N-terminal region is likely to have an extended, flexible conformation because it contains a large number of proline residues (Barnes et al. 1990). In addition, the high proline content of DNA ligase I (approximately 9%) results in anomalous mobility during SDS-polyacrylamide gel electrophoresis, giving DNA ligase I an apparent molecular mass of 125 kDa (Tomkinson et al. 1990).

The catalytic region of DNA ligase I contains six motifs (Fig. 17.2) that are conserved among the nucleotidyl transferase family that includes mRNA capping enzymes and RNA ligases in addition to DNA ligases (Shuman and Schwer 1995). Motif I contains the active site lysine residue to which the AMP (or GMP) residue is covalently via a phosphoramidite bond (Fig. 17.2). This residue was initially identified by sequencing an adenylylated tryptic peptide from bovine DNA ligase I (Tomkinson et al. 1991). Using this sequence, it was possible to predict the position of the putative active site lysine residues in DNA ligases, RNA ligases and mRNA capping enzymes. As expected, substitution of Lys568, the lysine residue that binds the AMP moiety in human DNA ligase I (Fig. 17.2), prevents formation of the enzyme-AMP complex and therefore abolishes enzymatic activity (Kodama et al. 1991; Tomkinson et al. 1991). Interestingly, the mutated DNA ligase I enzyme identified in the DNA ligase I-deficient individual has a lysine residue instead of a glutamic acid at position 566 (Barnes et al. 1992). This amino acid change two residues away from the active site lysine markedly reduces formation of the enzyme-AMP intermediate.

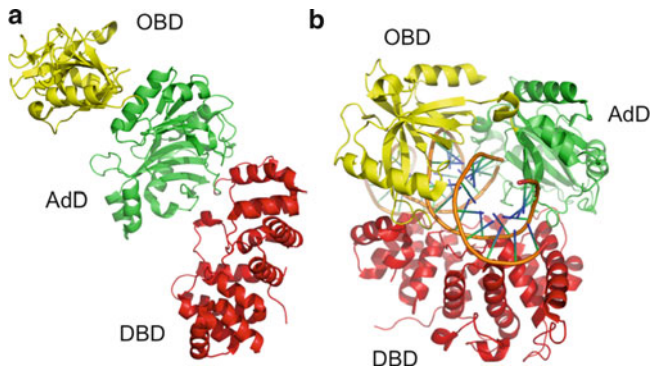


Fig. 17.3 DNA binding induces a large conformational change in the DNA ligase catalytic region. (a) In the crystal structure of *Sulfolobus solfataricus* DNA ligase obtained in the absence of DNA (Pascal et al. 2006), the DNA binding (DBD, shown in red), adenylation domain (AdD, green), and OB-fold (OBD, yellow) domains are in an extended conformation with relatively few contacts between the domains. (b) In the crystal structure of human DNA ligase I in complex with nicked DNA (Pascal et al. 2004), the DNA binding (DBD, red), adenylation domain (AdD, green) and OB-fold (OBD, yellow) domains encircle the nicked DNA forming a ring structure that is stabilized by each domains interacting with the DNA and by contacts between the DNA binding and OB-fold domains

The first nucleotidyl transferase structure to be determined was that of the DNA ligase encoded by bacteriophage T7 (Subramanya et al. 1996). This was shortly followed by the structure of the RNA-capping enzyme from PCBV-1 (Hakansson et al. 1997). These structures revealed the existence of the two domains: an adenylation/guanylation domain, which contains conserved motifs I through V, and an oligomer binding-fold (OB) domain (OBD) containing motif VI. In 2004, Pascal et al. successfully crystallized the catalytic C-terminal region of DNA ligase I (residues 233–919) bound to a nicked DNA substrate (Pascal et al. 2004). This structure revealed several novel features. Firstly, it showed that nicked DNA is encircled by DNA ligase I during catalysis (Fig. 17.3b), suggesting that the catalytic domain undergoes a large conformational change when it engages a nick. Secondly, it showed that the catalytic region of the larger eukaryotic DNA ligases contains a DNA binding domain in addition to the adenylation and OB-fold domains that make up the catalytic core (Fig. 17.2). Thus, the adenylation/guanylation and OB-fold domains constitute the conserved catalytic core of nucleotidyl transferases with the DNA binding domain (DBD) being a characteristic feature of eukaryotic DNA ligases (Pascal et al. 2004).

The DNA binding domain of DNA ligase I spans residues 262–534 (Fig. 16.2). As shown in Fig. 17.3b, these residues fold into 12 α -helices that exhibit a twofold symmetry (Pascal et al. 2004). Due to the symmetry of the DBD, the interaction with DNA occurs via one tight reverse turn of two α -helices and an extended loop formation. These loops and helices arrange to create a relatively flat surface of approximately 2,000 \AA^2 that interacts almost exclusively with the phosphodiester

backbone of the DNA substrate. The DBD interacts with the minor groove of the DNA backbone on both sides of the nick, explaining how DNA ligase I binds to DNA in a sequence-independent binding manner and why chemicals that bind to the minor groove of DNA, such as distamycin, inhibit DNA ligase activity (Montecucco et al. 1991). Notably, the DBD stimulates the weak DNA joining activity of the DNA ligase I catalytic core containing the AdD and OBD when added in *trans*, indicating that contacts between the DBD and both AdD and the OBD observed in the crystal structure stabilize the folding of the catalytic core around the DNA nick (Pascal et al. 2004).

