# The Mechanisms of DNA Replication

*Edited by David Stuart*

# THE MECHANISMS OF DNA REPLICATION

Edited by **David Stuart**

**The Mechanisms of DNA Replication**
http://dx.doi.org/10.5772/3433
Edited by David Stuart

**Contributors**

Andrey Aleksandrovich Grach, Lynne Cox, Penelope Mason, Christophe Thiriet, Angélique Galvani, Agustino Martinez-Antonio, Laura Espindola-Serna, César Quiñones-Valles, Susan Forsburg, Sarah Sabatinos, Radmila Capkova Frydrychova, James Mason, Naoki Sato, Takashi Moriyama, Apolonija Bedina Zavec, Amine Aloui, Herve Seligmann, Yoshizumi Ishino, Takeo Kubota, Maria Vittoria Di Tomaso, Alba Guarne, Lindsay Matthews, David Stuart, Douglas Maya, Mari Cruz Muñoz-Centeno, Macarena Morillo-Huesca, Sebastian Chavez, Lidia Delgado Ramos, Dianne C. Daniel, Edward Johnson, Ayuna Dagdanova, Thomas Melendy

**Notice**

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

# We are IntechOpen,
# the first native scientific
# publisher of Open Access books

**3,250+**
Open access books available

**106,000+**
International authors and editors

**112M+**
Downloads

**151**
Countries delivered to

Our authors are among the
**Top 1%**
most cited scientists

**12.2%**
Contributors from top 500 universities

CLARIVATE ANALYTICS
**BOOK CITATION INDEX**
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

## Interested in publishing with us?
## Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

# Meet the editor

Dr. David Stuart is an associate professor in the department of biochemistry at the University of Alberta in Edmonton, Canada. He holds an MSc in Kinesiology from the University of Waterloo and a PhD in Biochemistry from the University of Alberta. He conducted post-doctoral research in cell cycle regulation under the supervision of Dr. Curt Wittenberg at the Scripps Institute in San Diego California. Dr. Stuart's research is focused on mechanisms that regulate DNA replication and recombination during meiosis. His major investigations focus on the role of Cyclin Dependent Kinase and Dbf4 Dependent Kinase in activating DNA replication and linking DNA replication and recombination. He also pursues extensive investigations into the mechanisms that regulate gene expression during meiosis.

# Contents

# Preface

DNA replication is a fundamental part of the life cycle of all organisms. Not surprisingly many aspects of this process display profound conservation across organisms in all domains of life. Successful duplication of the genetic material can decide the life or death of an organism. Hence, the integrity of the DNA replication process is paramount and any defects or errors can lead to a myriad of problems ranging from cell death and developmental failure to increased propensity for cancer.

The importance of accurately regulating the initiation and progression of DNA synthesis is reflected in the complexity involved in assembling the molecular machines that carry out chromosomal DNA synthesis. Chapters by Ishino & Ishino and Martinez-Antonio et al. discuss the process of DNA replication in bacteria and archaea and reveal aspects of the process that are conserved, and aspects that are unique when compared to eukaryotes.

The large size of eukaryotic chromosomes presents challenges to accomplishing accurate and timely DNA replication required for cell proliferation. The molecular machines that drive DNA unwinding and chromosomal DNA synthesis are assembled in a multi-step process that allows for many layers of potential regulation to ensure that DNA replication is initiated accurately and only when appropriate. Many of these mechanisms serve double duty to ensure that DNA replication is initiated only once in any given cell cycle. This is essential to ensure that all portions of the genome are replicated but that none are over-replicated which could lead to the formation of structures at risk for breakage or inappropriate recombination.

The assembly and activity of the DNA helicases and "replisome" that unwinds chromosomal DNA and drives DNA replication are reviewed and discussed in chapters by Stuart, Fisk et al., and Daniel, et al. The assembly of these fantastic DNA replication machines depends upon highly specific and exquisitely regulated protein-protein interactions achieved by specific interaction domains and a subset of these important interaction domains and mechanisms are reviewed in chapters by Matthews & Guarne and Zavec.

The Integrity of chromosomal DNA replication is a high priority for cells and there are many mechanisms devoted to ensuring that damage to chromosomes is limited during the duplication processes. The intra S-phase checkpoint and mechanisms that retain integrity of the replication forks in the face of conditions that lead to pausing or stalling of the replication process is discussed by Sabatinos & Forsburg who also present a model for the consequences of replication fork collapse during conditions when fork stalling or pausing occurs globally during the replication process. Cox & Mason describe the current state of understanding of the WRN helicase that functions in mammalian cells with emphasis on the effect

of loss of function mutations in WRN that lead to Werners Syndrome, a disorder that reca-
pitulates cellular aging.

Cellular DNA is not "naked" but is wrapped and folded into complex three-dimensional
structures through its interaction with histone and other chromosomal proteins that com-
prise chromatin. The histone proteins are subject to an array of post-translational modifica-
tions that include acetylation, methylation, ubiquitination, and phosphorylation. The DNA-
protein complex that is chromatin can exist in a range of structures varying in the degree of
condensation and modification state of the proteins. Not surprisingly the state of the chro-
matin has significant effects on the replication of the DNA, influencing the selection of start
sites for DNA replication, the rate of fork progression and extent of fork pausing, as well as
having effects on DNA repair and recombination. Chapters by Kubota et al., Aloui et al, Di
Tomaso et al., Maya et al., and Galvani & Thiriet review aspects of the relationship of DNA
replication to chromatin structure and epigenetic regulation.

Not all segments of chromosomal DNA are the same even within the same cell. Some re-
gions of the chromosomes have unique characteristics required to carry out a particular
function. The ends or telomeres of eukaryotic chromosomes are particularly interesting as
they present a problem of how to fully replicate both strands without a loss of genetic
information. The end replication problem and mechanisms that solve the problem are de-
scribed in chapters by Grach and by Frydrychova and Mason.

This volume outlines and reviews the current state of knowledge on several key aspects of
the DNA replication process. This is a critical process in both normal growth and develop-
ment and in relation to a broad variety of pathological conditions including cancer. Under-
standing and defining the molecular mechanisms that drive and regulate DNA replication
will offer insight into the fundamental process that allows cellular life and proliferation. Ad-
ditionally, these insights will ultimately offer the hope of controlling diseases like cancer
that deregulate DAN replication and cell proliferation.

**David Stuart**
Associate Professor
Department of Biochemistry
University of Alberta
Edmonton, Alberta
Canada

# Machines that Drive DNA Replication

# Pulling the Trigger to Fire Origins of DNA Replication

David Stuart

Additional information is available at the end of the chapter

## 1. Introduction

DNA replication is a fundamental aspect of cell biology. The process is essential for chromosome doubling and segregation during cell division. Additionally, the DNA replication program can be manipulated to allow a reduction in ploidy as occurs during meiosis or an increase in ploidy as observed in endo-cycles during some developmental processes [1]. The importance of the integrity of the chromosome duplication process is inherently obvious. In somatic cells failure to replicate prevents cell division or leads to a catastrophic reductional division and cell death. Less drastic defects in DNA replication can appear as problems leading to gene amplification, chromosome breaks or chromosome missegregation [2]. These can manifest as birth defects or increased susceptibility to cancer [3]. The integrity of the DNA replication process is ensured partly by DNA repair mechanisms and checkpoint controls. However, the primary mechanism that safeguards the DNA replication process is the complex and multi-step process that leads to the assembly and activation of an active replication complex at chromosomal origins of DNA replication.

The assembly and activation of DNA replication complexes on eukaryotic chromosomes is critically dependent upon two cell cycle regulated protein kinase complexes; Cyclin Dependent Kinase (CDK) and Dbf4 Dependent Kinase (DDK). These protein kinases phosphorylate multiple protein substrates that play roles in assembling a replisome through promoting specific protein-protein interactions that recruit essential components to the complex and stabilize the assembled complex. Additionally, CDK and DDK play roles in the activation of the DNA replication complex and its helicase activity [4].

This chapter will review the key regulatory roles played by CDK and DDK activity in promoting timely assembly of DNA replication complexes. The focus of the article will be on the budding yeast *Saccharomyces cerevisae* where the assembly and activation of origins of DNA replication has been extensively studied. However, the yeast system will be compared and

contrasted with other eukaryotes in order to emphasize universal features of the process and highlight unique characteristics of DNA replication in different organisms and cell types.

## 2. Origins of replication: Where it all starts

DNA replication is a fundamental aspect of cellular proliferation. Bacterial cells with relatively small chromosomes initiate DNA replication from a single well-defined site on each chromosome referred to as oriC [5]. Eukaryotic chromosomes can be from 10 to 1000 times larger than bacterial chromosomes. In order to completely replicate so much chromosomal DNA within a timely fashion that will allow proliferation, eukaryotic cells employ multiple sites on each chromosome that act as origins for the initiation of DNA replication. These sites are referred to as origins of DNA replication (ORIs). In most metazoans ORIs are poorly defined in the sense that they lack a specific consensus DNA sequence but appear to localize to large regions of a chromosome and are defined by the structure of the chromatin and modification state of the histones and chromatin proteins rather than by specific DNA sequences [6-8]. Indeed, even in the single celled fission yeast *Schizosaccharomyces pombe* DNA replication initiates from relatively broad chromosomal regions [9, 10]. The budding yeast and particularly *Saccharomyces cerevisiae* differs from other eukaryotes in this regard. Autonomously Replicating Sequences (ARS) were first identified in *S. cerevisiae* chromosomal DNA in 1979 [11]. When incorporated into plasmid DNA an ARS sequence allowed for efficient replication and maintenance of the extrachromosomal plasmid. Characterization of ARSs revealed specific DNA sequence elements that act as ORIs reviewed by [12]. These sequences are about 100 – 150 basepairs in length and are composed of elements referred to as A, B1, B2, other sequence elements referred to as B3 and C are sometimes present [13]. The A module harbors an AT-rich 11 basepair ARS Consensus Sequence (ACS). Together the A and B1 element contribute to the formation of a binding site for Origin Recognition Complex (ORC) proteins [14], discussed in the next section. The B2 sequence module contains a double stranded DNA unwinding element (DUE). This sequence is where unwinding of the double helical DNA initiates to create a replication bubble [15, 16]. The B3 element acts as a binding site for the transcription factor Abf1 and excludes nucleosome occupancy of the origin sites [17]. The C element has transcription factor binding sites that may stimulate the utilization of some ORIs but are not essential for ORI function [12, 18].

Although there are specific sequence determinants for *S. cerevisiae* origins of replication, even in this yeast not all ORIs are equal. Significant heterogeneity exists among ORIs in the frequency with which they are activated and utilized [19]. Indeed, there are some origin sequences in the *S. cerevisae* genome that are not utilized and appear to be dormant [20]. In addition to the frequency of activation there is a distinct temporal order to ORI activation with a subset of origins being activated at early times in S-phase and others being activated later in S-phase [21, 22]. DNA combing studies with *S. cerevisiae* have revealed that at the single molecule level origin activation is highly stochastic with different sets of ORIs being activated in each cell cycle [19, 23]. Indeed while there are approximately 700 potential ORIs in the *S. cerevisiae* nuclear genome only about 200 are activated in any given S-phase. Recent genome-

wide studies investigating origin activation combined with mathematical modeling have suggested that replication timing can be explained by a stochastic mechanism [24-27]. The basis for the differential frequency of ORI activation and temporal regulation has been argued to be due to a limited availability of some essential activators [28-31]. In the case of *S. cerevisiae* over expression of Dbf4, the activating subunit of the Dbf4 dependent kinase (DDK) along with the Cdk substrates Sld2, Sld3 and their binding partner Dbp11 allow early activation of late firing ORIs [28]. Since Dbf4, Sld2, Sld3 do not remain associated with the replication complex once it has been activated, it has been proposed that once an origin fires, the limiting subunits are released from the complex and can then interact with another ORI and trigger its activation. In this scenario ORIs with the highest affinity for the rate limiting factors will have the highest probability of being activated and will have a high probability of being activated at early times in S-phase. ORIs with a lower affinity for the rate limiting factors will fire after those factors have been released from other ORIs. Hence a temporal order of ORI activation can be created. These models propose that the rate limiting activators of DNA replication have a higher affinity for some ORIs than others [28]. This differential affinity may be due to structural aspects of the chromatin in which the ORI is embedded as well as modification of the chromatin proteins by acetylation, methylation, and potentially other post-translational events [32-34]. Further, there is evidence that ORI usage can be influenced by the presence of nearby transcriptional units [35-37].

## 3. Assembly of the pre-RC: Orc marks the spot

The model of specific chromosomal locations acting as sequence specific sites for binding of protein complexes to initiate DNA replication is conserved across organisms from eukaryotes to prokaryotes and archaea. However, as already described there is no conservation of DNA sequences that act as ORIs across organisms. Indeed, even in *S. cerevisiae*, which has well defined ORIs the sequence of the origins of replication are rather degenerate with only the core ACS being well conserved. In other eukaryotic organisms ORIs display little similarity beyond being rich in AT sequences. Although the DNA sequences that act as sites for initiation of DNA replication are not conserved among eukaryotes the protein complex that binds to ORIs, the Origin Recognition Complex (ORC) is well conserved across eukaryotes and archaea [38-40]. The conserved ORC complex is a hetero-hexamer composed of six subunits Orc1 to Orc6. This complex binds directly to the chromosomal DNA. The *S. cerevisiae* Orc1-6 proteins bind as a hetero-hexamer to the ORI sequence constitutively throughout the cell cycle with Orc1, Orc2, Ocr4, and Orc5 making direct contact with the A and B1 sequence ORI DNA sequence [41-43]. In contrast metazoans and even the fission yeast *S. pombe* display regulated binding of the ORC complex to the chromosomal ORI sites. In particular the Orc1 subunit dissociates from the chromatin in G2-phase and re-associates with the complex in G1 [31, 44]. In *D. melanogaster* and human cells Orc1 is subject to degradation by the Anaphase Promoting Complex (APC) in G2-phase [44-48]. As Orc1-6 is required for DNA replication initiated at ORIs, the regulated binding of Orc in metazoans provides an additional layer of regulation that may be used to control the initiation of DNA replication.

The Orc1-6 proteins act as a marker of chromosomal ORI sites and a platform for the assembly of replication complexes. Orc1-6 does not perform this function in an entirely static fashion. Rather successful initiation of DNA replication requires that the Orc1-6 be capable of binding and hydrolyzing ATP, reviewed by [49]. The Orc1 and Orc5 subunits possess nucleotide-binding motifs, Orc1 has conserved Walker A and Walker B motifs and Orc5 has a Walker A motif and a questionable Walker B sequence [50]. Both Orc1 and Orc5 can bind DNA but only Orc1 displays ATPase activity and while mutations that inactivate the Orc1 Walker A sequence cause defects in DNA replication, mutations to the Orc5 Walker A sequence do not [50-52]. In yeast this activity is essential to allow Orc1-6 to bind specifically to chromosomal ORI DNA and to load other replication complex components on to the ORI [43, 50]. Site-specific binding of Orc1-6 to ORI DNA requires the ability to bind ATP; however ATP hydrolysis is not required, suggesting that ATP binding modulates Orc1 structure and its ability to complex with both DNA and other Orc subunits [50]. In contrast ATP hydrolysis is strictly required for the loading of other replication complex proteins and the formation of a functional DNA replication complex [50-52].

DNA replication is essential for developmental processes as well as for somatic cell proliferation. It is frequently the case that the cell cycle is altered or modified from the canonical form it takes in mature cells to achieve specific developmental aims. Orc1-6 is essential for DNA replication in many developmental contexts. Mutations in human Orc1 and Orc4 proteins are responsible for Meier-Gorlin syndrome, a developmental disorder characterized by primary dwarfism, microcephaly, developmental abnormalities of ear and patella [53, 54]. Additionally, Orc3 is essential for neuronal development and maturation [55]. However, there is some diversity in the regulation of Orc1-6 during developmental. For example endo-reduplication in *D. melanogaster* does not require Orc1 [56, 57]. The developmental regulation of Orc binding to chromatin may be influenced by changes in chromatin modification that occur during development since changes in chromatin acetylation have been associated with and shown to regulate the transition to endo-reduplication and the redistribution of Orc proteins during development [58]. And, while Orc1-6 and DNA replication is essential for premeiotic DNA replication, the requirements for these proteins and the mechanism by which they are organized to promote the initiation may differ between mitotic and meiotic S-phases [9].

## 4. Assembly of the pre-RC: Enter the helicase

The chromatin bound Orc1-6 acts as a nucleation site for the construction of a replication complex (RC). This begins with the assembly of a pre-Replicative Complex (pre-RC). The pre-RC is the multi-protein complex assembled on to ORIs in G1-phase prior to the initiation of DNA replication in S-phase. The base of the pre-RC is the chromatin bound Orc1-6, which acts as a landing pad for the assembly of a series of other protein factors required to assembly a replication fork and initiate bidirectional DNA synthesis. A key requirement for processive DNA synthesis is a dsDNA helicase that can unwind the chromosomal DNA. The Orc1-6 itself has no helicase activity but is essential for recruitment of the replicative helicase to origins of DNA replication. The replicative helicase in *S. cerevisiae* is the minichromosome maintenance

complex (Mcm2-7). The Mcm complex is a hetero-hexamer composed of the subunits Mcm2 – Mcm7 [59-61]. The Mcm subunits interact with each other in a 1:1 ratio to form a ring-like structure that initially binds by wrapping around the DNA such that the double helix passes through the rings central channel. Extensive investigation using biochemical characterization and mutagenesis studies have revealed that the Mcm ring structure has a subunit assembly with the order Mcm5 – Mcm3 – Mcm7 – Mcm4 – Mcm6 – Mcm2 [62]. Sub-complexes of the full Mcm2-7 ring can exist in vivo and in vitro and indeed a trimer composed of Mcm4 – Mcm6 – Mcm7 has ATPas activity and can unwind duplex DNA in vitro [63, 64]. Multiple potential ATPase active sites are formed by interactions between the Mcm subunits: however, only the ATPase activity catalyzed by sites formed by Mcm3 – Mcm7 and Mcm7 – Mcm4 are essential for the helicase activity of the Mcm2-7 holo-complex [64, 65].

In G1 phase of the cell cycle the Mcm2-7 complex is recruited and loaded on to Orc1-6 bound ORI sequences. The helicase is loaded on to the B2 sequence element as a pair of hexamers arranged on the DNA in a head – to – head orientation [66, 67]. The helicase initially assembles on to the DNA as an open complex with a central channel; the ring can be closed around the DNA helix by an ATP dependent conformational change (Figure 1). This involves ATP binding to the Mcm2 – Mcm5 subunits and acting as a "switch" that closes the open gate around the duplex DNA [68].
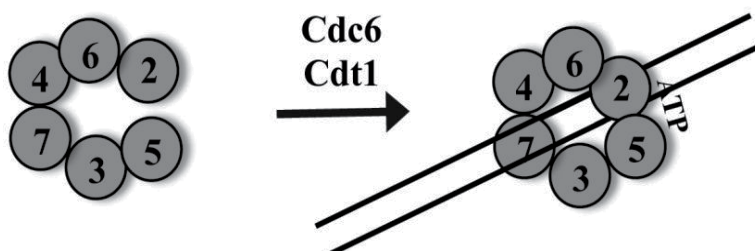


**Figure 1.** The Mcm2-7 hexamer assembles as an open complex that can be closed through ATP binding. The Mcm2-7 subunits can assemble with each other and in the presence of ATP the complex can assume a ring conformation. In vivo the hexamer is loaded on to Orc1-6 bound ORI duplex DNA. This loading is dependent upon the loading factors Cdc6 and Cdt1. The hexamer can be closed loosely around the duplex through binding to ATP.

Loading Mcm2-7 on to the Orc1-6 bound ORI DNA is accomplished through the combined action of the ATPase activity inherent to the chromatin bound Orc1-6 complex and interaction with the AAA$^+$ ATPase loading factor Cdc6. An additional protein required for loading of the Mcm complex is Cdt1, which was first identified in *S. pombe*, but subsequently functional homologs were discovered in *S. cerevisiae*, *X. laevis*, *D. melanogaster*, and mammalian cells [69-73]. The carboxyl-terminus of Cdt1 binds to the Mcm2 and Mcm6 subunits and these contacts are essential for recruitment of the functional Mcm2-7 helicase to Orc1-6 bound origins of DNA replication [74]. ATP hydrolysis catalyzed by both Orc1-6 and the Orc bound Cdc6 stimulate the recruitment of multiple Cdt1-Mcm2-7 complexes [75]. This allows two hexameric Mcm2-7 rings to bind the ORI in a head-to-head orientation, with the dsDNA running through a central channel in the complex [67, 76]. The double hexamers can slide on the duplex DNA

creating the potential to load multimers of double hexamer structures at a single ORI. This may explain why the number of double hexamers loaded on to the DNA can greatly exceed the number of origins that are activated in the subsequent S-phase [77]. Following loading of the Mcm2-7 complexes Cdt1, and Cdc6 are released and do not remain at the ORI as the replication complex continues to assemble [78].

Association of the Mcm2-7 complex with Orc1-6 is a tightly regulated process. In *S. pombe*, Cdt1 mRNA accumulates in the G1 and early S-phase of the cell cycle and in both *S. pombe* and mammalian cells the abundance of the Cdt1 protein is regulated through its destruction by the ubiquitin-proteosome system [71, 73]. In contrast the abundance of Cdt1 protein in *S. cerevisiae* does not fluctuate throughout the cell cycle [69, 79]. In metazoans Cdt1 binding to Mcm2-7 and recruitment to Orc1-6 is negatively regulated by the protein geminin [80]. No protein with a similar function to geminin has been identified in yeast; however, recruitment of *S. cerevisiae* Cdt1-Mcm2-7 complexes to Orc1-6 are negatively regulated by phosphorylation of Orc subunits by Cyclin Dependent Kinase (Cdk) activity [81]. This is an important mechanism to ensure that ORIs are loaded and licensed only once in each cell cycle. Additionally, the gene encoding the loader *CDC6* is transcriptionally regulated such that the mRNA accumulates exclusively during G1 and early S-phase [82]. The Cdc6 protein itself accumulates only in late G1 and early S-phase and is targeted for degradation outside of G1-phase by the Skip1-Cdc53-F box protein (SCF) mediated ubiqutin-proteosome complex [83]. The rigorous regulation applied to Cdc6 and Cdt1 ensures that the Orc1-6 complexes can only be loaded with the replicative DNA helicase machinery in G1 and early S-phase. This is essential to avoid the possibility of origin re-licensing during a cell cycle, which could lead to over replication of some segments of the genome, unscheduled changes in ploidy, the formation of structures that could be at risk for damage, and inappropriate recombination leading to chromosome damage and instability [2, 84].

## 5. Activating the pre-RC: DDK and CDK usher in the replication complex

Loading the Mcm2-7 helicase complex on to an Orc1-6 bound ORI creates a pre-RC, which licenses the origin and provides the potential for it to be activated or "fired" in S-phase. However, activation of the Mcm2-7 complex and unwinding of the DNA depends upon the further ordered addition of the protein factors Sld3, Cdc45, Sld2, Dpb11, the GINS complex (composed of Psf1, Psf2, Psf3, and Sld5], Mcm10, the replicative DNA polymerases Polε, Polδ, and Polα-primase, along with numerous accessory factors. The addition of these factors to the ORI bound Orc1-6 – Mcm2-7 is dependent upon the activity of two protein kinases DDK and CDK.

DDK (Dbf4 Dependant Kinase) is composed of a catalytic subunit, Cdc7 and an activating subunit, Dbf4 [4]. DDK is essential for the initiation of DNA replication and loss of function mutations in either subunit are lethal resulting in a G1 – S-phase arrest characterized by "dumbbell" morphology in *S. cerevisiae* [85, 86]. DDK is an acidiophilc protein kinase [87]. It phosphorylates serine/threonine residues and displays a preference for phosphorylating

serine or threonine residues that are followed by an acidic aspartic acid or glutamic acid residue [88-90]. Additionally, DDK will phosphorylate serine or theronine residues that precede a serine or threonine that has been phosphorylated by another kinase. This is the case with the DDK substrate protein Mer2 where phosphorylation of a serine residue by Cdk1 acts as a priming event to allow phosphorylation by DDK [88, 91]. In S. *cerevisae* the catalytic subunit Cdc7 does not fluctuate in abundance through the cell cycle; however the kinase activity associated with the protein significantly increases in late G1 and S-phase [92]. The kinase activity associated with Cdc7 is regulated primarily through the interaction of Cdc7 with its positively acting regulatory subunit Dbf4. While the abundance of Cdc7 is relatively constant through the cell cycle, Dbf4 displays a striking accumulation in late G1 and early S-phase and rapidly disappears following the completion of DNA replication [93]. The accumulation of Dbf4 in late G1 and S-phase is accounted for in part by transcriptional regulation; the gene is expressed exclusively in late G1 and S-phase [85], and by regulated destruction of Dbf4 by the ubiqutin-proteosome system [94]. Binding of Dbf4 to Cdc7 leads to a conformational shift in the structure of the inert Cdc7 monomer, that stabilizes the active state of the enzyme [95]. Dbf4 displays localization to ORIs [96]. This localization is driven by sequence motifs in Dbf4 that bind specifically to Orc2, Orc3, and to Mcm4 [97, 98]. Contacts with Mcm4 are particularly critical to achieve recruitment of DDK to the pre-RC. Thus, while Cdc7 possesses the catalytic kinase activity, Dbf4 is required to activate the enzyme and target its kinase activity to the appropriate substrates.

The second protein kinase required for conversion of the pre-RC into an active DNA replication complex is CDK. The enzyme is composed of a catalytic subunit Cdk1 (formerly known as Cdc28 in *S. cerevisiae*) that can be activated by association with a cyclin. Like Cdc7, the monomeric Cdk1 has little associated kinase activity [99]. Also similar to Cdc7 the abundance of Cdk1 does not vary appreciable through the cell cycle; however its associated kinase activity fluctuates from very low levels in early G1 to peak levels occurring in M-phase [100, 101]. Binding to an activating cyclin subunit triggers a conformational change in Cdk1 that reveals the active site and promotes the enzymes protein kinase activity [102]. *S. cerevisiae* expresses 9 Cdk1 activating cyclins that promote Cdk1 kinase activity in different phases of the cell cycle. Cln1, Cln2, and Cln3 are required for budding and events in G1 phase, Cln1 and Cln2 accumulate in late G1 and early S-phase while Cln3 is expressed throughout the cell cycle. Clb1, Clb2, Clb3, and Clb4 accumulate in G2 and M-phases, and promote events in G2 and mitosis [103]. Clb5-Cdk1, and Clb6-Cdk1 are the predominant Cdk complexes that promote the initiation of DNA replication during a normal cell cycle in *S. cerevisiae*. *CLB5* and *CLB6* are transcriptionally regulated such that their mRNAs accumulates in late G1 and S-phase. The Clb5 and Clb6 proteins begin to accumulate in late G1-phase [104-106]. Clb6 is targeted for destruction by the SCF and degraded early in S-phase whereas Clb5 persists into G2-phase [107]. Owing to its destruction early in S-phase Clb6-Cdk1 influences only early firing ORIs whereas Clb5-Cdk1 can regulate both early and later firing ORIs [107, 108]. Among the cyclin subunits Clb5 and Clb6 are the most effective at triggering ORI activation and henceforth I will refer to them as S-Cdk. Their effectiveness in activating DNA replication is in part due to the timing of their accumulation; however, even if other cyclins are expressed in late G1 and early S-phase they cannot activate DNA replication as effectively as S-Cdk [109-112]. Both Clb5

and Clb6 have a hydrophobic patch on their surfaces with an MRAIL sequence motif that allows them to interact with target proteins that have Arg–x–Leu or Lys-x-Leu sequences [111, 113, 114]. Whereas DDK physically interacts with the Mcm2-7 complex and this interaction is essential for conversion of a pre-RC to an active replicative complex, there is no evidence that Cdk must bind to the pre-RC in order to drive its conversion to an active complex. Clb5 can bind to Orc6 and does so following the initiation of DNA replication but this is a mechanism to prevent re-licensing and reactivation of ORIs rather than to promote their initial activation in S-phase [115].

## 6. Activating the licensed origins: All aboard the helicase train

The first additional components to interact with the loaded and licensed pre-RC are Sld3, its partner Sld7 and Cdc45 [116-118]. These factors associate with early firing ORIs and bind to the Mcm2-7 complex in G1 phase. Sld3 was originally identified in a genetic screen designed to isolate mutations that were synthetically lethal in an *S. cerevisiae* strain that harbored a temperature sensitive mutant allele of the DNA polymerase ε binding protein *DPB11* [119]. *CDC45* was discovered through its genetic interactions with *MCM5* and *MCM7* mutants [120]. Mutations in either *CDC45* or *SLD3* that cause loss of function prevent DNA replication and are thus lethal [116, 118]. Chromatin immunoprecipitation and in vitro reconstitution experiments indicate that the binding of Sld3 and Cdc45 to ORIs in G1-phase is relatively weak [121, 122]. DDK activity and binding of DDK to the pre-RC is required for the stable recruitment of Sld3 and Cdc45 both in vitro [121], and in vivo [116, 123, 124]. In addition, Sld3 and Cdc45 are required for each others interaction with the ORI bound Mcm2-7 complex.

Association of Cdc45, Sld3 and its partner Sld7 with ORIs is dependent upon DDK [29, 121]. Neither Sld3-Sld7 nor Cdc45 are directly phosphorylated by DDK rather Mcm2, Mcm4 and potentially Mcm6 are the critical S-phase substrates for DDK [89, 98, 125]. Indeed, modification of the structural architecture of the Mcm2-7 complex is likely the critical function for DDK in the activation of DNA replication since a mutation of Mcm5 that changes proline 83 to leucine alters the structure of the Mcm2-7 complex and allows cells lacking DDK to survive and replicate their DNA [122, 126, 127]. Additionally, DDK binds to the Mcm2-7 complex through interactions with a docking domain in Mcm4 and mutations in the Mcm can bypass the requirement for DDK [98, 125]. The initial interaction of DDK with the Mcm2-7 complex is dependent upon prior phosphorylation of at least Mcm4 and Mcm6 by yet to be identified protein kinases [89, 90].

The binding of Cdc45, Sld3 and Sld7 is a pre-requisite for the further assembly and conversion of the pre-RC to an active replication complex (RC). Following the loading of these factors Cdk activity is required. Accumulating S-Cdks interact with both Sld2 and Sld3 through RxL motifs in the substrate proteins [113-115, 128]. This leads to phosphorylation of Sld2 and Sld3 at multiple sites [129, 130]. The multi-site phosphorylation of Sld2 leads to a conformational change in the protein that allows the additional phosphorylation of threonine 84, which does not reside within a canonical Cdk recognition motif [131]. Phosphorylation of T[84] allows Sld2

to interact with Dpb11 a protein originally identified based upon its interactions with the replicative DNA polymerase, Polε [132]. Dpb11 has BRCT repeat domains at both its amino-terminal and carboxyl-terminal regions [133]. These sequence motifs function as phospho-peptide binding domains [134] allowing the phosphorylated Sld2 to bind the carboxyl-terminal BRCT phosphopeptide binding domain of Dpb11 [119, 129, 130]. Similarly phosphorylation of Sld3 allows Sld3 to bind the amino-terminal BRCT repeat of Dpb11 thus recruiting the Sld2-Sld3-Dpb11 complex to the Mcm2-7 complex and origin of replication [129, 130]. Dpb11 binds Polε, the leading strand replicative DNA polymerase in *S. cerevisiae* [132]. The interaction of Dpb11 with DNA Polε is not Cdk dependent but binding to phosphorylated Sld2 and Sld3 allows recruitment of the entire complex to the licensed ORI [135].

Although Sld2 and Sld3 are not the only components of the replication complex that can be phosphorylated by Cdk1 they are the critical substrates since phosphomimetic mutations in Sld2 and fusion of Sld3 with Dpb11 can bypass the need for Cdk1 activity to initiate DNA synthesis [129, 130].

The binding of Sld2 and Sld3 to the pre-RC allows the recruitment of GINS to the Mcm2-7 hexamer. GINS is a protein complex composed of Psf1, Psf2, Psf3 and Sld5 and is named after the number based names of its components Go, Ichi, Ni, San (Japanese for 5, 1, 2, 3). Sld5 was identified in a genetic screen for mutants that displayed synthetic lethality when combined with a thermo-sensitive *dpb11* allele [116]. Subsequent investigations reveled partners of Sld5 (Psf1, Psf2, Psf3) that formed a complex required for initiation and DNA strand elongation during DNA replication [136]. GINS associates with Cdc45 at the ORI and its recruitment leads to stable engagement of Cdc45 with the Mcm2-7 complex. In vitro Cdc45 and GINS strongly stimulate the ATPase and DNA unwinding activity of Mcm2-7 complex [137]. There is evidence that Cdc45 makes specific contacts with Mcm2 while GINS binds to Mcm5, when GINS and Cdc45 bind one another this tightly closes the Mcm2-7 rings "gate" with DNA trapped within the central channel of the Mcm ring structure reviewed by [59]. There is no evidence that Cdk phosphorylates either Cdc45 or GINS or regulates their activity, the primary role played by the Cdk appears to be in promoting their recruitment to the chromatin bound Mcm2-7 complex. The binding of the additional components including GINS results in conversion of the pre-RC into the CMG (Cdc45/Mcm2-7/GINS) complex, this is also referred to as the pre-initiation complex (pre-IC) [138]. While Sld2, Sld3 and Sld7 are released from the complex following stable engagement of Cdc45 and GINS, both of the latter factors remain associated with the Mcm2-7 and are required for elongation of the nascent DNA strands following the initiation of DNA synthesis [136, 139].

Mcm10 is an additional factor required for assembly of a functional replisome and conversion of the pre-IC to an RC. Mcm10 was originally identified in a screen similar to that used for the identification of other *S. cerevisiae MCM* genes [140, 141]. Homologs of *MCM10* can be found from yeast to humans [142, 143]. Mcm10 is an abundant chromatin bound protein that interacts with all six subunits of the Mcm2-7 complex and localizes to origins of DNA replication [141, 142, 144]. Mcm10 has a critical role in conversion of the pre-RC to an active RC as it makes contacts with DNA Polα and the CMG complex components [145-147]. It is certain that Mcm10 plays a role in stabilizing the Mcm2-7 complex with DNA Polα [148]; however its precise role

in the initial recruitment of DNA polymerases or their accessory factors to the replisome is not entirely clear.

The accumulation and action of DDK and CDK set in motion the assembly and conversion of the pre-RC to an activated RC. The use of two independent kinases to achieve this goal allows tight regulation over the assembly and activation process. Since both kinases are required to activate and "fire" the ORI it seems that there are in fact two triggers that can be pulled independently. For the initiation of DNA replication to take place both triggers must be pulled with the correct timing.
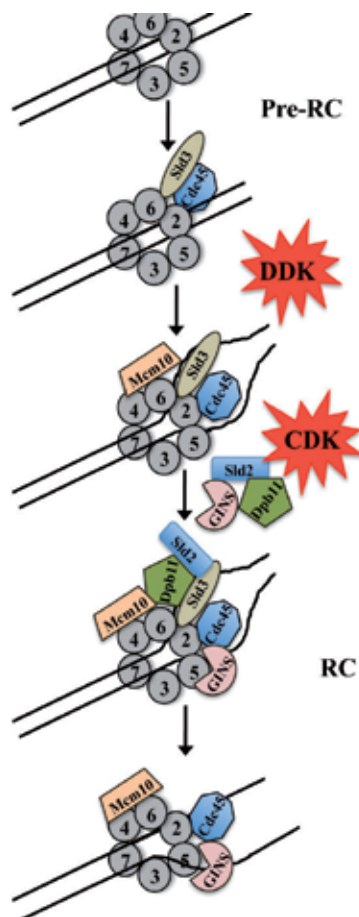


**Figure 2.** DDK and CDK promote assembly and activation of replication complexes at chromosomal origins of DNA replication. Sld3 and Cdc45 associate loosely to the ORI bound Mcm2-7 hexamer in G1-phase. Phosphorylation of the Mcm subunits by DDK promote tight binding by Cdc45 and Sld3, Mcm10 may associate with the complex at this time and plays an important role in unwinding of the ORI DNA duplex. CDK phosphorylation of Sld3, and Sld2 recruit Sld2, Dpb11, Polε and GINS to the Mcm2-7 complex. GINS binding increases the helicase activity of the Mcm2-7 hexamer allowing unwinding of duplex DNA. The association of GINS also marks a transition when Mcm2-7 binding to duplex DNA changes to binding such that a single strand is retained in the central channel, while the other strand is moved to the external surface of the complex.

## 7. The business end: Polymerases at the origin

The final critical steps of origin firing are the recruitment of the replicative polymerases, unwinding of the dsDNA and initiation of DNA synthesis. While all cells encode multiple different DNA polymerases the enzymes with the most well characterized roles in nuclear chromosomal DNA replication are DNA Polε, DNA Polδ, and DNA Polα – primase. DNA Polε acts as the leading strand DNA polymerase for nuclear DNA replication in *S. cerevisiae* [149]. Through its interaction with Dpb11 it is recruited to the pre-RC complex following Cdk1 mediated phosphorylation of Sld2, and Sld3. DNA Polδ is the major lagging strand DNA polymerase in *S. cerevisiae* [150]. Although DNA Polδ plays a key role in nuclear DNA replication it is currently unclear how this enzyme is recruited to the nascent RC. DNA Polα-primase is essential for the initiation of DNA replication as primase synthesizes RNA primers that Polα extends with short DNA oligonucleotides on the unwound ORI DNA providing primers for DNA Polε and DNA Polδ. [151, 152]. Mcm10 binds DNA Polα and this DNA polymerase may be initially recruited to the Mcm2-7 complex through these interactions. The primase polypeptide forms a complex with the carboxyl-terminus of Polα allowing the two to be incorporated into the growing replisome simultaneously [153]. Following or perhaps concurrent with recruitment of the replicative DNA polymerases there is a reorganization of the complex as it undergoes conversion from a pre-IC to RC. During this process Dpb11, Sld2 and Sld3 are ejected from the complex while Polε remains bound. Within the RC, DNA Polε makes contacts with Mrc1 that help to retain it within the complex [154]. It is currently unclear how Mrc1 is recruited to the complex upon conversion to a nascent RC or whether unwinding of the ORI DNA is required. Polα makes contacts initially with Mcm10 and once incorporated into the RC, it makes further contacts with Ctf4 a component of Replication Factor C (RFC), these contacts help stabilize the binding of Polα to the complex [155, 156]. During the remodeling of the pre-RC into an activated RC several accessory proteins: Replication Factor C (RFC), Proliferating Cell Nuclear Antigen (PCNA), and Replication Protein A (RPA) are added to the complex. The mechanism that leads to recruitment of these accessory proteins has not been determined. It may be that they simply recognize and bind to the protein-DNA structure formed by the initial unwinding of the ORI DNA. Owing to its ssDNA binding capability RPA associates with the RC once unwinding of the ORI DNA is underway; here it assists in stabilizing the nascent replication bubble and provides access for the replicative DNA polymerases [157]. All three subunits of DNA Polδ make contact with PCNA and these interactions are essential for processive lagging strand DNA synthesis [158]. These factors influence the processivity and integrity of DNA synthesis.

