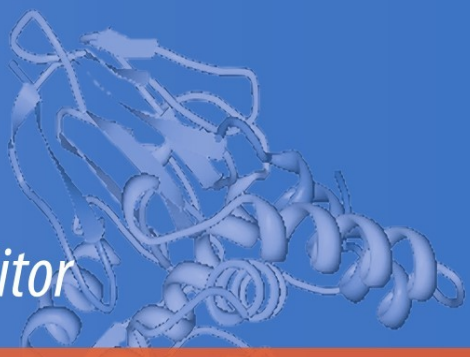
Subcellular Biochemistry 62



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 MarkusWyss

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 Rybakin

Volume 49 **Lipids in Health and Disease**
Edited by Peter J. Quinn and XiaoyuanWang

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

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 postgraduate 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

Chapter 1
Composition and Dynamics of the Eukaryotic Replisome: A Brief Overview

Stuart MacNeill

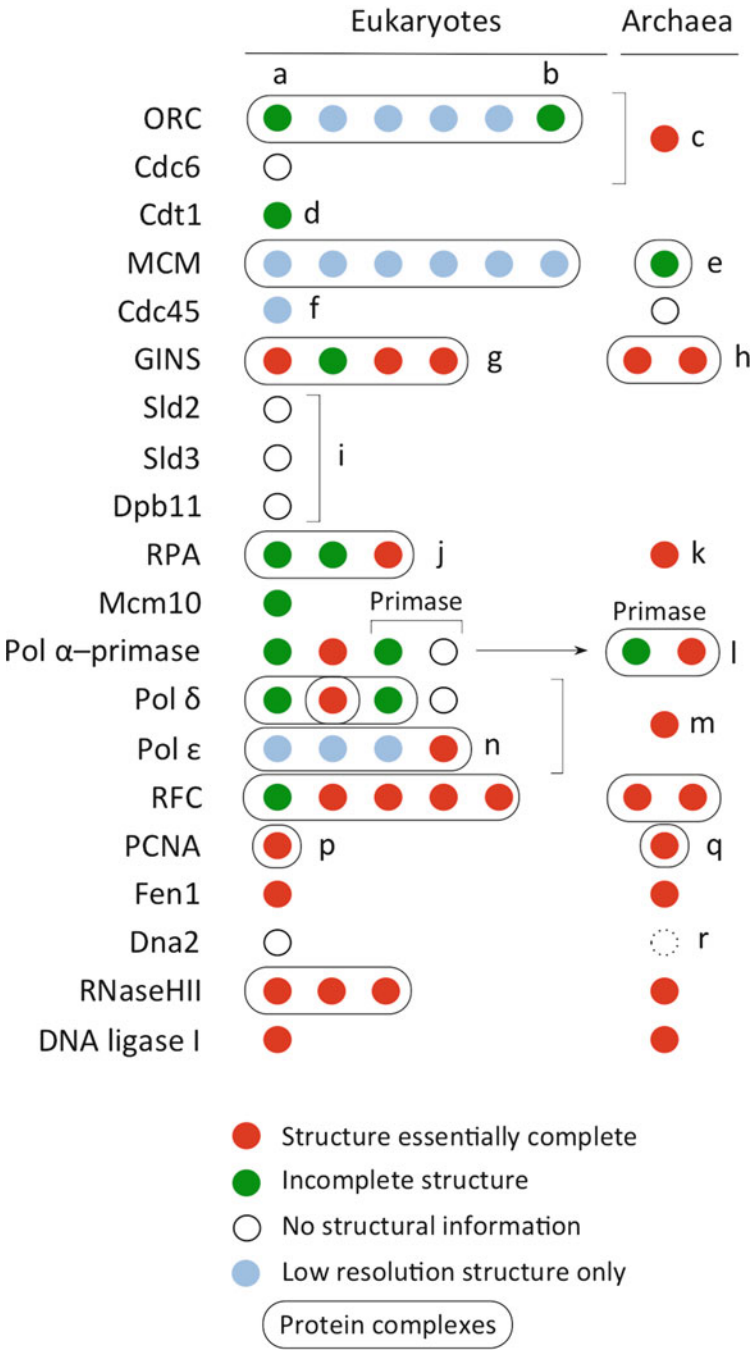
Abstract High-fidelity chromosomal DNA replication is vital for maintaining the integrity of the genetic material in all forms of cellular life. In eukaryotic cells, around 40–50 distinct conserved polypeptides are essential for chromosome replication, the majority of which are themselves component parts 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

S. MacNeill (ed.), *The Eukaryotic Replisome: A Guide to Protein Structure and Function*, Subcellular Biochemistry 62, DOI 10.1007/978-94-007-4572-8_1,


Eukaryotes
Archaea
a
b
ORC
Cdc6
Cdt1
MCM
Cdc45
GINS
Sld2
Sld3
Dpb11
RPA
Mcm10
Pol α–primase
Pol δ
Pol ε
RFC
PCNA
Fen1
Dna2
RNaseHII
DNA ligase I
c
d
e
f
g
h
i
j
k
Primase
Primase
l
m
n
p
q
r
Structure essentially complete
Incomplete structure
No structural information
Low resolution structure only
Protein complexes

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 Cvetic 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 $Gins15_2Gins23_2$ tetrameric GINS has been solved – other archaeal GINS complexes are $Gins15_4$ 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 been 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 Pol1, 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 progress 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, RNAseHII 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 RNAseHII 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 RNaseHII 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 replicative 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 heterohexamer in eukaryotes, is homohexeric 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, Vassylyev 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

Guernsey 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, 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

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 HII: 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

Watase 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

S. MacNeill (ed.), *The Eukaryotic Replisome: A Guide to Protein Structure and Function*, Subcellular Biochemistry 62, DOI 10.1007/978-94-007-4572-8_2,

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 opithokont 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 (Bapteste 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 castellani* 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 tricornutum* 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