The DNA ligase I adenylation domain, which spans from residue 535–747, contains conserved motifs I, III, IIIa, IV and V. These five motifs contribute to the surface of nucleotide binding pocket (Fig. 17.3b). Tryptophan 742 of motif V provides co-factor specificity by sterically excluding GTP. Furthermore, Arg573 and Glu621 of motifs I and III respectively, stabilize the hydroxyl groups on the ribose sugar of AMP via hydrogen bonding interactions (Pascal et al. 2004). As mentioned above, one of the two mutant *lig1* alleles identified in the individual with DNA ligase I deficiency encodes a polypeptide in which Glu566 of motif I is replaced by a lysine residue (Barnes et al. 1992). From the structure of DNA ligase I, it is evident that the Glu566 residue contributes to the specific interaction with ATP by forming a hydrogen-bond with the N6 of the adenine moiety (Pascal et al. 2004). Replacement of Glu566 with a positively charged lysine residues disrupts this, providing an explanation as to why the mutant polypeptide is defective in the first step of the ligation reaction, formation of the covalent enzyme-adenylate intermediate.

The major structural feature of the OB-fold domain is a β -barrel and similar to the other two domains, it also interacts with the minor groove of the DNA (Pascal et al. 2004). Contacts between AdD and OBD are critical for correctly positioning these domains when they engage a DNA nick. During catalysis, the AdD forms a salt bridge with the OBD via Asp570 and Arg871, stabilizing the ligase catalytic domains in a conformation in which they fully encircle the DNA nick. This positioning of the AdD and OBD creates a surface that binds to and distorts the nicked DNA. Notably, the phenylalanine residues at positions 635 and 872 of the AdD and OBD are forced into the minor groove both 3' and 5' to the nick. As a result of these interactions, the DNA duplex upstream of the nick duplex assumes an A-form structure and the nick is opened up for ligation (Pascal et al. 2004). Notably, the DNA binding site downstream of the nick is specific for B-form DNA, explaining why ligase I is not active on nicks within A-form duplexes formed by RNA duplexes and RNA-DNA hybrids (Pascal et al. 2004). This ability to discriminate against duplexes containing ribonucleotides 5' to the nick presumably prevents premature joining of Okazaki fragments before the RNA primer has been removed. The maternally inherited mutant *lig1* allele in the individual with DNA ligase I-deficiency encodes a polypeptide in which the arginine 771 within the OBD is replaced by a tryptophan residue (Barnes et al. 1992). This mutant enzyme has markedly reduced catalytic activity and is, as expected, defective in step 2 of the ligation reaction (Fig. 17.1), transfer of the AMP moiety from the ligase to the 5' phosphate termini of the DNA nick (Prigent et al. 1994).

Although eukaryotic DNA ligases have not been crystallized without nicked DNA, the structure of an ATP-dependent DNA ligase from the archaeal organism *Sulfolobus solfataricus* has been determined in the absence of DNA (Pascal et al. 2006). As shown in Fig. 17.3a, the catalytic region of the archaeal enzyme has the same three domain organization as eukaryotic DNA ligases but, in the absence of DNA, the three domains are arranged in an extended conformation. The major difference between the extended (Fig. 17.3a) and closed (Fig. 17.3b) conformations of the three domains is the position of the OBD domain relative to the other two domains (Pascal et al. 2006). Thus, the OBD undergoes a large change in conformation during the nicked DNA-dependent transition from the extended to the closed form with interactions between the OBD and DBD playing key roles in stabilizing the closed form. Based on structures of smaller DNA ligases, it appears likely that the OBD also undergoes conformational changes when the enzyme interacts with ATP to form the enzyme-adenylate, possibly reorienting the OBD to expose a DNA binding surface (Odell et al. 2000; Subramanya et al. 1996).

While the unstructured N-terminal region of DNA ligase I is dispensable for catalytic activity *in vitro*, it does appear to be essential for cellular function and viability (Mackenney et al. 1997; Petrini et al. 1995). As mentioned previously, this region contains a bipartite nuclear localization signal located between residues 111 and 179 and a sequence that is required for targeting to replication factories, residues 2–9 (Cardoso et al. 1997; Montecucco et al. 1995). In addition, the N-terminal region is phosphorylated on several serine/threonine residues (Fig. 17.2) by casein kinase II and cyclin-dependent kinases during cell cycle progression (Ferrari et al. 2003; Frouin et al. 2002; Prigent et al. 1992). This results in a hyperphosphorylated form of DNA ligase I in M-phase cells. It appears likely that these phosphorylation events regulate the participation of DNA ligase I in DNA replication because phosphorylation site mutants fail to correct the DNA replication defect of DNA ligase I-deficient 46BR cells (Soza et al. 2009; Vijayakumar et al. 2009).