Unwinding the ORI DNA to provide ssDNA as template for the DNA polymerases and to construct bidirectional replication forks is accomplished by the activated Mcm2-7 hexamer in concert with associated proteins Cdc45, GINS, Mcm10 and the replicative DNA polymerases. In vitro the Mcm2-7 hexamer unwinds DNA by tracking along a single strand while displacing the other strand [65, 159]. Achieving this end requires that the dsDNA initially bound be melted and locally unwound allowing release of one strand to the outside surface of the complex and retaining the other within the central channel of the hexamer. Although the molecular details

of this process remain unclear some of the current models to explain ORI unwinding by Mcm2-7 have been recently reviewed in detail [59].

Sld2, Sld3, and Mcm10 all display some ability to bind ssDNA and it has been speculated that they might participate in the initially melting of the dsDNA, allowing the Mcm2-7 rings to undergo conformational change such that they close around one of strands of the melted duplex. Mcm10 may be a real candidate for this role based upon its stable incorporation into the RC and ability to bind ssDNA [160]. Determining the precise mechanism and timing of ORI DNA unwinding will await higher resolution structural and biochemical analysis.

## 8. Who's on first? Ordered action of DDK and CDK in the activation of ORIs

The assembly of a preRC and its conversion first to an RC and then an active replication fork is a multistep process that requires the activity of both DDK and CDK. Multiple investigations have been performed to determine the order in which DDK and CDK act at the ORIs to trigger their activation. Genetic studies with *S. cerevisiae* have suggested that DDK cannot complete its function without prior S-Cdk activation implying either that Cdk must act before DDK or that DDK performs a multiple functions at the pre-RC and that some of them require Cdk activity for completion [161]. In *X. laevis* egg extracts DDK can complete its essential function in the absence of Cdk activity, however Cdk cannot perform its vital function in the absence of DDK [162, 163]. Recent investigations using an *S. cerevisiae* in vitro DNA replication system suggest that assembly and activation of origins of replication require that DDK act before Cdk but that completion of DDKs essential functions require Cdk activity [90, 121]. The apparent conflict in these results may reflect differences between DNA replication control in somatic cells and eggs. Additionally, some of the differences may stem from the limitations inherent to both genetic and in vitro biochemical experimental systems. Redundant systems and limits to the speed with which activities can be activated and inactivated in vivo place limits on genetic approaches to understanding the specific requirements for DDK and CDK. While in vitro it may be difficult to accurately recapitulate the in vivo environment. For example, Cdk activity increases during G1-phase in a graded fashion both in total kinase activity and kinase specificity. Relatively low levels of Cdk activity are sufficient to activate DNA replication and elevated levels of Cdk activity that accumulate in S, G2, and M-phases prevent licensing and activation of origins by promoting destruction of Cdc6, nuclear export of Mcm2-7 components and by binding to Orc6 and excluding recruitment of Mcm to ORIs [164, 165]. It is possible that low levels of Cdk activity are required prior to DDK initiating its function. Indeed phosphorylation of Mcm4 and Mcm6 is a prerequisite for DDK binding to the pre-RC and further inducing activation. It has been proposed that phosphorylation of Mcm4 by G1-Cdk activity may be required to allow DDK to bind to the Mcm2-7 complex [89].

## 9. Conclusion

DNA replication is a fundamental aspect of cellular proliferation and development. Many aspects of this process are well conserved not only within the domain of eukaryotes but also across bacteria and archaea. The multi-step assembly and activation of origins of DNA replication is more complicated and more rigorously regulated in eukayotes than it is in either prokaryotes or archaea. This complexity stems in part from the size of the eukaryotic genomes that necessitates multiple origins of replication on each chromosome. Additionally, multiple layers of regulation act as a safeguard that ensures each origin of DNA replication is activated only once in each cell cycle. This is crucial to prevent over replication, amplification of chromosomal segments and chromosome instability.

The initiation of DNA replication in *S. cerevisiae* has served as an exceptional model owing to the genetic and biochemical accessibility of this organism. Our current understanding of the steps leading to the initiation of DNA replication in *S. cerevisiae* can be summarized as follows. Orc1-6 bound ORI sequences act as a binding site for Cdc6, which in conjunction with Cdt1 recruits Mcm2-7 hexamers to the ORI. DDK is recruited to this structure by virtue of the affinity of Dbf4 for docking domains in Mcm4. DDK phosphorylates the Mcm2–7 helicase, promoting the recruitment of Sld3 and Cdc45. Next, S-CDK-dependent phosphorylation of Sld2 and Sld3 leads to their binding Dpb11 and recruitment of the complex, along with GINS and Polε to the pre-RC thus forming a CMG complex. These proteins then serve to both recruit Mcm10 and fully activate the Mcm2–7 helicase, which uses ATP hydrolysis to melt the origin DNA. Polα-primase and Polδ can then be loaded on to the ssDNA at the unwound ORI, leading to the formation of a complete replisome with accessory proteins such as PCNA, Mrc1, RFC, RPA, and topisomerase. The helicase activity of the Mcm2-7 hexamers then drives bidirectional dsDNA unwinding and replication fork movement along the chromosome allowing the synthesis of new DNA.

Initiating DNA replication is a serious event for a cell. The chromosomal DNA is rarely more at risk of damage than when it is being unwound and copied. During this processes single stranded DNA is revealed and the fork structures with the potential for breakage and recombination are formed. The requirement for two protein kinases, DDK and CDK, that perform non-redundant functions in the assembly and activation of replication complexes suggests that there are in fact two triggers that must be pulled to fire the origin. The requirement for two different kinases that are independently regulated and that each have distinct substrate specificity allows the initiation of DNA replication to be regulated with exquisite sensitivity. Perhaps rather than considering these two kinases as triggers they should really be though of as a double failsafe mechanism where each trigger must be pulled with the appropriate timing to allow DNA replication to proceed.

Despite our general understanding of this process many aspects of its molecular basis remain to be elucidated. How are Sld3 and Cdc45 initially recruited to the pre-RC? How does the Mcm2-7 helicase melt ORI DNA and what is the mechanism by which it is converted to a machine that directionally tracks along and unwinds dsDNA? Does DDK travel with the Mcm2-7 complex along the DNA? How are DNA Polδ and the accessory proteins RFC, and

PCNA recruited to the replication fork? It is likely that a combination of genetic analysis, biochemistry and high-resolution structure analysis will be required to answer these questions.

## Acknowledgements

## Author details

David Stuart

Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada

## References

[1]  Larkins BA, Dilkes BP, Dante RA, Coelho CM, Woo YM, Liu Y. Investigating the hows and whys of DNA endoreduplication. J Exp Bot. 2001;52(355):183-92.

[2]  Green BM, Finn KJ, Li JJ. Loss of DNA replication control is a potent inducer of gene amplification. Science. 2010;329(5994):943-6.

[3]  Pfau SJ, Amon A. Chromosomal instability and aneuploidy in cancer: from yeast to man. EMBO Rep. 2012;13(6):515-27.

[4]  Johnston LH, Masai H, Sugino A. First the CDKs, now the DDKs. Trends Cell Biol. 1999;9(7):249-52.

[5]  Baker TA, Wickner SH. Genetics and enzymology of DNA replication in *Escherichia coli*. Annu Rev Genet. 1992;26:447-77.

[6]  Falaschi A, Giacca M, Zentilin L, Norio P, Diviacco S, Dimitrova D, et al. Searching for replication origins in mammalian DNA. Gene. 1993;135(1-2):125-35.

[7]  Hamlin JL, Mesner LD, Lar O, Torres R, Chodaparambil SV, Wang L. A revisionist replicon model for higher eukaryotic genomes. J Cell Biochem. 2008;105(2):321-9.

[8]  Norio P, Kosiyatrakul S, Yang Q, Guan Z, Brown NM, Thomas S, et al. Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. Mol Cell. 2005;20(4):575-87.

[9]  Heichinger C, Penkett CJ, Bahler J, Nurse P. Genome-wide characterization of fission yeast DNA replication origins. EMBO J. 2006;25(21):5171-9.

[10]  Cotobal C, Segurado M, Antequera F. Structural diversity and dynamics of genomic replication origins in *Schizosaccharomyces pombe*. EMBO J. 2010;29(5):934-42.

[11]  Stinchcomb DT, Struhl K, Davis RW. Isolation and characterisation of a yeast chromosomal replicator. Nature. 1979;282(5734):39-43.

[12]  Dhar MK, Sehgal S, Kaul S. Structure, replication efficiency and fragility of yeast ARS elements. Res Microbiol. 2012;163(4):243-53.

[13]  Marahrens Y, Stillman B. A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science. 1992;255(5046):817-23.

[14]  Bell SP, Stillman B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature. 1992;357(6374):128-34.

[15]  Natale DA, Schubert AE, Kowalski D. DNA helical stability accounts for mutational defects in a yeast replication origin. Proc Natl Acad Sci U S A. 1992;89(7):2654-8.

[16]  Zou L, Stillman B. 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. 2000;20(9):3086-96.

[17]  Miyake T, Loch CM, Li R. Identification of a multifunctional domain in autonomously replicating sequence-binding factor 1 required for transcriptional activation, DNA replication, and gene silencing. Mol Cell Biol. 2002;22(2):505-16.

[18]  Newlon CS, Theis JF. The structure and function of yeast ARS elements. Curr Opin Genet Dev. 1993;3(5):752-8.

[19]  Czajkowsky DM, Liu J, Hamlin JL, Shao Z. DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. J Mol Biol. 2008;375(1):12-9.

[20]  Santocanale C, Sharma K, Diffley JF. Activation of dormant origins of DNA replication in budding yeast. Genes Dev. 1999;13(18):2360-4.

[21]  Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, et al. Replication dynamics of the yeast genome. Science. 2001;294(5540):115-21.

[22]  Friedman KL, Brewer BJ, Fangman WL. Replication profile of *Saccharomyces cerevisiae* chromosome VI. Genes Cells. 1997;2(11):667-78.

[23]  Patel PK, Arcangioli B, Baker SP, Bensimon A, Rhind N. DNA replication origins fire stochastically in fission yeast. Mol Biol Cell. 2006;17(1):308-16.

[24]  Luo H, Li J, Eshaghi M, Liu J, Karuturi RK. Genome-wide estimation of firing efficiencies of origins of DNA replication from time-course copy number variation data. BMC Bioinformatics. 2010;11:247.

[25] Spiesser TW, Klipp E, Barberis M. A model for the spatiotemporal organization of DNA replication in *Saccharomyces cerevisiae*. Mol Genet Genomics. 2009;282(1):25-35.

[26] Barberis M, Spiesser TW, Klipp E. Replication origins and timing of temporal replication in budding yeast: how to solve the conundrum? Curr Genomics. 2010;11(3): 199-211.

[27] Yang SC, Rhind N, Bechhoefer J. Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. Mol Syst Biol. 2010;6:404.

[28] Mantiero D, Mackenzie A, Donaldson A, Zegerman P. Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. EMBO J. 2011;30(23):4805-14.

[29] Tanaka S, Nakato R, Katou Y, Shirahige K, Araki H. Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. Curr Biol. 2011;21(24):2055-63.

[30] Patel PK, Kommajosyula N, Rosebrock A, Bensimon A, Leatherwood J, Bechhoefer J, et al. The Hsk1(Cdc7) replication kinase regulates origin efficiency. Mol Biol Cell. 2008;19(12):5550-8.

[31] Wu PY, Nurse P. Establishing the program of origin firing during S phase in fission Yeast. Cell. 2009;136(5):852-64.

[32] Vogelauer M, Rubbi L, Lucas I, Brewer BJ, Grunstein M. Histone acetylation regulates the time of replication origin firing. Mol Cell. 2002;10(5):1223-33.

[33] Aparicio JG, Viggiani CJ, Gibson DG, Aparicio OM. The Rpd3-Sin3 histone deacetylase regulates replication timing and enables intra-S origin control in *Saccharomyces cerevisiae*. Mol Cell Biol. 2004;24(11):4769-80.

[34] Knott SR, Viggiani CJ, Tavare S, Aparicio OM. Genome-wide replication profiles indicate an expansive role for Rpd3L in regulating replication initiation timing or efficiency, and reveal genomic loci of Rpd3 function in *Saccharomyces cerevisiae*. Genes Dev. 2009;23(9):1077-90.

[35] Kohzaki H, Ito Y, Murakami Y. Context-dependent modulation of replication activity of *Saccharomyces cerevisiae* autonomously replicating sequences by transcription factors. Mol Cell Biol. 1999;19(11):7428-35.

[36] Murakami Y, Ito Y. Transcription factors in DNA replication. Front Biosci. 1999;4:D824-33.

[37] Sequeira-Mendes J, Diaz-Uriarte R, Apedaile A, Huntley D, Brockdorff N, Gomez M. Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS Genet. 2009;5(4):e1000446.

[38] Gavin KA, Hidaka M, Stillman B. Conserved initiator proteins in eukaryotes. Science. 1995;270(5242):1667-71.

[39]  Duncker BP, Chesnokov IN, McConkey BJ. The origin recognition complex protein family. Genome Biol. 2009;10(3):214.

[40]  Wigley DB. ORC proteins: marking the start. Curr Opin Struct Biol. 2009;19(1):72-8.

[41]  Clarey MG, Erzberger JP, Grob P, Leschziner AE, Berger JM, Nogales E, et al. Nucleo-tide-dependent conformational changes in the DnaA-like core of the origin recognition complex. Nat Struct Mol Biol. 2006;13(8):684-90.

[42]  Lee DG, Bell SP. Architecture of the yeast origin recognition complex bound to origins of DNA replication. Mol Cell Biol. 1997;17(12):7159-68.

[43]  Speck C, Chen Z, Li H, Stillman B. ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. Nat Struct Mol Biol. 2005;12(11):965-71.

[44]  Araki M, Wharton RP, Tang Z, Yu H, Asano M. Degradation of origin recognition complex large subunit by the anaphase-promoting complex in Drosophila. EMBO J. 2003;22(22):6115-26.

[45]  Mendez J, Zou-Yang XH, Kim SY, Hidaka M, Tansey WP, Stillman B. Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. Mol Cell. 2002;9(3):481-91.

[46]  Ohta S, Tatsumi Y, Fujita M, Tsurimoto T, Obuse C. The ORC1 cycle in human cells: II. Dynamic changes in the human ORC complex during the cell cycle. J Biol Chem. 2003;278(42):41535-40.

[47]  Tatsumi Y, Ohta S, Kimura H, Tsurimoto T, Obuse C. The ORC1 cycle in human cells: I. cell cycle-regulated oscillation of human ORC1. J Biol Chem. 2003;278(42):41528-34.

[48]  Li CJ, DePamphilis ML. Mammalian Orc1 protein is selectively released from chroma-tin and ubiquitinated during the S-to-M transition in the cell division cycle. Mol Cell Biol. 2002;22(1):105-16.

[49]  Li H, Stillman B. The origin recognition complex: a biochemical and structural view. Subcell Biochem. 2012;62:37-58.

[50]  Klemm RD, Austin RJ, Bell SP. Coordinate binding of ATP and origin DNA regulates the ATPase activity of the origin recognition complex. Cell. 1997;88(4):493-502.

[51]  Lee DG, Bell SP. ATPase switches controlling DNA replication initiation. Curr Opin Cell Biol. 2000;12(3):280-5.

[52]  Lee DG, Makhov AM, Klemm RD, Griffith JD, Bell SP. Regulation of origin recognition complex conformation and ATPase activity: differential effects of single-stranded and double-stranded DNA binding. EMBO J. 2000;19(17):4774-82.

[53]  Hossain M, Stillman B. Meier-Gorlin syndrome mutations disrupt an Orc1 CDK inhibitory domain and cause centrosome reduplication. Genes Dev. 2012;26(16): 1797-810.

[54] Guernsey DL, Matsuoka M, Jiang H, Evans S, Macgillivray C, Nightingale M, et al. Mutations in origin recognition complex gene ORC4 cause Meier-Gorlin syndrome. Nat Genet. 2011;43(4):360-4.

[55] Cappuccio I, Colapicchioni C, Santangelo V, Sale P, Blandini F, Bonelli M, et al. The origin recognition complex subunit, ORC3, is developmentally regulated and supports the expression of biochemical markers of neuronal maturation in cultured cerebellar granule cells. Brain Res. 2010;1358:1-10.

[56] Asano M. Endoreplication: the advantage to initiating DNA replication without the ORC? Fly (Austin). 2009;3(2):173-5.

[57] Park SY, Asano M. The origin recognition complex is dispensable for endoreplication in Drosophila. Proc Natl Acad Sci U S A. 2008;105(34):12343-8.

[58] Aggarwal BD, Calvi BR. Chromatin regulates origin activity in Drosophila follicle cells. Nature. 2004;430(6997):372-6.

[59] Vijayraghavan S, Schwacha A. The eukaryotic mcm2-7 replicative helicase. Subcell Biochem. 2012;62:113-34.

[60] Forsburg SL. Eukaryotic MCM proteins: beyond replication initiation. Microbiol Mol Biol Rev. 2004;68(1):109-31.

[61] Tye BK. MCM proteins in DNA replication. Annu Rev Biochem. 1999;68:649-86.

[62] Bochman ML, Schwacha A. The Mcm complex: unwinding the mechanism of a replicative helicase. Microbiol Mol Biol Rev. 2009;73(4):652-83.

[63] Bochman ML, Schwacha A. 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. 2007;282(46):33795-804.

[64] Davey MJ, Indiani C, O'Donnell M. Reconstitution of the Mcm2-7p heterohexamer, subunit arrangement, and ATP site architecture. J Biol Chem. 2003;278(7):4491-9.

[65] Bochman ML, Bell SP, Schwacha A. Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability. Mol Cell Biol. 2008;28(19):5865-73.

[66] Chang F, May CD, Hoggard T, Miller J, Fox CA, Weinreich M. High-resolution analysis of four efficient yeast replication origins reveals new insights into the ORC and putative MCM binding elements. Nucleic Acids Res. 2011;39(15):6523-35.

[67] Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JF. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell. 2009;139(4):719-30.

[68] Bochman ML, Schwacha A. The *Saccharomyces cerevisiae* Mcm6/2 and Mcm5/3 ATPase active sites contribute to the function of the putative Mcm2-7 'gate'. Nucleic Acids Res. 2010;38(18):6078-88.

[69] Devault A, Vallen EA, Yuan T, Green S, Bensimon A, Schwob E. Identification of Tah11/ Sid2 as the ortholog of the replication licensing factor Cdt1 in *Saccharomyces cerevisiae*. Curr Biol. 2002;12(8):689-94.

[70] Whittaker AJ, Royzman I, Orr-Weaver TL. Drosophila double parked: a conserved, essential replication protein that colocalizes with the origin recognition complex and links DNA replication with mitosis and the down-regulation of S phase transcripts. Genes Dev. 2000;14(14):1765-76.

[71] Nishitani H, Lygerou Z, Nishimoto T, Nurse P. The Cdt1 protein is required to license DNA for replication in fission yeast. Nature. 2000;404(6778):625-8.

[72] Maiorano D, Moreau J, Mechali M. XCDT1 is required for the assembly of pre-replicative complexes in Xenopus laevis. Nature. 2000;404(6778):622-5.

[73] Nishitani H, Taraviras S, Lygerou Z, Nishimoto T. The human licensing factor for DNA replication Cdt1 accumulates in G1 and is destabilized after initiation of S-phase. J Biol Chem. 2001;276(48):44905-11.

[74] Liu C, Wu R, Zhou B, Wang J, Wei Z, Tye BK, et al. Structural insights into the Cdt1-mediated MCM2-7 chromatin loading. Nucleic Acids Res. 2012;40(7):3208-17.

[75] Takara TJ, Bell SP. Multiple Cdt1 molecules act at each origin to load replication-competent Mcm2-7 helicases. EMBO J. 2011;30(24):4885-96.

[76] Evrin C, Clarke P, Zech J, Lurz R, Sun J, Uhle S, et al. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proc Natl Acad Sci U S A. 2009;106(48):20240-5.

[77] Donovan S, Harwood J, Drury LS, Diffley JF. Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. Proc Natl Acad Sci U S A. 1997;94(11):5611-6.

[78] Randell JC, Bowers JL, Rodriguez HK, Bell SP. Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. Mol Cell. 2006;21(1):29-39.

[79] Jacobson MD, Munoz CX, Knox KS, Williams BE, Lu LL, Cross FR, et al. Mutations in *SID2*, a novel gene in *Saccharomyces cerevisiae*, cause synthetic lethality with sic1 deletion and may cause a defect during S phase. Genetics. 2001;159(1):17-33.

[80] Yanagi K, Mizuno T, You Z, Hanaoka F. Mouse geminin inhibits not only Cdt1-MCM6 interactions but also a novel intrinsic Cdt1 DNA binding activity. J Biol Chem. 2002;277(43):40871-80.

[81] Chen S, Bell SP. CDK prevents Mcm2-7 helicase loading by inhibiting Cdt1 interaction with Orc6. Genes Dev. 2011;25(4):363-72.

[82] Zhou C, Jong AY. Mutation analysis of *Saccharomyces cerevisiae CDC6* promoter: defining its UAS domain and cell cycle regulating element. DNA Cell Biol. 1993;12(4): 363-70.

[83]  Perkins G, Drury LS, Diffley JF. Separate SCF(*CDC4*) recognition elements target Cdc6 for proteolysis in S phase and mitosis. EMBO J. 2001;20(17):4836-45.

[84]  Tanny RE, MacAlpine DM, Blitzblau HG, Bell SP. Genome-wide analysis of re-replication reveals inhibitory controls that target multiple stages of replication initiation. Mol Biol Cell. 2006;17(5):2415-23.

[85]  Chapman JW, Johnston LH. The yeast gene, *DBF4*, essential for entry into S phase is cell cycle regulated. Exp Cell Res. 1989;180(2):419-28.

[86]  Kitada K, Johnston LH, Sugino T, Sugino A. Temperature-sensitive *cdc7* mutations of *Saccharomyces cerevisiae* are suppressed by the *DBF4* gene, which is required for the G1/S cell cycle transition. Genetics. 1992;131(1):21-9.

[87]  Mok J, Kim PM, Lam HY, Piccirillo S, Zhou X, Jeschke GR, et al. Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. Sci Signal. 2010;3(109):ra12.

[88]  Wan L, Niu H, Futcher B, Zhang C, Shokat KM, Boulton SJ, et al. Cdc28-Clb5 (CDK-S) and Cdc7-Dbf4 (DDK) collaborate to initiate meiotic recombination in yeast. Genes Dev. 2008;22(3):386-97.

[89]  Randell JC, Fan A, Chan C, Francis LI, Heller RC, Galani K, et al. Mec1 is one of multiple kinases that prime the Mcm2-7 helicase for phosphorylation by Cdc7. Mol Cell. 2010;40(3):353-63.

[90]  Francis LI, Randell JC, Takara TJ, Uchima L, Bell SP. Incorporation into the prereplicative complex activates the Mcm2-7 helicase for Cdc7-Dbf4 phosphorylation. Genes Dev. 2009;23(5):643-54.

[91]  Sasanuma H, Hirota K, Fukuda T, Kakusho N, Kugou K, Kawasaki Y, et al. Cdc7-dependent phosphorylation of Mer2 facilitates initiation of yeast meiotic recombination. Genes Dev. 2008;22(3):398-410.

[92]  Jackson AL, Pahl PM, Harrison K, Rosamond J, Sclafani RA. Cell cycle regulation of the yeast Cdc7 protein kinase by association with the Dbf4 protein. Mol Cell Biol. 1993;13(5):2899-908.

[93]  Oshiro G, Owens JC, Shellman Y, Sclafani RA, Li JJ. Cell cycle control of Cdc7p kinase activity through regulation of Dbf4p stability. Mol Cell Biol. 1999;19(7):4888-96.

[94]  Ferreira MF, Santocanale C, Drury LS, Diffley JF. Dbf4p, an essential S phase-promoting factor, is targeted for degradation by the anaphase-promoting complex. Mol Cell Biol. 2000;20(1):242-8.

[95]  Hughes S, Elustondo F, Di Fonzo A, Leroux FG, Wong AC, Snijders AP, et al. Crystal structure of human CDC7 kinase in complex with its activator DBF4. Nat Struct Mol Biol. 2012;19(11):1101-7.

[96]   Dowell SJ, Romanowski P, Diffley JF. Interaction of Dbf4, the Cdc7 protein kinase regulatory subunit, with yeast replication origins in vivo. Science. 1994;265(5176): 1243-6.

[97]   Duncker BP, Shimada K, Tsai-Pflugfelder M, Pasero P, Gasser SM. An N-terminal domain of Dbf4p mediates interaction with both origin recognition complex (ORC) and Rad53p and can deregulate late origin firing. Proc Natl Acad Sci U S A. 2002;99(25): 16087-92.

[98]   Sheu YJ, Stillman B. Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. Mol Cell. 2006;24(1):101-13.

[99]   Wittenberg C, Reed SI. Control of the yeast cell cycle is associated with assembly/ disassembly of the Cdc28 protein kinase complex. Cell. 1988;54(7):1061-72.

[100]  Mendenhall MD, Hodge AE. Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev. 1998;62(4):1191-243.

[101]  Nasmyth K. Control of the yeast cell cycle by the Cdc28 protein kinase. Curr Opin Cell Biol. 1993;5(2):166-79.

[102]  Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massague J, et al. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature. 1995;376(6538):313-20.

[103]  Andrews B, Measday V. The cyclin family of budding yeast: abundant use of a good idea. Trends Genet. 1998;14(2):66-72.

[104]  Kuhne C, Linder P. A new pair of B-type cyclins from *Saccharomyces cerevisiae* that function early in the cell cycle. EMBO J. 1993;12(9):3437-47.

[105]  Schwob E, Nasmyth K. *CLB5* and *CLB6*, a new pair of B cyclins involved in DNA replication in *Saccharomyces cerevisiae*. Genes Dev. 1993;7(7A):1160-75.

[106]  Epstein CB, Cross FR. *CLB5*: a novel B cyclin from budding yeast with a role in S phase. Genes Dev. 1992;6(9):1695-706.

[107]  Jackson LP, Reed SI, Haase SB. Distinct mechanisms control the stability of the related S-phase cyclins Clb5 and Clb6. Mol Cell Biol. 2006;26(6):2456-66.

[108]  Donaldson AD, Raghuraman MK, Friedman KL, Cross FR, Brewer BJ, Fangman WL. *CLB5*-dependent activation of late replication origins in *S. cerevisiae*. Mol Cell. 1998;2(2): 173-82.

[109]  Donaldson AD. The yeast mitotic cyclin Clb2 cannot substitute for S phase cyclins in replication origin firing. EMBO Rep. 2000;1(6):507-12.

[110]  Hu F, Gan Y, Aparicio OM. Identification of Clb2 residues required for Swe1 regulation of Clb2-Cdc28 in *Saccharomyces cerevisiae*. Genetics. 2008;179(2):863-74.

[111]  Cross FR, Jacobson MD. Conservation and function of a potential substrate-binding domain in the yeast Clb5 B-type cyclin. Mol Cell Biol. 2000;20(13):4782-90.

[112]  DeCesare JM, Stuart DT. Among B-type cyclins only *CLB5* and *CLB6* promote premeiotic S phase in Saccharomyces cerevisiae. Genetics. 2012;190(3):1001-16.

[113]  Koivomagi M, Valk E, Venta R, Iofik A, Lepiku M, Morgan DO, et al. Dynamics of Cdk1 substrate specificity during the cell cycle. Mol Cell. 2011;42(5):610-23.

[114]  Loog M, Morgan DO. Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. Nature. 2005;434(7029):104-8.

[115]  Wilmes GM, Archambault V, Austin RJ, Jacobson MD, Bell SP, Cross FR. 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. 2004;18(9):981-91.

[116]  Kamimura Y, Tak YS, Sugino A, Araki H. Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in *Saccharomyces cerevisiae*. EMBO J. 2001;20(8):2097-107.

[117]  Tanaka T, Umemori T, Endo S, Muramatsu S, Kanemaki M, Kamimura Y, et al. Sld7, an Sld3-associated protein required for efficient chromosomal DNA replication in budding yeast. EMBO J. 2011;30(10):2019-30.

[118]  Hopwood B, Dalton S. Cdc45p assembles into a complex with Cdc46p/Mcm5p, is required for minichromosome maintenance, and is essential for chromosomal DNA replication. Proc Natl Acad Sci U S A. 1996;93(22):12309-14.

[119]  Kamimura Y, Masumoto H, Sugino A, Araki H. Sld2, which interacts with Dpb11 in *Saccharomyces cerevisiae*, is required for chromosomal DNA replication. Mol Cell Biol. 1998;18(10):6102-9.

[120]  Hennessy KM, Lee A, Chen E, Botstein D. A group of interacting yeast DNA replication genes. Genes Dev. 1991;5(6):958-69.

[121]  Heller RC, Kang S, Lam WM, Chen S, Chan CS, Bell SP. Eukaryotic origin-dependent DNA replication in vitro reveals sequential action of DDK and S-CDK kinases. Cell. 2011;146(1):80-91.

[122]  Sclafani RA, Tecklenburg M, Pierce A. The *mcm5-bob1* bypass of Cdc7p/Dbf4p in DNA replication depends on both Cdk1-independent and Cdk1-dependent steps in Saccharomyces cerevisiae. Genetics. 2002;161(1):47-57.

[123]  Aparicio OM, Stout AM, Bell SP. Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. Proc Natl Acad Sci U S A. 1999;96(16):9130-5.

[124]  Kanemaki M, Labib K. Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. EMBO J. 2006;25(8):1753-63.

[125] Sheu YJ, Stillman B. The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. Nature. 2010;463(7277):113-7.

[126] Hardy CF, Dryga O, Seematter S, Pahl PM, Sclafani RA. *mcm5/cdc46-bob1* bypasses the requirement for the S phase activator Cdc7p. Proc Natl Acad Sci U S A. 1997;94(7): 3151-5.

[127] Hoang ML, Leon RP, Pessoa-Brandao L, Hunt S, Raghuraman MK, Fangman WL, et al. Structural changes in Mcm5 protein bypass Cdc7-Dbf4 function and reduce replication origin efficiency in *Saccharomyces cerevisiae*. Mol Cell Biol. 2007;27(21): 7594-602.

[128] Masumoto H, Muramatsu S, Kamimura Y, Araki H. S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. Nature. 2002;415(6872):651-5.

[129] Tanaka S, Umemori T, Hirai K, Muramatsu S, Kamimura Y, Araki H. CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. Nature. 2007;445(7125):328-32.

[130] Zegerman P, Diffley JF. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. Nature. 2007;445(7125):281-5.

[131] Tak YS, Tanaka Y, Endo S, Kamimura Y, Araki H. A CDK-catalysed regulatory phosphorylation for formation of the DNA replication complex Sld2-Dpb11. EMBO J. 2006;25(9):1987-96.

[132] Araki H, Leem SH, Phongdara A, Sugino A. Dpb11, which interacts with DNA polymerase II(epsilon) in *Saccharomyces cerevisiae*, has a dual role in S-phase progression and at a cell cycle checkpoint. Proc Natl Acad Sci U S A. 1995;92(25):11791-5.

[133] Bork P, Hofmann K, Bucher P, Neuwald AF, Altschul SF, Koonin EV. A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. FASEB J. 1997;11(1):68-76.

[134] Williams RS, Lee MS, Hau DD, Glover JN. Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. Nat Struct Mol Biol. 2004;11(6):519-25.

[135] Muramatsu S, Hirai K, Tak YS, Kamimura Y, Araki H. CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol (epsilon}, and GINS in budding yeast. Genes Dev. 2010;24(6):602-12.

[136] Takayama Y, Kamimura Y, Okawa M, Muramatsu S, Sugino A, Araki H. GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. Genes Dev. 2003;17(9):1153-65.

[137] Ilves I, Petojevic T, Pesavento JJ, Botchan MR. Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. Mol Cell. 2010;37(2):247-58.

[138] Sclafani RA, Holzen TM. Cell cycle regulation of DNA replication. Annu Rev Genet. 2007;41:237-80.

[139] Tercero JA, Longhese MP, Diffley JF. A central role for DNA replication forks in checkpoint activation and response. Mol Cell. 2003;11(5):1323-36.

[140] Maine GT, Sinha P, Tye BK. Mutants of *S. cerevisiae* defective in the maintenance of minichromosomes. Genetics. 1984;106(3):365-85.

[141] Merchant AM, Kawasaki Y, Chen Y, Lei M, Tye BK. A lesion in the DNA replication initiation factor Mcm10 induces pausing of elongation forks through chromosomal replication origins in *Saccharomyces cerevisiae*. Mol Cell Biol. 1997;17(6):3261-71.

[142] Homesley L, Lei M, Kawasaki Y, Sawyer S, Christensen T, Tye BK. Mcm10 and the MCM2-7 complex interact to initiate DNA synthesis and to release replication factors from origins. Genes Dev. 2000;14(8):913-26.

[143] Izumi M, Yanagi K, Mizuno T, Yokoi M, Kawasaki Y, Moon KY, et al. 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. 2000;28(23):4769-77.

[144] Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, et al. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat Cell Biol. 2006;8(4):358-66.

[145] Warren EM, Huang H, Fanning E, Chazin WJ, Eichman BF. Physical interactions between Mcm10, DNA, and DNA polymerase alpha. J Biol Chem. 2009;284(36): 24662-72.

[146] van Deursen F, Sengupta S, De Piccoli G, Sanchez-Diaz A, Labib K. Mcm10 associates with the loaded DNA helicase at replication origins and defines a novel step in its activation. EMBO J. 2012;31(9):2195-206.

[147] Kanke M, Kodama Y, Takahashi TS, Nakagawa T, Masukata H. Mcm10 plays an essential role in origin DNA unwinding after loading of the CMG components. EMBO J. 2012;31(9):2182-94.

[148] Ricke RM, Bielinsky AK. Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. Mol Cell. 2004;16(2):173-85.

[149] Pursell ZF, Isoz I, Lundstrom EB, Johansson E, Kunkel TA. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science. 2007;317(5834):127-30.

[150] Larrea AA, Lujan SA, Nick McElhinny SA, Mieczkowski PA, Resnick MA, Gordenin DA, et al. Genome-wide model for the normal eukaryotic DNA replication fork. Proc Natl Acad Sci U S A. 2010;107(41):17674-9.

[151] Santocanale C, Foiani M, Lucchini G, Plevani P. The isolated 48,000-dalton subunit of yeast DNA primase is sufficient for RNA primer synthesis. J Biol Chem. 1993;268(2): 1343-8.

[152] Plevani P, Foiani M, Valsasnini P, Badaracco G, Cheriathundam E, Chang LM. Poly-peptide structure of DNA primase from a yeast DNA polymerase-primase complex. J Biol Chem. 1985;260(11):7102-7.

[153] Kilkenny ML, De Piccoli G, Perera RL, Labib K, Pellegrini L. A conserved motif in the C-terminal tail of DNA polymerase alpha tethers primase to the eukaryotic replisome. J Biol Chem. 2012;287(28):23740-7.

[154] Lou H, Komata M, Katou Y, Guan Z, Reis CC, Budd M, et al. Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint. Mol Cell. 2008;32(1):106-17.

[155] Gambus A, van Deursen F, Polychronopoulos D, Foltman M, Jones RC, Edmondson RD, et al. A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome. EMBO J. 2009;28(19):2992-3004.

[156] Lee C, Liachko I, Bouten R, Kelman Z, Tye BK. Alternative mechanisms for coordinating polymerase alpha and MCM helicase. Mol Cell Biol. 2010;30(2):423-35.

[157] Fanning E, Klimovich V, Nager AR. A dynamic model for replication protein A (RPA) function in DNA processing pathways. Nucleic Acids Res. 2006;34(15):4126-37.

[158] Acharya N, Klassen R, Johnson RE, Prakash L, Prakash S. PCNA binding domains in all three subunits of yeast DNA polymerase delta modulate its function in DNA replication. Proc Natl Acad Sci U S A. 2011;108(44):17927-32.

[159] Costa A, Ilves I, Tamberg N, Petojevic T, Nogales E, Botchan MR, et al. The structural basis for MCM2-7 helicase activation by GINS and Cdc45. Nat Struct Mol Biol. 2011;18(4):471-7.

[160] Watase G, Takisawa H, Kanemaki MT. Mcm10 plays a role in functioning of the eukaryotic replicative DNA helicase, Cdc45-Mcm-GINS. Curr Biol. 2012;22(4):343-9.

[161] Nougarede R, Della Seta F, Zarzov P, Schwob E. Hierarchy of S-phase-promoting factors: yeast Dbf4-Cdc7 kinase requires prior S-phase cyclin-dependent kinase activation. Mol Cell Biol. 2000;20(11):3795-806.

[162] Jares P, Blow JJ. Xenopus cdc7 function is dependent on licensing but not on XORC, XCdc6, or CDK activity and is required for XCdc45 loading. Genes Dev. 2000;14(12): 1528-40.

[163] Walter JC. Evidence for sequential action of cdc7 and cdk2 protein kinases during initiation of DNA replication in Xenopus egg extracts. J Biol Chem. 2000;275(50): 39773-8.

[164] Oikonomou C, Cross FR. Rising cyclin-CDK levels order cell cycle events. PLoS One. 2011;6(6):e20788.

[165] Ikui AE, Archambault V, Drapkin BJ, Campbell V, Cross FR. Cyclin and cyclin-dependent kinase substrate requirements for preventing rereplication reveal the need for concomitant activation and inhibition. Genetics. 2007;175(3):1011-22.

# Replicative Helicases as the Central Organizing Motor Proteins in the Molecular Machines of the Elongating Eukaryotic Replication Fork

John C. Fisk, Michaelle D. Chojnacki and
Thomas Melendy

Additional information is available at the end of the chapter

## 1. Introduction

Major processes in the cell often involve the coordinated and efficient assembly of macro-molecular complexes; such examples include: RNA transcription, DNA replication, trans-lation, and cellular motion. These processes can be likened to miniature forms of machines, so-called "molecular machines" with multiple components and motors at their heart driving the systems. This term has been used by several researchers, which equate many of life's inner workings as homologous to machines; albeit much more efficient than their macro-type counterparts [104]. In 1998, Bruce Alberts wrote an elegant article for *Cell* noting the inherent beauty of molecular biology's machines, praising them and stating that as with all machines these macromolecular complexes must in turn contain an assortment of moving parts that act in a highly coordinated fashion with each other [1]. One such studied process is DNA replication, which has been extensively studied since the discovery of the DNA double helix. Due to the biological necessity for duplica-tion of the genetic material, and the intricate link between the faithful replication of the genomic blueprint and its mismanagement leading to cancer, it is difficult to envision a process more important to human health than the study of DNA replication. The motor that drives the molecular machine that is DNA replication is the replicative DNA heli-case. Replicative DNA helicases are well known as the motors that drive DNA replication forks along the DNA strands. But in more recent years it is becoming evident that repli-cative helicases also coordinate the necessary associations and dissociations of the various DNA replication complexes that need to act at the elongating replication fork. Here we

---

will review the current knowledge of how the molecular motors, replicative DNA helicases, coordinate the actions of the molecular machines that are elongating eukaryotic DNA replication forks.

## 2. Phases of DNA replication

The replication of DNA during the Synthesis (S) Phase of the cell is generally differentiated into distinct stages. The first is the binding and **recognition of the origin** of replication by origin binding proteins. For cellular replication in eukaryotes, these proteins are the Origin Recognition Complex (ORC) proteins, many of which belong to AAA+ family of cellular ATPases [20, 97]. To begin activation of the origin (i.e. - **licensing**), two other proteins must act to make origins competent, Cdc6 and Cdt1 [5, 112]. These two proteins in turn are regulated by phosphorylation by Cdc7/Dbf4 as well as by geminin (in metazoans). The presence of ORC/Cdc6/Cdt1 are necessary for recruitment of the next set of vital DNA replication proteins, the minichromosome maintenance (MCM) proteins, which are components of the replicative DNA helicase [70, 115]. For many years, the MCM complex was proposed to be the replicative helicase; but both *in vitro* and *in vivo* studies could not verify that the MCM complex was in fact the DNA helicase necessary for eukaryotic replication [53, 68, 137]. However, it was well established that the six 'core' MCMs, MCM2-7, were essential for DNA replication and that their deletion was lethal in yeasts [125]. Additionally, MCMs appear to associate with chromatin just prior to S Phase, and dissociate from the chromatin as S Phase progresses, consistent with that of a DNA replication helicase [24, 117]. Only recently was it discovered that the MCM complex appears to be an incomplete DNA helicase, in that several additional proteins recruited during origin activation appear to be required to make up the DNA helicase holoenzyme. Cdc45 and the GINS (in Japanese Go-Ichi-Ni-San, which stands for the numbers 5-1-2-3 in the subunits Sld5, Psf1, Psf2, and Psf3) complex appear to make up the CMG (Cdc45-MCM-GINS), the complex multisubunit eukaryotic helicase [91], required for initiation of DNA replication. In spite of this elucidation of the CMG, the step-wise recruitment of these helicase components, and the complex nature of the post-translational modification steps required to reconstitute a functional CMG replicative DNA helicase, has severely constrained the ability to carry out detailed biochemical analyses of the eukaryotic DNA replication fork.

The formation of an active pre-replication complex at the origin, and the subsequent formation and activation of the CMG replicative DNA helicase allows for the recruitment of DNA polymerase $\alpha$ primase, which is necessary for the synthesis of RNA primers and a short DNA extension of those primers. Also recruited is RPA, the major ssDNA binding complex necessary to prevent the re-annealing of the DNA duplex [132], and topoisomerase I, which resolves the compression of the DNA helix caused by progression of the replication fork along the DNA duplex (**Initiation of DNA replication**). Following primer synthesis, the clamp loader, RFC, is loaded at the 5' end of the primers, and RFC in turn loads the DNA polymerase processivity factor, PCNA. Due to the 5'->3' nature of DNA

replication, synthesis occurs continuously on the leading strand through the recruitment and activity of DNA polymerase ε [106] and discontinuously on the lagging strand by DNA polymerase δ extension of the repeated primers laid down by DNA polymerase α primase [98]. The components of polymerase δ are also often found, not surprisingly, associated with the proteins involved in "processing" the lagging strand Okazaki fragments, namely those proteins involved in removing the RNA primers (see below) [65, 103]. During this **elongation phase of DNA replication** is when the majority of DNA synthesis occurs. However, while the ORC complex and other components of the origin recognition/licensing machinery are dispensable following origin firing [26], the heart of the DNA replication apparatus remains associated with the replicative helicase, the molecular motor that is actively unwinding the DNA duplex. How the replicative helicase interacts with components of the elongation machinery is probably the least understood remaining aspect of DNA replication and is the focus of this review. Many other components are implicated in the elongation phase of eukaryotic DNA replication, such as Mrc1 (Claspin), which has been suggested to be involved in linking the helicase to the polymerases and has been found to be involved in the "uncoupling" of these two aspects of the fork during the DNA damage response [6, 56], and for regulating fork progression during uncompromised DNA synthesis [44, 78, 118, 122].

Following elongation, the RNA primers and the RNA-DNA linkages are removed through the actions of the flap endonuclease-1 (FEN1) nuclease and/or Pif1 helicase and Dna2 nuclease, assisted by RPA and DNA polymerase δ [74, 105, 108]. Following the removal of the primers, gaps are filled in, apparently by the action of the DNA polymerase δ and its cofactors, and the final DNA strands are ligated by DNA ligase I into long uninterrupted DNA chains. The removal of all the primers, filling of the subsequent gaps, and the final ligation of the products represent the completion of S-phase.

## 3. Model systems for elongation of DNA replication

As mentioned previously, eukaryotic cellular DNA replication is highly complicated, and only recently has the replicative DNA helicase finally been identified as MCM2-7 complexed with Cdc45 and the GINS complex (CMG) [91]; furthermore, the complex nature of assembly and regulation of this CMG replicative helicase has limited the ability to study the eukaryotic replication fork biochemically. However, early mechanistic studies of eukaryotic DNA replication were largely carried out using the small DNA tumor virus SV40 and to a lesser extent the papillomaviruses. What makes these viruses ideal models for the mechanistic study of eukaryotic DNA replication? One reason lies in their small genome size. To facilitate their duplication, these viruses make the most of their small number of ORFs by combining multiple replication functions into one or two proteins, and relying primarily on the host cell DNA replication machinery (see Table 1). In addition, the lack of these viruses utilizing the once-and-only-once per S Phase regulation of DNA replication means that their DNA replication systems were not subject to the complicated and constraining regulatory systems that control replication of cellular DNA.

SV40 DNA replication is driven by a single viral protein, SV40 large T-antigen (Tag), a protein that combines all the core DNA replication functions of the cellular initiation and origin activation proteins listed above for eukaryotic DNA replication. Tag recognizes and binds to the SV40 origin of replication, melts the DNA helix surrounding the origin, and establishes itself into a double hexameric structure. Tag then recruits the cellular DNA replication factors: RPA, topoisomerase I, and polymerase $\alpha$ primase. These four replication factors are all that is required for the initiation of SV40 DNA replication through the initial synthesis of RNA-DNA primers. Following these initiation events, the clamp loader, RFC, and the polymerase processivity factor, PCNA, are recruited and loaded, which leads to the binding and activity of DNA polymerase $\delta$, which extends both lagging and leading strands in this viral DNA replication system. As in the mammalian system, Okazaki fragments are processed by FEN1, DNA helicase 2, and DNA ligase 1, completing synthesis of the viral DNA genomes. It was the early studies of this viral DNA synthesis system that elucidated these basic mechanisms of how eukaryotic DNA replication is carried out.

| Replication Step/Function | Mammalian | SV40 | Papillomavirus (PV) |
|---|---|---|---|
| Origin Recognition/Initiator | Orc complex (2-6) | T-antigen (Tag) | E2/E1 |
| pre-RC | Cdc6, Cdt1, Cdc45, Geminin, MCM10, Sld2(RecQL4), Sld3, Dpb11(TopBP1) | Tag | E2/E1 |
| Helicase | MCM 2-7, GINS, Cdc45 | Tag | E1 |
| SSB | RPA | RPA | RPA |
| Torsional relaxation | Topoisomerase I | Topoisomerase I | Topoisomerase I |
| Clamp loader | RFC | RFC | RFC |
| Processivity factor | PCNA | PCNA | PCNA |
| DNA polymerases | DNA pol $\alpha$ primase, DNA pol $\delta$, DNA pol $\varepsilon$ | DNA pol $\alpha$ primase, DNA pol $\delta$ | DNA pol $\alpha$ primase, DNA pol $\delta$ |
| Accessory factors | Mrc1(Claspin) | None | None? |

**Table 1.** Known and Proposed Components of the DNA replication complex

Similar findings were also found for another virus family, the papillomaviruses, which have also proven to be an apt model for cellular DNA replication mechanisms due to

their dependence on the host replication machinery. Initial studies were carried out in the bovine version BPV-1, and later corroborated with several human HPV isotypes. In general, papillomaviruses follow the same mode and progression of events found in SV40, except for the need for two viral proteins instead of the single Tag protein required for SV40. In addition, PV appears to require other cellular factors that SV40 does not [73, 80, 87], which to date remain unidentified. In papillomavirus DNA replication, the E2 protein assists and directs faithful viral origin recognition of E1 [79, 90, 110, 126], while E1 itself serves the role of the replicative DNA helicase, melting the DNA around the origin of replication and establishing itself as a double hexameric helicase. In a fashion similar to that of SV40 Tag, E1 also acts to recruit the cellular DNA replication proteins to the PV DNA replication fork [36, 113, 131]. E1 itself is a weak origin binding protein, but can bind to and unwind DNA even in the absence of E2 at high E1 concentrations, even on DNA without an apparent E1 binding sequence and is therefore relatively nonspecific without E2 [66]. Furthermore, following establishment of the double hexamer, the E2 protein is purportedly absent from subsequent steps of DNA replication, indicating E1 is the only viral protein implicated in the actual HPV elongating DNA fork [72]. Otherwise, these two small DNA viruses display very similar mechanisms of replication, especially during the elongation phase. So why rely on two very similar viruses as models and not just SV40? One reason is that by comparing and contrasting the DNA replication mechanisms in two subtly different systems, one gains further insight into the mechanisms of DNA replication. In specific aspects of DNA replication, one or the other virus might provide a more apt reflection of the mechanism of cellular DNA replication. Another reason lies in the diseases each virus causes and the implications for antiviral research. Although SV40 Tag is a potent transforming agent for cell culture due to its ability to inactivate p53, Rb protein, and many other components of the cell, SV40 itself does not appear to readily cause tumors in humans. Conversely, human papillomaviruses are the major cause of cervical, anogenital, and oral cancers and represent the major cause of infectious-agent-induced cancers in humans. These viruses represent historically important and still valuable models for DNA replication and can still be used to elucidate hitherto unknown mechanisms of mammalian DNA replication. Furthermore, the replicative DNA helicases of these viral DNA replication systems still provide the best biochemical system for investigating the role of DNA helicases in the elongation stage of eukaryotic DNA replication.

## 4. Replicative DNA helicases

When the structure of the DNA double helix was first proposed, one of the major questions concerning the replication of dsDNA was how the duplex would be opened to facilitate reading of the base sequence encoded by the DNA. The first such discovered protein that could carry out this function was the prokaryotic helicase of *E. coli*, discovered in 1976. All known helicases use the energy from NTPs to drive the remodeling of their substrate nucleic acids [75, 85]. Helicases are grouped into six superfamilies (SF1-SF6) and all

possess typical Walker A and B motifs involved in NTP binding and hydrolysis. The motor proteins of the macromolecular machines at DNA replication forks are all AAA+ module-containing helicases, which function to unwind the DNA helix and to drive the replication machinery along the DNA template. Another common characteristic of replicative helicases is that most form higher order oligomeric structures to facilitate their functions as DNA helicases at DNA replication forks. The MCM complex of the CMG cellular helicase, SV40 Tag, and PV E1 all form hexamers. Both Tag and E1 have been recently crystalized in their hexameric forms, which has contributed significantly to elucidating how these helicases function in splitting the DNA helix [32, 38, 71]. Further, Tag and E1, and later MCMs, were shown to form dimers of two hexamers [34, 36], which are presumed to act in bridging the two DNA replication forks, holding them together during replication fork progression, and creating a system whereby the template DNA is threaded through the DNA replication machinery in both directions simultaneously.

Various models have been proposed for how DNA helicases unwind the DNA helix. Some early proposals included the monomers binding to the DNA backbone and essentially rolling one DNA strand away from the other using the circular nature of the hexamer. Other models included a hexamer 'embracing' ssDNA, excluding it from its partner, or two hexamers acting at a distance pumping dsDNA through their central pore. Some studies indicate the double hexamers stay associated during elongation, and this led to a double hexameric DNA pumping mechanism that pumps dsDNA through the central pore somehow splitting the helix [42]. The more recent structural studies of the BPV1 E1 helicase bound to DNA, ATP, and ADP indicate an intricate hybrid model whereby the E1 hexamer pumps ssDNA through each central pore in a staircase type mechanism as ATP is bound and hydrolyzed by each subsequent E1 monomer [32, 33, 109]. In this model E1 uses the ATP binding/hydrolysis-induced conformational changes of the individual monomers to drive each nucleotide base of the enclosed ssDNA template through the central pore, displacing the hybridized (lagging-strand template) DNA strand freeing it to be available as a template for lagging strand DNA synthesis [32]. Although the model for helicase action based on the SV40 Tag structure was not the same, the Tag structure was done in the absence of ssDNA, and the structural information on the Tag hexamer would be consistent with a helicase model similar to that of E1.

## 5. Helicase interactions with replication proteins that initiate elongation

As stated previously, DNA replication proteins commonly recruited by both of these viral replicative helicases are: RPA, topoisomerase I, and DNA polymerase $\alpha$ primase. In this section, we will look closer at the individual and combinatorial interactions between the helicase and these necessary DNA replication factors that are intimately involved in both the initiation and elongation stages of DNA replication. In many cases, studies have focused on specific interactions, often detailed down to specific amino acid residues required for recruitment of these factors. Various groups have used the powerful ability to investigate the interactions of these factors with the viral helicases both *in vitro* and *in vivo*, to elegantly

demonstrate the importance of these molecular contacts. For each of these three DNA replication factors, we will look into the extensive work that has been performed in the SV40 system with Tag, then in the PV system with E1. Following this, we will briefly touch on the mammalian system, highlighting some of the similarities between the viral and the mammalian host systems.

## 6. Helicase interactions with replication proteins that initiate elongation: Topoisomerase I

The unwinding action of the DNA replication fork driving along the DNA helix creates torsional stress and overwound DNA that must be relieved to allow replication to proceed. Topoisomerases are enzymes that help relieve this stress and aid in maintaining chromosome structure and integrity by modifying DNA topology, and resolving specific DNA structures that arise from cellular processes such as DNA repair, replication, transcription, recombination and chromosome compaction [13]. These processes result in compression (positive supercoiling) of the DNA helix and the entanglement of DNA segments and chromosomal regions that can lead to cytotoxic or mutagenic breaks in the DNA if left unmanaged [127]. Hence, topoisomerases play a vital role in living cells, particularly during DNA replication.

Enzymatically, topoisomerases act through the action of a nucleophilic tyrosine; the enzyme cleaves one or more DNA strands and generates an enzyme-DNA complex that serves to prevent the release of nicked or broken DNA that could possibly result in chromosome damage [127]. After passage of one or more DNA strands through this transient break(s), the topoisomerase re-ligates the strands leaving the original DNA sequence intact. Though all topoisomerases have this feature in common, topoisomerases are separated into two classes, type I and type II, depending on whether they cleave one or two strands of DNA, respectively [127]. Type I topoisomerases act on one strand, and generally pass a single DNA strand through the transient break, while type II topoisomerases break both DNA strands and generally pass dsDNA through the transient break. Type I topoisomerases generally work in front of replication or transcription forks, to relax positive supercoils in a highly processive manner; while type II topoisomerases are involved in untangling intertwined duplex DNA such as that found in newly replicated molecules or during chromosome resolution during cell division [30].

Topoisomerases have roles in each of the major replicative phases: initiation, fork progression and termination. During DNA replication in eukaryotes, topoisomerases have been observed to bind directly to the replication origin to aid in activation in the initiation phase [45, 127]. During strand synthesis, topoisomerases are required to alleviate compression of the DNA helix caused by positive supercoiling that results from DNA unwinding, which is mediated by replicative helicases [127]. Topoisomerases are also required for daughter strand resolution. Eukaryotes rely on topoisomerase I (topo I) to fulfill the initiation and elongation functions during DNA replication [127].

Human topo I is an ATP-independent, 100-kDa monomeric protein capable of relaxing positive or negative superhelical twists by making a transient single-strand break, thus relieving the tension generated by the replicative helicases during the DNA-unwinding process [61, 127, 135]. Topo I can be divided into four domains: the highly charged $NH_2$ –terminal domain; the conserved core domain; a short, positively charged linker domain, which links the N-terminal domain to the core domain; and the highly conserved COOH-terminal domain, which contains the active-site tyrosine [116]. Due to the topologically constrained nature of a circular dsDNA molecule, it is no surprise that topo I is required for the replication of the genomes of small circular double-stranded DNA viruses. The role of topo I in DNA replication of the small DNA circular DNA viruses was first noted when it was observed that the extent of DNA replication in SV40 DNA replication *in vitro* was limited by the level of topoisomerase activity; addition of topo I to crude extracts stimulated SV40 DNA replication *in vitro* [51]. This effect could have been due to an enhanced rate of chain elongation resulting from an increased efficiency of unlinking of the parental DNA strands [135], or merely due to the presence of limited levels of topo I in the extracts used. Ultimately it was shown that the DNA replication of SV40 and PV both require topo I [134, 136].

While the role of topoisomerases in DNA replication had always been presumed to be due to their need to resolve topological constraint, more recent studies have indicated that topo I plays additional, highly specific, roles in DNA replication of the small DNA viruses, SV40 and PV. Topo I appears to be involved in the very earliest stages of DNA replication, namely origin recognition. It is evident that topo I is stably associated with the initiation complex and is one of the first cellular proteins to be recruited to the initiation machinery [11,45]. Topo I was shown to preferentially associate with the fully formed Tag double hexamer initiation complexes and to be recruited to the initiation complex prior to the beginning of unwinding [11]. This stable association of topo I with Tag results in an increased specificity of Tag for duplex unwinding at the origin by inhibiting unwinding at non-origin sites [39]. Perhaps for this reason, topo I was observed to be required at initiation to stimulate DNA replication *in vitro*, and was shown to have no effect on replication if introduced during the elongation phase, indicating it enhanced the synthesis of fully replicated DNA molecules by forming essential interactions with Tag and enabling initiation [45,11]. In contrast, topo I specifically enhances origin binding of PV E1 several-fold, but has no effect on non-origin binding [14]. After origin binding, E1 recruits topo I to the replication fork through direct protein interactions and the relaxation activity is strongly enhanced [14,4]. This enhancement of topo I is critical to relax the supercoiling created by the progressing replication fork during the elongation phase of DNA replication. Notably, although topo I plays a significant role in where Tag unwinds the DNA, topo I does not activate origin binding or unwinding and does not structurally distort the DNA [39]. Nonetheless, the similarities in these findings indicate that topo I plays an active role in origin recognition/specificity for the replication of both of these small DNA viral systems. Moreover, following initiation, the topo I-helicase complex remains stably associated and moves with the replication fork during DNA replication [45].

Topoisomerases have been proposed to act together with DNA helicases as "swivelases", tightly coordinating DNA duplex unwinding with the topoisomerase relaxing activity during DNA replication [15, 30, 61]. With the progression of the replication fork and unwinding of duplex DNA, topo I is needed to release the torsion created by the progressing replication fork [37]. Optimally topo I should be present and its activity regulated to suit the pace of the helicase [37]. This suggested that there might be direct interactions between the helicases and topo I, and that might be modulation of function due to these interactions. The early finding that topo I was localized at SV40 DNA replication forks supported this concept [4], as did evidence that topo I played an important role in the elongation phase of SV40 DNA replication. Reports of the interactions between SV40 TAg and E1 with topo I were also consistent with the swivelase model [15, 133]. The demonstration that E1 stimulates the enzymatic activity of topo I up to seven-fold and that SV40 TAg also stimulates topo I activity (R. Clower and T. Melendy, unpublished results) provided the first evidence of the cooperative nature of this interaction predicted by the swivelase model [15]. Based on these studies it is clear that the viral helicases interact productively with topo I at DNA replication forks forming active coordinated swivelase molecular machines.

The physical interactions between the viral helicases and topo I have been investigated. In 1996, it was found that two independent regions of Tag, one N-terminal and one C-terminal, bind to the cap region of topo I (see Fig. 1), and binding can take place while DNA-bound. Similarly, for PV E1 it was also observed that topo I binds two distinct regions within E1, within E1's DNA binding domain (DBD) and at the C-terminus [15, 45]. The E1 C-terminal region was shown to enhance topo I relaxation activity, and to a lesser extent, so did a truncation that included the DBD with additional sequence, flanking either side of the DBD [15]. More detailed studies identified mutants in the DNA binding domain of Tag that were unable to unwind the DNA and were partially defective in their association with topo I, suggesting that this interaction maybe important for proper unwinding of viral DNA at replication forks [114]. More recently, four specific amino acid residues within the C-terminal domain of Tag when mutated were shown to exhibit decreased topo I binding and to abolish SV40 DNA replication *in vitro* and to have dramatic effects on virus production *in vivo* [61]. These were the critical results that conclusively demonstrated the vital nature of the helicase-topo I interaction for DNA replication. Though first only believed to be involved in the relaxation of overwound DNA during replication fork progression, topo I has proven to be an integral part of the entire replication process in SV40 DNA replication, including critical roles in initiation and even in RNA-DNA primer synthesis in the elongation phase [37, 60, 61, 123]. Though less well-studied, topo I has been observed to be similarly important in these stages of PV DNA replication. These viral systems are vital models for eukaryotic DNA replication, and as of yet these biochemical studies cannot be recapitulated for cellular DNA replication. The only evidence to date of corroboration of these findings for chromosomal DNA replication is the co-purification of topo I with the GINS-MCM complex [39].
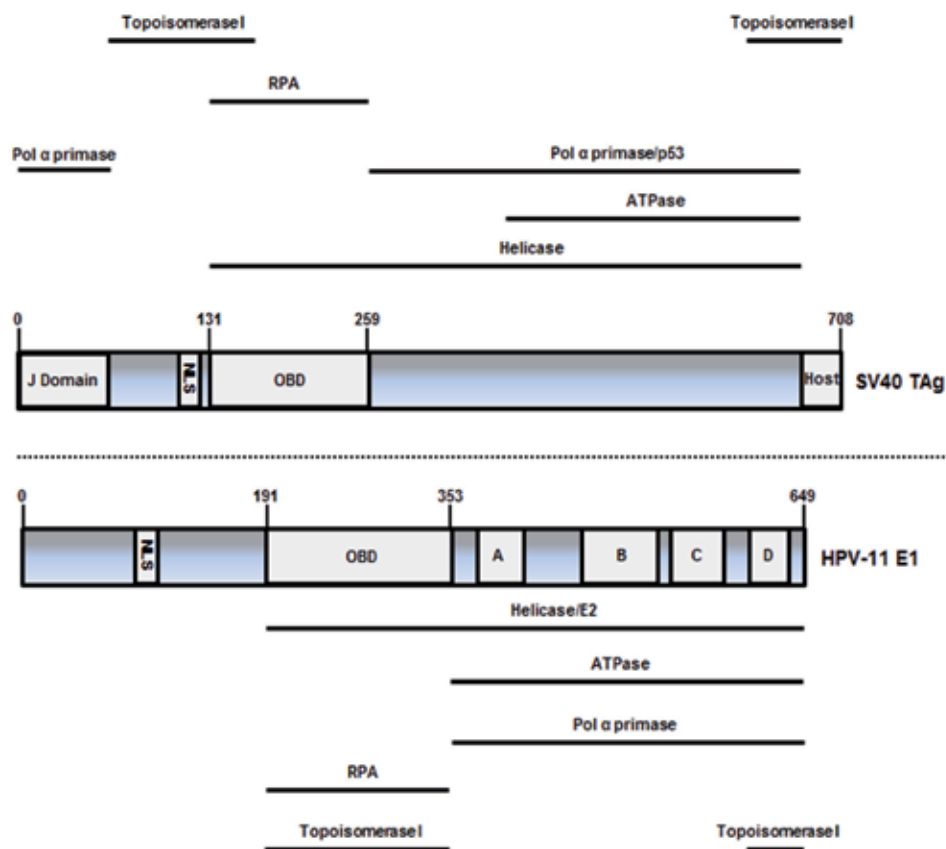
**Figure 1. General replication domains of the SV40 T-antigen and papillomavirus E1 helicases.** The known domains for SV40 Tag (upper) and HPV-11 E1 (lower) are indicated by horizontal lines. Four domains in E1 have limited homology with SV40 T-antigen (regions A-D). A nuclear localization signal has only been elucidated for BPV-1 E1, therefore the HPV-11 NLS is currently only speculated to be in the analogous sequence area.

## 7. Helicase interactions with replication proteins that initiate elongation: Replication protein A

One of the first proteins identified as necessary for eukaryotic DNA replication is arguably also one of the most important DNA binding proteins in the cell, the ssDNA binding complex, Replication Protein A (RPA). RPA is a heterotrimeric complex conserved in all eukaryotes, and also shows strong homology to the ssDNA binding proteins of archaebacteria [57, 59]. The human RPA complex is comprised of three subunits, RPA70, RPA32, and RPA14, and the complex binds to ssDNA with extremely high affinity (approximately $10^{-9}$ to $10^{-10}$ M [62]), showing much lower affinity for dsDNA. RPA binds ssDNA with a distinct polarity, in a 5'->3' orientation [22, 51]. Like SSB [132], RPA is required for DNA replication *in vivo*; knockdown of the largest RPA subunit, RPA70, using siRNA results in inhibition of DNA synthesis [25]. The presence of RPA

as the ssDNA binding protein is critical in keeping the DNA double helix from reannealing during DNA replication, as well as protecting the exposed ssDNA from nuclease attack. And while other non-related ssDNA binding proteins (such as *E. coli* or T4 SSB) can support some of these functions (such as ssDNA stabilization and stimulation of the processive DNA polymerases) RPA is specifically required for the early initiation steps of replication, including primer synthesis and stimulation of the DNA polymerase activity of DNA polymerase $\alpha$ primase [10, 81]. RPA is also involved in many DNA recombination and DNA repair pathways, acting as a central coordinator of DNA metabolism [52, 132].

RPA exhibits several DNA binding states. RPA70 has three ssDNA binding sites or oligonucleotide binding (OB) domains and RPA32 has one OB domain [8, 121]. When only RPA70 interacts, this is a lower affinity compacted state, binding to only 8-10 nts. When all four OB domains bind, this represents a higher affinity extended mode that spans ~30 nts [7]. The ability of other proteins to facilitate these binding modes in turn impact the binding of RPA to ssDNA, either covering or exposing various stretches of ssDNA. Since several other proteins bind to RPA through its OB domains, this facilitates a model in which RPA cooperatively hands off and orients the binding of each DNA replication protein through increasing affinity with the subsequent factor [64, 89, 138].

## 7.1. RPA loading onto ssDNA by replicative DNA helicases

RPA plays many roles in the initial steps of elongation as well as throughout DNA replication. Due to its role in ssDNA stabilization, RPA is one of first proteins required following the unwinding of dsDNA. The critical question here is how this process is coordinated in relation to the double hexameric helicase. The RPA heterotrimer itself makes direct contact with the helicase, be it MCM, SV40 Tag, or PV E1 [3, 43, 77, 95, 101, 130]. The first such studied interaction was through Tag, which interacts with RPA through the helicase's origin binding domain (OBD) (Figure 1). The importance of this interaction is implied by the absolute necessity for RPA for SV40 replication, RPA cannot be replaced by ssDNA binding proteins from *E. coli* or even RPA from *S. cerevisiae* [11, 58, 88]. In turn, RPA interacts with Tag through both its RPA70 and possibly to a lesser degree its RPA32 subunits. In PV DNA replication, the E1 helicase interaction with RPA is also critical for viral DNA replication. E1 directly binds to RPA through its largest subunit, RPA70, but does not appear to bind to RPA32 or RPA14 (unlike Tag which binds RPA70 and RPA32) [43, 52, 69, 77]. Similar to the SV40 system, RPA binds to the PV E1 helicase through its major dsDNA binding domain (Figure 1, Fisk JC and T. Melendy, unpublished data).

Evaluation of the multiple interactions between RPA, E1 and ssDNA in various combinations led to development of a novel model for how DNA helicases may 'load' ssDNA binding proteins onto ssDNA being displaced through helicase action [77]. RPA binds well to the E1 protein, but only in the absence of free ssDNA. When RPA was prebound to short (~10 nt) stretches of ssDNA, thereby adopting the short compacted form of RPA, it still bound to E1 as well as RPA not bound to DNA. However, when RPA was bound to longer ssDNA templates (~30 nt or longer), consistent with RPA being in its fully-engaged extended form, RPA would no longer bind to E1. This implied a 'releasing mechanism' by which the E1-

RPA interaction would be released upon RPA binding to ssDNA in RPA's extended form. Based on this data, a model was developed in which free, non-ssDNA-bound RPA is bound by E1. As the E1 helicase unwinds the dsDNA, producing ssDNA, it positions the RPA to bind to the newly exposed ssDNA, releasing RPA from the helicase complex (see Figure 2). As the helicase progresses, subsequent helicase monomers bring subsequent RPA molecules to the ssDNA continuously displaced by helicase action [77]. Very similar results were later shown for SV40 T-antigen, leading to a nearly identical model for RPA placement onto ssDNA during SV40 T-antigen helicase progression [9, 54]. Of course, this simplified model does not take into account topo I or polymerase $\alpha$ primase interactions, but it does suggest how the newly produced ssDNA can be rapidly coated with RPA to prevent reannealing or hairpin formation, and to protect from nuclease attack.
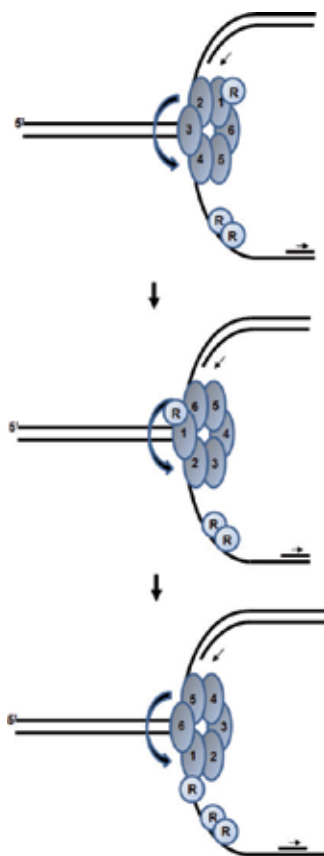


**Figure 2. Generalized model for RPA deposition by replicative helicases.** For simplicity, only a singular replication fork is shown. Free, unbound RPA interacts with a monomer of the helicase. As the helicase rotates relative to the DNA helix, the RPA bound monomer comes in juxtaposition to the freshly unwound ssDNA,. A 'hand off' occurs, whereby RPA binds to the ssDNA released by the helicase. Upon RPA binding ssDNA in the 'elongated' (~30 bp-bound) form, it can no longer be bound by the helicase monomer. This 'release' action leaves RPA bound to the newly exposed ssDNA, and allows the helicase to progress along the DNA template. As additional ssDNA is exposed this process is repeated, creating an array of RPA coating the ssDNA lagging strand template. (Adapted from [78].).

## 8. Helicase interactions with proteins that initiate elongation: DNA polymerase alpha-primase

In bacteria and the T4 bacteriophage, the importance of the primase is clear as they are linked physically to the helicase, which is necessary for efficient lagging strand synthesis [19, 92]. In the T7 bacteriophage, this is even more evident as the primase is actually fused to the functional hexameric helicase [31, 99, 102]. In a more complex fashion, in the mammalian system, GINS/ctf4 are required to link the helicase to the catalytic core of DNA polymerase (pol) alpha [40, 120, 140]. Clearly the interaction between primase and the helicase machinery is conserved throughout evolution.

Pol $\alpha$ primase was the first eukaryotic polymerase discovered in 1957 and was thought to be the only replicative DNA polymerase. The later discovery of the proofreading and highly processive polymerases δ and ε indicated that this was not the case [49, 50]. Pol $\alpha$ primase is a heterotetrameric complex comprised of a large p180 catalytic subunit, the regulatory p68 "B" subunit, and the two primase subunits of p55 and p49. Pol $\alpha$ primase is critical for first synthesizing an approximately 10 nt RNA primer, followed by a short ~20-30 nt DNA extension [23, 41, 119]. Polymerase switching then occurs on this RNA/DNA primer through the action of the eukaryotic clamp-loading complex, RFC, which loads the eukaryotic sliding clamp, PCNA, and then a processive DNA polymerase (DNA pol δ or ε) for synthesis of both leading and lagging DNA strands [124, 128]. RFC is integral here, as it competes with RPA for the end of the primer, disrupting the RPA-pol $\alpha$ interaction and allowing polymerase switching [138]. As with many of the core aspects of eukaryotic DNA replication, the functions of pol $\alpha$ primase were largely elucidated using the SV40 system. Pol $\alpha$ primase is absolutely essential for SV40 DNA replication *in vitro* [94]. Tag interacts with three subunits of pol $\alpha$ primase [17, 27-29, 48, 100, 129]. Recent work has demonstrated the importance of the Tag-p68 interaction for facilitating priming in both cell-free systems and in monkey cell culture [46, 139]. Mutations in Tag that abrogate the Tag-p68 interaction, but do not affect the interactions with p180 or primase, severely decrease priming (in the SV40 monopolymerase assay, which uses a plasmid with the SV40 ori, and purified Tag, pol $\alpha$ primase, RPA, and topo I, all of which are subjects of this review [86]). The amino residues in SV40 Tag shown to be critical for interaction with pol $\alpha$ p68 are H395, R548, K550, and K616, all of which are highly conserved between Tags from other polyomaviruses [139]. Interestingly, the helicase activity of Tag is dispensable for primosome activity (stimulation of priming by pol $\alpha$ primase), indicating that this effect on priming is likely due to the protein-protein interaction between the helicase and the pol, and not some indirect role of DNA helicase action [47]. In general, the Tag-pol $\alpha$ primase interaction mediates a process that allows the helicase to hand off the pol-primase to the ssDNA to enable primer synthesis [16, 35, 88]. Obviously there would also need to be interplay between the helicase and the two proteins competing for the exposed ssDNA, RPA and pol $\alpha$ primase.

The interactions between E1 and pol $\alpha$ primase show some differences between those found with SV40 Tag. Early studies indicated that the p180 catalytic subunit interacted with the N-terminal half of E1, while the p68 subunit interacted with the C-terminus of the helicase [18,

83]. A later study then looked closely at the role of the E1 interaction with pol $\alpha$ primase in regards to supporting HPV-11 DNA replication *in vitro* [2]. This study confirmed the earlier findings by indicating that E1 interacts with the pol p68 regulatory subunit through its C-terminal half (Figure 1). The presence of E2, whose trans-activation domain binds a similar region of E1, stimulates the E1-p68 interaction; but E2 and p68 nonetheless compete for [2, 83]. This is consistent with a step-wise mechanism whereby E2 helps E1 assemble into a functional helicase, which is then recognized by p68 of the pol $\alpha$ primase complex. No interaction with the pol p180 subunit was detected in these latter studies. Whether this was due to subtle differences in the BPV-1 E1 used in the earlier study and HPV-11 E1 in the later studies has not been further investigated. Regardless, PV E1 appears to interact with pol $\alpha$ primase within the same E1 domain as the ATPase/helicase function. Further studies are necessary to determine if similar residues in E1 as those in SV40 Tag facilitate the binding to pol $\alpha$ primase; these studies may be beneficial as E1 may only use this subunit to bind and recruit pol $\alpha$ primase to the viral replication machinery.

## 9. Interactions between replication proteins that initiate elongation: coordination

While the earlier sections have alluded to interplay between the multiple cellular replication factors that interact with the viral helicases during DNA replication, the complexity of the interplay between these interactions is what truly epitomizes the term Molecular Machines.

### 9.1. RPA's involvement in de-repression of priming

While the interaction of the viral helicases with RPA has been shown to have a direct effect, apparently through the placement of RPA on the ssDNA being displaced by helicase action, this interaction has also been shown to play another vital role in DNA replication: de-repression of priming. RPA binds directly to pol $\alpha$ primase [10, 28, 96], and can stimulate the fidelity and processivity of pol $\alpha$ primase activity [10, 81]. However, when RPA is present in excess, which it is in human cell nuclei [76], RPA strongly represses synthesis of primers by pol $\alpha$-primase, likely due to the high affinity of RPA out-competing pol $\alpha$ primase for the ssDNA template [16, 88]. While Tag and pol $\alpha$ primase are required for correct initiation of SV40 DNA replication [27, 130], and the interaction between Tag and pol $\alpha$-primase is sufficient for stimulation of RNA/DNA primer synthesis by pol $\alpha$-primase on ssDNA [16], these are insufficient for efficient primer synthesis when there is competition with ssDNA binding proteins. Tag can de-repress primer synthesis by pol $\alpha$-primase, but only when the ssDNA template is coated by RPA, and not by other ssDNA binding proteins or evolutionarily divergent RPAs [88]. The interaction between Tag and RPA is vital for de-repression of priming [88, 111]. E1 has similarly been shown to interact with RPA, and RPA is required for PV DNA replication (and RPA cannot be replaced by other ssDNA binding proteins in PV DNA replication). So while the E1-RPA interaction has not been shown to be essential for priming de-repression during PV DNA replication, this is nonetheless likely to be the case.

## 9.2. Topo I's involvement in priming

Similarly, in addition to its roles in origin recognition/specificity and release of DNA helix compression during elongation, another role for topo I was elucidated when it was observed that topo I induces pol $\alpha$-primase to synthesize larger amounts of primers with higher molecular weight [60]. In this study, Tag mutants that failed to bind topo I normally did not participate in the synthesis of expected amounts of primers or large molecular weight DNA molecules, indicating that the association of topo I with the C-terminal Tag binding site is required for these processes. Whether this is due to a direct effect on Tag function at the replication fork, or due to an indirect effect on pol $\alpha$-primase through Tag (analogous to the effect of the RPA-Tag effect on priming by pol $\alpha$-primase described above) is unclear. Additionally, topo I was shown to bind directly to RPA, and RPA binds directly to pol $\alpha$-primase, and can stimulate its DNA polymerase activity. It is unclear whether or not RPA may be influencing the interaction of Topo I with pol $\alpha$-primase, or vice versa [60]. However these interactions are integrated, the binding of topo I to the helicase domain of Tag significantly enhances the synthesis of DNA-RNA primers and their extension by pol $\alpha$-primase.

## 9.3. Helicase interactions with other proteins involved in elongation

What of helicase interaction with the other proteins involved in DNA replication elongation? In the model systems of SV40 and PV little has been elucidated about any direct interactions. Of the proteins involved in elongation, very little is known about the role of helicase interaction with pol δ, RFC, PCNA, or the proteins involved in primer removal: RNaseH, DNA2, Fen I, or DNA ligase I. In the accepted model of SV40 DNA replication, the first primers synthesized by pol a primase on the two strands at the origin become the primers for the leading strand of the opposite fork [124]. After recruitment of RFC, PCNA and pol δ, the leading strand polymerase continuously tracks along behind the helicase action. Since the helicase, in this case Tag, unwinds dsDNA at the relatively slow rate of approximately 200 bp/min [93] while pol δ/RFC/PCNA polymerizes at about 80 nts/sec [12], it is reasonable to speculate that the slower speed of the helicase limits the polymerase in such a way to coordinate the entire machinery mechanism. However, the speeds of polymerases are often assayed on artificial templates, and this rate for pol δ/RFC/PCNA is faster than the measured rate of eukaryotic replication forks (~ 2 kb/min). Conversely the measured speed of Tag is far slower than the measured rate for eukaryotic replication forks. It is likely that coordination between the various factors and complexes involved in the replication fork lead to the final replication fork rate that is not dependent on any one factor, but is a characteristic of the coordinated complex. Indeed, it is critical that these machines are tightly regulated; without a tight molecular machine at the fork, there would be wild exposure of ssDNA via the helicase leading to DNA damage signaling. It should also be noted here that DNA pol ε is not needed in SV40 DNA replication [141]. This finding may be due to the lack of a need for two replicative helicases to duplicate small virus genomes. Alternately, DNA pol ε and TopBP1 (Dbp11) play roles in initiation in mammalian replication; this role may be dispensable or even inter-

fere with the Tag/E1 initiator functions [82, 84]. In *E. coli* DNA replication it has been shown that the tau subunit actually links the leading strand DNA polymerase to the replicative helicase, dnaB [63] (which tracks along the lagging strand template, unlike the case for the SV40 and PV replicative helicases, that track along the leading strand template). It remains possible that these viral replicative helicases may have heretofore unobserved interactions with additional cellular factors involved in the elongation stages of replication that play important roles in DNA replication. This is a potential area for future study.