17.5 DNA Ligase I: Protein Interactions

Proteins are directed to participate in complex DNA transactions, such as DNA replication, by specific protein-protein interactions. Proliferating cell nuclear antigen (PCNA), the eukaryotic homotrimeric DNA sliding clamp (see Chap. 15, this volume) that functions as a processivity factor for the replicative DNA polymerases, was the first DNA ligase I-interacting protein to be identified (Levin et al. 1997), and mediates DNA ligase I interaction with the rest of the eukaryotic replisome. Residues 2–9 within the non-catalytic N-terminal region of DNA ligase I constitute the major PCNA binding site within DNA ligase I (Montecucco et al. 1998). Notably, this same sequence, which is homologous to a PCNA-interacting protein motif, or ‘PIP box,’ that has been identified in a large number of proteins, is required for the recruitment of DNA ligase I to replication factories (Montecucco et al. 1998). Amino acid changes that disrupt PCNA binding abolish both the recruitment of

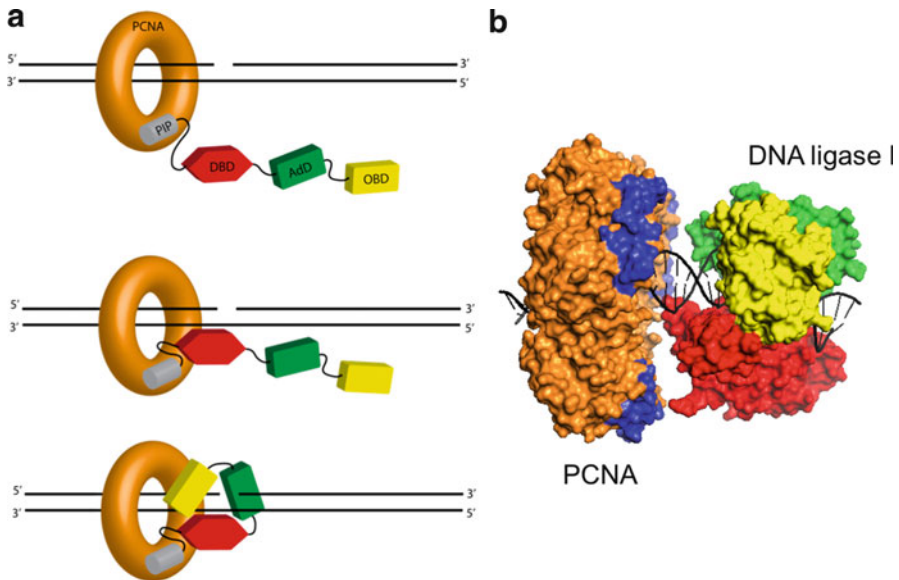


Fig. 17.4 Models for the interaction between PCNA and DNA ligase I on nicked DNA. **(a)** DNA ligase I initially engages a PCNA trimer (*orange*) by an interaction between the DNA ligase I PIP box (*grey*) and the interdomain connector loop (IDCL) region of the one of PCNA subunits. This docking facilitates a lower affinity interaction between the DNA ligase I DNA binding domain (*red*) and the PCNA trimer. At this stage, the DNA ligase I catalytic region remains in an extended conformation. This complex of PCNA and DNA ligase I slides freely along the duplex until it encounters a nick. The catalytic region of DNA ligase I then encircles the DNA nick. Interactions with the face of the PCNA trimer are likely to facilitate the transition of the catalytic region from the extended conformation into the compact ring structure. Domains of DNA ligase I are coloured as Figure 17.3. **(b)** A theoretical space-filling model of the double ring structure formed by PCNA (shown predominantly in *orange* but with the IDCL highlighted in blue; PDB: 1AXC) and the catalytic region of DNA ligase I (PDB: 1X9N) on nicked duplex DNA

DNA ligase I to replication factories and the correction of the DNA replication defect in 46BR cells, demonstrating the critical role of this interaction in the sub-nuclear targeting of DNA ligase I and the efficient joining of Okazaki fragments (Levin et al. 2000; Montecucco et al. 1998). Since the PIP box motif binds to the interdomain connector loop of PCNA (Vijayakumar et al. 2007), it appears likely that, when the flexible N-terminal region of DNA ligase I initially binds to the interdomain connector loop of PCNA trimer, the catalytic region remains in an extended conformation (Fig. 17.4a) (Pascal et al. 2006).

DNA ligase I stably interacts with PCNA trimers that are topologically linked to duplex DNA with only one molecule of DNA ligase I bound per PCNA trimer (Levin et al. 1997). This suggests that the other potential binding sites are occluded because of dynamic conformational changes in DNA ligase I as a consequence of its flexible, extended structure. Alternatively, it is possible that the initial docking of DNA ligase I with a PCNA trimer via the PIP box facilitates lower affinity interactions that extend

the protein-protein interaction interface. In support of this idea, the DBD binds weakly to the subunit-subunit interface region of homotrimeric PCNA (Song et al. 2009). Interestingly, the DBD also mediates the interaction with human Rad9-Rad1-Hus1, a heterotrimeric clamp involved in cell cycle checkpoints (Song et al. 2009) and heterotrimeric PCNA from *Sulfolobus solfataricus* (Pascal et al. 2006). Given the similarity in size and shape between the PCNA ring and the ring structure formed when the catalytic region of DNA ligase I engages a DNA nick (Fig. 17.4b), the PCNA ring may facilitate the transition of the catalytic region of DNA ligase I from the extended conformation to the compact ring structure. It has been proposed that the DBD, which provides the majority of the DNA binding affinity, serves as a pivot during this transition after initial docking via the PIP box (Fig. 17.4a). Unlike studies on the interaction between DNA ligase I and the heterotrimeric Rad9-Rad1-Hus1 checkpoint DNA sliding clamp (Song et al. 2007; Wang et al. 2006), there are contradictory reports as to whether the interaction with PCNA stimulates nick joining by DNA ligase I (Levin et al. 1997, 2004; Tom et al. 2001).