## 9.4. Extrapolation to the cellular chromosomal replication fork

The cellular 'replicative helicase' is still poorly defined. Some have designated the human CMG helicase (a large 11 subunit complex comprising Cdc45 and the MCMs and GINS sub-complexes [91]) to be the replicative helicase, while others have designated the RPC, the "replisome progression complex", comprised of the CMG in complex with Mrc1 (Claspin), Tof1 (Tim or Timeless), Csm3 (Swi3/Tipin), Ctf4 (And-1), and the FACT heterodimer (Spt16, and Pob3 (SSRP1) as the 'true replicative helicase' [39]. This study found that MCM10 and topo I associate weakly with this RPC complex, although it is unclear with which specific subunit. It is unknown if the MCM helicase itself interacts with topo I; however, considering the elaborate number of regulatory subunits now known in the eukaryotic helicase supercomplex, this may not be necessary, and may be unlikely. The GINS complex of CMG can bind to and directly stimulate the activity of pol $\alpha$-primase [21]. A later study showed that the Ctf4 subunit couples the MCMs to pol $\alpha$-primase and the Mrc1 subunit interacts with polymerase $\epsilon$ [40]. Other studies have found that both Mcm10 and Cdc45 interact with pol $\alpha$-primase and also found that loss of Mcm10 in yeast led to uncoupling of the MCMs from pol $\alpha$-primase and resulted in large stretches of ssDNA, a potent DNA damage signal [67, 107]. In human cells, Mcm10 has been suggested to interact with and regulate pol $\alpha$-primase levels and prevent inappropriate induction of DNA damage [14]. RPA interacts with many components of the RPC, including Mcm3-7, Cdc45, and Claspin (Mrc1) and requires Mcm for chromatin localization [95]. It is intriguing that only RPA appears to directly interact with the Mcms in eukaryotes; this may be due to the intimate linkage with ssDNA and the helicase machine and the highest priority of multicellular organisms to prevent the aberrant signaling of DNA damage through ssDNA coating by RPA. Additionally, in the absence of the RPC interacting protein Mcm10 or in the presence of a mutant zinc finger bearing Mcm10, RPA is also prevented from loading [55]. In general, the major components of the elongation machinery interact with the replicative helicase in eukaryotes through multiple layers of regulation as the RPC complex, a feature that is nonexistent in the simplified machinery presented by these small DNA viral systems. These viral factories simplify the entire complex by using their own central multifunctional helicases. But this simplification has led to the ability to use these viral systems as models where the biochemical nature and functions of these important interactions that occur at the interface of initiation and elongation can be studied.

## 10. Conclusion

Replicative DNA helicases, modeled by the SV40 and PV DNA replication systems, play complex roles coordinating the multiple actions of multiple DNA replication factors at eukaryotic replication forks. Their interactions with topo I are involved in origin recognition/specificity, DNA helix decompression function, and primer synthesis. Their interactions with pol $\alpha$-primase are vital for primer synthesis. Their interactions with RPA are involved in loading of RPA onto ssDNA, and de-repression of priming on RPA-coated ssDNA. And the complex interplay between all these factors is intricate, highly-regulated, and appears to be coordinated at least in large part, through the action of the replicative helicases.

Using this wealth of knowledge about the viral replication forks, we have assembled a likely model of replication elongation using the viral helicases as the central molecular machine at the fork. For ease of the various steps of elongation, only a single helicase is pictured in this model (Figure 3). Following assembly of the replication machinery at the viral origin, there is a very intricate four-way interaction comprised of the helicase, topo I, RPA and pol $\alpha$ primase. Topo I has two interactions with helicase; one within the N-terminal half of the helicase and one within the C-terminus. Through these interactions the topo I-helicase interaction assists in helicase origin recognition and creates the swivelase; a machine that couples the unwinding of the DNA duplex with relaxation of torsional stress. During elongation, topo I is likely in front of the helicase to facilitate the easing of positive supercoiling, likely through interactions with the helicase N-terminus. The helicase encircles the leading strand of the newly unwound DNA, actively pumping the leading strand template through the central channel of the helicase and away from the lagging strand replication machinery. While the leading strand template is bound to the central channel and the helicase domain, the lagging strand template is therefore left relatively unprotected. To facilitate a protective role at this point, the OBD of the helicase binds to free RPA, which swings into place as the helicase turns, actively loading RPA onto the lagging strand template. This serves in the role of nuclease protection, as well as preventing aberrant ssDNA structures. However, this coating of the lagging strand template is counterproductive to the process of priming. Therefore, at regular intervals roughly equivalent to the length an Okazaki fragment, the helicase interacts with pol $\alpha$ primase and RPA to facilitate the placement of the pol $\alpha$ primase onto the template, possibly while simultaneously removing RPA in a localized fashion, so that pol $\alpha$ primase can synthesize the RNA-DNA primer. It is intriguing to speculate that it is through this regular placement that Okazaki fragments are placed and spaced; primarily through helicase action and its protein-protein interactions with the primase. Although given the size of eukaryotic Okazaki fragments, it is likely that interactions with histones may play a role as well. The coordinated and highly regulated roles of the multi-subunit DNA helicase in modulating the proteins and their protein-protein interactions involved in the late initiation and elongation stages of DNA replication clearly play a central organizing role in the molecular machine that is the eukaryotic DNA replication fork.
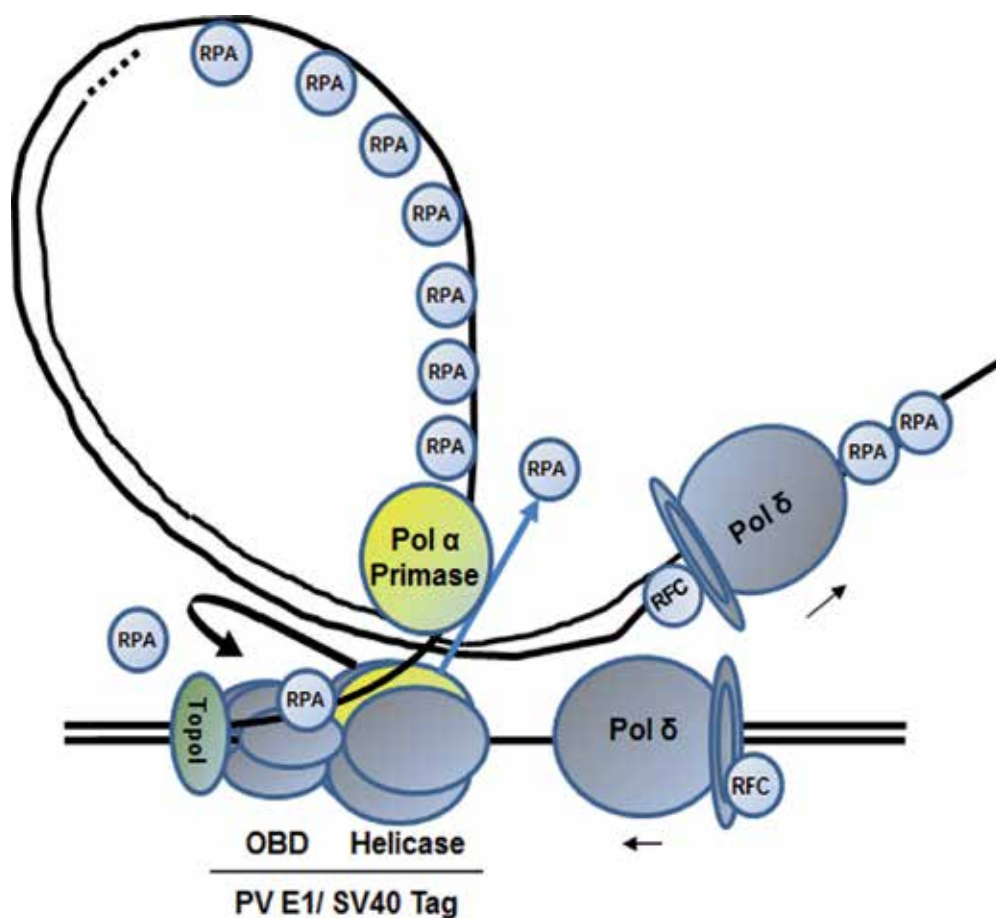
**Figure 3. Proposed Model for the PV/SV40 DNA Replication Fork.** Using the proposed helicase model presented in [33], the replicative helicase is shown oriented with the N-terminal OBD facing towards the unwound dsDNA. For simplicity, only one of the two hexamers is shown. The interaction of topo I with the OBD of the helicase both assists it in origin binding/specificity, and targets it to the incoming dsDNA, where topoisomerase action is vital for replication fork progression. The interaction of RPA within the OBD is involved in the process of directing loading RPA onto the ssDNA newly exposed by helicase action. The interaction of the helicase domain with pol α primase stimulates primer synthesis; and the interaction of the helicase with RPA allows for pol α primase to synthesize primers even in the presence of RPA, through localized RPA removal or 'priming de-repression'. As each primer is synthesized, RFC, in coordination with RPA, loads PCNA and DNA pol δ onto the 3' DNA end to allow for processive DNA synthesis. The various interactions of the helicase with topo I, RPA, and DNA pol α primase, as well as other interactions between the cellular factors themselves, coordinately the complex interplay necessary for replication fork function.

## Author details

John C. Fisk, Michaelle D. Chojnacki and Thomas Melendy

Department of Microbiology & Immunology, University at Buffalo School of Medicine & Biomedical Sciences, Buffalo, NY, USA

# References

[1] Alberts, B. 1998. The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell 92:291-294.

[2] Amin, A. A., S. Titolo, A. Pelletier, D. Fink, M. G. Cordingley, and J. Archambault. 2000. Identification of domains of the HPV11 E1 protein required for DNA replication in vitro. Virology 272:137-150.

[3] Arunkumar, A. I., V. Klimovich, X. Jiang, R. D. Ott, L. Mizoue, E. Fanning, and W. J. Chazin. 2005. Insights into hRPA32 C-terminal domain--mediated assembly of the simian virus 40 replisome. Nature Structural & Molecular Biology 12:332-339.

[4] Avemann, K., R. Knippers, T. Koller, and J. M. Sogo. 1988. Camptothecin, a specific inhibitor of type I DNA topoisomerase, induces DNA breakage at replication forks. Molecular and Cellular Biology 8:3026-3034.

[5] Bell, S. P., and A. Dutta. 2002. DNA replication in eukaryotic cells. Annual Review of Biochemistry 71:333-374.

[6] Bjergbaek, L., J. A. Cobb, M. Tsai-Pflugfelder, and S. M. Gasser. 2005. Mechanistically distinct roles for Sgs1p in checkpoint activation and replication fork maintenance. Embo J 24:405-417.

[7] Blackwell, L. J., J. A. Borowiec, and I. A. Mastrangelo. 1996. Single-stranded-DNA binding alters human replication protein A structure and facilitates interaction with DNA-dependent protein kinase. Molecular and Cellular Biology 16:4798-4807.

[8] Bochkarev, A., and E. Bochkareva. 2004. From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. Curr Opin Struct Biol 14:36-42.

[9] Bochkareva, E., D. Martynowski, A. Seitova, and A. Bochkarev. 2006. Structure of the origin-binding domain of simian virus 40 large T antigen bound to DNA. Embo J 25:5961-5969.

[10] Braun, K. A., Y. Lao, Z. He, C. J. Ingles, and M. S. Wold. 1997. Role of protein-protein interactions in the function of replication protein A (RPA): RPA modulates the activity of DNA polymerase alpha by multiple mechanisms. Biochemistry 36:8443-8454.

[11] Brill, S. J., and B. Stillman. 1989. Yeast replication factor-A functions in the unwinding of the SV40 origin of DNA replication. Nature 342:92-95.

[12] Burgers, P. M., and K. J. Gerik. 1998. Structure and processivity of two forms of Saccharomyces cerevisiae DNA polymerase delta. The Journal of Biological Chemistry 273:19756-19762.

[13] Champoux, J. J. 2001. DNA topoisomerases: structure, function, and mechanism. Annual Review of Biochemistry 70:369-413.

[14]  Chattopadhyay, S., and A. K. Bielinsky. 2007. Human Mcm10 regulates the catalytic subunit of DNA polymerase-alpha and prevents DNA damage during replication. Molecular Biology of the Cell 18:4085-4095.

[15]  Clower, R. V., J. C. Fisk, and T. Melendy. 2006. Papillomavirus E1 protein binds to and stimulates human topoisomerase I. Journal of Virology 80:1584-1587.

[16]  Collins, K. L., and T. J. Kelly. 1991. Effects of T antigen and replication protein A on the initiation of DNA synthesis by DNA polymerase alpha-primase. Molecular and Cellular Biology 11:2108-2115.

[17]  Collins, K. L., A. A. Russo, B. Y. Tseng, and T. J. Kelly. 1993. The role of the 70 kDa subunit of human DNA polymerase alpha in DNA replication. Embo J 12:4555-4566.

[18]  Conger, K. L., J. S. Liu, S. R. Kuo, L. T. Chow, and T. S. Wang. 1999. Human papillomavirus DNA replication. Interactions between the viral E1 protein and two subunits of human dna polymerase alpha/primase. The Journal of Biological Chemistry 274:2696-2705.

[19]  Corn, J. E., and J. M. Berger. 2006. Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions. Nucleic Acids Research 34:4082-4088.

[20]  Davey, M. J., D. Jeruzalmi, J. Kuriyan, and M. O'Donnell. 2002. Motors and switches: AAA+ machines within the replisome. Nat Rev Mol Cell Biol 3:826-835.

[21]  De Falco, M., E. Ferrari, M. De Felice, M. Rossi, U. Hubscher, and F. M. Pisani. 2007. The human GINS complex binds to and specifically stimulates human DNA polymerase alpha-primase. EMBO Reports 8:99-103.

[22]  de Laat, W. L., E. Appeldoorn, K. Sugasawa, E. Weterings, N. G. Jaspers, and J. H. Hoeijmakers. 1998. DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. Genes & Development 12:2598-2609.

[23]  Denis, D., and P. A. Bullock. 1993. Primer-DNA formation during simian virus 40 DNA replication in vitro. Molecular and Cellular Biology 13:2882-2890.

[24]  Diffley, J. F., and K. Labib. 2002. The chromosome replication cycle. Journal of Cell Science 115:869-872.

[25]  Dodson, G. E., Y. Shi, and R. S. Tibbetts. 2004. DNA replication defects, spontaneous DNA damage, and ATM-dependent checkpoint activation in replication protein A-deficient cells. The Journal of Biological Chemistry 279:34010-34014.

[26]  Donovan, S., J. Harwood, L. S. Drury, and J. F. Diffley. 1997. Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. Proceedings of the National Academy of Sciences of the United States of America 94:5611-5616.

[27]  Dornreiter, I., W. C. Copeland, and T. S. Wang. 1993. Initiation of simian virus 40 DNA replication requires the interaction of a specific domain of human DNA polymerase alpha with large T antigen. Molecular and Cellular Biology 13:809-820.

[28] Dornreiter, I., L. F. Erdile, I. U. Gilbert, D. von Winkler, T. J. Kelly, and E. Fanning. 1992. Interaction of DNA polymerase alpha-primase with cellular replication protein A and SV40 T antigen. Embo J 11:769-776.

[29] Dornreiter, I., A. Hoss, A. K. Arthur, and E. Fanning. 1990. SV40 T antigen binds directly to the large subunit of purified DNA polymerase alpha. Embo J 9:3329-3336.

[30] Duguet, M. 1997. When helicase and topoisomerase meet! Journal of Cell Science 110 ( Pt 12):1345-1350.

[31] Egelman, E. H., X. Yu, R. Wild, M. M. Hingorani, and S. S. Patel. 1995. Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases. Proceedings of the National Academy of Sciences of the United States of America 92:3869-3873.

[32] Enemark, E. J., and L. Joshua-Tor. 2006. Mechanism of DNA translocation in a replicative hexameric helicase. Nature 442:270-275.

[33] Enemark, E. J., and L. Joshua-Tor. 2008. On helicases and other motor proteins. Curr Opin Struct Biol 18:243-257.

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

[35] Fanning, E., V. Klimovich, and A. R. Nager. 2006. A dynamic model for replication protein A (RPA) function in DNA processing pathways. Nucleic Acids Research 34:4126-4137.

[36] Fouts, E. T., X. Yu, E. H. Egelman, and M. R. Botchan. 1999. Biochemical and electron microscopic image analysis of the hexameric E1 helicase. The Journal of Biological Chemistry 274:4447-4458.

[37] Gai, D., R. Roy, C. Wu, and D. T. Simmons. 2000. Topoisomerase I associates specifically with simian virus 40 large-T-antigen double hexamer-origin complexes. Journal of Virology 74:5224-5232.

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

[39] Gambus, A., R. C. Jones, A. Sanchez-Diaz, M. Kanemaki, F. van Deursen, R. D. Edmondson, and K. Labib. 2006. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat Cell Biol 8:358-366.

[40] Gambus, A., F. van Deursen, D. Polychronopoulos, M. Foltman, R. C. Jones, R. D. Edmondson, A. Calzada, and K. Labib. 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.

[41]  Garg, P., and P. M. Burgers. 2005. DNA polymerases that propagate the eukaryotic DNA replication fork. Crit Rev Biochem Mol Biol 40:115-128.

[42]  Gomez-Lorenzo, M. G., M. Valle, J. Frank, C. Gruss, C. O. Sorzano, X. S. Chen, L. E. Donate, and J. M. Carazo. 2003. Large T antigen on the simian virus 40 origin of replication: a 3D snapshot prior to DNA replication. Embo J 22:6205-6213.

[43]  Han, Y., Y. M. Loo, K. T. Militello, and T. Melendy. 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. Journal of Virology 73:4899-4907.

[44]  Hodgson, B., A. Calzada, and K. Labib. 2007. Mrc1 and Tof1 regulate DNA replication forks in different ways during normal S phase. Molecular Biology of the Cell 18:3894-3902.

[45]  Hu, Y., R. V. Clower, and T. Melendy. 2006. Cellular topoisomerase I modulates origin binding by bovine papillomavirus type 1 E1. Journal of Virology 80:4363-4371.

[46]  Huang, H., B. E. Weiner, H. Zhang, B. E. Fuller, Y. Gao, B. M. Wile, K. Zhao, D. R. Arnett, W. J. Chazin, and E. Fanning. 2010. Structure of a DNA polymerase alpha-primase domain that docks on the SV40 helicase and activates the viral primosome. The Journal of Biological Chemistry 285:17112-17122.

[47]  Huang, H., K. Zhao, D. R. Arnett, and E. Fanning. 2010. A specific docking site for DNA polymerase {alpha}-primase on the SV40 helicase is required for viral primosome activity, but helicase activity is dispensable. The Journal of Biological Chemistry 285:33475-33484.

[48]  Huang, S. G., K. Weisshart, I. Gilbert, and E. Fanning. 1998. Stoichiometry and mechanism of assembly of SV40 T antigen complexes with the viral origin of DNA replication and DNA polymerase alpha-primase. Biochemistry 37:15345-15352.

[49]  Hubscher, U., G. Maga, and S. Spadari. 2002. Eukaryotic DNA polymerases. Annual Review of Biochemistry 71:133-163.

[50]  Hubscher, U., H. P. Nasheuer, and J. E. Syvaoja. 2000. Eukaryotic DNA polymerases, a growing family. Trends in Biochemical Sciences 25:143-147.

[51]  Iftode, C., and J. A. Borowiec. 2000. 5' --> 3' molecular polarity of human replication protein A (hRPA) binding to pseudo-origin DNA substrates. Biochemistry 39:11970-11981.

[52]  Iftode, C., Y. Daniely, and J. A. Borowiec. 1999. Replication protein A (RPA): the eukaryotic SSB. Crit Rev Biochem Mol Biol 34:141-180.

[53]  Ishimi, Y. 1997. A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex. The Journal of Biological Chemistry 272:24508-24513.

[54] Jiang, X., V. Klimovich, A. I. Arunkumar, E. B. Hysinger, Y. Wang, R. D. Ott, G. D. Guler, B. Weiner, W. J. Chazin, and E. Fanning. 2006. Structural mechanism of RPA loading on DNA during activation of a simple pre-replication complex. Embo J 25:5516-5526.

[55] Kanke, M., Y. Kodama, T. S. Takahashi, T. Nakagawa, and H. Masukata. 2012. Mcm10 plays an essential role in origin DNA unwinding after loading of the CMG components. Embo J 31:2182-2194.

[56] Katou, Y., Y. Kanoh, M. Bando, H. Noguchi, H. Tanaka, T. Ashikari, K. Sugimoto, and K. Shirahige. 2003. S-phase checkpoint proteins Tof1 and Mrc1 form a stable rep-lication-pausing complex. Nature 424:1078-1083.

[57] Kelly, T. J., P. Simancek, and G. S. Brush. 1998. Identification and characterization of a single-stranded DNA-binding protein from the archaeon Methanococcus janna-schii. Proceedings of the National Academy of Sciences of the United States of Amer-ica 95:14634-14639.

[58] Kenny, M. K., S. H. Lee, and J. Hurwitz. 1989. Multiple functions of human single-stranded-DNA binding protein in simian virus 40 DNA replication: single-strand sta-bilization and stimulation of DNA polymerases alpha and delta. Proceedings of the National Academy of Sciences of the United States of America 86:9757-9761.

[59] Kerr, I. D., R. I. Wadsworth, L. Cubeddu, W. Blankenfeldt, J. H. Naismith, and M. F. White. 2003. Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of Sulfolobus SSB protein. Embo J 22:2561-2570.

[60] Khopde, S., R. Roy, and D. T. Simmons. 2008. The binding of topoisomerase I to T antigen enhances the synthesis of RNA-DNA primers during simian virus 40 DNA replication. Biochemistry 47:9653-9660.

[61] Khopde, S., and D. T. Simmons. 2008. Simian virus 40 DNA replication is dependent on an interaction between topoisomerase I and the C-terminal end of T antigen. Jour-nal of Virology 82:1136-1145.

[62] Kim, C., B. F. Paulus, and M. S. Wold. 1994. Interactions of human replication protein A with oligonucleotides. Biochemistry 33:14197-14206.

[63] Kim, S., H. G. Dallmann, C. S. McHenry, and K. J. Marians. 1996. Coupling of a repli-cative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement. Cell 84:643-650.

[64] Kowalczykowski, S. C. 2000. Initiation of genetic recombination and recombination-dependent replication. Trends in Biochemical Sciences 25:156-165.

[65] Kunkel, T. A., and P. M. Burgers. 2008. Dividing the workload at a eukaryotic repli-cation fork. Trends Cell Biol 18:521-527.

[66] Kuo, S. R., J. S. Liu, T. R. Broker, and L. T. Chow. 1994. Cell-free replication of the human papillomavirus DNA with homologous viral E1 and E2 proteins and human cell extracts. The Journal of Biological Chemistry 269:24058-24065.

[67] Lee, C., I. Liachko, R. Bouten, Z. Kelman, and B. K. Tye. 2010. Alternative mechanisms for coordinating polymerase alpha and MCM helicase. Molecular and Cellular Biology 30:423-435.

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

[69] Lee, S. H., and D. K. Kim. 1995. The role of the 34-kDa subunit of human replication protein A in simian virus 40 DNA replication in vitro. The Journal of Biological Chemistry 270:12801-12807.

[70] Lei, M., and B. K. Tye. 2001. Initiating DNA synthesis: from recruiting to activating the MCM complex. Journal of Cell Science 114:1447-1454.

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

[72] Lin, B. Y., A. M. Makhov, J. D. Griffith, T. R. Broker, and L. T. Chow. 2002. Chaperone proteins abrogate inhibition of the human papillomavirus (HPV) E1 replicative helicase by the HPV E2 protein. Molecular and Cellular Biology 22:6592-6604.

[73] Liu, J. S., S. R. Kuo, A. M. Makhov, D. M. Cyr, J. D. Griffith, T. R. Broker, and L. T. Chow. 1998. Human Hsp70 and Hsp40 chaperone proteins facilitate human papillomavirus-11 E1 protein binding to the origin and stimulate cell-free DNA replication. The Journal of Biological Chemistry 273:30704-30712.

[74] Liu, Y., H. I. Kao, and R. A. Bambara. 2004. Flap endonuclease 1: a central component of DNA metabolism. Annual Review of Biochemistry 73:589-615.

[75] Lohman, T. M. 1992. Escherichia coli DNA helicases: mechanisms of DNA unwinding. Molecular Microbiology 6:5-14.

[76] Loo, Y. M., and T. Melendy. 2000. The majority of human replication protein A remains complexed throughout the cell cycle. Nucleic Acids Research 28:3354-3360.

[77] Loo, Y. M., and T. Melendy. 2004. Recruitment of replication protein A by the papillomavirus E1 protein and modulation by single-stranded DNA. Journal of Virology 78:1605-1615.

[78] Lou, H., M. Komata, Y. Katou, Z. Guan, C. C. Reis, M. Budd, K. Shirahige, and J. L. Campbell. 2008. Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint. Molecular Cell 32:106-117.

[79] Lusky, M., J. Hurwitz, and Y. S. Seo. 1993. Cooperative assembly of the bovine papilloma virus E1 and E2 proteins on the replication origin requires an intact E2 binding site. The Journal of Biological Chemistry 268:15795-15803.

[80] Ma, T., N. Zou, B. Y. Lin, L. T. Chow, and J. W. Harper. 1999. Interaction between cyclin-dependent kinases and human papillomavirus replication-initiation protein E1 is required for efficient viral replication. Proceedings of the National Academy of Sciences of the United States of America 96:382-387.

[81] Maga, G., I. Frouin, S. Spadari, and U. Hubscher. 2001. Replication protein A as a "fidelity clamp" for DNA polymerase alpha. The Journal of Biological Chemistry 276:18235-18242.

[82] Makiniemi, M., T. Hillukkala, J. Tuusa, K. Reini, M. Vaara, D. Huang, H. Pospiech, I. Majuri, T. Westerling, T. P. Makela, and J. E. Syvaoja. 2001. BRCT domain-containing protein TopBP1 functions in DNA replication and damage response. The Journal of Biological Chemistry 276:30399-30406.

[83] Masterson, P. J., M. A. Stanley, A. P. Lewis, and M. A. Romanos. 1998. A C-terminal helicase domain of the human papillomavirus E1 protein binds E2 and the DNA polymerase alpha-primase p68 subunit. Journal of Virology 72:7407-7419.

[84] Masumoto, H., A. Sugino, and H. Araki. 2000. Dpb11 controls the association between DNA polymerases alpha and epsilon and the autonomously replicating sequence region of budding yeast. Molecular and Cellular Biology 20:2809-2817.

[85] Matson, S. W., and K. A. Kaiser-Rogers. 1990. DNA helicases. Annual Review of Biochemistry 59:289-329.

[86] Matsumoto, T., T. Eki, and J. Hurwitz. 1990. Studies on the initiation and elongation reactions in the simian virus 40 DNA replication system. Proceedings of the National Academy of Sciences of the United States of America 87:9712-9716.

[87] Melendy, T., J. Sedman, and A. Stenlund. 1995. Cellular factors required for papillomavirus DNA replication. Journal of Virology 69:7857-7867.

[88] Melendy, T., and B. Stillman. 1993. An interaction between replication protein A and SV40 T antigen appears essential for primosome assembly during SV40 DNA replication. The Journal of Biological Chemistry 268:3389-3395.

[89] Mer, G., A. Bochkarev, W. J. Chazin, and A. M. Edwards. 2000. Three-dimensional structure and function of replication protein A. Cold Spring Harbor Symposia on Quantitative Biology 65:193-200.

[90] Mohr, I. J., R. Clark, S. Sun, E. J. Androphy, P. MacPherson, and M. R. Botchan. 1990. Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. Science 250:1694-1699.

[91] Moyer, S. E., P. W. Lewis, and M. R. Botchan. 2006. Isolation of the Cdc45/Mcm2-7/ GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase.

Proceedings of the National Academy of Sciences of the United States of America 103:10236-10241.

[92] Mueser, T. C., J. M. Hinerman, J. M. Devos, R. A. Boyer, and K. J. Williams. 2010. Structural analysis of bacteriophage T4 DNA replication: a review in the Virology Journal series on bacteriophage T4 and its relatives. Virol J 7:359.

[93] Murakami, Y., and J. Hurwitz. 1993. DNA polymerase alpha stimulates the ATP-dependent binding of simian virus tumor T antigen to the SV40 origin of replication. The Journal of Biological Chemistry 268:11018-11027.

[94] Murakami, Y., C. R. Wobbe, L. Weissbach, F. B. Dean, and J. Hurwitz. 1986. Role of DNA polymerase alpha and DNA primase in simian virus 40 DNA replication in vitro. Proceedings of the National Academy of Sciences of the United States of America 83:2869-2873.

[95] Nakaya, R., J. Takaya, T. Onuki, M. Moritani, N. Nozaki, and Y. Ishimi. 2010. Identification of proteins that may directly interact with human RPA. Journal of Biochemistry 148:539-547.

[96] Nasheuer, H. P., D. von Winkler, C. Schneider, I. Dornreiter, I. Gilbert, and E. Fanning. 1992. Purification and functional characterization of bovine RP-A in an in vitro SV40 DNA replication system. Chromosoma 102:S52-59.

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

[98] Nick McElhinny, S. A., D. A. Gordenin, C. M. Stith, P. M. Burgers, and T. A. Kunkel. 2008. Division of labor at the eukaryotic replication fork. Molecular Cell 30:137-144.

[99] Notarnicola, S. M., K. Park, J. D. Griffith, and C. C. Richardson. 1995. A domain of the gene 4 helicase/primase of bacteriophage T7 required for the formation of an active hexamer. The Journal of Biological Chemistry 270:20215-20224.

[100] Ott, R. D., C. Rehfuess, V. N. Podust, J. E. Clark, and E. Fanning. 2002. Role of the p68 subunit of human DNA polymerase alpha-primase in simian virus 40 DNA replication. Molecular and Cellular Biology 22:5669-5678.

[101] Park, C. J., J. H. Lee, and B. S. Choi. 2005. Solution structure of the DNA-binding domain of RPA from Saccharomyces cerevisiae and its interaction with single-stranded DNA and SV40 T antigen. Nucleic Acids Research 33:4172-4181.

[102] Patel, S. S., and M. M. Hingorani. 1993. Oligomeric structure of bacteriophage T7 DNA primase/helicase proteins. The Journal of Biological Chemistry 268:10668-10675.

[103] Pavlov, Y. I., and P. V. Shcherbakova. 2010. DNA polymerases at the eukaryotic fork-20 years later. Mutation Research 685:45-53.

[104]  Piccolino, M. 2000. Biological machines: from mills to molecules. Nat Rev Mol Cell Biol 1:149-153.

[105]  Pike, J. E., R. A. Henry, P. M. Burgers, J. L. Campbell, and R. A. Bambara. 2010. An alternative pathway for Okazaki fragment processing: resolution of fold-back flaps by Pif1 helicase. The Journal of Biological Chemistry 285:41712-41723.

[106]  Pursell, Z. F., I. Isoz, E. B. Lundstrom, E. Johansson, and T. A. Kunkel. 2007. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science 317:127-130.

[107]  Ricke, R. M., and A. K. Bielinsky. 2004. Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. Molecular Cell 16:173-185.

[108]  Rossi, M. L., V. Purohit, P. D. Brandt, and R. A. Bambara. 2006. Lagging strand replication proteins in genome stability and DNA repair. Chem Rev 106:453-473.

[109]  Sanders, C. M., O. V. Kovalevskiy, D. Sizov, A. A. Lebedev, M. N. Isupov, and A. A. Antson. 2007. Papillomavirus E1 helicase assembly maintains an asymmetric state in the absence of DNA and nucleotide cofactors. Nucleic Acids Research 35:6451-6457.

[110]  Sanders, C. M., and A. Stenlund. 1998. Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalysed by a transcription factor. Embo J 17:7044-7055.

[111]  Schneider, C., K. Weisshart, L. A. Guarino, I. Dornreiter, and E. Fanning. 1994. Species-specific functional interactions of DNA polymerase alpha-primase with simian virus 40 (SV40) T antigen require SV40 origin DNA. Molecular and Cellular Biology 14:3176-3185.

[112]  Sclafani, R. A., and T. M. Holzen. 2007. Cell cycle regulation of DNA replication. Annu Rev Genet 41:237-280.

[113]  Sedman, J., and A. Stenlund. 1998. The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities. J Virol 72:6893-6897.

[114]  Simmons, D. T., T. Melendy, D. Usher, and B. Stillman. 1996. Simian virus 40 large T antigen binds to topoisomerase I. Virology 222:365-374.

[115]  Speck, C., Z. Chen, H. Li, and B. Stillman. 2005. ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. Nature Structural & Molecular Biology 12:965-971.

[116]  Stewart, L., G. C. Ireton, and J. J. Champoux. 1996. The domain organization of human topoisomerase I. The Journal of Biological Chemistry 271:7602-7608.

[117]  Stillman, B. 2005. Origin recognition and the chromosome cycle. FEBS Letters 579:877-884.

[118]   Szyjka, S. J., C. J. Viggiani, and O. M. Aparicio. 2005. Mrc1 is required for normal progression of replication forks throughout chromatin in S. cerevisiae. Molecular Cell 19:691-697.

[119]   Taljanidisz, J., R. S. Decker, Z. S. Guo, M. L. DePamphilis, and N. Sarkar. 1987. Initiation of simian virus 40 DNA replication in vitro: identification of RNA-primed nascent DNA chains. Nucleic Acids Research 15:7877-7888.

[120]   Tanaka, H., Y. Katou, M. Yagura, K. Saitoh, T. Itoh, H. Araki, M. Bando, and K. Shirahige. 2009. Ctf4 coordinates the progression of helicase and DNA polymerase alpha. Genes Cells 14:807-820.

[121]   Theobald, D. L., R. M. Mitton-Fry, and D. S. Wuttke. 2003. Nucleic acid recognition by OB-fold proteins. Annu Rev Biophys Biomol Struct 32:115-133.

[122]   Tourriere, H., G. Versini, V. Cordon-Preciado, C. Alabert, and P. Pasero. 2005. Mrc1 and Tof1 promote replication fork progression and recovery independently of Rad53. Molecular Cell 19:699-706.

[123]   Trowbridge, P. W., R. Roy, and D. T. Simmons. 1999. Human topoisomerase I promotes initiation of simian virus 40 DNA replication in vitro. Mol Cell Biol 19:1686-1694.

[124]   Tsurimoto, T., T. Melendy, and B. Stillman. 1990. Sequential initiation of lagging and leading strand synthesis by two different polymerase complexes at the SV40 DNA replication origin. Nature 346:534-539.

[125]   Tye, B. K. 1999. MCM proteins in DNA replication. Annual Review of Biochemistry 68:649-686.

[126]   Ustav, M., and A. Stenlund. 1991. Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames. Embo J 10:449-457.

[127]   Vos, S. M., E. M. Tretter, B. H. Schmidt, and J. M. Berger. 2011. All tangled up: how cells direct, manage and exploit topoisomerase function. Nat Rev Mol Cell Biol 12:827-841.

[128]   Waga, S., and B. Stillman. 1994. Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication in vitro. Nature 369:207-212.

[129]   Weisshart, K., H. Forster, E. Kremmer, B. Schlott, F. Grosse, and H. P. Nasheuer. 2000. Protein-protein interactions of the primase subunits p58 and p48 with simian virus 40 T antigen are required for efficient primer synthesis in a cell-free system. The Journal of Biological Chemistry 275:17328-17337.

[130]   Weisshart, K., P. Taneja, and E. Fanning. 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. Journal of Virology 72:9771-9781.

[131]   Wilson, V. G., M. West, K. Woytek, and D. Rangasamy. 2002. Papillomavirus E1 proteins: form, function, and features. Virus Genes 24:275-290.

[132] Wold, M. S. 1997. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annual Review of Biochemistry 66:61-92.

[133] Wun-Kim, K., R. Upson, W. Young, T. Melendy, B. Stillman, and D. T. Simmons. 1993. The DNA-binding domain of simian virus 40 tumor antigen has multiple functions. Journal of Virology 67:7608-7611.

[134] Yang, L., R. Li, I. J. Mohr, R. Clark, and M. R. Botchan. 1991. Activation of BPV-1 replication in vitro by the transcription factor E2. Nature 353:628-632.

[135] Yang, L., M. S. Wold, J. J. Li, T. J. Kelly, and L. F. Liu. 1987. Roles of DNA topoisomerases in simian virus 40 DNA replication in vitro. Proceedings of the National Academy of Sciences of the United States of America 84:950-954.

[136] Yang, L., M. S. Wold, J. J. Li, T. J. Kelly, and L. F. Liu. 1987. Roles of DNA topoisomerases in simian virus 40 DNA replication in vitro. Proceedings of the National Academy of Sciences of the United States of America 84:950-954.

[137] You, Z., Y. Komamura, and Y. Ishimi. 1999. Biochemical analysis of the intrinsic Mcm4-Mcm6-mcm7 DNA helicase activity. Molecular and Cellular Biology 19:8003-8015.

[138] Yuzhakov, A., Z. Kelman, J. Hurwitz, and M. O'Donnell. 1999. Multiple competition reactions for RPA order the assembly of the DNA polymerase delta holoenzyme. Embo J 18:6189-6199.

[139] Zhou, B., D. R. Arnett, X. Yu, A. Brewster, G. A. Sowd, C. L. Xie, S. Vila, D. Gai, E. Fanning, and X. S. Chen. 2012. Structural basis for the interaction of a hexameric replicative helicase with the regulatory subunit of human DNA polymerase alpha-primase. The Journal of Biological Chemistry.

[140] Zhu, W., C. Ukomadu, S. Jha, T. Senga, S. K. Dhar, J. A. Wohlschlegel, L. K. Nutt, S. Kornbluth, and A. Dutta. 2007. Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication. Genes & Development 21:2288-2299.

[141] 141.Zlotkin, T., G. Kaufmann, Y. Jiang, M. Y. Lee, L. Uitto, J. Syvaoja, I. Dornreiter, E. Fanning, and T. Nethanel. 1996. DNA polymerase epsilon may be dispensable for SV40- but not cellular-DNA replication. Embo J 15:2298-2305.