DNA ligase I also functionally interacts with two other DNA replication proteins, replication protein A (RPA), the heterotrimeric complex that binds to single stranded DNA (see Chap. 10, this volume, and Ranalli et al. 2002), and replication factor C (RFC), a heteropentameric complex that loads PCNA onto DNA (see Chap. 14, this volume, and Levin et al. 2004). Although a direct physical interaction between RPA and DNA ligase I has not been demonstrated, RPA specifically stimulates the rate of catalysis by DNA ligase I (Ranalli et al. 2002). In contrast to RPA, RFC inhibits DNA joining by DNA ligase I (Levin et al. 2004). This interaction and inhibition, which involves the large subunit of RFC, p140/RFC1, is abolished by replacement of the four serine residues that are phosphorylated in DNA ligase I with glutamic acid residues (Vijayakumar et al. 2009). Notably, the inhibition of DNA ligase I by RFC can also be alleviated by inclusion of PCNA in the reaction, providing that DNA ligase I has a functional PIP box (Vijayakumar et al. 2009). Unlike the interaction with RFC, DNA ligase I binding to PCNA is not modulated by phosphorylation (Vijayakumar et al. 2009). Thus, the failure of the phosphorylation site mutant of DNA ligase I to complement the replication defect in DNA ligase I-deficient cells may be due to the disrupted interaction with RFC (Vijayakumar et al. 2009). Although these studies indicate that physical and functional interactions among DNA ligase I, RFC and PCNA are critical for DNA replication, the mechanisms by which these interactions contribute to Okazaki fragment processing and joining are not fully understood.

17.6 Concluding Remarks

The catalytic region of DNA ligase I encircles and ligates the nicks between adjacent Okazaki fragments during lagging strand DNA synthesis. Although there is compelling evidence that interactions with PCNA and almost certainly RFC are critical for the specific participation of DNA ligase I in DNA replication, further

studies are needed to delineate the precise molecular mechanisms by which these interactions contribute to the co-ordinated processing and joining of Okazaki fragments and how these interactions are regulated by phosphorylation of DNA ligase I. The cancer predisposition exhibited by a mouse model of DNA ligase I-deficiency highlight the importance of this co-ordination and regulation in preventing genome instability during DNA replication.

References

- Barnes DE, Johnston LH, Kodama K, Tomkinson AE, Lasko DD, Lindahl T (1990) Human DNA ligase I cDNA: cloning and functional expression in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 87:6679–6683
- Barnes DE, Tomkinson AE, Lehmann AR, Webster AD, Lindahl T (1992) Mutations in the DNA ligase I gene of an individual with immunodeficiencies and cellular hypersensitivity to DNA-damaging agents. Cell 69:495–503
- Bentley D, Selfridge J, Millar JK, Samuel K, Hole N, Ansell JD, Melton DW (1996) DNA ligase I is required for fetal liver erythropoiesis but is not essential for mammalian cell viability. Nat Genet 13:489–491
- Bentley DJ, Harrison C, Ketchen AM, Redhead NJ, Samuel K, Waterfall M, Ansell JD, Melton DW (2002) DNA ligase I null mouse cells show normal DNA repair activity but altered DNA replication and reduced genome stability. J Cell Sci 115:1551–1561
- Cardoso MC, Joseph C, Rahn HP, Reusch R, Nadal-Ginard B, Leonhardt H (1997) Mapping and use of a sequence that targets DNA ligase I to sites of DNA replication *in vivo*. J Cell Biol 139:579–587
- Chen J, Tomkinson AE, Ramos W, Mackey ZB, Danehower S, Walter CA, Schultz RA, Besterman JM, Husain I (1995) Mammalian DNA ligase III: molecular cloning, chromosomal localization, and expression in spermatocytes undergoing meiotic recombination. Mol Cell Biol 15:5412–5422
- Ellenberger T, Tomkinson AE (2008) Eukaryotic DNA ligases: structural and functional insights. Annu Rev Biochem 77:313–338
- Ferrari G, Rossi R, Arosio D, Vindigni A, Biamonti G, Montecucco A (2003) Cell cycle-dependent phosphorylation of human DNA ligase I at the cyclin-dependent kinase sites. J Biol Chem 278:37761–37767
- Frouin I, Montecucco A, Biamonti G, Hubscher U, Spadari S, Maga G (2002) Cell cycle-dependent dynamic association of cyclin/Cdk complexes with human DNA replication proteins. EMBO J 21:2485–2495
- Gao Y, Katyal S, Lee Y, Zhao J, Rehg JE, Russell HR, McKinnon PJ (2011) DNA ligase III is critical for mtDNA integrity but not Xrcc1-mediated nuclear DNA repair. Nature 471:240–244
- Hakansson K, Doherty AJ, Shuman S, Wigley DB (1997) X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. Cell 89:545–553
- Harrison C, Ketchen AM, Redhead NJ, O'Sullivan MJ, Melton DW (2002) Replication failure, genome instability, and increased cancer susceptibility in mice with a point mutation in the DNA ligase I gene. Cancer Res 62:4065–4074
- Henderson LM, Arlett CF, Harcourt SA, Lehmann AR, Broughton BC (1985) Cells from an immunodeficient patient (46BR) with a defect in DNA ligation are hypomutable but hypersensitive to the induction of sister chromatid exchanges. Proc Natl Acad Sci U S A 82:2044–2048
- Johnston LH, Nasmyth KA (1978) *Saccharomyces cerevisiae* cell cycle mutant *cdc9* is defective in DNA ligase. Nature 274:891–893

- Kodama K, Barnes DE, Lindahl T (1991) *In vitro* mutagenesis and functional expression in *Escherichia coli* of a cDNA encoding the catalytic domain of human DNA ligase I. *Nucleic Acids Res* 19:6093–6099
- Lasko DD, Tomkinson AE, Lindahl T (1990) Mammalian DNA ligases. Biosynthesis and intracellular localization of DNA ligase I. *J Biol Chem* 265:12618–12622
- Lehman IR (1974) DNA ligase: structure, mechanism, and function. *Science* 186:790–797
- Lehmann AR, Willis AE, Broughton BC, James MR, Steingrimsdottir H, Harcourt SA, Arlett CF, Lindahl T (1988) Relation between the human fibroblast strain 46BR and cell lines representative of Bloom's syndrome. *Cancer Res* 48:6343–6347
- Levin DS, Bai W, Yao N, O'Donnell M, Tomkinson AE (1997) An interaction between DNA ligase I and proliferating cell nuclear antigen: implications for Okazaki fragment synthesis and joining. *Proc Natl Acad Sci U S A* 94:12863–12868
- Levin DS, McKenna AE, Motycka TA, Matsumoto Y, Tomkinson AE (2000) Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair. *Curr Biol* 10:919–922
- Levin DS, Vijayakumar S, Liu X, Bermudez VP, Hurwitz J, Tomkinson AE (2004) A conserved interaction between the replicative clamp loader and DNA ligase in eukaryotes: implications for Okazaki fragment joining. *J Biol Chem* 279:55196–55201
- Mackenny VJ, Barnes DE, Lindahl T (1997) Specific function of DNA ligase I in simian virus 40 DNA replication by human cell-free extracts is mediated by the amino-terminal non-catalytic domain. *J Biol Chem* 272:11550–11556
- Montecucco A, Fontana M, Foche F, Lestingi M, Spadari S, Ciarrocchi G (1991) Specific inhibition of human DNA ligase adenylation by a distamycin derivative possessing antitumor activity. *Nucleic Acids Res* 19:1067–1072
- Montecucco A, Savini E, Weighardt F, Rossi R, Ciarrocchi G, Villa A, Biamonti G (1995) The N-terminal domain of human DNA ligase I contains the nuclear localization signal and directs the enzyme to sites of DNA replication. *EMBO J* 14:5379–5386
- Montecucco A, Rossi R, Levin DS, Gary R, Park MS, Motycka TA, Ciarrocchi G, Villa A, Biamonti G, Tomkinson AE (1998) DNA ligase I is recruited to sites of DNA replication by an interaction with proliferating cell nuclear antigen: identification of a common targeting mechanism for the assembly of replication factories. *EMBO J* 17:3786–3795
- Nasmyth KA (1977) Temperature-sensitive lethal mutants in the structural gene for DNA ligase in the yeast *Schizosaccharomyces pombe*. *Cell* 12:1109–1120
- Odell M, Sriskanda V, Shuman S, Nikolov DB (2000) Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining. *Mol Cell* 6:1183–1193
- Okano S, Lan L, Caldecott KW, Mori T, Yasui A (2003) Spatial and temporal cellular responses to single-strand breaks in human cells. *Mol Cell Biol* 23:3974–3981
- Okano S, Lan L, Tomkinson AE, Yasui A (2005) Translocation of XRCC1 and DNA ligase IIIalpha from centrosomes to chromosomes in response to DNA damage in mitotic human cells. *Nucleic Acids Res* 33:422–429
- Pascal JM, O'Brien PJ, Tomkinson AE, Ellenberger T (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* 432:473–478
- Pascal JM, Tsodikov OV, Hura GL, Song W, Cotner EA, Classen S, Tomkinson AE, Tainer JA, Ellenberger T (2006) A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. *Mol Cell* 24:279–291
- Petrini JH, Huwiler KG, Weaver DT (1991) A wild-type DNA ligase I gene is expressed in Bloom's syndrome cells. *Proc Natl Acad Sci U S A* 88:7615–7619
- Petrini JH, Xiao Y, Weaver DT (1995) DNA ligase I mediates essential functions in mammalian cells. *Mol Cell Biol* 15:4303–4308
- Prigent C, Lasko DD, Kodama K, Woodgett JR, Lindahl T (1992) Activation of mammalian DNA ligase I through phosphorylation by casein kinase II. *EMBO J* 11:2925–2933
- Prigent C, Satoh MS, Daly G, Barnes DE, Lindahl T (1994) Aberrant DNA repair and DNA replication due to an inherited enzymatic defect in human DNA ligase I. *Mol Cell Biol* 14:310–317