# The MCM and RecQ Helicase Families: Ancient Roles in DNA Replication and Genomic Stability Lead to Distinct Roles in Human Disease

Dianne C. Daniel*, Ayuna V. Dagdanova and
Edward M. Johnson

Additional information is available at the end of the chapter

## 1. Introduction

### 1.1. Rationale for comparison of MCM and RecQ helicase families

DNA helicases are currently organized into superfamilies based on their sequence structures and 3-D conformations. Within each superfamily, there are members that have further evolved for specialized functions [1]. There is conservation of RecQ proteins from bacteria to humans. Whereas bacteria have one RecQ helicase, humans have evolved at least five different proteins [2]. The RecQ members belong to the helicase Superfamily II, and as such have the characteristic Rec fold [1]. In this chapter, we will focus on RecQ family members WRN, BLM and RECQL4 (RecQ protein-like 4), which is also referred to in the literature as RECQ4. Eukaryotic and archaeal MCMs belong to the helicase Superfamily VI, and have the AAA+ (ATPases associated with diverse cellular activities) fold [1, 3, 4, 5]. Both Rec and AAA+ folds are based on the ancestral ASCE (additional strand conserved E) fold or an alpha-beta-alpha domain necessary for nucleoside triphosphate binding and catalysis [1, 6, 7]. A rationale for comparison of the RecQ and MCM family members relates to the importance of their activities for genomic integrity. The WRN and BLM proteins as well as other members of the RecQ family are characterized by this feature [8]. Both WRN and BLM are involved in DNA repair and a role for WRN in telomere homeostasis in humans is well established [2, 9]. MCM2-7 proteins, along with cofactors, are thought to function as the eukaryotic replicative helicase [10]. MCM8 [11, 12] and MCM9 [13, 14] are more recently discovered and their roles are less well defined. Although data point to a role for MCM8 in DNA replication, that role may be specialized in higher organisms. In human cells, MCM10 is recruited to chromoso-

mal domains before they replicate and studies in yeast suggest a role in DNA replication, but not as a helicase [15-18]. Members of each family are essential for chromosome homeostasis. When replication forks stall, there may be involvement of members of each family. These proteins have interlocking functions since, for example, a stalled replication fork with attendant MCM proteins can lead to a DNA double-strand break (DSB), which requires RecQ proteins for repair [19].

## 2. RecQ and MCM family structures

Breaks in double-strand DNA can occur during DNA replication, at specific loci (e.g., at telomeres) and during meiosis. RecQ helicases are implicated in DNA repair based on their involvement in such processes as DNA end resection, branch migration, D-loop processing, Holliday Junction (HJ) and double Holliday junction (dHJ) resolution [2]. RecQ proteins function at multiple steps, both early and late, during repair of DSB [19]. RecQ helicases travel in a 3′ to 5′ direction on ssDNA [20]. The RecQ proteins have an ancient lineage based on an ancestral ASCE fold ($\alpha\beta\alpha$ domain) of distant relation to P-loop NTPase folds. RecQ structural domains include a conserved core helicase domain for binding and hydrolysis of nucleoside triphosphate that is equivalent to the Walker A and Walker B boxes seen in MCM proteins [21]. They have a helicase and RNAase D C-terminal (HRDC) domain thought to mediate structure-specific nucleic acid binding, double HJ dissolution and protein-protein interactions. They also have a RecQ C-terminal (RQC) domain thought to mediate interactions with other proteins, structure-specific nucleic acid binding and metal cofactor binding [22]. Acidic regions present in many RecQ proteins aid in protein-protein interactions. There are also nuclear localization signals in some RecQ proteins (e.g., *H. sapiens* WRN and BLM). There are two RecQ members with an exonuclease domain, one of which is *H. sapiens* WRN. Two members have been functionally characterized as having an N-terminal strand exchange domain, one of which is *H. sapiens* BLM [2].

RECQL4 has been reported to have ssDNA binding and DNA strand-annealing activities. In this single study, RECQL4 did not display substrate unwinding or resolution of substrates resembling replication or recombination intermediates [23]. Recognizable HRDC and RQC domains that are important in BLM activity are missing in RECQL4 [22, 24]. As observed, the ssDNA annealing activity would allow RECQL4 to function during synthesis-dependent strand annealing (SDSA) along with another helicase. RECQL4 could help direct pathway choice during HR in DSB repair through aiding ssDNA annealing activity in non-homologous end joining (NHEJ) [19]. Thus, RECQL4 is similar to other RecQ helicases in its core helicase domain, but its function as a helicase is unclear [23, 25, 26].

AAA+ proteins, including MCMs, have a core molecular motor. Like the RecQ proteins, the AAA+ fold is also based on the ancestral ASCE fold. Acquisition of a catalytic glutamate (Fig. 1) to initiate efficient hydrolysis of ATP marked the emergence of the ASCE division from the ancestral P-loop fold [7]. For further discussion of the glutamate "switch" in AAA+ proteins, see reference [27]. Mechanisms of action are diverse, although members are typi-

cally oligomeric ring assemblies with inter-subunit communication. The central AAA+ motor has been adapted in evolution through structural changes to the core module and through domains added either N- or C-terminal to the AAA+ core. Activities are facilitated by recognition of protein partners functioning in these diverse events. Thus, AAA+ proteins display a variety of macromolecular remodelling events that are energy-driven by nucleotide hydrolysis thought to be occurring throughout what is typically a hexameric complex assembly [28, 29]. The conserved Walker A and Walker B motifs within the central module mediate ATP-binding and hydrolysis [7, 30-32]. MCM proteins have two active site motifs, the P-loop domain and the lid. Motifs in the P-loop include Walker A, Walker B and Sensor 1. The lid domain contains the arginine finger and Sensor 2. A catalytic site is created by a dimer interface that employs a *cis* P-loop from one subunit and a *trans* lid from the adjacent subunit [4, 33, 34]. A similar catalytic site created at the interfaces between adjacent monomers is also characteristic of the RecQ ATPase core [21].
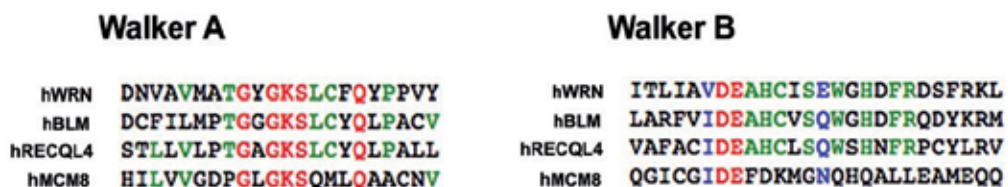


**Figure 1.** Comparison of conserved ATPase motifs in RecQ proteins WRN, BLM and RECQL4 to the Walker A and Walker B boxes of MCM8. MCM8 was chosen as the MCM for comparison because it has a canonical GKS Walker A [12] and the signature MCM IDEFDKM Walker B ATP-binding domains. In the Walker B motif, note the conserved structural features and the conserved DE motif (containing aspartate, D, and the catalytic glutamate, E).

# 3. RecQ and MCM helicases: association with disease and aging

### 3.1. BLM and Bloom syndrome, WRN and Werner syndrome, RECQL4 and Rothmund-Thomson syndrome

As a group, mutations in the RecQ helicases lead to adult segmental progeria with abnormalities in development, predisposition to cancer and acceleration of aging processes. Three of the RecQ family members are associated with rare autosomal recessive diseases [19]. These disorders Werner syndrome (WS) [35], Bloom syndrome (BS) [36] and Rothmund-Thomson syndrome (RTS) [37, 38] are caused by mutations in the genes coding for WRN, BLM and RECQL4, respectively. RTS is a heterogeneous disorder with about 60% of cases resulting from mutations in the *RECQL4* gene [37]. Mutations in *RECQL4* can also lead to two other disease phenotypes [39], but only RTS will be discussed here. The RecQ deficiency diseases are associated with cancer predisposition and several characteristics of aging [8, 20, 26, 40]. In BS cells, there is a 10-fold elevation in frequency of homologous recombination (HR), and reciprocal exchanges occur between homologous chromosomes and sister chromatids [41, 42]. WS cells, on the other hand, display large

chromosome deletions and an increase in illegitimate recombination [43]. A higher frequency of chromosomal aberrations is reported for cells from RTS [44]. These deficiencies thus provide hints as to the cellular activities of these three helicases. Clinical features of these diseases, as referenced above, are as follows.

BS manifests in pleiotropic phenotypes such as growth retardation leading to proportional dwarfism, erythema with light sensitivity, skin lesions with hypo- and hyperpigmentation, immunodeficiency, susceptibility to type II diabetes, male infertility, female sub-fertility, reports of mental retardation, cancer predisposition (all types but at an earlier age of diagnosis than in the normal population).

WS leads to short stature and early onset age-related diseases, including greying hair, alopecia, bilateral cataracts, osteoporosis, arteriosclerosis, atherosclerosis, skin atrophy, hypogonadism, type II diabetes mellitus and susceptibility to tumors, especially those of mesenchymal origin (sarcomas).

RTS manifests as early growth deficiency, congenital bone defects, poikiloderma, cataracts, greying hair, alopecia, hypogonadism, and some increased susceptibility to cancer, especially osteogenic sarcomas.

### 3.2. MCMs and genomic stability

The MCM proteins 2-7 are necessary for DNA replication in yeast [45, 46], and this basic life function extends in evolution to a single MCM protein in archaea [47-49]. MCM2-7 are also essential for replication in *Xenopus* [50, 51] and have been proposed as a licensing factor for initiation of eukaryotic replication [52, 53]. The MCM proteins 2-8 have an identical Walker B-box motif of IDEFDKM. The MCM2-7 complex is enigmatic in that MCMs 4, 6 and 7 function alone as a heterohexameric helicase [54, 55]. For a discussion of individual MCM subunit arrangements and activities, see the references [33, 34]. MCM2-7 have now been shown to have helicase activity *in vitro* [56], and to be components of a holo-helicase Cdc45/MCM2-7/GINS (CMG) complex [57, 58]. The MCMs require a clamp-loading factor to assemble as a multimeric ring on DNA, and this function is fulfilled in known cases by the protein Cdc6 [59-64] although the regulation of this step in the formation of the CMG complex proceeds through multiple pathways [57, 58]. Various papers have dealt with the function of MCM proteins in DNA replication [10, 46, 65-68], and regarding their processive mechanism of DNA unwinding [27, 56, 58, 69, 70] and only certain lingering, disease-related questions will be considered here. A summary statement can be made regarding known relationships between MCM and RecQ helicases. Members of the MCM protein family are essential for the life-creating process of DNA replication, whereas members of the RecQ family are essential for the life-prolonging maintenance of the genome.

Due to their essential roles, it is not surprising that there are few diseases directly ascribed to defects in MCMs 2-7. This does, however, bring up an unresolved MCM enigma: there are more MCM proteins than are required to form initiation complexes at cellular origins active within a given round of DNA replication [71]. In addition, MCM proteins in human cells remain at peak levels in G2 phase of the cell cycle, after DNA replication is complete [59, 72, 73]

This high copy number could be an aspect of securing sufficient protein quantities for basic function. A role for MCM proteins in transcription has also been proposed, and this remains to be defined [74]; for further discussion see reference [46]. Redundancy of functions among MCM members is also a factor to be considered. Intriguingly, knockdown of MCM8 [75], not even present in yeast [12], retards S-phase approximately 25% in human cells in culture [75], suggesting a specialized function in higher eukaryotes critical for basic replication. It has been suggested that the excess MCMs license dormant origins, which are used under conditions of stress upon normally functional origins [71]. An overview provides a discussion of recent work connecting dormant origins of replication and tumor suppression based on the role of dormant origins during fork restart after repair of DNA damage [76].

Mutations in MCM family members have been studied in yeast and mouse models, and they confirm an essential role for MCM2-7 in DNA replication [34, 45, 77, 78]. A human MCM4 mutation (destabilizing the MCM2-7 complex) was recently reported concurrently by two groups who studied consanguineous families, and the resulting phenotype was found to be associated with immune deficiency (NK cells), adrenal insufficiency and short stature [79, 80]. Patient fibroblasts showed chromosome fragility [79]. The susceptibility of these patients to cancer is not currently known [80]. An *MCM8* disruption and alternative splice form have been noted in hepatic carcinoma [11] and choriocarcinoma [12], respectively. Although not necessarily a cause of disease, the MCM proteins may be useful tools in diagnoses [81-85]. Elevated levels of MCM proteins 2-7 have been observed in several cancers [81, 84, 85]. In contrast, reduced levels of MCM8 mRNA have been reported in colon carcinoma [12]. Nuclear MCM7 is a good marker for proliferating cells [86], and MCMs 2, 5 or 7 may be an alternative to the Ki-67 marker to distinguish certain hyperproliferative disorders [83].

### 3.2.1. A structural domain deleted from MCM8 and present in WRN and BLM

A brief discussion of MCM8 is included in this section because it contains a motif that may be structurally similar to one found in BLM and WRN and is linked to neoplasia. Human MCM8 has a splice variant that results in a 16 amino acid (aa) deletion in a location between the Zn finger of MCM8 and its Walker A box [12], Fig. 2. Thus far this deletion has been detected by various different groups only in cases of choriocarcinoma. MCM8 with this same sequence deleted (Fig. 2) has, however, been detected in several higher eukaryotes other than humans, suggesting that the variant does have, or perhaps lacks, a function. That function is as yet unknown, but clues to it may be gleaned from a comparison of the MCM8 deletion with sequences from WRN and BLM. These RecQ proteins have a counterpart 16 aa domain with partial sequence homology and notable structural homology, as denoted in Fig. 2. This sequence in the RecQ proteins is located in a different orientation to that of MCM8 with regard to the helicase Walker A and B boxes (Fig. 2). In each case the first 8 aa of this sequence are highly charged and, in WRN, BLM and MCM8, contain a polar S. The 9th aa in this deletion is a conserved C. The remaining 7 aa contain a preponderance of aromatic and hydrophobic residues. This configuration of charged and aromatic residues is characteristic of known single-stranded nucleic acid binding proteins [87]. Among MCMs, this 16 aa domain is identifiable in the single MCM of *Sulfolobus solfataricus*, in which it has been implicated as a single-stranded DNA-

binding "finger" [88]. Mutations of the positively charged amino acids strongly reduce single-stranded DNA binding of this MCM. In contrast, in BLM the polar S residue is thought to be involved in ATP binding [89]. Because of the aromatic nature of a portion of this domain, binding to one or more DNA nucleotide bases may be involved as a common link between functions of these 16 aa in RECQ and MCM helicases.



```
hWRN       ESYYQEIGRAGRDGLQSSCHVLWAPA
hBLM       EGYYQESGRAGRDGEISHCLLFYTYH
hRECQL4    ESYVQAVGRAGRDGQPAHCHLFLQPQ
hMCM8 del. ----------GSRNKNDKCMFLLYIE
```

**hWRN  (RecA fold)**

**hMCM8  (AAA+ fold)**
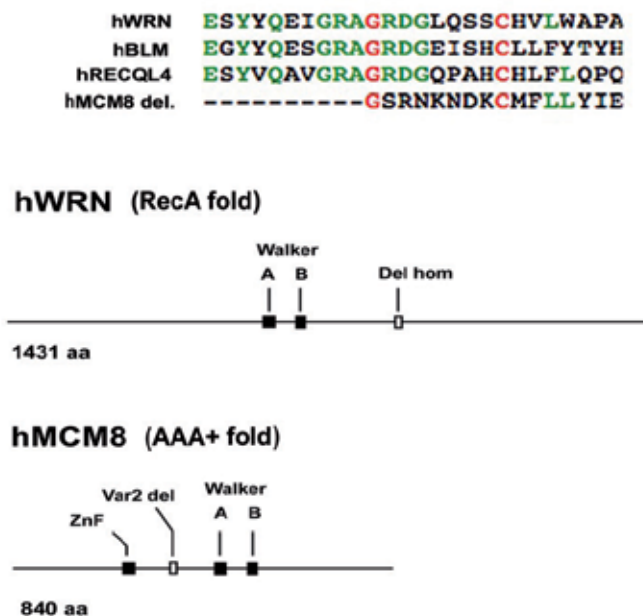
**Figure 2.** Comparison of placement of selected structural motifs in human RecQ and MCM8 helicases. MCM8 has a splice variant that results in a 16 aa deletion in choriocarcinoma [12]. The Walker A and B boxes and the MCM Zn finger (ZnF) domains are indicated by filled rectangles. The MCM8 variant 2 deletion (Var2 del) is indicated by an open rectangle, as is its partially-homologous counterpart (del hom) in WRN and BLM. This conservatively structured domain consists of an N-terminal highly charged sequence followed by a conserved C and an aromatic-hydrophobic sequence. The open box is located in a different orientation relative to Walker A and B boxes in WRN vs. MCM8. The positions of black and white boxes are approximately to scale.

## 4. Supportive roles for WRN, BLM helicases and RECQL4 during replication elongation

During normal metabolism, such as in mitochondrial respiration, endogenous reactive oxygen species (ROS) are produced that lead to oxidative DNA modifications. In addition to endogenous mutagens, there are also environmental mutagens that damage the DNA [90]. Furthermore, during replication and transcription, duplex DNA is transiently opened, and there is an opportunity for non-B DNA structures to form in the genomic DNA [40]. RecQ helicases act on recombination intermediates, on preferred substrates including those resembling G-quadruplex DNA [91, 92], D-loops [93], HJ [94] and double HJ [95]. RecQ helicases

may play a regulatory role in both pro- and anti-HR events. They function at the interface of HR with the stressed replication fork and may affect repair pathway choice. There is little evidence for a specific clear cut role, but for a discussion of proposed mechanistic models of RecQ protein function, see the references [19, 20].

## 4.1. BLM and RECQL4

In BS cells, there are S-phase defects in DNA replication involving abnormal replication inter‐mediates, and replication elongation is slower [96-98]. These cells don't recover well from in‐duced fork stalling and accumulate DSB [99, 100]. In human cells, RECQL4 interacts with the MCM2-7 complex, Cdc45 and GINS. The interaction is facilitated by MCM10 [25]. BLM and RECQL4 are at their highest levels during S phase. At stalled replication forks, BLM physical‐ly associates with Rad51 and p53, and BLM and p53 function synergistically in HR [101]. BLM colocalizes in foci with PCNA and with the BASC (BRCA1-associated genome surveillance complex) [102]. BLM is phosphorylated by ATR [103]. When replication forks are stalled by use of hydroxyurea (HU), BLM colocalizes with Chk1 and p53BP1 foci. Chk1 is required for BLM and 53BP1 foci formation. Thus Chk1 may recruit BLM to stalled forks [104]. This impli‐cates RecQ proteins as DNA damage checkpoint mediators in response to stalled forks. *In vi‐tro* studies with use of substrates similar to Okazaki fragments showed BLM stimulation of flap endonuclease [105] (a protein that functions in lagging strand synthesis [106]). BLM func‐tions to prevent the association of homologous sequences in the displaced flap DNA of the Okazaki fragment and the sister chromatid [107-110]. D-loops are formed when a ssDNA tail invades a homologous duplex, and BLM has a preference for dissociating D-loops with a 5′ in‐vaded end suggesting a selection of recombination intermediates that are not extended by polymerase [93, 111, 112]. BLM is able to disrupt the initial Rad51 filament formation step to destabilize recombinase-nucleoprotein filaments. D-loops are susceptible to BLM activity when Rad51 is in an inactive form (ADP-bound) [112, 113]. BLM physically associates with CAF-1 (chromatin assembly factor I) largest subunit, and the colocalization of these two pro‐teins occurs at sites of DNA synthesis. BLM inhibits CAF-1 function in chromatin assembly during DNA repair *in vitro*, and inhibits its mobilization after damage induction *in vivo* [114]. Mammalian WRN and BLM interact [115], and they both interact with RPA [116-118] and p53 [119-121]. Based on coimmunoprecipitation there is limited BLM and WRN interaction, but they may function in the same pathway during HR [20]. BLM helicase activity is stimulated by its binding to the RPA70 kDa subunit [116]. In mouse spermatocytes during meiotic pro‐phase, BLM and RPA are nuclear colocalized [122]. This suggests a potential role for these proteins in resolution of recombination intermediates during meiosis [8]. BLM has a prefer‐ence for unwinding G-quadruplex structures versus HJ [92]. Both mammalian WRN and BLM bind to G-quadruplex structures, which are roadblocks to polymerases [123, 124].

Aberrant replication intermediates arise in cells lacking WRN and BLM [97, 125]. Such unre‐solved replication or recombination structures lead to incomplete chromosome segregation. BLM, topoisomerase 3 alpha (Topo3$\alpha$) and BLAP75/RMI1 (for BLM-associated polypeptide/ RecQ-mediated genome instability) or a BLM-Topo3$\alpha$-BLAP75/RMI1 complex localizes to resulting anaphase bridges [126]. A helicase known as PICH arrives first, followed by the

resolution activity of BLM [127]. As helicases unwind duplex DNA, torsional stress produced in the DNA may require relief through topoisomerase activity, and such activity finally decatenates interlocked DNA molecules [128, 129]. *In vitro* studies show that BLM can partner with Topo3$\alpha$ to resolve dHJ and prevent sequence exchange through resolution of this recombination intermediate [95]. The double-junction dissolution reaction requires the HRDC domain of BLM [24]. An additional protein, BLAP75/DMI1 mediates formation of the "dissolvasome" (BLM, Topo3$\alpha$ and BLAP75/RMI1) [130-132]. Mammalian BLAP18/RMI2 has also been found to be part of this complex [133].

## 4.2. WRN

Over 50 distinct *WRN* mutations have been reported, most of which lead to premature termination of translation [19]. Recent missense mutations in the exonuclease domain in one patient compromised protein stability [134]. Most mutations in WS patients occur in the WRN C-terminal domain, which could disrupt the WRN/p53 interaction [20, 134]. Such premature termination could also disrupt the Del hom sequence shown in Fig. 2. No WS mutations have been reported that eliminate only helicase or only exonuclease activity. Both activities are compromised in the development of WS [134]. WS fibroblasts undergo replicative senescence prematurely [135-139]. Telomere defects in WS cells suggest WRN activity in human telomere homeostasis [19]. Telomeres are needed to avoid loss of genetic material. They are important for chromosome end replication and for protection of the ends from enzymatic attack [140]. Human telomeres contain 5 to 20 kb of the repetitive sequence TTAGGG [141]. At the terminal there are 100-200 bp of 3' ssDNA overhang. This overhang can anneal with telomere DNA to form a stable D-loop leading to a structure referred to as a 't-loop' [142, 143]. Alternatively, this free unannealed end may form G-quadruplex DNA [144]. Human WRN functions in lagging-strand synthesis, and in the replication of telomeric G-rich DNA ends [145]. *C. elegans* WRN-1 can disrupt D-loops [146] and human WRN can prevent aberrant recombination [147]. WRN 3' to 5' exonuclease is stimulated by the interacting Ku70/Ku86 complex supporting a role for WRN in DNA repair [148]. Evidence suggests that in the absence of telomerase, WRN and BLM have a role in the ALT (Alternate Lengthening of Telomeres) pathway for telomere maintenance [149, 150]. In biochemical experiments, WRN releases a 3' invading tail from a telomeric type D-loop by coordinated WRN helicase and exonuclease activities [149].

WRN and BLM catalytic activities are comparable except for 3' to 5' exonuclease activity of WRN [151-153], which BLM lacks. On dsDNA and on RNA-DNA hybrids, the WRN exonuclease activity degrades a 3' recessed end. This activity can remove only one mismatched NT at the end of the recessed 3' DNA and can initiate exonuclease activity from a gap or nick [154, 155]. The exonuclease activity of WRN can degrade abnormal DNA structures suggesting that WRN helicase and exonuclease activities are involved in resolution of aberrant DNA structures at stalled forks [156]. Human WRN interacts with proteins involved in DNA replication. WRN coimmunoprecipitates with PCNA and topoisomerase 1 [157]. WRN functionally and physically interacts with RPA [117], and it functionally interacts with DNA polymerase delta [152]. WS cells accumulate recombination intermediates that impede cell

growth [158]. In cells treated with HU, WRN colocalizes with RPA foci and is thought to dissociate recombination intermediates at the stalled forks [94, 147]. WRN stimulates polymerase delta activity in the absence of its processivity factor PCNA. This suggests a role for WRN in recruiting polymerase delta for replication restart at blocked or collapsed forks [152]. RPA can stimulate the processivity of WRN. The stimulation of WRN by RPA is due to protein-protein interactions as opposed to enhanced ATPase activity [117, 118].

# 5. Unification of BLM, WRN, RECQL4 and MCM2-7 activities in DNA replication and recombination/repair

## 5.1. BLM: Role in DNA damage response with a complex role in inhibiting or promoting HR

BLM is found mostly in fine granules throughout the nucleoplasm at highest levels during S and G2 phases of the cell cycle. Its focal localization is in PML nuclear bodies (PML-NB). The name PML derives from the promyelocytic leukemia protein, PML [159-162]. This protein forms the structural groundwork of the PML bodies, which store various nuclear proteins [163]. These PML-NB store repair proteins (e.g., Topo3$\alpha$, MRN and p53) and may be involved in sensing DNA damage [163]. By regulating the availability of repair proteins, response can be directed to DNA damage sites. Trafficking of proteins to the PML-NB is regulated by sumoylation [164]. The sumoylation pathway involves E1, E2 and E3 enzymes, which regulate respectively, SUMO activation, SUMO conjugation and targeting of specific substrates for sumoylation through ligation [165]. BLM contains a motif for SUMO binding that would facilitate its integration into this repair protein storage network [166]. In addition, BLM is SUMO-1 and SUMO-2 modified [167]. When mutants are prepared in which the SUMO-binding sequences are deleted, BLM cannot localize to PML-NB [168]. When mutants are prepared that do not allow BLM localization to PML-NB, there is about a two-fold increase in sister chromatid exchange. These findings indicate that there is a need for BLM-SUMO interaction in order for BLM to localize to PML-NB, and that BLM activity, such as its accumulation at stalled replication forks, may be regulated by this specific localization [168].

At sites of DSB, repair foci form. A central player, H2AX, is phosphorylated when DSBs are induced, and this phosphorylation involves ATM, ATR and DNA-PK. Over one million bp are then marked by phosphorylated H2AX ($\gamma$H2AX) on each side of the break [112, 168-171]. $\gamma$H2AX recruits additional repair proteins to the damage site [172]. In studies where normal S-phase cells are treated with DNA damaging agents (HU, UV and cross-linking agents), BLM responds by leaving the PML-NBs to relocate to repair foci and colocalize with the marker $\gamma$H2AX [101, 173]. BLM interacts physically and functionally with $\gamma$H2AX as well as with ATM and ATR [103, 173-175]. In damage that is S-phase specific, BLM associates with the complex ATR/CHK1/53BP1, which gathers at repair foci as an early response to the damage. Based on kinetic studies, BLM may facilitate BRCA1 and MRN complex localization at repair foci in S phase [173], which may involve BLM regulation of p53 in these foci [101]. This early function of BLM at repair foci may allow for BML to influence the choice of repair pathways

and facilitate a BLM function in anti-recombination at stalled forks that perhaps involves SU-MO regulation [19].

As discussed in the previous section, BLM interacts with the recombinase, Rad51. Rad51 catalyzes the pairing of a ssDNA tail and a homologous stretch of dsDNA to promote strand exchange early in the HR pathway [111]. Following ionizing radiation, Rad51 foci contained BLM [159]. Rad51 functions in HR, and localizes to ssDNA when DSB are induced [176]. BML can displace the recombinase from the ssDNA filament [112], which can be viewed as an anti-recombinagenic function [19]. In Rad51-associated D-loops, BLM can interact with Rad51 and unwind DNA in front of the polymerase [177]. This could favor SDSA leading to pro-recombinagenic function [19, 177]. BLM, however, also functions in resolution of G-quadruplex DNA structure and has higher binding affinity for it compared to HJ. BLM helicase activity is required for resolution of this structure [92]. Whereas the BLM/Rad51 interaction would represent an early event in a DNA damage/repair process, the formation of HJ, on the other hand, is a late event in HR. As discussed in section 5.1, the BLM-Topo3$\alpha$-BLAP75/RMI1 complex functions to "dissolve" dHJ by convergent fork migration to generate non-cross-over products [178]. This is facilitated by Topo3$\alpha$ relief of superhelicity and by its ability to cleave and rejoin one strand of a DNA duplex. In the absence of BLM, Topo3$\alpha$ activity would involve break and rejoining activities instead of dissolution, which could lead to crossover events and an increase in sister chromatid exchange in BS cells [19]. BLM that is mutated to be unable to interact with Topo3$\alpha$ can only partially rescue the frequency of sister chromatid exchange [168]. Thus, together in a complex, BLM and Topo3$\alpha$ achieve dissolution of a recombinogenic intermediate. BLM has been proposed to regulate ploidy based on a role along with other members of this complex in resolution of anaphase bridges [126]. BLM has been found in complexes with mismatch repair protein MLH1 [179, 180]. It has been found to be present in large complexes containing not only BLM-Topo3a-BLAF75, but also additional factors. These additional factors could include several BLAF factors, proteins from the FA (Fanconi anemia) pathway, RPA, and mismatch repair protein MLH1. [181, 182]. The interactive role of these pathway components remains to be determined.

## 5.2. WRN: Helicase and exonuclease activities in concert

When WRN biochemical activities were compared using nonhydrolyzable ATP$\gamma$S to inhibit only WRN helicase activity or aa substitutions to eliminate only the WRN exonuclease activity, the exonuclease activity was shown to function in degradation of the leading strand on replication fork-type substrates and in degradation of the annealed telomere overhangs on substrates resembling D-loop structures. WRN binding proteins were inhibitory [149, 183]. In addition, WRN was found to degrade ssDNA substrates longer than 40 nt with dependence upon the helicase activity [183]. WRN activities were also explored using WS fibroblasts. In WS cells, WRN and the enzyme telomerase are able to reverse the phenotype of excess chromosome fusion. In these cells, the anaphase bridges were missing telomere DNA. Dominant negative telomerase was not able to rescue the phenotype indicating that a stable telomere length was needed for rescue [184]. Telomeres are stabilized by a complex of proteins that bind DNA, known as the shelterin complex [185]. This complex consists of TRF1, TRF2 (double-strand DNA binding proteins), and POT1 (single-strand DNA binding protein) as well as

adaptor proteins. The coordinate action of these proteins in the presence of telomerase is needed to regulate telomere length. Without telomerase, telomeres shorten in length each cell division [185]. In S phase, WRN colocalizes with TRF2 [145, 186]. TRF1 and TRF2 limit WRN exonuclease activity on synthetic telomere D-loops [149]. WRN helicase is stimulated by single-stranded DNA binding proteins, RPA or POT1. These proteins modulate WRN exonuclease degradation of the 3′ overhang [149, 187]. The hypothesis is that WRN could unwind D-loops to facilitate leading strand synthesis through telomeric DNA, and there is also the possibility that WRN prevents interchromosomal interactions between telomeres [19].

### 5.2.1. WRN activities in response to G-quadruplex structures in the lagging strand

At the telomere, the G-rich strand is duplicated by lagging-strand synthesis. It may assume a G-quadruplex structure, which would interrupt the replication fork. Experimentally, when inhibiting WRN by use of overexpression of a dominant-negative WRN, the helicase deficiency leads to loss almost entirely affecting the sister telomere on the lagging strand [145]. WRN interaction with FEN-1 (the flap endonuclease involved in the processing of Okazaki fragments, [188]) could assist in the maturation of these fragments [105, 189]. Data suggest that WRN may function in concert with lagging strand synthesis perhaps in unwinding complex structure at the telomere. WRN interacts physically with DNA polymerase delta to stimulate its activity [190]. WRN can prevent stalling of polymerases delta at telomere sequence *in vitro* [191]. WRN stimulation of polymerase delta happens only in the absence of PCNA, which suggests that WRN has a role at stalled forks rather than in the regulation of processive DNA synthesis [19]. WRN, as opposed to BLM, is specific for G-quadruplex structure in the trinucleotide repeat of Fragile X syndrome [192].

### 5.2.2. WRN and Ku in suppression of aberrant recombination at telomeres

Proteins involved in the NHEJ path for repair of DSBs that are found at telomeres include Ku heterodimer, DNA-PK, MRN complex and ATM [193]. Through interaction with NBS1, WRN colocalizes with two of these components, Ku and the MRN complex [194]. Under unstable conditions, Ku can suppress sister chromatid exchange at telomeres [195]. Mouse knockout studies show that WRN normally suppresses aberrant recombination at telomeres [196].

### 5.2.3. WRN activities in base excision repair and interstrand cross link repair by HR

WS cells show increased sensitivity to alkylating agents and to agents that increase ROS. Human fibroblasts in which WRN has been knocked down respond to oxidative stress with an increase in DNA damage [197]. These observations support a role for WRN in repair of such damage [198]. DNA damage resulting from oxidation, alkylation, methylation and deamination is repaired by the base excision repair (BER) pathway. WRN interacts with proteins in this pathway. In addition, WRN helicase activity stimulates and is stimulated by polymerase beta activity [199]. There is evidence that WRN activity in BER is regulated by PARP-1, but the reader is referred to references [200, 201]. BRCA1 and BRCA2 play an important role in DSB repair in the HR subpathway [202, 203]. HR is a pathway for repairing interstrand crosslinks (ICL). Cells deficient in either WRN or BRCA1 are hypersensitive to

induction of ICLs [204, 205]. A physical interaction of WRN with BRCA1 enhances the activities of WRN [206]. In addition, data from WRN and/or BRCA1 knockdown studies indicate that BRCA1 may act cooperatively with WRN in HR during ICL repair [206].

## 5.3. RECQL4

Data from *Recql4⁻/⁻* mice indicate a role for mouse RECQL4 in genomic stability and in the promoting cohesion between sister chromatids [207]. Depleting the *X. laevis* homologue of RTS, xRTS, in *Xenopus* egg extracts led to reduced DNA synthesis and an inhibition of RPA stabilization of ssDNA prior to polymerase loading at unwound origins. The addition of purified human RECQL4 could reverse this effect [208]. A nonhelical region in the N terminus of xRTS could be important in initiation of DNA replication based on its interaction with the Cut5 protein and the importance of xRTS in loading DNA polymerase alpha onto chromatin [209]. The N terminus of xRTS is not homologous to that of RECQL4 in mammals, however, and this role may not be conserved [19]. RECQL4, along with Ctf4 and MCM10, has been shown to be required for stable association of the CMG complex in human cells. In this study, Cut5/TopBP1 was not required for CMG stabilization [210]. In the *Xenopus* replication model, RECQL4 binds to chromatin that has been processed to resemble DSBs with a dependence on RPA and ATM activity [211]. Chromatin immunoprecipitation experiments show that RECQL4 functions on DNA in close association with Ku [212]. In HeLa cells human RECQL4 forms a complex with Rad51 and colocalizes with Rad51 foci formed after treatment with etoposide [212]. PML-NB contain a portion of RECQL4 [212]. It is also found in the nucleolus [213]. When using a T7 phage display screen, RECQL4 was found to interact with PARP-1, and the association influenced the nuclear localization of RECQL4. PARP-1 has a role in RECQL4 movement to the nucleolus from the nucleoplasm. When oxidative stress is induced, as opposed to other types of damage induction, RECQL4 increasingly localizes in the nucleolus. This trafficking is inhibited by inhibition of PARP [213]. RTS cells decrease proliferation and synthesis of DNA when exposed to hydrogen peroxide [214]. Lack of proper response to ROS could lead to premature aging as see in RTS. PARP is also involved in an end-joining pathway of DNA repair [215, 216]. The role of RECQL4 in its interaction with PARP-1 is not known (for further discussion see reference [19]).

## 5.4. MCM2-7

Replication fork stalling, can lead to DSB and chromosomal rearrangements. An S phase checkpoint is triggered by these events and there is a block to elongation. When this occurs, the proteins Mrc1, Tof1, and Csm3 (M/T/C complex) interact with the MCM2-7 complex to stabilize the replication fork. When the M/T/C complex is missing, the replisome continues, but synthesis stops. This may be partially due to loss of DNA polymerase epsilon from the fork [217-219]. Studies in yeast provide insight into these activities. The M/T/C complex associates with MCM2-7 [218, 220-222] and also with polymerase epsilon [221, 223]. These physical interactions permit communication between polymerase epsilon and the MCM complex [223]. The M/T/C complex may be part of the normal replication fork protein entourage [218, 220, 221, 224]. Mrc1 and Cdc45 coimmunoprecipitate, indicating an interaction of Mrc1 with the The Replication Progression Complex core, which includes Cdc45, the MCMs and GINS [223]. In each cell cycle Mrc loads onto replication origins along with the

polymerases. This occurs after the Replication Progression Complex forms. Mrc migrates with the replication forks. [218, 219, 224-226]. Tof1, like Mrc, also coimmunoprecipitates with Cdc45 [227]. The exact mechanism of action of the M/T/C complex is not known. In a yeast study, the Tof1 homologue could switch regulation between pro- and anti-recombination activities in a site-specific manner [228]. Data indicate that a Mrc1 and MCM6-C terminal interaction senses alkylated DNA damage [221]. The other two subunits Tof1 and Csm3, may function to sense other types of damage. Although the helicase domains of MCM2-7 are conserved, the N and C terminals are divergent. Other negative regulators could differentially bind to these regions to regulate powering of elongation by the MCM2-7 helicase during times of stress [10]. A future question relates to the extent to which leading or lagging strand polymerase arrest is associated with formation of ssDNA, fork regression and formation of abnormal DNA structures [66]. These data indicate a functional connection between the MCM proteins, which act at stalled forks, and the RecQ proteins, which facilitate repair of the resulting damage.

## 6. Summary

BLM, WRN and RECQL4 act during events that stress the advancing replication fork providing relief through DNA damage repair and through resolution of aberrant replication/recombination intermediates, including those present at the telomere. At checkpoint, the replication proteins at a stalled fork are held stable through communication that occurs due to proteins that bind and signal to both the MCM complex and polymerase(s). This would allow repair proteins such as WRN and BLM helicases and RECQL4 to resolve the stress and thus aid in fork restart. Advancing our knowledge of the RecQ and MCM family activities and the mechanisms and signalling behind these activities will increase our understanding of cancer and aging and perhaps enlighten us regarding how to accommodate these challenges to human health.