- Ranalli TA, DeMott MS, Bambara RA (2002) Mechanism underlying replication protein a stimulation of DNA ligase I. *J Biol Chem* 277:1719–1727
- Shuman S, Schwer B (1995) RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl transferases. *Mol Microbiol* 17:405–410
- Simsek D, Furda A, Gao Y, Artus J, Brunet E, Hadjantonakis AK, Van Houten B, Shuman S, McKinnon PJ, Jasin M (2011) Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair. *Nature* 471:245–248
- Soderhall S, Lindahl T (1976) DNA ligases of eukaryotes. *FEBS Lett* 67:1–8
- Song W, Levin DS, Varkey J, Post S, Bermudez VP, Hurwitz J, Tomkinson AE (2007) A conserved physical and functional interaction between the cell cycle checkpoint clamp loader and DNA ligase I of eukaryotes. *J Biol Chem* 282:22721–22730
- Song W, Pascal J, Ellenberger T, Tomkinson AE (2009) The DNA binding domain of human DNA ligase I interacts with both nicked DNA and the DNA sliding clamps, PCNA and hRad9-hRad1-hHus1. *DNA Repair (Amst)* 8:912–919
- Soza S, Leva V, Vago R, Ferrari G, Mazzini G, Biamonti G, Montecucco A (2009) DNA ligase I deficiency leads to replication-dependent DNA damage and impacts cell morphology without blocking cell cycle progression. *Mol Cell Biol* 29:2032–2041
- Subramanya HS, Doherty AJ, Ashford SR, Wigley DB (1996) Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell* 85:607–615
- Teo IA, Arlett CF, Harcourt SA, Priestley A, Broughton BC (1983) Multiple hypersensitivity to mutagens in a cell strain (46BR) derived from a patient with immuno-deficiencies. *Mutat Res* 107:371–386
- Tom S, Henricksen LA, Park MS, Bambara RA (2001) DNA ligase I and proliferating cell nuclear antigen form a functional complex. *J Biol Chem* 276:24817–24825
- Tomkinson AE, Lasko DD, Daly G, Lindahl T (1990) Mammalian DNA ligases. Catalytic domain and size of DNA ligase I. *J Biol Chem* 265:12611–12617
- Tomkinson AE, Totty NF, Ginsburg M, Lindahl T (1991) Location of the active site for enzyme-adenylate formation in DNA ligases. *Proc Natl Acad Sci U S A* 88:400–404
- Tomkinson AE, Vijayakumar S, Pascal JM, Ellenberger T (2006) DNA ligases: structure, reaction mechanism, and function. *Chem Rev* 106:687–699
- Vijayakumar S, Chapados BR, Schmidt KH, Kolodner RD, Tainer JA, Tomkinson AE (2007) The C-terminal domain of yeast PCNA is required for physical and functional interactions with Cdc9 DNA ligase. *Nucleic Acids Res* 35:1624–1637
- Vijayakumar S, Dziegielewska B, Levin DS, Song W, Yin J, Yang A, Matsumoto Y, Bermudez VP, Hurwitz J, Tomkinson AE (2009) Phosphorylation of human DNA ligase I regulates its interaction with replication factor C and its participation in DNA replication and DNA repair. *Mol Cell Biol* 29:2042–2052
- Wang W, Lindsey-Boltz LA, Sancar A, Bambara RA (2006) Mechanism of stimulation of human DNA ligase I by the Rad9-Rad1-Hus1 checkpoint complex. *J Biol Chem* 281:20865–20872
- Webster AD, Barnes DE, Arlett CF, Lehmann AR, Lindahl T (1992) Growth retardation and immunodeficiency in a patient with mutations in the DNA ligase I gene. *Lancet* 339:1508–1509
- Wei YF, Robins P, Carter K, Caldecott KW, Papin DJC, Yu G-L, Wang R-P, Shell BK, Nash RA, Schar P, Barnes DE, Haseltine WA, Lindahl T (1995) Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and genetic recombination. *Mol Cell Biol* 15:3206–3216

Index

A

- AAA+ protein, 5, 8, 41, 62, 65, 91, 101, 102, 114, 118, 119, 241, 262–264, 275
- Accessory subunit, 163, 218, 224, 232, 233, 238, 244, 246
- AdD. *See* Adenylation domain (AdD)
- Adenylation domain (AdD), 321, 331, 334, 335
- Archaea, 3, 5, 8, 9, 11, 12, 20, 21, 25–28, 30, 31, 42, 59–62, 67, 89–91, 97, 106, 117, 149, 152, 153, 173, 202, 226, 238, 241, 261, 282, 283, 315
- ATPase, 8, 39, 41–43, 45, 48, 62, 90, 101–106, 114, 117–119, 122, 125, 128, 147, 151, 261, 262, 270
- ATP binding, 8, 41, 48, 63, 65, 101–103, 114, 115, 119, 260, 263, 264, 267–271, 274
- ATP hydrolysis, 43, 45, 62, 118, 122, 124, 246, 270–273, 275