## Acknowledgments

## Author details

Dianne C. Daniel*, Ayuna V. Dagdanova and Edward M. Johnson

*Address all correspondence to: danieldc@evms.edu

Department of Microbiology and Molecular Cell Biology, Eastern Virginia Medical School, Norfolk, Virginia, USA

# References

[1]  Berger JM. SnapShot: nucleic acid helicases and translocases. Cell. 2008 Sep 5;134(5): 888- e1. 18775318

[2]  Bernstein KA, Gangloff S, Rothstein R. The RecQ DNA helicases in DNA repair. Annu Rev Genet. 2010;44:393-417. 21047263

[3]  Kelman Z, Hurwitz J. Structural lessons in DNA replication from the third domain of life. Nat Struct Biol. 2003 Mar;10(3):148-50. 12605215

[4]  Erzberger JP, Berger JM. Evolutionary relationships and structural mechanisms of AAA+ proteins. Annu Rev Biophys Biomol Struct. 2006;35:93-114. 16689629

[5]  Koonin EV. 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. 1993 Jun 11;21(11):2541-7. 8332451

[6]  Iyer LM, Leipe DD, Koonin EV, Aravind L. Evolutionary history and higher order classification of AAA+ ATPases. J Struct Biol. 2004 Apr-May;146(1-2):11-31. 15037234

[7]  Leipe DD, Koonin EV, Aravind L. Evolution and classification of P-loop kinases and related proteins. J Mol Biol. 2003 Oct 31;333(4):781-815. 14568537

[8]  Mohaghegh P, Hickson ID. DNA helicase deficiencies associated with cancer predisposition and premature ageing disorders. Hum Mol Genet. 2001 Apr;10(7):741-6. 11257107

[9]  Brosh RM, Jr., Bohr VA. Human premature aging, DNA repair and RecQ helicases. Nucleic Acids Res. 2007;35(22):7527-44. 18006573

[10]  Bochman ML, Schwacha A. The Mcm complex: unwinding the mechanism of a replicative helicase. Microbiol Mol Biol Rev. 2009 Dec;73(4):652-83. 19946136

[11]  Gozuacik D, Chami M, Lagorce D, Faivre J, Murakami Y, Poch O, Biermann E, Knippers R, Brechot C, Paterlini-Brechot P. Identification and functional characterization of a new member of the human Mcm protein family: hMcm8. Nucleic Acids Res. 2003 Jan 15;31(2):570-9. 12527764

[12]  Johnson EM, Kinoshita Y, Daniel DC. A new member of the MCM protein family encoded by the human MCM8 gene, located contrapodal to GCD10 at chromosome band 20p12.3-13. Nucleic Acids Res. 2003 Jun 1;31(11):2915-25. 12771218

[13]  Lutzmann M, Maiorano D, Mechali M. Identification of full genes and proteins of MCM9, a novel, vertebrate-specific member of the MCM2-8 protein family. Gene. 2005 Dec 5;362:51-6. 16226853

[14]  Yoshida K. Identification of a novel cell-cycle-induced MCM family protein MCM9. Biochem Biophys Res Commun. 2005 Jun 3;331(2):669-74. 15850810

[15] Fien K, Cho YS, Lee JK, Raychaudhuri S, Tappin I, Hurwitz J. Primer utilization by DNA polymerase alpha-primase is influenced by its interaction with Mcm10p. J Biol Chem. 2004 Apr 16;279(16):16144-53. 14766746

[16] Ricke RM, Bielinsky AK. Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. Mol Cell. 2004 Oct 22;16(2):173-85. 15494305

[17] Sawyer SL, Cheng IH, Chai W, Tye BK. Mcm10 and Cdc45 cooperate in origin activation in Saccharomyces cerevisiae. J Mol Biol. 2004 Jul 2;340(2):195-202. 15201046

[18] Yang X, Gregan J, Lindner K, Young H, Kearsey SE. Nuclear distribution and chromatin association of DNA polymerase alpha-primase is affected by TEV protease cleavage of Cdc23 (Mcm10) in fission yeast. BMC Mol Biol. 2005;6:13. 15941470

[19] Ouyang KJ, Woo LL, Ellis NA. Homologous recombination and maintenance of genome integrity: cancer and aging through the prism of human RecQ helicases. Mech Ageing Dev. 2008 Jul-Aug;129(7-8):425-40. 18430459

[20] Bachrati CZ, Hickson ID. RecQ helicases: suppressors of tumorigenesis and premature aging. Biochem J. 2003 Sep 15;374(Pt 3):577-606. 12803543

[21] Singleton MR, Dillingham MS, Wigley DB. Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem. 2007;76:23-50. 17506634

[22] Killoran MP, Keck JL. Sit down, relax and unwind: structural insights into RecQ helicase mechanisms. Nucleic Acids Res. 2006;34(15):4098-105. 16935877

[23] Macris MA, Krejci L, Bussen W, Shimamoto A, Sung P. Biochemical characterization of the RECQ4 protein, mutated in Rothmund-Thomson syndrome. DNA Repair (Amst). 2006 Feb 3;5(2):172-80. 16214424

[24] Wu L, Chan KL, Ralf C, Bernstein DA, Garcia PL, Bohr VA, Vindigni A, Janscak P, Keck JL, Hickson ID. The HRDC domain of BLM is required for the dissolution of double Holliday junctions. Embo J. 2005 Jul 20;24(14):2679-87. 15990871

[25] Xu X, Rochette PJ, Feyissa EA, Su TV, Liu Y. MCM10 mediates RECQ4 association with MCM2-7 helicase complex during DNA replication. Embo J. 2009 Oct 7;28(19): 3005-14. 19696745

[26] Chu WK, Hickson ID. RecQ helicases: multifunctional genome caretakers. Nat Rev Cancer. 2009 Sep;9(9):644-54. 19657341

[27] Mogni ME, Costa A, Ioannou C, Bell SD. The glutamate switch is present in all seven clades of AAA+ protein. Biochemistry. 2009 Sep 22;48(37):8774-5. 19702328

[28] Snider J, Thibault G, Houry WA. The AAA+ superfamily of functionally diverse proteins. Genome Biol. 2008;9(4):216. 18466635

[29] Davey MJ, Jeruzalmi D, Kuriyan J, O'Donnell M. Motors and switches: AAA+ machines within the replisome. Nat Rev Mol Cell Biol. 2002 Nov;3(11):826-35. 12415300

[30] Ogura T, Wilkinson AJ. AAA+ superfamily ATPases: common structure--diverse function. Genes Cells. 2001 Jul;6(7):575-97. 11473577

[31] Saraste M, Sibbald PR, Wittinghofer A. The P-loop--a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci. 1990 Nov;15(11):430-4. 2126155

[32] Walker JE, Saraste M, Runswick MJ, Gay NJ. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. Embo J. 1982;1(8):945-51. 6329717

[33] Davey MJ, Indiani C, O'Donnell M. Reconstitution of the Mcm2-7p heterohexamer, subunit arrangement, and ATP site architecture. J Biol Chem. 2003 Feb 14;278(7): 4491-9. 12480933

[34] Schwacha A, Bell SP. Interactions between two catalytically distinct MCM subgroups are essential for coordinated ATP hydrolysis and DNA replication. Mol Cell. 2001 Nov;8(5):1093-104. 11741544

[35] Yu CE, Oshima J, Fu YH, Wijsman EM, Hisama F, Alisch R, Matthews S, Nakura J, Miki T, Ouais S, Martin GM, Mulligan J, Schellenberg GD. Positional cloning of the Werner's syndrome gene. Science. 1996 Apr 12;272(5259):258-62. 8602509

[36] Bloom D. Congenital telangiectatic erythema resembling lupus erythematosus in dwarfs; probably a syndrome entity. AMA Am J Dis Child. 1954 Dec;88(6):754-8. 13206391

[37] Kitao S, Lindor NM, Shiratori M, Furuichi Y, Shimamoto A. Rothmund-thomson syndrome responsible gene, RECQL4: genomic structure and products. Genomics. 1999 Nov 1;61(3):268-76. 10552928

[38] Siitonen HA, Sotkasiira J, Biervliet M, Benmansour A, Capri Y, Cormier-Daire V, Crandall B, Hannula-Jouppi K, Hennekam R, Herzog D, Keymolen K, Lipsanen-Nyman M, Miny P, Plon SE, Riedl S, Sarkar A, Vargas FR, Verloes A, Wang LL, Kaariainen H, Kestila M. The mutation spectrum in RECQL4 diseases. Eur J Hum Genet. 2009 Feb;17(2):151-8. 18716613

[39] Hanada K, Hickson ID. Molecular genetics of RecQ helicase disorders. Cell Mol Life Sci. 2007 Sep;64(17):2306-22. 17571213

[40] Karow JK, Wu L, Hickson ID. RecQ family helicases: roles in cancer and aging. Curr Opin Genet Dev. 2000 Feb;10(1):32-8. 10679384

[41] German J. Bloom syndrome: a mendelian prototype of somatic mutational disease. Medicine (Baltimore). 1993 Nov;72(6):393-406. 8231788

[42] German J. Bloom's syndrome. Dermatol Clin. 1995 Jan;13(1):7-18. 7712653

[43] Shen JC, Loeb LA. The Werner syndrome gene: the molecular basis of RecQ helicase-deficiency diseases. Trends Genet. 2000 May;16(5):213-20. 10782115

[44] Vasseur F, Delaporte E, Zabot MT, Sturque MN, Barrut D, Savary JB, Thomas L, Thomas P. Excision repair defect in Rothmund Thomson syndrome. Acta Derm Venereol. 1999 Mar;79(2):150-2. 10228638

[45] Tye BK. MCM proteins in DNA replication. Annu Rev Biochem. 1999;68:649-86. 10872463

[46] Forsburg SL. Eukaryotic MCM proteins: beyond replication initiation. Microbiol Mol Biol Rev. 2004 Mar;68(1):109-31. 15007098

[47] Chong JP, Hayashi MK, Simon MN, Xu RM, Stillman B. A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. Proc Natl Acad Sci U S A. 2000 Feb 15;97(4):1530-5. 10677495

[48] Kelman Z, Lee JK, Hurwitz J. The single minichromosome maintenance protein of Methanobacterium thermoautotrophicum DeltaH contains DNA helicase activity. Proc Natl Acad Sci U S A. 1999 Dec 21;96(26):14783-8. 10611290

[49] Tye BK. Insights into DNA replication from the third domain of life. Proc Natl Acad Sci U S A. 2000 Mar 14;97(6):2399-401. 10716976

[50] Madine MA, Swietlik M, Pelizon C, Romanowski P, Mills AD, Laskey RA. The roles of the MCM, ORC, and Cdc6 proteins in determining the replication competence of chromatin in quiescent cells. J Struct Biol. 2000 Apr;129(2-3):198-210. 10806069

[51] Romanowski P, Madine MA, Rowles A, Blow JJ, Laskey RA. The Xenopus origin recognition complex is essential for DNA replication and MCM binding to chromatin. Curr Biol. 1996 Nov 1;6(11):1416-25. 8939603

[52] Blow JJ, Hodgson B. Replication licensing--defining the proliferative state? Trends Cell Biol. 2002 Feb;12(2):72-8. 11849970

[53] Madine MA, Khoo CY, Mills AD, Musahl C, Laskey RA. The nuclear envelope prevents reinitiation of replication by regulating the binding of MCM3 to chromatin in Xenopus egg extracts. Curr Biol. 1995 Nov 1;5(11):1270-9. 8574584

[54] Ishimi Y. A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex. J Biol Chem. 1997 Sep 26;272(39):24508-13. 9305914

[55] Lee JK, Hurwitz J. Processive DNA helicase activity of the minichromosome maintenance proteins 4, 6, and 7 complex requires forked DNA structures. Proc Natl Acad Sci U S A. 2001 Jan 2;98(1):54-9. 11136247

[56] Bochman ML, Schwacha A. The Mcm2-7 complex has in vitro helicase activity. Mol Cell. 2008 Jul 25;31(2):287-93. 18657510

[57] Boos D, Frigola J, Diffley JF. Activation of the replicative DNA helicase: breaking up is hard to do. Curr Opin Cell Biol. 2012 Jun;24(3):423-30. 22424671

[58] Kang YH, Galal WC, Farina A, Tappin I, Hurwitz J. Properties of the human Cdc45/Mcm2-7/GINS helicase complex and its action with DNA polymerase epsilon in roll-

ing circle DNA synthesis. Proc Natl Acad Sci U S A. 2012 Apr 17;109(16):6042-7. 22474384

[59] Kinoshita Y, Johnson EM. Site-specific loading of an MCM protein complex in a DNA replication initiation zone upstream of the c-MYC gene in the HeLa cell cycle. J Biol Chem. 2004 Aug 20;279(34):35879-89. 15190069

[60] Schepers A, Diffley JF. Mutational analysis of conserved sequence motifs in the budding yeast Cdc6 protein. J Mol Biol. 2001 May 11;308(4):597-608. 11350163

[61] Nevis KR, Cordeiro-Stone M, Cook JG. Origin licensing and p53 status regulate Cdk2 activity during G(1). Cell Cycle. 2009 Jun 15;8(12):1952-63. 19440053

[62] Coverley D, Pelizon C, Trewick S, Laskey RA. Chromatin-bound Cdc6 persists in S and G2 phases in human cells, while soluble Cdc6 is destroyed in a cyclin A-cdk2 dependent process. J Cell Sci. 2000 Jun;113 ( Pt 11):1929-38. 10806104

[63] Alexandrow MG, Hamlin JL. Cdc6 chromatin affinity is unaffected by serine-54 phosphorylation, S-phase progression, and overexpression of cyclin A. Mol Cell Biol. 2004 Feb;24(4):1614-27. 14749377

[64] Shin JH, Grabowski B, Kasiviswanathan R, Bell SD, Kelman Z. Regulation of minichromosome maintenance helicase activity by Cdc6. J Biol Chem. 2003 Sep 26;278(39):38059-67. 12837750

[65] Daniel DC, Johnson EM. Addressing the enigma of MCM8 in DNA replication. In: Kušić-Tišma J, ed. Fundamental Aspects of DNA Replication: INTECH 2011:37-52. DOI: 10.5772/21177. http://www.intechopen.com/books/fundamental-aspects-of-dna-replication/addressing-the-enigma-of-mcm8-in-dna-replication

[66] Forsburg SL. The MCM helicase: linking checkpoints to the replication fork. Biochem Soc Trans. 2008 Feb;36(Pt 1):114-9. 18208397

[67] Masai H, You Z, Arai K. Control of DNA replication: regulation and activation of eukaryotic replicative helicase, MCM. IUBMB Life. 2005 Apr-May;57(4-5):323-35. 16036617

[68] Costa A, Onesti S. The MCM complex: (just) a replicative helicase? Biochem Soc Trans. 2008 Feb;36(Pt 1):136-40. 18208401

[69] Brewster AS, Wang G, Yu X, Greenleaf WB, Carazo JM, Tjajadi M, Klein MG, Chen XS. Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase. Proc Natl Acad Sci U S A. 2008 Dec 23;105(51):20191-6. 19073923

[70] Makarova KS, Koonin EV, Kelman Z. The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. Biol Direct. 2012;7:7. 22329974

[71] Woodward AM, Gohler T, Luciani MG, Oehlmann M, Ge X, Gartner A, Jackson DA, Blow JJ. Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. J Cell Biol. 2006 Jun 5;173(5):673-83. 16754955

[72] Claycomb JM, MacAlpine DM, Evans JG, Bell SP, Orr-Weaver TL. Visualization of replication initiation and elongation in Drosophila. J Cell Biol. 2002 Oct 28;159(2): 225-36. 12403810

[73] Hirai K, Shirakata M. Replication licensing of the EBV oriP minichromosome. Curr Top Microbiol Immunol. 2001;258:13-33. 11443858

[74] Yankulov K, Todorov I, Romanowski P, Licatalosi D, Cilli K, McCracken S, Laskey R, Bentley DL. MCM proteins are associated with RNA polymerase II holoenzyme. Mol Cell Biol. 1999 Sep;19(9):6154-63. 10454562

[75] Volkening M, Hoffmann I. Involvement of human MCM8 in prereplication complex assembly by recruiting hcdc6 to chromatin. Mol Cell Biol. 2005 Feb;25(4):1560-8. 15684404

[76] Klotz-Noack K, Blow JJ. A role for dormant origins in tumor suppression. Mol Cell. 2011 Mar 4;41(5):495-6. 21362544

[77] Gomez EB, Catlett MG, Forsburg SL. Different phenotypes in vivo are associated with ATPase motif mutations in Schizosaccharomyces pombe minichromosome maintenance proteins. Genetics. 2002 Apr;160(4):1305-18. 11973289

[78] You Z, Komamura Y, Ishimi Y. Biochemical analysis of the intrinsic Mcm4-Mcm6-mcm7 DNA helicase activity. Mol Cell Biol. 1999 Dec;19(12):8003-15. 10567526

[79] Gineau L, Cognet C, Kara N, Lach FP, Dunne J, Veturi U, Picard C, Trouillet C, Eidenschenk C, Aoufouchi S, Alcais A, Smith O, Geissmann F, Feighery C, Abel L, Smogorzewska A, Stillman B, Vivier E, Casanova JL, Jouanguy E. Partial MCM4 deficiency in patients with growth retardation, adrenal insufficiency, and natural killer cell deficiency. J Clin Invest. 2012 Mar 1;122(3):821-32. 22354167

[80] Hughes CR, Guasti L, Meimaridou E, Chuang CH, Schimenti JC, King PJ, Costigan C, Clark AJ, Metherell LA. MCM4 mutation causes adrenal failure, short stature, and natural killer cell deficiency in humans. J Clin Invest. 2012 Mar 1;122(3):814-20. 22354170

[81] Chatrath P, Scott IS, Morris LS, Davies RJ, Rushbrook SM, Bird K, Vowler SL, Grant JW, Saeed IT, Howard D, Laskey RA, Coleman N. Aberrant expression of minichromosome maintenance protein-2 and Ki67 in laryngeal squamous epithelial lesions. Br J Cancer. 2003 Sep 15;89(6):1048-54. 12966424

[82] Davidson EJ, Morris LS, Scott IS, Rushbrook SM, Bird K, Laskey RA, Wilson GE, Kitchener HC, Coleman N, Stern PL. Minichromosome maintenance (Mcm) proteins, cyclin B1 and D1, phosphohistone H3 and in situ DNA replication for functional analysis of vulval intraepithelial neoplasia. Br J Cancer. 2003 Jan 27;88(2):257-62. 12610511

[83] Freeman A, Morris LS, Mills AD, Stoeber K, Laskey RA, Williams GH, Coleman N. Minichromosome maintenance proteins as biological markers of dysplasia and malignancy. Clin Cancer Res. 1999;5(8):2121-32.

[84] Gonzalez MA, Pinder SE, Callagy G, Vowler SL, Morris LS, Bird K, Bell JA, Laskey RA, Coleman N. Minichromosome maintenance protein 2 is a strong independent prognostic marker in breast cancer. J Clin Oncol. 2003 Dec 1;21(23):4306-13. 14645419

[85] Hunt DP, Freeman A, Morris LS, Burnet NG, Bird K, Davies TW, Laskey RA, Coleman N. Early recurrence of benign meningioma correlates with expression of minichromosome maintenance-2 protein. Br J Neurosurg. 2002 Feb;16(1):10-5. 11928726

[86] Khalili K, Del Valle L, Muralidharan V, Gault WJ, Darbinian N, Otte J, Meier E, Johnson EM, Daniel DC, Kinoshita Y, Amini S, Gordon J. Puralpha Is Essential for Postnatal Brain Development and Developmentally Coupled Cellular Proliferation As Revealed by Genetic Inactivation in the Mouse. Mol Cell Biol. 2003 Oct 1;23(19): 6857-75. 12972605

[87] Bergemann AD, Johnson EM. The HeLa Pur factor binds single-stranded DNA at a specific element conserved in gene flanking regions and origins of DNA replication. Mol Cell Biol. 1992;12:1257-65.

[88] Liu W, Pucci B, Rossi M, Pisani FM, Ladenstein R. Structural analysis of the Sulfolobus solfataricus MCM protein N-terminal domain. Nucleic Acids Res. 2008 Jun; 36(10):3235-43. 18417534

[89] Rong SB, Valiaho J, Vihinen M. Structural basis of Bloom syndrome (BS) causing mutations in the BLM helicase domain. Mol Med. 2000 Mar;6(3):155-64. 10965492

[90] Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993 Apr 22;362(6422):709-15. 8469282

[91] Sun H, Karow JK, Hickson ID, Maizels N. The Bloom's syndrome helicase unwinds G4 DNA. J Biol Chem. 1998 Oct 16;273(42):27587-92. 9765292

[92] Huber MD, Lee DC, Maizels N. G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. Nucleic Acids Res. 2002 Sep 15;30(18): 3954-61. 12235379

[93] van Brabant AJ, Ye T, Sanz M, German IJ, Ellis NA, Holloman WK. Binding and melting of D-loops by the Bloom syndrome helicase. Biochemistry. 2000 Nov 28;39(47):14617-25. 11087418

[94] Karow JK, Constantinou A, Li JL, West SC, Hickson ID. The Bloom's syndrome gene product promotes branch migration of holliday junctions. Proc Natl Acad Sci U S A. 2000 Jun 6;97(12):6504-8. 10823897

[95] Wu L, Hickson ID. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. Nature. 2003 Dec 18;426(6968):870-4. 14685245

[96] Hand R, German J. A retarded rate of DNA chain growth in Bloom's syndrome. Proc Natl Acad Sci U S A. 1975 Feb;72(2):758-62. 1054854

[97] Lonn U, Lonn S, Nylen U, Winblad G, German J. An abnormal profile of DNA replication intermediates in Bloom's syndrome. Cancer Res. 1990 Jun 1;50(11):3141-5. 2110504

[98] Han H, Hurley LH. G-quadruplex DNA: a potential target for anti-cancer drug design. Trends Pharmacol Sci. 2000 Apr;21(4):136-42. 10740289

[99] Davies SL, North PS, Hickson ID. Role for BLM in replication-fork restart and suppression of origin firing after replicative stress. Nat Struct Mol Biol. 2007 Jul;14(7):677-9. 17603497

[100] Rao VA, Conti C, Guirouilh-Barbat J, Nakamura A, Miao ZH, Davies SL, Sacca B, Hickson ID, Bensimon A, Pommier Y. Endogenous gamma-H2AX-ATM-Chk2 checkpoint activation in Bloom's syndrome helicase deficient cells is related to DNA replication arrested forks. Mol Cancer Res. 2007 Jul;5(7):713-24. 17634426

[101] Sengupta S, Linke SP, Pedeux R, Yang Q, Farnsworth J, Garfield SH, Valerie K, Shay JW, Ellis NA, Wasylyk B, Harris CC. BLM helicase-dependent transport of p53 to sites of stalled DNA replication forks modulates homologous recombination. Embo J. 2003 Mar 3;22(5):1210-22. 12606585

[102] Wang Y, Cortez D, Yazdi P, Neff N, Elledge SJ, Qin J. BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. Genes Dev. 2000 Apr 15;14(8):927-39. 10783165

[103] Davies SL, North PS, Dart A, Lakin ND, Hickson ID. Phosphorylation of the Bloom's syndrome helicase and its role in recovery from S-phase arrest. Mol Cell Biol. 2004 Feb;24(3):1279-91. 14729972

[104] Sengupta S, Robles AI, Linke SP, Sinogeeva NI, Zhang R, Pedeux R, Ward IM, Celeste A, Nussenzweig A, Chen J, Halazonetis TD, Harris CC. Functional interaction between BLM helicase and 53BP1 in a Chk1-mediated pathway during S-phase arrest. J Cell Biol. 2004 Sep 13;166(6):801-13. 15364958

[105] Brosh RM, Jr., Driscoll HC, Dianov GL, Sommers JA. Biochemical characterization of the WRN-FEN-1 functional interaction. Biochemistry. 2002 Oct 8;41(40):12204-16. 12356323

[106] Kao HI, Veeraraghavan J, Polaczek P, Campbell JL, Bambara RA. On the roles of Saccharomyces cerevisiae Dna2p and Flap endonuclease 1 in Okazaki fragment processing. J Biol Chem. 2004 Apr 9;279(15):15014-24. 14747468

[107] Bartos JD, Wang W, Pike JE, Bambara RA. Mechanisms by which Bloom protein can disrupt recombination intermediates of Okazaki fragment maturation. J Biol Chem. 2006 Oct 27;281(43):32227-39. 16950766

[108] Sharma S, Otterlei M, Sommers JA, Driscoll HC, Dianov GL, Kao HI, Bambara RA, Brosh RM, Jr. 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. 2004 Feb;15(2):734-50. 14657243

[109]    Sharma S, Sommers JA, Gary RK, Friedrich-Heineken E, Hubscher U, Brosh RM, Jr. The interaction site of Flap Endonuclease-1 with WRN helicase suggests a coordination of WRN and PCNA. Nucleic Acids Res. 2005;33(21):6769-81. 16326861

[110]    Sharma S, Sommers JA, Wu L, Bohr VA, Hickson ID, Brosh RM, Jr. Stimulation of flap endonuclease-1 by the Bloom's syndrome protein. J Biol Chem. 2004 Mar 12;279(11):9847-56. 14688284

[111]    Bachrati CZ, Borts RH, Hickson ID. Mobile D-loops are a preferred substrate for the Bloom's syndrome helicase. Nucleic Acids Res. 2006;34(8):2269-79. 16670433

[112]    Bugreev DV, Yu X, Egelman EH, Mazin AV. Novel pro- and anti-recombination activities of the Bloom's syndrome helicase. Genes Dev. 2007 Dec 1;21(23):3085-94. 18003860

[113]    Bugreev DV, Brosh RM, Jr., Mazin AV. RECQ1 possesses DNA branch migration activity. J Biol Chem. 2008 Jul 18;283(29):20231-42. 18495662

[114]    Jiao R, Bachrati CZ, Pedrazzi G, Kuster P, Petkovic M, Li JL, Egli D, Hickson ID, Stagljar I. Physical and functional interaction between the Bloom's syndrome gene product and the largest subunit of chromatin assembly factor 1. Mol Cell Biol. 2004 Jun;24(11):4710-9. 15143166

[115]    von Kobbe C, Karmakar P, Dawut L, Opresko P, Zeng X, Brosh RM, Jr., Hickson ID, Bohr VA. Colocalization, physical, and functional interaction between Werner and Bloom syndrome proteins. J Biol Chem. 2002 Jun 14;277(24):22035-44. 11919194

[116]    Brosh RM, Jr., Li JL, Kenny MK, Karow JK, Cooper MP, Kureekattil RP, Hickson ID, Bohr VA. Replication protein A physically interacts with the Bloom's syndrome protein and stimulates its helicase activity. J Biol Chem. 2000 Aug 4;275(31):23500-8. 10825162

[117]    Brosh RM, Jr., Orren DK, Nehlin JO, Ravn PH, Kenny MK, Machwe A, Bohr VA. Functional and physical interaction between WRN helicase and human replication protein A. J Biol Chem. 1999 Jun 25;274(26):18341-50. 10373438

[118]    Shen JC, Gray MD, Oshima J, Loeb LA. Characterization of Werner syndrome protein DNA helicase activity: directionality, substrate dependence and stimulation by replication protein A. Nucleic Acids Res. 1998 Jun 15;26(12):2879-85. 9611231

[119]    Blander G, Kipnis J, Leal JF, Yu CE, Schellenberg GD, Oren M. Physical and functional interaction between p53 and the Werner's syndrome protein. J Biol Chem. 1999 Oct 8;274(41):29463-9. 10506209

[120]    Spillare EA, Robles AI, Wang XW, Shen JC, Yu CE, Schellenberg GD, Harris CC. p53-mediated apoptosis is attenuated in Werner syndrome cells. Genes Dev. 1999 Jun 1;13(11):1355-60. 10364153

[121] Wang XW, Tseng A, Ellis NA, Spillare EA, Linke SP, Robles AI, Seker H, Yang Q, Hu P, Beresten S, Bemmels NA, Garfield S, Harris CC. Functional interaction of p53 and BLM DNA helicase in apoptosis. J Biol Chem. 2001 Aug 31;276(35):32948-55. 11399766

[122] Walpita D, Plug AW, Neff NF, German J, Ashley T. Bloom's syndrome protein, BLM, colocalizes with replication protein A in meiotic prophase nuclei of mammalian spermatocytes. Proc Natl Acad Sci U S A. 1999 May 11;96(10):5622-7. 10318934

[123] Li JL, Harrison RJ, Reszka AP, Brosh RM, Jr., Bohr VA, Neidle S, Hickson ID. Inhibition of the Bloom's and Werner's syndrome helicases by G-quadruplex interacting ligands. Biochemistry. 2001 Dec 18;40(50):15194-202. 11735402

[124] Popuri V, Bachrati CZ, Muzzolini L, Mosedale G, Costantini S, Giacomini E, Hickson ID, Vindigni A. The Human RecQ helicases, BLM and RECQ1, display distinct DNA substrate specificities. J Biol Chem. 2008 Jun 27;283(26):17766-76. 18448429

[125] Poot M, Hoehn H, Runger TM, Martin GM. Impaired S-phase transit of Werner syndrome cells expressed in lymphoblastoid cell lines. Exp Cell Res. 1992 Oct;202(2): 267-73. 1327851

[126] Chan KL, North PS, Hickson ID. BLM is required for faithful chromosome segregation and its localization defines a class of ultrafine anaphase bridges. Embo J. 2007 Jul 25;26(14):3397-409. 17599064

[127] Baumann C, Korner R, Hofmann K, Nigg EA. PICH, a centromere-associated SNF2 family ATPase, is regulated by Plk1 and required for the spindle checkpoint. Cell. 2007 Jan 12;128(1):101-14. 17218258

[128] Johnson FB, Lombard DB, Neff NF, Mastrangelo MA, Dewolf W, Ellis NA, Marciniak RA, Yin Y, Jaenisch R, Guarente L. Association of the Bloom syndrome protein with topoisomerase IIIalpha in somatic and meiotic cells. Cancer Res. 2000 Mar 1;60(5): 1162-7. 10728666

[129] Wu L, Davies SL, North PS, Goulaouic H, Riou JF, Turley H, Gatter KC, Hickson ID. The Bloom's syndrome gene product interacts with topoisomerase III. J Biol Chem. 2000 Mar 31;275(13):9636-44. 10734115

[130] Bussen W, Raynard S, Busygina V, Singh AK, Sung P. Holliday junction processing activity of the BLM-Topo IIIalpha-BLAP75 complex. J Biol Chem. 2007 Oct 26;282(43):31484-92. 17728255

[131] Raynard S, Zhao W, Bussen W, Lu L, Ding YY, Busygina V, Meetei AR, Sung P. Functional role of BLAP75 in BLM-topoisomerase IIIalpha-dependent holliday junction processing. J Biol Chem. 2008 Jun 6;283(23):15701-8. 18390547

[132] Wu L, Bachrati CZ, Ou J, Xu C, Yin J, Chang M, Wang W, Li L, Brown GW, Hickson ID. BLAP75/RMI1 promotes the BLM-dependent dissolution of homologous recombination intermediates. Proc Natl Acad Sci U S A. 2006 Mar 14;103(11):4068-73. 16537486

[133]  Singh TR, Ali AM, Busygina V, Raynard S, Fan Q, Du CH, Andreassen PR, Sung P, Meetei AR. BLAP18/RMI2, a novel OB-fold-containing protein, is an essential component of the Bloom helicase-double Holliday junction dissolvasome. Genes Dev. 2008 Oct 15;22(20):2856-68. 18923083

[134]  Huang S, Lee L, Hanson NB, Lenaerts C, Hoehn H, Poot M, Rubin CD, Chen DF, Yang CC, Juch H, Dorn T, Spiegel R, Oral EA, Abid M, Battisti C, Lucci-Cordisco E, Neri G, Steed EH, Kidd A, Isley W, Showalter D, Vittone JL, Konstantinow A, Ring J, Meyer P, Wenger SL, von Herbay A, Wollina U, Schuelke M, Huizenga CR, Leistritz DF, Martin GM, Mian IS, Oshima J. The spectrum of WRN mutations in Werner syndrome patients. Hum Mutat. 2006 Jun;27(6):558-67. 16673358

[135]  Davis T, Singhrao SK, Wyllie FS, Haughton MF, Smith PJ, Wiltshire M, Wynford-Thomas D, Jones CJ, Faragher RG, Kipling D. Telomere-based proliferative lifespan barriers in Werner-syndrome fibroblasts involve both p53-dependent and p53-independent mechanisms. J Cell Sci. 2003 Apr 1;116(Pt 7):1349-57. 12615976

[136]  Epstein CJ, Martin GM, Schultz AL, Motulsky AG. Werner's syndrome a review of its symptomatology, natural history, pathologic features, genetics and relationship to the natural aging process. Medicine (Baltimore). 1966 May;45(3):177-221. 5327241

[137]  Faragher RG, Kill IR, Hunter JA, Pope FM, Tannock C, Shall S. The gene responsible for Werner syndrome may be a cell division "counting" gene. Proc Natl Acad Sci U S A. 1993 Dec 15;90(24):12030-4. 8265666

[138]  Schulz VP, Zakian VA, Ogburn CE, McKay J, Jarzebowicz AA, Edland SD, Martin GM. Accelerated loss of telomeric repeats may not explain accelerated replicative decline of Werner syndrome cells. Hum Genet. 1996 Jun;97(6):750-4. 8641691

[139]  Tahara H, Tokutake Y, Maeda S, Kataoka H, Watanabe T, Satoh M, Matsumoto T, Sugawara M, Ide T, Goto M, Furuichi Y, Sugimoto M. Abnormal telomere dynamics of B-lymphoblastoid cell strains from Werner's syndrome patients transformed by Epstein-Barr virus. Oncogene. 1997 Oct 16;15(16):1911-20. 9365237

[140]  Sandell LL, Zakian VA. Loss of a yeast telomere: arrest, recovery, and chromosome loss. Cell. 1993 Nov 19;75(4):729-39. 8242745

[141]  Nurnberg P, Thiel G, Weber F, Epplen JT. Changes of telomere lengths in human intracranial tumours. Hum Genet. 1993 Mar;91(2):190-2. 8462979

[142]  Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H, de Lange T. Mammalian telomeres end in a large duplex loop. Cell. 1999 May 14;97(4):503-14. 10338214

[143]  Murti KG, Prescott DM. Telomeres of polytene chromosomes in a ciliated protozoan terminate in duplex DNA loops. Proc Natl Acad Sci U S A. 1999 Dec 7;96(25):14436-9. 10588723

[144]  Williamson JR, Raghuraman MK, Cech TR. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. Cell. 1989 Dec 1;59(5):871-80. 2590943

[145] Crabbe L, Verdun RE, Haggblom CI, Karlseder J. Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. Science. 2004 Dec 10;306(5703): 1951-3. 15591207

[146] Hyun M, Bohr VA, Ahn B. Biochemical characterization of the WRN-1 RecQ helicase of Caenorhabditis elegans. Biochemistry. 2008 Jul 15;47(28):7583-93. 18558712

[147] Constantinou A, Tarsounas M, Karow JK, Brosh RM, Bohr VA, Hickson ID, West SC. Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest. EMBO Rep. 2000 Jul;1(1):80-4. 11256630

[148] Cooper MP, Machwe A, Orren DK, Brosh RM, Ramsden D, Bohr VA. Ku complex interacts with and stimulates the Werner protein. Genes Dev. 2000 Apr 15;14(8):907-12. 10783163

[149] Opresko PL, Otterlei M, Graakjaer J, Bruheim P, Dawut L, Kolvraa S, May A, Seidman MM, Bohr VA. The Werner syndrome helicase and exonuclease cooperate to resolve telomeric D loops in a manner regulated by TRF1 and TRF2. Mol Cell. 2004 Jun 18;14(6):763-74. 15200954

[150] Bhattacharyya S, Sandy A, Groden J. Unwinding protein complexes in ALTernative telomere maintenance. J Cell Biochem. 2010 2010;109(1):7-15. 19911388

[151] Huang S, Li B, Gray MD, Oshima J, Mian IS, Campisi J. The premature ageing syndrome protein, WRN, is a 3'-->5' exonuclease. Nat Genet. 1998 Oct;20(2):114-6. 9771700

[152] Kamath-Loeb AS, Johansson E, Burgers PM, Loeb LA. Functional interaction between the Werner Syndrome protein and DNA polymerase delta. Proc Natl Acad Sci U S A. 2000 Apr 25;97(9):4603-8. 10781066

[153] Shen JC, Gray MD, Oshima J, Kamath-Loeb AS, Fry M, Loeb LA. Werner syndrome protein. I. DNA helicase and dna exonuclease reside on the same polypeptide. J Biol Chem. 1998 Dec 18;273(51):34139-44. 9852073

[154] Huang S, Beresten S, Li B, Oshima J, Ellis NA, Campisi J. Characterization of the human and mouse WRN 3'-->5' exonuclease. Nucleic Acids Res. 2000 Jun 15;28(12): 2396-405. 10871373

[155] Kamath-Loeb AS, Shen JC, Loeb LA, Fry M. Werner syndrome protein. II. Characterization of the integral 3' --> 5' DNA exonuclease. J Biol Chem. 1998 Dec 18;273(51): 34145-50. 9852074

[156] Shen JC, Loeb LA. Werner syndrome exonuclease catalyzes structure-dependent degradation of DNA. Nucleic Acids Res. 2000 Sep 1;28(17):3260-8. 10954593

[157] Lebel M, Spillare EA, Harris CC, Leder P. The Werner syndrome gene product copurifies with the DNA replication complex and interacts with PCNA and topoisomerase I. J Biol Chem. 1999 Dec 31;274(53):37795-9. 10608841

[158]  Prince PR, Emond MJ, Monnat RJ, Jr. Loss of Werner syndrome protein function promotes aberrant mitotic recombination. Genes Dev. 2001 Apr 15;15(8):933-8. 11316787

[159]  Bischof O, Kim SH, Irving J, Beresten S, Ellis NA, Campisi J. Regulation and localization of the Bloom syndrome protein in response to DNA damage. J Cell Biol. 2001 Apr 16;153(2):367-80. 11309417

[160]  Dutertre S, Ababou M, Onclercq R, Delic J, Chatton B, Jaulin C, Amor-Gueret M. Cell cycle regulation of the endogenous wild type Bloom's syndrome DNA helicase. Oncogene. 2000 May 25;19(23):2731-8. 10851073

[161]  Sanz MM, Proytcheva M, Ellis NA, Holloman WK, German J. BLM, the Bloom's syndrome protein, varies during the cell cycle in its amount, distribution, and co-localization with other nuclear proteins. Cytogenet Cell Genet. 2000;91(1-4):217-23. 11173860

[162]  Zhong S, Hu P, Ye TZ, Stan R, Ellis NA, Pandolfi PP. A role for PML and the nuclear body in genomic stability. Oncogene. 1999 Dec 23;18(56):7941-7. 10637504

[163]  Dellaire G, Ching RW, Ahmed K, Jalali F, Tse KC, Bristow RG, Bazett-Jones DP. Promyelocytic leukemia nuclear bodies behave as DNA damage sensors whose response to DNA double-strand breaks is regulated by NBS1 and the kinases ATM, Chk2, and ATR. J Cell Biol. 2006 Oct 9;175(1):55-66. 17030982

[164]  Matunis MJ, Zhang XD, Ellis NA. SUMO: the glue that binds. Dev Cell. 2006 Nov; 11(5):596-7. 17084352

[165]  Johnson ES. Protein modification by SUMO. Annu Rev Biochem. 2004;73:355-82. 15189146

[166]  Shen TH, Lin HK, Scaglioni PP, Yung TM, Pandolfi PP. The mechanisms of PML-nuclear body formation. Mol Cell. 2006 Nov 3;24(3):331-9. 17081985

[167]  Eladad S, Ye TZ, Hu P, Leversha M, Beresten S, Matunis MJ, Ellis NA. Intra-nuclear trafficking of the BLM helicase to DNA damage-induced foci is regulated by SUMO modification. Hum Mol Genet. 2005 May 15;14(10):1351-65. 15829507

[168]  Hu P, Beresten SF, van Brabant AJ, Ye TZ, Pandolfi PP, Johnson FB, Guarente L, Ellis NA. Evidence for BLM and Topoisomerase IIIalpha interaction in genomic stability. Hum Mol Genet. 2001 Jun 1;10(12):1287-98. 11406610

[169]  Paull TT, Rogakou EP, Yamazaki V, Kirchgessner CU, Gellert M, Bonner WM. A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. Curr Biol. 2000 Jul 27-Aug 10;10(15):886-95. 10959836

[170]  Pilch DR, Sedelnikova OA, Redon C, Celeste A, Nussenzweig A, Bonner WM. Characteristics of gamma-H2AX foci at DNA double-strand breaks sites. Biochem Cell Biol. 2003 Jun;81(3):123-9. 12897845

[171]  Rogakou EP, Boon C, Redon C, Bonner WM. Megabase chromatin domains involved in DNA double-strand breaks in vivo. J Cell Biol. 1999 Sep 6;146(5):905-16. 10477747

[172] Bassing CH, Alt FW. H2AX may function as an anchor to hold broken chromosomal DNA ends in close proximity. Cell Cycle. 2004 Feb;3(2):149-53. 14712078

[173] Davalos AR, Campisi J. Bloom syndrome cells undergo p53-dependent apoptosis and delayed assembly of BRCA1 and NBS1 repair complexes at stalled replication forks. J Cell Biol. 2003 Sep 29;162(7):1197-209. 14517203

[174] Beamish H, Kedar P, Kaneko H, Chen P, Fukao T, Peng C, Beresten S, Gueven N, Purdie D, Lees-Miller S, Ellis N, Kondo N, Lavin MF. Functional link between BLM defective in Bloom's syndrome and the ataxia-telangiectasia-mutated protein, ATM. J Biol Chem. 2002 Aug 23;277(34):30515-23. 12034743

[175] Franchitto A, Pichierri P. Bloom's syndrome protein is required for correct relocalization of RAD50/MRE11/NBS1 complex after replication fork arrest. J Cell Biol. 2002 Apr 1;157(1):19-30. 11916980

[176] Raderschall E, Golub EI, Haaf T. Nuclear foci of mammalian recombination proteins are located at single-stranded DNA regions formed after DNA damage. Proc Natl Acad Sci U S A. 1999 Mar 2;96(5):1921-6. 10051570

[177] Adams MD, McVey M, Sekelsky JJ. Drosophila BLM in double-strand break repair by synthesis-dependent strand annealing. Science. 2003 Jan 10;299(5604):265-7. 12522255

[178] Plank JL, Wu J, Hsieh TS. Topoisomerase IIIalpha and Bloom's helicase can resolve a mobile double Holliday junction substrate through convergent branch migration. Proc Natl Acad Sci U S A. 2006 Jul 25;103(30):11118-23. 16849422

[179] Langland G, Kordich J, Creaney J, Goss KH, Lillard-Wetherell K, Bebenek K, Kunkel TA, Groden J. The Bloom's syndrome protein (BLM) interacts with MLH1 but is not required for DNA mismatch repair. J Biol Chem. 2001 Aug 10;276(32):30031-5. 11325959

[180] Pedrazzi G, Perrera C, Blaser H, Kuster P, Marra G, Davies SL, Ryu GH, Freire R, Hickson ID, Jiricny J, Stagljar I. Direct association of Bloom's syndrome gene product with the human mismatch repair protein MLH1. Nucleic Acids Res. 2001 Nov 1;29(21):4378-86. 11691925

[181] Meetei AR, Sechi S, Wallisch M, Yang D, Young MK, Joenje H, Hoatlin ME, Wang W. A multiprotein nuclear complex connects Fanconi anemia and Bloom syndrome. Mol Cell Biol. 2003 May;23(10):3417-26. 12724401

[182] Yin J, Sobeck A, Xu C, Meetei AR, Hoatlin M, Li L, Wang W. BLAP75, an essential component of Bloom's syndrome protein complexes that maintain genome integrity. Embo J. 2005 Apr 6;24(7):1465-76. 15775963

[183] Machwe A, Xiao L, Lloyd RG, Bolt E, Orren DK. Replication fork regression in vitro by the Werner syndrome protein (WRN): holliday junction formation, the effect of leading arm structure and a potential role for WRN exonuclease activity. Nucleic Acids Res. 2007;35(17):5729-47. 17717003

[184]  Crabbe L, Jauch A, Naeger CM, Holtgreve-Grez H, Karlseder J. Telomere dysfunction as a cause of genomic instability in Werner syndrome. Proc Natl Acad Sci U S A. 2007 Feb 13;104(7):2205-10. 17284601

[185]  de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev. 2005 Sep 15;19(18):2100-10. 16166375

[186]  Opresko PL, von Kobbe C, Laine JP, Harrigan J, Hickson ID, Bohr VA. Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases. J Biol Chem. 2002 Oct 25;277(43):41110-9. 12181313

[187]  Opresko PL, Mason PA, Podell ER, Lei M, Hickson ID, Cech TR, Bohr VA. POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. J Biol Chem. 2005 Sep 16;280(37):32069-80. 16030011

[188]  Kao HI, Campbell JL, Bambara RA. Dna2p helicase/nuclease is a tracking protein, like FEN1, for flap cleavage during Okazaki fragment maturation. J Biol Chem. 2004 Dec 3;279(49):50840-9. 15448135

[189]  Sharma S, Sommers JA, Brosh RM, Jr. In vivo function of the conserved non-catalytic domain of Werner syndrome helicase in DNA replication. Hum Mol Genet. 2004 Oct 1;13(19):2247-61. 15282207

[190]  Szekely AM, Chen YH, Zhang C, Oshima J, Weissman SM. Werner protein recruits DNA polymerase delta to the nucleolus. Proc Natl Acad Sci U S A. 2000 Oct 10;97(21):11365-70. 11027336

[191]  Kamath-Loeb AS, Loeb LA, Johansson E, Burgers PM, Fry M. Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)n trinucleotide repeat sequence. J Biol Chem. 2001 May 11;276(19):16439-46. 11279038

[192]  Fry M, Loeb LA. Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. J Biol Chem. 1999 Apr 30;274(18):12797-802. 10212265

[193]  Riha K, Heacock ML, Shippen DE. The role of the nonhomologous end-joining DNA double-strand break repair pathway in telomere biology. Annu Rev Genet. 2006;40:237-77. 16822175

[194]  Cheng WH, von Kobbe C, Opresko PL, Arthur LM, Komatsu K, Seidman MM, Carney JP, Bohr VA. Linkage between Werner syndrome protein and the Mre11 complex via Nbs1. J Biol Chem. 2004 May 14;279(20):21169-76. 15026416

[195]  Celli GB, Denchi EL, de Lange T. Ku70 stimulates fusion of dysfunctional telomeres yet protects chromosome ends from homologous recombination. Nat Cell Biol. 2006 Aug;8(8):885-90. 16845382

[196]  Laud PR, Multani AS, Bailey SM, Wu L, Ma J, Kingsley C, Lebel M, Pathak S, DePinho RA, Chang S. Elevated telomere-telomere recombination in WRN-deficient, telo-

mere dysfunctional cells promotes escape from senescence and engagement of the ALT pathway. Genes Dev. 2005 Nov 1;19(21):2560-70. 16264192

[197] Harrigan JA, Wilson DM, 3rd, Prasad R, Opresko PL, Beck G, May A, Wilson SH, Bohr VA. The Werner syndrome protein operates in base excision repair and cooperates with DNA polymerase beta. Nucleic Acids Res. 2006;34(2):745-54. 16449207

[198] Blank A, Bobola MS, Gold B, Varadarajan S, D DK, Meade EH, Rabinovitch PS, Loeb LA, Silber JR. The Werner syndrome protein confers resistance to the DNA lesions N3-methyladenine and O6-methylguanine: implications for WRN function. DNA Repair (Amst). 2004 Jun 3;3(6):629-38. 15135730

[199] Harrigan JA, Opresko PL, von Kobbe C, Kedar PS, Prasad R, Wilson SH, Bohr VA. The Werner syndrome protein stimulates DNA polymerase beta strand displacement synthesis via its helicase activity. J Biol Chem. 2003 Jun 20;278(25):22686-95. 12665521

[200] Li B, Navarro S, Kasahara N, Comai L. Identification and biochemical characterization of a Werner's syndrome protein complex with Ku70/80 and poly(ADP-ribose) polymerase-1. J Biol Chem. 2004 Apr 2;279(14):13659-67. 14734561

[201] von Kobbe C, Harrigan JA, Schreiber V, Stiegler P, Piotrowski J, Dawut L, Bohr VA. Poly(ADP-ribose) polymerase 1 regulates both the exonuclease and helicase activities of the Werner syndrome protein. Nucleic Acids Res. 2004;32(13):4003-14. 15292449

[202] Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, Kyle S, Meuth M, Curtin NJ, Helleday T. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature. 2005 Apr 14;434(7035):913-7. 15829966

[203] Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C, Martin NM, Jackson SP, Smith GC, Ashworth A. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature. 2005 Apr 14;434(7035):917-21. 15829967

[204] Bohr VA, Souza Pinto N, Nyaga SG, Dianov G, Kraemer K, Seidman MM, Brosh RM, Jr. DNA repair and mutagenesis in Werner syndrome. Environ Mol Mutagen. 2001;38(2-3):227-34. 11746759

[205] Yun J, Zhong Q, Kwak JY, Lee WH. Hypersensitivity of Brca1-deficient MEF to the DNA interstrand crosslinking agent mitomycin C is associated with defect in homologous recombination repair and aberrant S-phase arrest. Oncogene. 2005 Jun 9;24(25):4009-16. 15782115

[206] Cheng WH, Kusumoto R, Opresko PL, Sui X, Huang S, Nicolette ML, Paull TT, Campisi J, Seidman M, Bohr VA. Collaboration of Werner syndrome protein and BRCA1 in cellular responses to DNA interstrand cross-links. Nucleic Acids Res. 2006;34(9):2751-60. 16714450

[207] Mann MB, Hodges CA, Barnes E, Vogel H, Hassold TJ, Luo G. Defective sister-chromatid cohesion, aneuploidy and cancer predisposition in a mouse model of type II Rothmund-Thomson syndrome. Hum Mol Genet. 2005 Mar 15;14(6):813-25. 15703196

[208] Sangrithi MN, Bernal JA, Madine M, Philpott A, Lee J, Dunphy WG, Venkitaraman AR. Initiation of DNA replication requires the RECQL4 protein mutated in Rothmund-Thomson syndrome. Cell. 2005 Jun 17;121(6):887-98. 15960976

[209] Matsuno K, Kumano M, Kubota Y, Hashimoto Y, Takisawa H. The N-terminal non-catalytic region of Xenopus RecQ4 is required for chromatin binding of DNA polymerase alpha in the initiation of DNA replication. Mol Cell Biol. 2006 Jul;26(13): 4843-52. 16782873

[210] Im JS, Ki SH, Farina A, Jung DS, Hurwitz J, Lee JK. Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins. Proc Natl Acad Sci U S A. 2009 Sep 15;106(37):15628-32. 19805216

[211] Kumata Y, Tada S, Yamanada Y, Tsuyama T, Kobayashi T, Dong YP, Ikegami K, Murofushi H, Seki M, Enomoto T. Possible involvement of RecQL4 in the repair of double-strand DNA breaks in Xenopus egg extracts. Biochim Biophys Acta. 2007 Apr;1773(4):556-64. 17320201

[212] Petkovic M, Dietschy T, Freire R, Jiao R, Stagljar I. The human Rothmund-Thomson syndrome gene product, RECQL4, localizes to distinct nuclear foci that coincide with proteins involved in the maintenance of genome stability. J Cell Sci. 2005 Sep 15;118(Pt 18):4261-9. 16141230

[213] Woo LL, Futami K, Shimamoto A, Furuichi Y, Frank KM. The Rothmund-Thomson gene product RECQL4 localizes to the nucleolus in response to oxidative stress. Exp Cell Res. 2006 Oct 15;312(17):3443-57. 16949575

[214] Werner SR, Prahalad AK, Yang J, Hock JM. RECQL4-deficient cells are hypersensitive to oxidative stress/damage: Insights for osteosarcoma prevalence and heterogeneity in Rothmund-Thomson syndrome. Biochem Biophys Res Commun. 2006 Jun 23;345(1):403-9. 16678792

[215] Wang M, Wu W, Wu W, Rosidi B, Zhang L, Wang H, Iliakis G. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Res. 2006;34(21):6170-82. 17088286

[216] Audebert M, Salles B, Calsou P. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. J Biol Chem. 2004 Dec 31;279(53):55117-26. 15498778

[217] Bailis JM, Luche DD, Hunter T, Forsburg SL. Minichromosome maintenance proteins interact with checkpoint and recombination proteins to promote s-phase genome stability. Mol Cell Biol. 2008 Mar;28(5):1724-38. 18180284

[218] Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, Labib K. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat Cell Biol. 2006 Apr;8(4): 358-66. 16531994

[219] Katou Y, Kanoh Y, Bando M, Noguchi H, Tanaka H, Ashikari T, Sugimoto K, Shirahige K. S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. Nature. 2003 Aug 28;424(6952):1078-83. 12944972

[220] Calzada A, Hodgson B, Kanemaki M, Bueno A, Labib K. Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. Genes Dev. 2005 Aug 15;19(16):1905-19. 16103218

[221] Komata M, Bando M, Araki H, Shirahige K. 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. 2009 Sep;29(18):5008-19. 19620285

[222] Nedelcheva MN, Roguev A, Dolapchiev LB, Shevchenko A, Taskov HB, Shevchenko A, Stewart AF, Stoynov SS. Uncoupling of unwinding from DNA synthesis implies regulation of MCM helicase by Tof1/Mrc1/Csm3 checkpoint complex. J Mol Biol. 2005 Apr 1;347(3):509-21. 15755447

[223] Lou H, Komata M, Katou Y, Guan Z, Reis CC, Budd M, Shirahige K, Campbell JL. Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint. Mol Cell. 2008 Oct 10;32(1):106-17. 18851837

[224] Szyjka SJ, Viggiani CJ, Aparicio OM. Mrc1 is required for normal progression of replication forks throughout chromatin in S. cerevisiae. Mol Cell. 2005 Sep 2;19(5):691-7. 16137624

[225] Alcasabas AA, Osborn AJ, Bachant J, Hu F, Werler PJ, Bousset K, Furuya K, Diffley JF, Carr AM, Elledge SJ. Mrc1 transduces signals of DNA replication stress to activate Rad53. Nat Cell Biol. 2001 Nov;3(11):958-65. 11715016

[226] Bjergbaek L, Cobb JA, Tsai-Pflugfelder M, Gasser SM. Mechanistically distinct roles for Sgs1p in checkpoint activation and replication fork maintenance. Embo J. 2005 Jan 26;24(2):405-17. 15616582

[227] Bando M, Katou Y, Komata M, Tanaka H, Itoh T, Sutani T, Shirahige K. Csm3, Tof1, and Mrc1 form a heterotrimeric mediator complex that associates with DNA replication forks. J Biol Chem. 2009 Dec 4;284(49):34355-65. 19819872

[228] Pryce DW, Ramayah S, Jaendling A, McFarlane RJ. Recombination at DNA replication fork barriers is not universal and is differentially regulated by Swi1. Proc Natl Acad Sci U S A. 2009 Mar 24;106(12):4770-5. 19273851

# DNA Replication in Archaea, the Third Domain of Life

Yoshizumi Ishino and Sonoko Ishino

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/53986

## 1. Introduction

The accurate duplication and transmission of genetic information are essential and crucially important for living organisms. The molecular mechanism of DNA replication has been one of the central themes of molecular biology, and continuous efforts to elucidate the precise molecular mechanism of DNA replication have been made since the discovery of the double helix DNA structure in 1953 [1]. The protein factors that function in the DNA replication process, have been identified to date in the three domains of life (Figure 1).
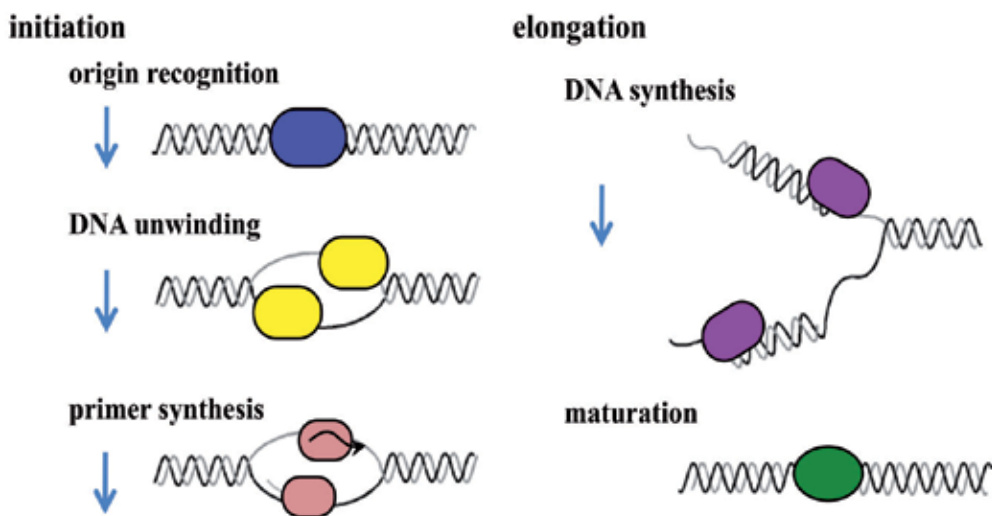


**Figure 1.** Stage of DNA replication

|  | Archaea | Eukaryota | Bacteria |
|---|---|---|---|
| initiation |  |  |  |
|     origin recognition | Cdc6/Orc1 | ORC | DnaA |
|   DNA unwinding | Cdc6/Orc1<br>MCM<br>GINS | Cdc6<br>Cdt1<br>MCM<br>GINS<br>Cdc45 | DnaC<br>DnaB |
|     primer synthesis | DNA primase | Pol α / primase | DnaG |
| elongation |  |  |  |
|   DNA synthesis | family B<br>DNA polymerase<br>(Pol B)<br>family D<br>DNA polymerase<br>(Pol D) | family B<br>DNA polymerase<br>(Pol δ) (Pol ε) | family C<br>DNA polymerase<br>(Pol III) |
|  | clamp loader<br>(RFC)<br>clamp<br>(PCNA) | clamp loader<br>(RFC)<br>clamp<br>(PCNA) | clamp loader<br>(γ-complex)<br>clamp<br>(β-clamp) |
|   maturation | Fen1<br>Dna2<br>DNA ligase | FEN1<br>DNA2<br>DNA ligase | Pol I<br>RNaseH<br>DNA ligase |

**Table 1.** The proteins involved in DNA replication from the three domains of life

Archaea, the third domain of life, is a very interesting living organism to study from the aspects of molecular and evolutionary biology. Rapid progress of whole genome sequence analyses has allowed us to perform comparative genomic studies. In addition, recent microbial ecology has revealed that archaeal organisms inhabit not only extreme environments, but also more ordinary habitats. In these situations, archaeal biology is among the most exciting of research fields. Archaeal cells have a unicellular ultrastructure without a nucleus, resembling bacterial cells, but the proteins involved in the genetic information processing pathways, including DNA replication, transcription, and translation, share strong similarities with those of eukaryotes. Therefore, most of the archaeal proteins were identified as homologues of many eukaryotic replication proteins, including ORC (origin recognition complex), Cdc6, GINS (Sld5-Psf1-Psf2-Psf3), MCM (minichromosome maintenance), RPA (replication protein A), PCNA (proliferating cell nuclear antigen), RFC (replication factor C), FEN1 (flap endonuclease 1), in addition to the eukaryotic primase, DNA polymerase, and DNA ligase; these are obviously different from bacterial proteins (Table 1) and these proteins were biochemically characterized [2-4]. Their similarities indicate that the DNA replication machi-

neries of Archaea and Eukaryota evolved from a common ancestor, which was different from that of Bacteria [5]. Therefore, the archaeal organisms are good models to elucidate the functions of each component of the eukaryotic type replication machinery complex. Genomic and comparative genomic research with archaea is made easier by the fact that the genome size and the number of genes of archaea are much smaller than those of eukaryotes. The archaeal replication machinery is probably a simplified form of that in eukaryotes. On the other hand, it is also interesting that the circular genome structure is conserved in Bacteria and Archaea and is different from the linear form of eukaryotic genomes. These features have encouraged us to study archaeal DNA replication, in the hopes of gaining fundamental insights into this molecular mechanism and its machinery from an evolutional perspective. The study of bacterial DNA replication at a molecular level started in about 1960, and then eukaryotic studies followed since 1980. Because Archaea was recognized as the third domain of life later, the archaeal DNA replication research became active after 1990. With increasing the available total genome sequences, the progress of research on archaeal DNA replication has been rapid, and the depth of our knowledge of archaeal DNA replication has almost caught up with those of the bacterial and eukaryotic research fields. In this chapter, we will summarize the current knowledge of DNA replication in Archaea.

## 2. Replication origin

The basic mechanism of DNA replication was predicted as "replicon theory" by Jacob et al. [6]. They proposed that an initiation factor recognizes the replicator, now referred to as a replication origin, to start replication of the chromosomal DNA. Then, the replication origin of *E. coli* DNA was identified as *oriC* (origin of chromosome). The archaeal replication origin was identified in the *Pyrococcus abyssi* in 2001 as the first archaeal replication origin. The origin was located just upstream of the gene encoding the Cdc6 and Orc1-like sequences in the *Pyrococcus* genome [7]. We discovered a gene encoding an amino acid sequence that bore similarity to those of both eukaryotic Cdc6 and Orc1, which are the eukaryotic initiators. After confirming that this protein actually binds to the *oriC* region on the chromosomal DNA we named the gene product Cdc6/Orc1 due to its roughly equal homology with regions of eukaryotic Orc1 and Cdc6, [7]. The gene consists of an operon with the gene encoding DNA polymerase D (it was originally called Pol II, as the second DNA polymerase from *Pyrococcus furiosus*) in the genome [8]. A characteristic of the *oriC* is the conserved 13 bp repeats, as predicted earlier by bioinformatics [9], and two of the repeats are longer and surround a predicted DUE (DNA unwinding element) with an AT-rich sequence in *Pyrococcus* genomes (Figure 2) [10]. The longer repeated sequence was designated as an ORB (Origin Recognition Box), and it was actually recognized by Cdc6/Orc1 in a *Sulfolobus solfataricus* study [11]. The 13 base repeat is called a miniORB, as a minimal version of ORB. A whole genome microarray analysis of *P. abyssi* showed that the Cdc6/Orc1 binds to the *oriC* region with extreme specificity, and the specific binding of the highly purified *P. furiosus* Cdc6/Orc1 to ORB and miniORB was confirmed *in vitro* [12]. It has to be noted that multiple origins were identified in the *Sulfolobus* genomes. It is now well recognized that *Sulfolobus* has three origins and

they work at the same time in the cell cycle [11, 13-16]. Analysis of the mechanism of how the multiple origins are utilized for genome replication is an interesting subject in the research field of archaeal DNA replication. The main questions are how the initiation of replication from multiple origins is regulated and how the replication forks progress after the collision of two forks from opposite directions.
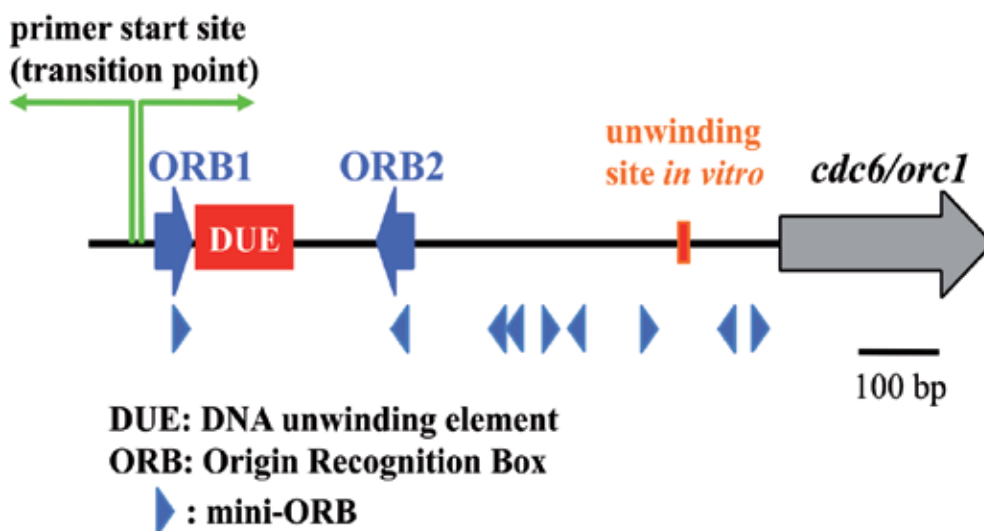


**Figure 2. The *oriC* region in *Pyrococcus* genome.** The region surrounding *oriC* is presented schematically. The ORB1 and ORB2 are indicated by large arrow, and the mini-ORB repeats are indicated by small arrowheads. DUE is indicated in red. The unwinding site, determined by *in vitro* analysis, is indicated in orange. The transition site is indicated by green arrows. The *cdc6/orc1* gene located in downstream is drawn by gray arrow.

## 3. How does Cdc6/Orc1 recognize *oriC*?

An important step in characterizing the initiation of DNA replication in Archaea is to understand how the Cdc6/Orc1 protein recognizes the *oriC* region. Based upon amino acid sequence alignments, the archaeal Cdc6/Orc1 proteins belong to the AAA+ family of proteins. The crystal structures of the Cdc6/Orc1 protein from *Pyrobaculum aerophilum* [17] and one of the two Cdc6/Orc1 proteins, ORC2 from *Aeropyrum pernix* (the two homologs in this organism are called ORC1 and ORC2 by the authors) [18] were determined. These Cdc6/Orc1 proteins consist of three structural domains. Domains I and II adopt a fold found in the AAA+ family proteins. A winged helix (WH) fold, which is present in a number of DNA binding proteins, is found in the domain III. There are four ORBs arranged in pairs on both sides of the DUE in the *oriC* region of *A. pernix*, and ORC1 binds to each ORB as a dimer. A mechanism was proposed in which ORC1 binds to all four ORBs to introduce a higher-order assembly for unwinding of the DUE with alterations in both topology and superhelicity [19]. Furthermore, the crystal structures of *S. solfataricus* Cdc6-1 and Cdc6-3 (two of the three

Cdc6/Orc1 proteins in this organism) forming a heterodimer bound to *ori2* DNA (one of the three origins in this organism) [20], and that of *A. pernix* ORC1 bound to an origin sequence [21] were determined. These studies revealed that both the N-terminal AAA+ ATPase domain (domain I+II) and C-terminal WH domain (domain III) contribute to origin DNA binding, and the structural information not only defined the polarity of initiator assembly on the origin but also indicated the induction of substantial distortion, which probably triggers the unwinding of the duplex DNA to start replication, into the DNA strands. These structural data also provided the detailed interaction mode between the initiator protein and the *oriC* DNA. Mutational analyses of the *Methanothermobacter thermautotrophicus* Cdc6-1 protein revealed the essential interaction between an arginine residue conserved in the archaeal Cdc6/Orc1 and an invariant guanine in the ORB sequence [22].

*P. furiosus* Cdc6/Orc1 is difficult to purify in a soluble form. A specific site in the *oriC* to start unwinding *in vitro*, was identified using the protein prepared by a denaturation-renaturation procedure recently [23]. As shown in Figure 2, the local unwinding site is about 670 bp away from the transition site between leading and lagging syntheses, which was determined earlier by an *in vivo* replication initiation point (RIP) assay [10]. Although the details of the replication machinery that must be established at the unwound site are not fully understood in Archaea, it is expected to minimally include MCM, GINS, primase, PCNA, DNA polymerase, and RPA, as described below. The following *P. furiosus* studies revealed that the ATPase activity of the Cdc6/Orc1 protein was completely suppressed by binding to DNA containing the ORB. Limited proteolysis and DNase I-footprint experiments suggested that the Cdc6/Orc1 protein changes its conformation on the ORB sequence in the presence of ATP. The physiological meaning of this conformational change has not been solved, but it should have an important function to start the initiation process [24] as in the case of bacterial DnaA protein. In addition, results from an *in vitro* recruiting assay indicated that MCM (Mcm protein complex), the replicative DNA helicase, is recruited onto the *oriC* region in a Cdc6/Orc1-dependent, but not ATP-dependent, manner [24], as described below. However, this recruitment is not sufficient for the unwinding function of MCM, and some other function remains to be identified for the functional loading of this helicase to promote the progression of the DNA replication fork.

## 4. MCM helicase

After unwinding of the *oriC* region, the replicative helicase needs to remain loaded to provide continuous unwinding of double stranded DNA (dsDNA) as the replication forks progress bidirectionally. The MCM protein complex, consisting of six subunits (Mcm2, 3, 4, 5, 6, and 7), is known to be the replicative helicase "core" in eukaryotic cells [25]. The MCM further interacts with Cdc45 and GINS, to form a ternary assembly referred to as the "CMG complex", that is believed to be the functional helicase in eukaryotic cells (Figure 3) [26]. However, this idea is still not universal for the eukaryotic replicative helicase.
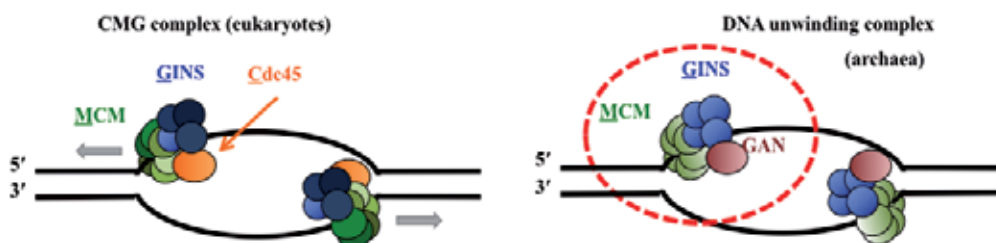
**Figure 3. DNA-Unwinding complex in eukaryotes and archaea.** The CMG complex is the replicative helicase for the template DNA unwinding reaction in eukaryotes. The archaeal genomes contain the homologs of the Mcm and Gins proteins, but a Cdc45 homolog has not been identified. Recent research suggests that a RecJ-like exonuclease GAN, which has weak sequence homology to that of Cdc45, may work as a helicase complex with MCM and GINS.

Most archaeal genomes appear to encode at least one Mcm homologue, and the helicase activities of these proteins from several archaeal organisms have been confirmed *in vitro* [27-31]. In contrast to the eukaryotic MCM, the archaeal MCMs, consist of a homohexamer or homo double hexamer, having distinct DNA helicase activity by themselves *in vitro*, and therefore, these MCMs on their own may function as the replicative helicase *in vivo*. The structure-function relationships of the archaeal Mcms have been aggressively studied using purified proteins and site-directed mutagenesis [32]. An early report using the ChIP method showed that the *P. abyssi* Mcm protein preferentially binds to the origin *in vivo* in exponentially growing cells [7, 12]. The *P. furiosus* MCM helicase does not display significant helicase activity *in vitro*. However, the DNA helicase activity was clearly stimulated by the addition of GINS (the Gins23-Gins51 complex), which is the homolog of the eukaryotic GINS complex (described below in more detail). This result suggests that MCM works with other accessory factors to form a core complex in *P. furiosus* similar to the eukaryotic CMG complex as described above [31].

Some archaeal organisms have more than two Cdc6/Orc1 homologs. It was found that the two Cdc6/Orc1 homologs, Cdc6-1 and Cdc6-2, both inhibit the helicase activity of MCM in *M. thermautotrophicus* [33. 34]. Similarly, Cdc6-1 inhibits MCM activity in *S. solfataricus* [35]. In contrast, the Cdc6-2 protein stimulates the helicase activity of MCM in *Thermoplasma acidophilum* [36]. Functional interactions between Cdc6/Orc1 and Mcm proteins need to be investigated in greater detail to achieve a more comprehensive understanding of the conservation and diversity of the initiation mechanism in archaeal DNA replication.

Another interesting feature of DNA replication initiation is that several archaea have multiple genes encoding Mcm homologs in their genomes. Based on the recent comprehensive genomic analyses, thirteen archaeal species have more than one *mcm* gene. However, many of the *mcm* genes in the archaeal genomes seem to reside within mobile elements, originating from viruses [37]. For example, two of the three genes in the *Thermococcus kodakarensis* genome are located in regions where genetic elements have presumably been integrated [38]. The establishment of a genetic manipulation system for *T. kodakarensis*, is the first for a hyperthermophilic euryarchaeon [39. 40], and is advantageous for investigating the function of these Mcm proteins. Two groups have recently performed gene disruption experiments for each *mcm* gene [41, 42]. These experiments revealed that the knock-out strains for *mcm1*

and *mcm2* were easily isolated, but *mcm3* could not be disrupted. Mcm3 is relatively abundant in the *T. kodakarensis* cells. Furthermore, an *in vitro* experiment using purified Mcm proteins showed that only Mcm3 forms a stable hexameric structure in solution. These results support the contention that Mcm3 is the main helicase core protein in the normal DNA replication process in *T. kodakarensis*.

The functions of the other two Mcm proteins remain to be elucidated. The genes for Mcm1 and Mcm2 are stably inherited, and their gene products may perform some important functions in the DNA metabolism in *T. kodakarensis.* The DNA helicase activity of the recombinant Mcm1 protein is strong *in vitro,* and a distinct amount of the Mcm1 protein is present in *T. kodakarensis* cells. Moreover, Mcm1 functionally interacts with the GINS complex from *T. kodakarensis* [42]*.* These observations strongly suggest that Mcm1 does participate in some aspect of DNA transactions, and may be substituted with Mcm3. Our immunoprecipitation experiments showed that Mcm1 co-precipitated with Mcm3 and GINS, although they did not form a heterohexameric complex [42], suggesting that Mcm1 is involved in the replisome or repairsome and shares some function in *T. kodakarensis* cells. Although western blot analysis could not detect Mcm2 in the extract from exponentially growing *T. kodakarensis* cells [42], a RT-PCR experiment detected the transcript of the *mcm2* gene in the cells (Ishino et al., unpublished). The recombinant Mcm2 protein also has ATPase and helicase activities *in vitro.* [41] Therefore, the *mcm2* gene is expressed under normal growth conditions and may work in some process with a rapid turn over. Further experiments to measure the efficiency of *mcm2* gene transcription by quantitative PCR, as well as to assess the stability of the Mcm2 protein in the cell extract, are needed. Phenotypic analyses investigating the sensitivities of the Δ*mcm1* and Δ*mcm2* mutant strains to DNA damage caused by various mutagens, as reported for other DNA repair-related genes in *T. kodakarensis* [43], may provide a clue to elucidate the functions of these Mcm proteins.

*Methanococcus maripaludis* S2 harbors four *mcm* genes in its genome, three of which seem to be derived from phage, a shotgun proteomics study detected peptides originating from three out of the four *mcm* gene products [44]. Furthermore, the four gene products co-expressed in *E. coli* cells were co-purified in the same fraction [45]. These results suggest that multiple Mcm proteins are functional in the *M. maripaludis* cells.

## 5. Recruitment of Mcm to the *oriC* region

Another important question is how MCM is recruited onto the unwound region of *oriC.* The detailed loading mechanism of the MCM helicase has not been elucidated. It is believed that archaea utilize divergent mechanisms of MCM helicase assembly at the *oriC* [46].

An *in vitro* recruiting assay showed that *P. furiosus* MCM is recruited to the *oriC* DNA in a Cdc6/Orc1-dependent manner [24]. This assay revealed that preloading Cdc6/Orc1 onto the ORB DNA resulted in a clear reduction in MCM recruitment to the *oriC* region, suggesting that free Cdc6/Orc1 is preferable as a helicase recruiter, to associate with MCM and bring it to *oriC*. It would be interesting to understand how the two tasks, origin recognition and

MCM recruiting, are performed by the Cdc6/Orc1 protein, because the WH domain, which primarily recognizes and binds ORB, also has strong affinity for the Mcm protein. The assembly of the Mcm protein onto the ORB DNA by the Walker A-motif mutant of *P. furiosus* Cdc6/Orc1 occurred with the same efficiency as the wild type Cdc6/Orc1. The DNA binding of *P. furiosus* Cdc6/Orc1 was not drastically different in the presence and absence of ATP, as in the case of the initiator proteins from *Archaeoglobus fulgidus* [28], *S. solfataricus* [11], and *A. pernix* [19]. Therefore, it is still not known whether the ATP binding and hydrolysis activity of Cdc6/Orc1 regulates the Mcm protein recruitment onto *oriC* in the cells.

One more important issue is the very low efficiency of the Mcm protein recruitment in the reported *in vitro* assay [24]. Quantification of the recruited Mcm protein by the *in vitro* assay showed that less than one Mcm hexamer was recruited to the ORB. The linear DNA containing ORB1 and ORB2, used in the recruiting assay, may not be suitable to reconstitute the archaeal DNA replication machinery and a template that more closely mimics the chromosomal DNA may be required. Additionally, it may be that as yet unidentified proteins are required to achieve efficient *in vitro* helicase loading in the *P. furiosus* cells. Finally, it will ultimately be necessary to construct a more defined *in vitro* replication system to analyze the regulatory functions of Cdc6/Orc1 precisely during replication initiation.

In *M. thermautotrophicus*, the Cdc6-2 proteins can dissociate the Mcm multimers [47]. The activity of Cdc6-2 might be required as the MCM helicase loader in this organism. The interaction between Cdc6/Orc1 and Mcm is probably general. However, the effect of Cdc6/Orc1 on the MCM helicase activity differs among various organisms, as described above. Some other protein factors may function in various archaea, for example a protein that is distantly related to eukaryotic Cdt1, which plays a crucial role during MCM loading in Eukaryota, exists in some archaeal organisms, although its function has not been characterized yet [14].

## 6. GINS

The eukaryotic GINS complex was originally identified in *Saccharomyces cerevisiae* as essential protein factor for the initiation of DNA replication [48]. GINS consists of four different proteins, Sld5, Psf1, Psf2, and Psf3 (therefore, GINS is an acronym for Japanese go-ichi-ni-san, meaning 5-1-2-3, after these four subunits). The amino acid sequences of the four subunits in the GINS complex share some conservation, suggesting that they are ancestral paralogs [49]. However, most of the archaeal genomes have only one gene encoding this family protein, and more interestingly, the Crenarchaeota and Euryarchaeota (the two major subdomains of Archaea) characteristically have two genes with sequences similar to Psf2 and Psf3, and Sld5 and Psf1, respectively referred to as Gins23 and Gins51 [31, 49]. A Gins homolog, designated as Gins23, was biochemically detected in *S. solfataricus* as the first Gins protein in Archaea, in a yeast two-hybrid screening for interaction partners of the Mcm protein, and another subunit, designated as Gins15, was identified by mass-spectrometry analysis of an immunoaffinity-purified native GINS from an *S. solfataricus* cell extract. [50]. The *S. solfataricus* GINS, composed of two proteins, Gins23 and Gins15, forms a tetrameric struc-

ture with a 2:2 molar ratio [50]. The GINS from *P. furiosus*, a complex of Gins23 and Gins51 with a 2:2 ratio, was identified as the first euryarchaeal GINS [31]. Gins51 was preferred over Gins15 because of the order of the name of GINS.