C

- Cancer predisposition, 339
- Catalytic subunit, 7, 27, 74, 76, 140, 159–162, 218–224, 228, 229, 238, 240, 244, 264
- Cdc1, 218, 227–230
- Cdc6, 3–5, 8, 11, 28, 42–44, 50, 59–67, 71, 72, 79, 80, 84, 89, 90, 95, 115, 116, 124, 136, 151, 198, 246, 262
- Cdc23, 200
- Cdc27, 218, 227–229, 250
- Cdc45, 3, 5, 6, 25, 27, 90, 107, 115, 116, 120–124, 127, 128, 136–141, 147, 152, 153, 198–200, 202, 250

- Cdc45-MCM-GINS (CMG) complex, 90, 147
- Cdm1, 218, 229, 230
- Cdt1, 3–5, 29, 41, 42, 44, 45, 50, 59, 61, 71–84, 89, 106, 115, 116, 124, 136, 198, 202
- Cell cycle, 1, 3–5, 8, 10, 42, 44, 45, 61, 62, 72–74, 89, 116, 124, 135, 136, 140, 142, 172, 173, 187–189, 198–200, 202, 213, 228, 232, 237, 243, 246, 251, 282, 302, 332, 336, 338
- Clamp loader, 8–10, 26, 27, 60, 198, 249, 259–275
- Coiled-coil domain, 75–77, 203
- Comparative genomics, 20, 21, 24, 25

D

- DNA helicase, 10, 39, 90, 120, 136, 141, 172, 204, 232
- DNA ligase I (Lig I), 8, 27, 60, 250, 286, 318, 327
- DNA polymerase α -primase (Pol α -primase), 6, 10, 172
- DNA polymerase δ (Pol δ), 6, 7, 10, 140, 172, 217
- DNA polymerase ϵ (Pol ϵ), 6, 7, 137, 172, 237
- DNA unwinding, 5, 6, 10, 25, 27, 90, 97–99, 105, 106, 114, 117, 118, 120–122, 124, 125, 128, 129, 187, 198, 209
- DNA unwinding element (DUE), 39, 92
- Double hexamer formation, 91, 128, 209
- Dpb2, 137, 241–242, 244–247, 251
- Dpb3, 26, 28, 137, 242–246
- Dpb4, 3, 26, 242–246, 251
- Dpb11, 3, 6, 12, 29, 127, 128, 136–139, 198, 243, 247, 251
- DUE. *See* DNA unwinding element (DUE)

E

Evolution, 4, 20, 27, 30, 31, 39, 49, 137, 145, 149–150, 160, 165, 173–175, 185, 204, 218, 230, 261

F

FEN1. *See* Flap endonuclease 1 (FEN1)
[Fe4S4], 161, 162

Flap endonuclease 1 (FEN1), 8, 9, 25, 27, 60, 163, 172, 228, 249, 250, 274, 285, 287, 288, 291, 303–305, 307, 308, 310, 313, 318–320

5'-Flap structure, 307, 310, 314

G

Gate model, 120, 121, 123

Geminin, 3–5, 29, 71–84

Genome instability, 183, 189, 221, 232, 250, 252, 296, 304, 333, 339

GIN5, 3, 28, 90, 115, 135, 198, 247

G-quadruplex, 181, 183–185

H

Helicase, 4, 24, 39, 59, 89, 113, 136, 160, 172, 198, 232, 247, 262, 282, 303

Helicase loader, 43

Helix-2 insert (H2I), 96, 103, 114, 115

Hexameric helicase, 91, 95, 97, 119, 120, 124, 128

I

Iron-sulphur cluster, 7, 160–162

L

Lagging strand, 1, 5, 7, 8, 27, 97, 158, 162, 172, 185, 198, 205, 211, 247–251, 266, 273, 286, 302, 303, 305, 332, 338

Last common eukaryotic ancestor (LCEA), 20, 21, 24, 27–31

Leading strand, 1, 5, 7, 27, 97, 117, 137, 158, 162, 172, 184, 198, 205, 211, 217, 247, 248, 250–252, 273, 302

Licensing, 4, 72–73, 78, 81, 83, 84, 136, 140, 198

M

MCM. *See* Minichromosome maintenance (MCM)

Minichromosome maintenance

(MCM), 3, 24, 43, 59, 72, 89, 114, 136, 198, 251

helicase, 5, 11, 12, 60, 67, 89, 90, 95, 105, 141, 151, 153, 202, 205, 251

Mcm2, 5, 25, 96, 114, 147, 198

Mcm3, 114, 117–119, 123, 147

Mcm4, 99, 114, 118, 123, 198, 202

Mcm5, 95, 96, 99, 114, 117–120, 123, 147, 202

Mcm6, 79, 80, 114, 118, 119, 202, 251

Mcm7, 5, 114, 118, 119, 202

Mcm8, 24–26, 28, 209

Mcm9, 24–26, 80, 209

Mcm10, 6, 24, 116, 141, 197

N

Nucleotide unpairing, 311, 313, 315–317

O

OB-fold/OB domain, 182, 185, 331, 334, 335
(found as OB fold domain)

Okazaki fragment maturation, 9, 249, 302, 303, 308, 318–322

Opisthokonta, 21, 25, 28, 29, 31

ORC. *See* Origin recognition complex (ORC)

Orc1/Cdc6, 3, 42, 44, 59–67, 151

Order-disorder transition, 315, 318

Origin recognition complex (ORC), 3–5, 8, 12, 19, 28, 29, 37–50, 59, 60, 65, 71, 80, 84, 89–91, 93, 95, 96, 107, 115, 124, 125, 198, 200, 202, 247

ORC1, 3, 4, 28, 39–45, 48, 49, 59–67, 79–81, 116, 247

ORC2, 28, 40–43, 48, 49, 60, 79, 202

ORC3, 29, 40–42, 48, 49, 60

ORC4, 4, 28, 40–42, 44, 47–49, 247

ORC5, 28, 40–43, 45, 46, 48, 49

ORC6, 3, 4, 28, 29, 39–44, 48, 49, 59, 80

P

PCNA. *See* Proliferating cell nuclear antigen (PCNA)

Phosphodiester bond formation, 328, 329

Phosphorylation, 6, 9, 29, 42, 46, 79, 82, 99, 124, 126, 128, 136, 138, 152, 173, 188, 189, 198–200, 202, 232, 242, 250, 251, 274, 291, 320, 331, 336, 338, 339

Phylogeny, 21
 PIP motif, 8, 267, 285–290, 296
 Pol2, 238–241, 244–248, 251
 Pol3, 7, 218, 228, 229, 248, 249
 Pol31, 227, 230
 Pol32, 250
 Proliferating cell nuclear antigen (PCNA), 3,
 25, 60, 78, 90, 148, 162,
 172, 198, 218, 241, 260,
 281–296, 302, 331
 phosphorylation, 9, 79, 291, 320, 331, 338
 SUMOylation, 282, 291, 293
 ubiquitylation, 8, 294
 Pre-LC. *See* Pre-loading complex (pre-LC)
 Pre-loading complex (pre-LC), 6, 127, 136,
 139, 247
 Pre-RC. *See* Pre-replicative
 complex (pre-RC)
 Pre-replicative complex (pre-RC), 1, 4–6, 11,
 38, 39, 42–44, 48, 71, 72, 79, 89,
 90, 93, 125, 127, 128, 136, 138,
 139, 198–200, 202, 209, 211, 212
 Primase, 6–8, 10, 25–27, 60, 97, 117, 136,
 142, 149–151, 157–167, 172,
 186–188, 198, 200, 217, 266, 302
 Primer, 6–8, 149, 158, 160–162, 165–167,
 172, 181, 187, 198, 218, 219, 221,
 231, 232, 237–239, 244, 248–250,
 252, 266, 274, 302, 304, 307, 312,
 314, 335
 Processivity, 8, 10, 99, 105, 187, 223, 244,
 246, 249, 259, 273, 275, 282,
 286, 336
 Processivity factor, 8, 10, 246, 275, 336
 Protein motor, 90
 Protein-protein interaction, 79, 165, 173, 175,
 179, 182, 187–189, 204, 226,
 318–319, 331, 336, 338
 Psf1, 3, 121, 137, 138, 140, 142–147, 149, 247
 Psf2, 121, 123, 137–141, 143–147, 149, 150,
 152
 Psf3, 121, 123, 137–140, 143–149

R
 RecQ, 187, 202, 319
 Replication factor C (RFC), 8–10, 27, 60, 172,
 186, 187, 198, 231, 249, 259–275,
 288, 289, 338
 Replication initiator, 43, 49
 Replication licensing, 4, 72–73, 136, 140
 Replication origin(s), 3–5, 10, 11, 38, 39, 43,
 44, 47, 60, 61, 67, 83, 91, 115, 128,
 136, 138–140, 142, 151, 172, 200,
 203, 212, 247

Replication protein A (RPA), 6, 10, 27–30,
 71–189, 198, 200, 205, 226, 231,
 251, 303, 338
 RPA1, 25, 173–182, 185–189
 RPA2, 28, 173–182, 185, 186, 188
 RPA3, 173, 174, 176–178, 180,
 181, 185
 Replisome progression complex (RPC), 6,
 122, 136, 137, 141, 142, 212
 RFC. *See* Replication factor C (RFC)
 RNase HIII, 8, 9
 RPA. *See* Replication protein A (RPA)
 RPC. *See* Replisome progression
 complex (RPC)

S

Single-stranded DNA binding protein (SSB),
 27, 173, 187, 189
 Sld2, 3, 6, 12, 29, 123, 124, 127–129,
 136–139, 198, 247, 251
 Sld3, 3, 6, 12, 29, 123, 124, 127–129,
 136–140, 198
 Sld5, 116, 121, 123, 137, 138, 140,
 142–149
 Sliding clamp, 8, 10, 25–27, 60, 172, 198,
 245, 259–261, 266, 267, 273, 282,
 283, 294, 318, 336, 338
 SSB. *See* Single-stranded DNA binding
 protein (SSB)
 Structure-specific nuclease, 305
 SUMO, 282, 293–296, 298
 Supergroup, 20–26, 28–31

T

T-antigen, 10, 105, 119, 124, 125, 172, 173,
 186, 187

U

Ubiquitin ligase, 48, 73, 74, 79, 82, 293

W

WHD. *See* Winged-helix (wH) domain
 (WHD)
 Winged-helix (wH) domain (WHD),
 42, 62, 77, 79–82

Z

Zinc binding module, 163
 Zinc finger, 174, 178, 201, 203, 205–207, 211,
 222, 241, 246, 331