The MCM2-7 hexamer was copurified in complex with Cdc45 and GINS from *Drosophila melanogaster* embryo extracts and *S. cerevisiae* lysates, and the "CMG (Cdc45-MCM2-7-GINS) complex" (Figure 3), as described above, should be important for the function of the replicative helicase. The CMG complex was also associated with the replication fork in *Xenopus laevis* egg extracts, and a large molecular machine, containing Cdc45, GINS, and MCM2-7, was proposed as the unwindosome to separate the DNA strands at the replication fork [51]. Therefore, GINS must be a critical factor for not only the initiation process, but also the elongation process in eukaryotic DNA replication. *S. solfataricus* GINS interacts with MCM and primase, suggesting that GINS is involved in the replisome. The concrete function of GINS in the replisome remains to be determined. No stimulation or inhibition of either the helicase or primase activity was observed by the interaction with *S. solfataricus* GINS *in vitro* [50]. On the other hand, the DNA helicase activity of *P. furiosus* MCM is clearly stimulated by the addition of the *P. furiosus* GINS complex, as described above [31].

In contrast to *S. solfataricus* and *P. furiosus*, which each express a Gins23 and Gins51, *Thermoplasma acidophilum* has a single Gins homolog, Gins51. The recombinant Gins51 protein from *T. acidophilum* was confirmed to form a homotetramer by gel filtration and electron microscopy analyses. Furthermore, a physical interaction between *T. acidophilum* Gins51 and Mcm was detected by a surface plasmon resonance analysis (SPR). Although the *T. acidophilum* Gins51 did not affect the helicase activity of its cognate MCM, when the equal ratio of each molecule was tested *in vitro* [52], an excess amount of Gins51 clearly stimulated the helicase activity (Ogino et al., unpublished). In the case of *T. kodakarensis*, the ATPase and helicase activities of MCM1 and MCM3 were clearly stimulated by *T. kodakarensis* GINS *in vitro*. It is interesting that the helicase activity of MCM1 was stimulated more than that of MCM3. Physical interactions between the *T. kodakarensis* Gins and Mcm proteins were also detected [53]. These reports suggested that the MCM-GINS complex is a common part of the replicative helicase in Archaea (Figure 3).

Recently, the crystal structure of the *T. kodakarensis* GINS tetramer, composed of Gins51 and Gins23 was determined, and the structure was conserved with the reported human GINS structures [53]. Each subunit of human GINS shares a similar fold, and assembles into the heterotetramer of a unique trapezoidal shape [54-56]. Sld5 and Psf1 possess the α-helical (A) domain at the N-terminus and the β-stranded domain (B) at the C-terminus (AB-type). On the other hand, Psf2 and Psf3 are the permuted version (BA-type). The backbone structure of each subunit and the tetrameric assembly of *T. kodakarensis* GINS are similar to those of human GINS. However, the location of the C-terminal B domain of Gins51 is remarkably different between the two GINS structures [53]. A homology model of the homotetrameric GINS from *T. acidophilum* was performed using the *T. kodakarensis* GINS crystal structure as a template. The Gins 51 protein has a long disordered region inserted between the A and B domains and this allows the conformation of the C-terminal domains to be more flexible.

This domain arrangement leads to the formation of an asymmetric homotetramer, rather than a symmetrical assembly, of the *T. kodakarensis* GINS [53].

The Cdc45 protein is ubiquitously distributed from yeast to human, supporting the notion that the formation of the CMG complex is universal in the eukaryotic DNA replication process. However, no archaeal homologue of Cdc45 has been identified. A recent report of bioinformatic analysis showed that the primary structure of eukaryotic Cdc45 and prokaryotic RecJ share a common ancestry [57]. Indeed, a homolog of the DNA binding domain of RecJ has been co-purified with GINS from *S. solfataricus* [50]. Our experiment detected the stimulation of the 5′-3′ exonuclease activity of the RecJ homologs from *P. furiosus* and *T. kodakarensis* by the cognate GINS complexes (Ishino et al., unpublished). The RecJ homolog from *T. kodakarensis* forms a stable complex with the GINS, and the 5′-3′ exonuclease activity is enhanced *in vitro*; therefore, the RecJ homolog was designated as GAN, from GINS-Associated Nuclease in a very recent paper [58]. Another related report found that the human Cdc45 structure obtained by the small angle X-ray scattering analysis (SAXS) is consistent with the crystallographic structure of the RecJ family members [59]. These current findings will promote further research on the structures and functions of the higher-order unwindosome in archaeal and eukaryotic cells (Figure 3).

# 7. Primase

To initiate DNA strand synthesis, a primase is required for the synthesis of a short oligonucleotide, as a primer. The DnaG and p48-p58 proteins are the primases in Bacteria and Eukaryota, respectively. The p48-p58 primase is further complexed with p180 and p70, to form DNA polymerase α-primase complex. The catalytic subunits of the eukaryotic (p48) and archaeal primases, share a little, but distinct sequence homology with those of the family X DNA polymerases [60]. The first archaeal primase was identified from *Methanococcus jannaschii*, as an ORF with a sequence similar to that of the eukaryotic p48. The gene product exhibited DNA polymerase activity and was able to synthesize oligonucleotides on the template DNA [61]. We characterized the p48-like protein (p41) from *P. furiosus*. Unexpectedly, the archaeal p41 protein did not synthesize short RNA by itself, but preferentially utilized deoxynucleotides to synthesize DNA strands up to several kilobases in length [62]. Furthermore, the gene neighboring the p41 gene encodes a protein with very weak similarity to the p58 subunit of the eukaryotic primase. The gene product, designated p46, actually forms a stable complex with p41, and the complex can synthesize a short RNA primer, as well as DNA strands of several hundred nucleotides *in vitro* [63]. The short RNA but not DNA primers were identified in *Pyrococcus* cells, and therefore, some mechanism to dominantly use RNA primers exists in the cells [10].

Further research on the primase homologs from *S. solfataricus* [64-66], *Pyrococcus horikoshii* [67-69], and *P. abyssi* [70] showed similar properties *in vitro*. Notably, p41 is the catalytic subunit, and the large one modulates the activity in the heterodimeric archaeal primases. The small and large subunits are also called PriS and PriL, respectively. The crystal structure of

the N-terminal domain of PriL complexed with PriS of *S. solfataricus* primase revealed that PriL does not directly contact the active site of PriS, and therefore, the large subunit may interact with the synthesized primer, to adjust its length to a 7-14 mer. The structure of the catalytic center is similar to those of the family X DNA polymerases. The 3'-terminal nucleotidyl transferase activity, detected in the *S. solfataricus* primase [64, 66], and the gap-filling and strand-displacement activities in the *P. abyssi* primase [70] also support the structural similarity between PriS and the family X DNA polymerases.

A unique activity, named PADT (template-dependent Polymerization Across Discontinuous Template), in the *S. solfataricus* PriSL complex was published very recently [71]. The activity may be involved in double-strand break repair in Archaea.

The archaeal genomes also encode a sequence similar to the bacterial type DnaG primase. The DnaG homolog from the *P. furiosus* genome was expressed in *E. coli*, but the protein did not show any primer synthesis activity *in vitro*, and thus the archaeal DnaG-like protein may not act as a primase in *Pyrococcus* cells (Fujikane et al. unpublished). The DnaG-like protein was shown to participate in RNA degradation, as an exosome component [72, 73]. However, a recent paper reported that a DnaG homolog from *S. solfataricus* actually synthesizes primers with a 13 nucleotide length [74]. It would be interesting to investigate if the two different primases share the primer synthesis for leading and lagging strand replication, respectively, in the *Sulfolobus* cells, as the authors suggested [74]. A proposed hypothesis about the evolution of PriSL and DnaG from the last universal common ancestor (LUCA) is interesting [71].

The *Sulfolobus* PriSL protein was shown to interact with Mcm through Gins23 [50]. This primase-helicase interaction probably ensures the coupling of DNA unwinding and priming during the replication fork progression [50]. Furthermore, the direct interaction between PriSL and the clamp loader RFC (described below) in *S. solfataricus* may regulate the primer synthesis and its transfer to DNA polymerase in archaeal cells [75].

## 8. Single-stranded DNA binding protein

The single-stranded DNA binding protein, which is called SSB in Bacteria and RPA in Archaea and Eukaryota, is an important factor to protect the unwound single-stranded DNA from nuclease attack, chemical modification, and other disruptions during the DNA replication and repair processes. SSB and RPA have a structurally similar domain containing a common fold, called the OB (oligonucleotide/oligosaccharide binding)-fold, although there is little amino acid sequence similarity between them [76]. The common structure suggests that the mechanism of single-stranded DNA binding is conserved in living organisms despite the lack of sequence similarity. *E. coli* SSB is a homotetramer of a 20 kDa peptide with one OB-fold, and the SSBs from *Deinococcus radiodurans* and *Thermus aquaticus* consist of a homodimer of the peptide containing two OB-folds. The eukaryotic RPA is a stable heterotrimer, composed of 70, 32, and 14 kDa proteins. RPA70 contains two tandem repeats of an OB-fold, which are responsible for the major interaction with a single-stranded DNA in its

central region. The N-terminal and C-terminal regions of RPA70 mediate interactions with RPA32 and also with many cellular or viral proteins [77, 78]. RPA32 contains an OB-fold in the central region [79-81], and the C-terminal region interacts with other RPA subunits and various cellular proteins [77, 78, 82, 83]. RPA14 also contains an OB-fold [77]. The eukaryotic RPA interacts with the SV40 T-antigen and the DNA polymerase $\alpha$-primase complex, and thus forms part of the initiation complex at the replication origin [84]. The RPA also stimulates Pol$\alpha$-primase activity and PCNA-dependent Pol $\delta$ activity [85, 86].

The RPAs from *M. jannaschii* and *M. thermautotrophicus* were reported in 1998, as the first archaeal single-stranded DNA binding proteins [87-89]. These proteins share amino acid sequence similarity with the eukaryotic RPA70, and contain four or five repeated OB-fold and one zinc-finger motif. The *M. jannaschii* RPA exists as a monomer in solution, and has single-strand DNA binding activity. On the other hand, *P. furiosus* RPA forms a complex consisting of three distinct subunits, RPA41, RPA32, and RPA14, similar to the eukaryotic RPA [90]. The *P. furiosus* RPA strikingly stimulates the RadA-promoted strand-exchange reaction *in vitro* [90].

While the euryarchaeal organisms have a eukaryotic-type RPA homologue, the crenarchaeal SSB proteins appear to be much more related to the bacterial proteins, with a single OB fold and a flexible C-terminal tail. However, the crystal structure of the SSB protein from *S. solfataricus* showed that the OB-fold domain is more similar to that of the eukaryotic RPAs, supporting the close relationship between Archaea and Eukaryota [91].

The RPA from *Methanosarcina acetivorans* displays a unique property. Unlike the multiple RPA proteins found in other archaea and eukaryotes, each subunit of the *M. acetivorans* RPAs, RPA1, RPA2, and RPA3, have 4, 2, and 2 OB-folds, respectively, and can act as a distinct single-stranded DNA-binding proteins. Furthermore, each of the three RPA proteins, as well as their combinations, clearly stimulates the primer extension activity of *M. acetivorans* DNA polymerase BI *in vitro*, as shown previously for bacterial SSB and eukaryotic RPA [92]. Architectures of SSB and RPA suggested that they are composed of different combinations of the OB fold. Bacterial and eukaryotic organisms contain one type of SSB or RPA, respectively. In contrast, archaeal organisms have various RPAs, composed of different organizations of OB-folds. A hypothesis that homologous recombination might play an important role in generating this diversity of OB-folds in archaeal cells was proposed, based on experiments characterizing the engineered RPAs with various OB-folds [93].

## 9. DNA polymerase

DNA polymerase catalyzes phosphodiester bond formation between the terminal 3'-OH of the primer and the $\alpha$-phosphate of the incoming triphosphate to extend the short primer, and is therefore the main player of the DNA replication process. Based on the amino acid sequence similarity, DNA polymerases have been classified into seven families, A, B, C, D, E, X, and Y (Table 2) [94-98].

The fundamental ability of DNA polymerases to synthesize a deoxyribonucleotide chain is widely conserved, but more specific properties, including processivity, synthesis accuracy,

and substrate nucleotide selectivity, differ depending on the family. The enzymes within the same family have basically similar properties. *E. coli* has five DNA polymerases, and Pol I, Pol II, and Pol III belong to families A, B, and C, respectively. Pol IV and Pol V are classified in family Y, as the DNA polymerases for translesion synthesis (TLS). In eukaryotes, the replicative DNA polymerases, Pol $\alpha$, Pol $\delta$, and Pol $\varepsilon$, belong to family B, and the translesion DNA polymerases, $\eta$, $\iota$, and $\kappa$, belong to family Y [99].

The most interesting feature discovered at the inception of this research area was that the archaea indeed have the eukaryotic Pol $\alpha$-like (Family B) DNA polymerases [100-102]. Members of the Crenarchaeota have at least two family B DNA polymerases [103, 104]. On the other hand, there is only one family B DNA polymerase in the Euryarchaeota. Instead, the euryarchaeal genomes encode a family D DNA polymerase, proposed as Pol D, which seems to be specific for these archaeal organisms and has never been found in other domains [95, 105]. The genes for family Y-like DNA polymerases are conserved in several, but not all, archaeal genomes. The role of each DNA polymerase in the archaeal cells is still not known, although the distribution of the DNA polymerases is getting clearer (Table 2) [106].

| | families of DNA polymerases | | | | | | |
| | A | B | C | D | E | X | Y |
|---|---|---|---|---|---|---|---|
| **Archaea** | | | | | | | |
| Crenarchaeota | | Pol BI, Pol BII Pol BIII | | | Pol E* | | Pol Y |
| Euryarchaeota | | Pol BI | | Pol D | Pol E* | | Pol Y |
| Korarchaeota | | Pol BI, Pol BII | | Pol D | | | |
| Aigarchaeota | | Pol BI, Pol BII | | Pol D | | | Pol Y |
| Thaumarchaeota | | Pol BI | | Pol D | | | Pol Y |
| **Bacteria** | Pol I | Pol II | Pol III | | | | Pol IV Pol V |
| **Eukaryota** | Pol θ Pol γ** | Pol α, Pol δ Pol ε, Pol ζ | | | | Pol β, Pol λ Pol μ, Pol σ | Pol η, Pol ι Pol κ |

* plasmid-encoded

** mitochondrial

**Table 2.** Distribution of DNA polymerases from seven families in the three domains of life.

The first family D DNA polymerase was identified from *P. furiosus,* by screening for DNA polymerase activity in the cell extract [107]. The corresponding gene was cloned, revealing that this new DNA polymerase consists of two proteins, named DP1 and DP2, and that the deduced amino acid sequences of these proteins were not conserved in the DNA polymerase families [8]. *P. furiosus* Pol D exhibits efficient strand extension activity and strong proofreading activity [8, 108]. Other family D DNA polymerases were also characterized by several groups [109-115]. The Pol D genes had been found only in Euryarchaeota. However, recent environmental genomics and cultivation efforts revealed novel phyla in Archaea: Thaumarchaeota, Korarchaeota, and Aigarchaeota, and their genome sequences harbor the genes encoding Pol D.

A genetic study on *Halobacterium* sp. NRC-1 showed that both Pol B and Pol D are essential for viability [116]. An interesting issue is to elucidate whether Pol B and Pol D work together at the replication fork for the synthesis of the leading and lagging strands, respectively. According to the usage of an RNA primer and the presence of strand displacement activity, Pol D may catalyze lagging strand synthesis [106, 114].

Thaumarchaeota and Aigarchaeota harbor the genes encoding Pol D and crenarchaeal Pol BII [117, 118], while Korarchaeota encodes Pol BI, Pol BII and Pol D [119]. Biochemical characterization of these gene products will contribute to research on the evolution of DNA polymerases in living organisms. A hypothesis that the archaeal ancestor of eukaryotes encoded three DNA polymerases, two distinct family B DNA polymerases and a family D DNA polymerase, which all contributed to the evolution of the eukaryotic replication machinery, consisting of Pol $\alpha$, $\delta$, and $\epsilon$, has been proposed [120].

A protein is encoded in the plasmid pRN1 isolated from a *Sulfolobus* strain [121]. This protein, ORF904 (named RepA), has primase and DNA polymerase activities in the N-terminal domain and helicase activity in the C-terminal domain, and is likely to be essential for the replication of pRN1 [122, 123]. The amino acid sequence of the N-terminal domain lacks homology to any known DNA polymerases or primases, and therefore, family E is proposed. Similar proteins are encoded by various archaeal and bacterial plasmids, as well as by some bacterial viruses [124]. Recently, one protein, tn2-12p, encoded in the plasmid pTN2 isolated from *Thermococcus nautilus,* was experimentally identified as a DNA polymerase in this family [125]. This enzyme is likely responsible for the replication of the plasmids. Further investigations of this family of DNA polymerases will be interesting from an evolutional perspective.

## 10. PCNA and RFC

The sliding clamp with the doughnut-shaped ring structure is conserved among living organisms, and functions as a platform or scaffold for proteins to work on the DNA strands. The eukaryotic and archaeal PCNAs form a homotrimeric ring structure [126, 127], which encircles the DNA strand and anchors many important proteins involved in DNA replication and repair (Figure 4). PCNA works as a processivity factor that retains the DNA poly-

merase on the DNA by binding it on one surface (front side) of the ring for continuous DNA strand synthesis in DNA replication (Figure 5). To introduce the DNA strand into the central hole of the clamp ring, a clamp loader is required to interact with the clamp and open its ring. The archaeal and eukaryotic clamp loader is called RFC (Figure 5). The most studied archaeal PCNA and RFC molecules to date are *P. furiosus* PCNA [128-132] and RFC [133-136]. The PCNA and RFC molecules are essential for DNA polymerase to perform processive DNA synthesis. The molecular mechanism of the clamp loading process has been actively investigated [137] (Figure 5). An intermediate PCNA-RFC-DNA complex, in which the PCNA ring is opened with out-of plane mode, was detected by a single particle analysis of electron microscopic images using *P. furiosus* proteins (Figure 6) [138]. The crystal structure of the complex, including the ATP-bound clamp loader, the ring-opened clamp, and the template-primer DNA, using proteins from bacteriophage T4, has recently been published [139], and our knowledge about the clamp loading mechanism is continuously progressing.
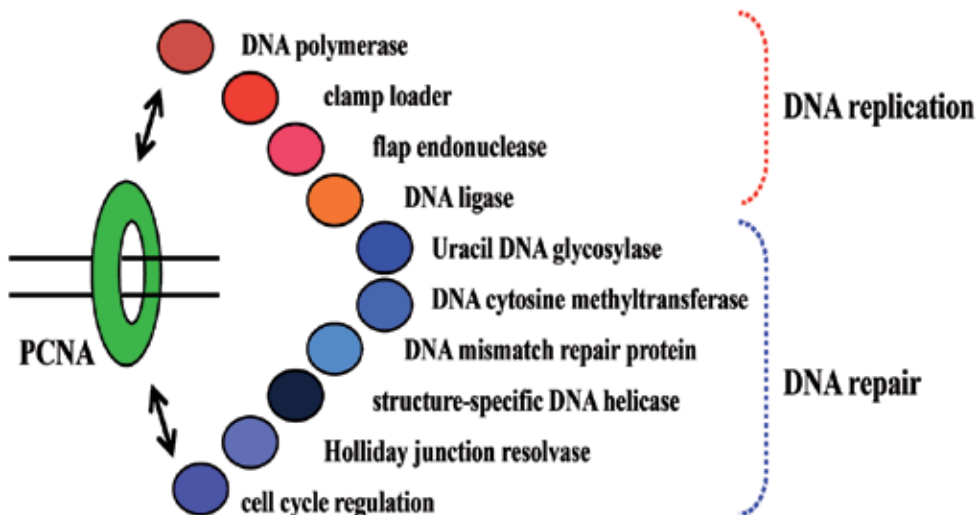


**Figure 4.** PCNA-interacting proteins

After clamp loading, DNA polymerase accesses the clamp and the polymerase-clamp complex performs processive DNA synthesis. Therefore, structural and functional analyses of the DNA polymerase-PCNA complex is the next target to elucidate the overall mechanisms of replication fork progression. The PCNA interacting proteins contain a small conserved sequence motif, called the PIP box, which binds to a common site on PCNA [140]. The PIP box consists of the sequence "Qxxhxxaa", where "x" represents any amino acid, "h" represents a hydrophobic residue (e.g. L, I or M), and "a" represents an aromatic residue (e.g. F, Y or W). Archaeal DNA polymerases have PIP box-like motifs in their sequences [141]). However, only a few studies have experimentally investigated the function of the motifs. The crystal structure of *P. furiosus* Pol B complexed with a monomeric PCNA mutant was determined,

and a convincing model of the polymerase-PCNA ring interaction was constructed [142]. This study revealed that a novel interaction is formed between a stretched loop of PCNA and the thumb domain of Pol B, in addition to the authentic PIP box. A comparison of the model structure with the previously reported structures of a family B DNA polymerase from RB69 phage, complexed with DNA [143, 144], suggested that the second interaction site plays a crucial role in switching between the polymerase and exonuclease modes, by inducing a PCNA-polymerase complex configuration that favors synthesis over editing. This putative mechanism for the fidelity control of replicative DNA polymerases is supported by experiments, in which mutations at the second interaction site enhanced the exonuclease activity in the presence of PCNA [144]. Furthermore, the three-dimensional structure of the DNA polymerase-PCNA-DNA ternary complex was analyzed by electron microscopic (EM) single particle analysis. This structural view revealed the entire domain configuration of the trimeric ring of PCNA and DNA polymerase, including the protein-protein or protein-DNA contacts. This architecture provides clearer insights into the switching mechanism between the editing and synthesis modes [145].
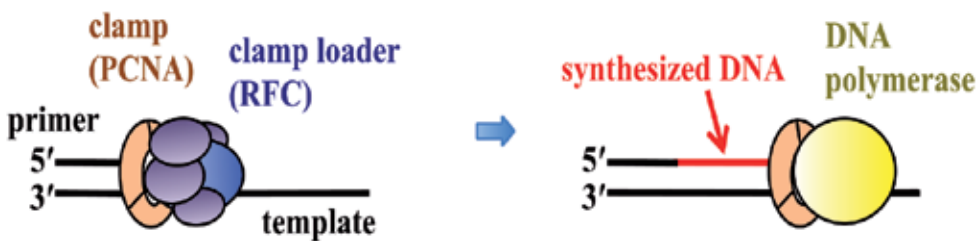


**Figure 5. Mechanisms of processive DNA synthesis** The clamp loader (RFC) tethers the clamp (PCNA) onto the primer terminus of the DNA strand. The clamp loader is then replaced by DNA polymerase, which can synthesize the DNA strand processively without falling off.

In contrast to most euryarchaeal organisms, which have a single PCNA homolog forming a homotrimeric ring structure, the majority of crenarchaea have multiple PCNA homologues, and they are capable of forming heterotrimeric rings for their functions [146, 147]. It is especially interesting that the three PCNAs, PCNA1, PCNA2, and PCNA3, specifically bind PCNA binding proteins, including DNA polymerases, DNA ligases, and FEN-1 endonuclease [147, 148]. Detailed structural studies of the heterologous PCNA from *S. solfataricus* revealed that the interaction modes between the subunits are conserved with those of the homotrimeric PCNAs [149, 150].

*T. kodakarensis* is the only euryarchaeal species that has two genes encoding PCNA homologs on the genome [38]. These two genes from the *T. kodakarensis* genome, and the highly purified gene products, PCNA1 and PCNA2, were characterized [151]. PCNA1 stimulated the DNA synthesis reactions of the two DNA polymerases, Pol B and Pol D, from *T. kodakarensis in vitro*. PCNA2 however only had an effect on Pol B. The *T. kodakarensis* strain with *pcna2* disruption was isolated, whereas gene disruption for *pcna1* was not possible. These results suggested that PCNA1 is essential for DNA replication, and PCNA2 may play a different role in *T. kodakarensis* cells. The sensitivities of the Δ*pcna2* mutant strain to ultraviolet

irradiation (UV), methyl methanesulfonate (MMS) and mitomycin C (MMC) were indistinguishable to those of the wild type strain. Both PCNA1 and PCNA2 form a stable ring structure and work as a processivity factor for *T. kodakarensis* Pol B *in vitro*. The crystal structures of the two PCNAs revealed the different interactions at the subunit-subunit interfaces [152].
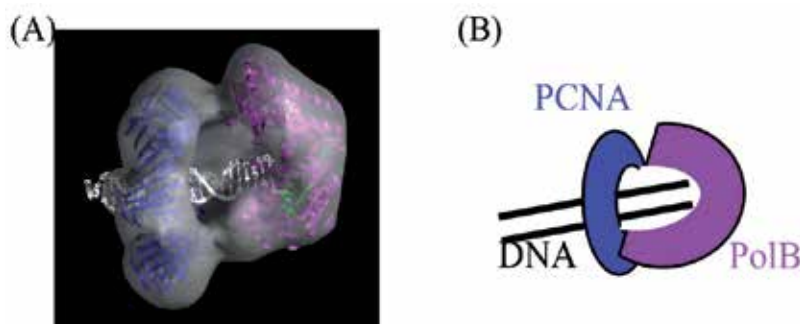


**Figure 6. Electron Microscopic Analysis of *P. furious* DNA polymerase-PCNA-DNA complex.** The complex in the editing mode of the DNA polymerase-PCNA-DNA ternary complex was shown. (A) Electron microscopic (EM) map of the complex is depicted by gray surface. DNA polymerase and PCNA are shown in a ribbon representation colored purple and blue, respectively. The DNA is colored white. The exonuclease active site is shown in a green ribbon. (B) Schematic view of the complex.

The RFC molecule is conserved as a pentameric complex in Eukaryota and Archaea. However, the eukaryotic RFC is a heteropentameric complex, consisting of five different proteins, RFC1 to 5, in which RFC1 is larger than the other four RFCs. On the other hand, the archaeal RFC consists of two proteins, RFCS (small) and RFCL (large), in a 4 to 1 ratio. A different form of RFC, consisting of three subunits, RFCS1, RFCS2, and RFCL, in a 3 to 1 to 1 ratio, was also identified from *M. acetivorans* [153]. The three subunits of RFC may represent an intermediate stage in the evolution of the more complex RFC in Eukaryota from the less complex RFC in Archaea [153, 154]. The subunit organization and the spatial distribution of the subunits in the *M. acetivorans* RFC complex were analyzed and compared with those of the *E. coli* γ-complex, which is also a pentamer consisting of three different proteins. These two clamp loaders adopt similar subunit organizations and spatial distributions, but the functions of the individual subunits are likely to be diverse [154].

## 11. DNA ligase

DNA ligase is essential to connect the Okazaki fragments of the discontinuous strand synthesis during DNA replication, and therefore, it universally exists in all living organisms. This enzyme catalyzes phosphodiester bond formation via three nucleotidyl transfer steps [155, 156]. In the first step, DNA ligase forms a covalent enzyme-AMP intermediate, by reacting with ATP or NAD$^+$ as a cofactor. In the second step, DNA ligase recognizes the substrate DNA, and the AMP is subsequently transferred from the ligase to the 5'-phosphate terminus of the DNA, to form a DNA-adenylate intermediate (AppDNA). In the final step,

the 5′-AppDNA is attacked by the adjacent 3′-hydroxy group of the DNA and a phospho-diester bond is formed. DNA ligases are grouped into two families, according to their requirement for ATP or NAD⁺ as a nucleotide cofactor in the first step reaction. ATP-dependent DNA ligases are widely found in all three domains of life, whereas NAD⁺-dependent DNA ligases exist mostly in Bacteria. Some halophilic archaea [157] and eukaryotic viruses [158] also have NAD⁺-dependent enzymes.

Three genes (*LIG1*, *LIG3* and *LIG4*) encoding ATP-dependent DNA ligases have been identified in the human genome to date and DNA ligase I (Lig I), encoded by *LIG1*, is a replicative enzyme that joins Okazaki fragments during DNA replication [156]. The first gene encoding a eukaryotic-like ATP-dependent DNA ligase was found in the thermophilic archaeon, *Desulfolobus ambivalens* [159]. Subsequent identifications of the DNA ligases from archaeal organisms revealed that these enzymes primarily use ATP as a cofactor. However, this classification may not be so strict. The utilization of NAD⁺, as well as ATP, as a cofactor has been observed in several DNA ligases, including those from *T. kodakarensis* [160], *T. fumicolans*, *P. abyssi* [161]), *Thermococcus* sp. NA1 [162], *T. acidophilum*, *Picrophilus torridus,* and *Ferroplasma acidophilum*, although ATP is evidently preferable in all of the cases [163] (Table 3). The dual co-factor specificity (ATP/NAD⁺) is an interesting feature of these DNA ligase enzymes and it will be enlightening to investigate the structural basis for this. Another dual co-factor specificity exists in the archaeal DNA ligases, which use ADP as well as ATP, as found in the enzymes from *A. pernix* [164] and *Staphylothermus marinus* [165], and in the case of *Sulfobococcus zilligii*, GTP is also the functional cofactor [166]. The DNA ligases from *P. horikoshii* [167] and *P. furiosus* [168] have a strict ATP preference (Table 3). Sufficient biochemical data have not been obtained to resolve the issue of dual co-factor specificity, and further biochemical and structural analyses are required.

| cofactor | | | |
| --- | --- | --- | --- |
| **ATP** | **ATP and ADP** | **ATP and NAD⁺** | **ATP, ADP, and GTP** |
| *Acidithiobacillus ferrooxidans* | *Aeropyrum pernix* | *Ferroplasma acidophilum* | *Sulfophobococcus zilligii* |
| *Ferroplasma acidarmanus* | *Staphylothermus marinus* | *Picrophilus torridus* | |
| *Methanothermobacterium thermoautotrophicum* | | *Thermococcus fumicolans* | |
| *Pyrococcus horikoshii* | | *Thermococcus kodakarensis* | |
| *Pyrococcus furiosus* | | *Thermococcus* sp. | |
| *Sulfolobus acidocaldarius* | | *Thermoplasma acidophilum* | |
| *Sulfolobus shibatae* | | | |
| *Thermococcus* sp. 1519 | | | |

**Table 3.** Cofactor dependency of the archaeal DNA ligases

The crystal structure of *P. furiosus* DNA ligase [169] was solved and the physical and functional interactions between the DNA ligase and PCNA was shown [168]. The detailed interaction mode between human Lig I and PCNA is somewhat unclear, because of several controversial reports [170-172]. The stimulatory effect of *P. furiosus* PCNA on the enzyme activity of the cognate DNA ligase was observed at a high salt concentration, at which a DNA ligase alone cannot bind to a nicked DNA substrate. Interestingly, the PCNA-binding site is located in the middle of the N-terminal DNA binding domain (DBD) of the *P. furiosus* DNA ligase, and the binding motif, QKSFF, which is proposed as a shorter version of the PIP box, is actually looped out from the protein surface [168]. Interestingly, this motif is located in the middle of the protein chain, rather than the N- or C-terminal region, where the PIP boxes are usually located. To confirm that this motif is conserved in the archaeal/eukaryotic DNA ligases, the physical and functional interactions between *A. pernix* DNA ligase and PCNA was analyzed and the interaction was shown to mainly depend on the phenylalanine 132 residue, which is located in the predicted region from the multiple sequence alignment of the ATP-dependent DNA ligases [173].

The crystal structure of the human Lig I, complexed with DNA, was solved as the first ATP-dependent mammalian DNA ligase, although the ligase was an N-terminal truncated form [174]. The structure comprises the N-terminal DNA binding domain, the middle adenylation domain, and the C-terminal OB-fold domain. The crystal structure of Lig I (residues 233 to 919) in complex with a nicked, 5'-adenylated DNA intermediate revealed that the enzyme redirects the path of the dsDNA, to expose the nick termini for the strand-joining reaction. The N-terminal DNA-binding domain works to encircle the DNA substrate like PCNA and to stabilize the DNA in a distorted structure, positioning the catalytic core on the nick. The crystal structure of the full length DNA ligase from *P. furiosus* revealed that the architecture of each domain resembles those of Lig I, but the domain arrangements strikingly differ between the two enzymes [168]. This domain rearrangement is probably derived from the "domain-connecting" role of the helical extension conserved at the C-termini in the archaeal and eukaryotic DNA ligases. The DNA substrate in the open form of Lig I is replaced by motif VI at the C-terminus, in the closed form of *P. furiosus* DNA ligase. Both the shapes and electrostatic distributions are similar between motif VI and the DNA substrate, suggesting that motif VI in the closed state mimics the incoming substrate DNA. The subsequently solved crystal structure of *S. solfataricus* DNA ligase is the fully open structure, in which the three domains are highly extended [175]. In this work, the *S. solfataricus* ligase-PCNA complex was also analyzed by SAXS. *S. solfataricus* DNA ligase bound to the PCNA ring still retains an open, extended conformation. The closed, ring-shaped conformation observed in the Lig I structure as described above is probably the active form to catalyze a DNA end-joining reaction, and therefore, it is proposed that the open-to-closed movement occurs for ligation, and the switch in the conformational change is accommodated by a malleable interface with PCNA, which serves as an efficient platform for DNA ligation [175]. After the publication of these crystal structures, the three-dimensional structure of the ternary complex, consisting of DNA ligase-PCNA-DNA, using the *P. furiosus* proteins was obtained by EM single particle analysis [176]. In the complex structure, the three domains of the crescent-shaped *P. furiosus* DNA ligase surround the central DNA duplex, encircled by the closed PCNA ring. The

relative orientations of the ligase domains remarkably differ from those of the crystal structures, and therefore, a large domain rearrangement occurs upon ternary complex formation. In the EM image model, the DNA ligase contacts PCNA at two sites, the conventional PIP box and a novel second contact in the middle adenylation domain. It is also interesting that a substantial DNA tilt from the PCNA ring axis is observed. Based on these structural analyses, a mechanism in which the PCNA binding proteins are bound and released sequentially. In fact, most of the PCNA binding proteins share the same binding sites in the interdomain connecting loop (IDCL) and the C-terminal tail of the PCNA. The structural features exclude the possibility that the three proteins contact the single PCNA ring simultaneously, because DNA ligase occupies two of the three subunits of the PCNA trimer. In the case of the RFC-PCNA-DNA complex structure obtained by the same EM technique, RFC entirely covers the PCNA ring, thus blocking the access of other proteins [138]. These ternary complexes appear to favor a mechanism involving the sequential binding and release of replication factors.

## 12. Flap endonuclease 1 (FEN1)

Efficient processing of Okazaki fragments to make a continuous DNA strand is essential for the lagging strand synthesis in asymmetric DNA replication. The primase-synthesized RNA/DNA primers need to be removed to join the Okazaki fragments into an intact continuous strand DNA. Flap endonuclease 1 (FEN1) is mainly responsible for this task. Okazaki fragment maturation is highly coordinated with continuous DNA synthesis, and the interactions of DNA polymerase, FEN1, and DNA ligase with PCNA allow these enzymes to act sequentially during the maturation process, as described above.

FEN1, a structure-specific 5′-endonuclease, specifically recognizes a dsDNA with an unannealed 5′-flap [177, 178]. In the eukaryotic Okazaki fragment processing system, 5′-flap DNA structures are formed by the strand displacement activity of DNA polymerase δ. Lig I seals the nick after the flapped DNA is cleaved by FEN1. These processing steps are facilitated by PCNA [179]. The interactions between eukaryotic FEN1 and PCNA have been well characterized [140, 171], and the stimulatory effect of PCNA on the FEN1 activity was also shown [180]. The crystal structure of the human FEN1-PCNA complex revealed three FEN1 molecules bound to each PCNA subunit of the trimer ring in different configurations [181]. Based on these structural analyses together with the description in the DNA ligase section, a flip-flop transition mechanism, which enables proteins to internally switch for different functions on the same DNA clamp are currently being considered.

The eukaryotic homologs of FEN1 were found in Archaea [182]. The crystal structures of FEN1 from *M. jannaschii* [183], *P. furiosus* [184], *P. horikoshii* [185], *A. fulgidus* [186], and *S. solfataricus* [150] have been determined. In addition, detailed biochemical studies were performed on *P. horikoshii* FEN1 [187, 188]. Thus, studies of the archaeal FEN1 proteins have provided important insights into the structural basis of the cleavage reaction of the flapped DNA. Our recent research showed that the flap endonuclease activity of *P. furiosus* FEN1 was stimulated by PCNA. Furthermore, the stimulatory effect of PCNA on the sequential

action of FEN1 and DNA ligase was observed *in vitro* (Kiyonari et al., unpublished). Based on these results, a model of the molecular switching mechanisms of the last steps of Okaza-ki-fragment maturation was constructed. The quaternary complex of FEN1-Lig-PCNA-DNA was also isolated for the EM single particle analysis. These studies will provide more con-crete image of the molecular mechanism.

## 13. Summary and perspectives

Research on the molecular mechanism of DNA replication has been a central theme of mo-lecular biology. Archaeal organisms became popular in the total genome sequencing age, as described above, and most of the DNA replication proteins are now equally understood by biochemical characterizations. In addition, the archaeal studies are especially interesting to understand the mechanisms by which cells live in extreme environmental conditions. Fur-thermore, it is also noteworthy that the proteins from the hyperthermophilic archaea are more stable than those from mesophilic organisms, and they are advantageous for the struc-tural and functional analyses of higher-ordered complexes, such as the replisome. Studies on the higher-ordered complexes, rather than single proteins, are essential for understand-ing each of the events involved in DNA metabolism, and the archaeal research will continu-ously contribute to the development and advancement of the DNA replication research field, as summarized in part in a recent review [189, 190].

In addition to basic molecular biology research, DNA replication proteins from thermo-philes have been quite useful reagents for gene manipulations, including genetic diagnosis, forensic DNA typing, and detection of bacterial and virus infections, as well as basic re-search. Numerous enzymes have been commercialized around the world, and are utilized daily. An example of the successful engineering of an archaeal DNA polymerase for PCR is the creation of the fusion protein between *P. furiosus* Pol B and a nonspecific dsDNA bind-ing protein, Sso7d, from *S. solfataricus*, by genetic engineering techniques [191]. The fusion DNA polymerase overcame the low processivity of the wild type Pol B by the high affinity Sso7d to the DNA strand. As another example, we successfully developed a novel proces-sive PCR method, using the archaeal Pol B with the help of a mutant PCNA [192, 193]. Sev-eral DNA sequencing technologies, referred to as "next-generation sequencing", have been developed [194, 195], and are now commercially available. Single-molecule detection, using dye-labeled modified nucleotides and longer read lengths, is now known as "third-genera-tion DNA sequencing" [196]. These technologies apply DNA polymerases or DNA ligases from various sources, indicating that these DNA replication enzymes are indispensable for the development of DNA manipulation technology. These facts prove that the progress of the basic research on the molecular biology of archaeal DNA replication will promote the development of the new technologies for genetic engineering.

## Author details

Yoshizumi Ishino and Sonoko Ishino

Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, Japan

## References

[1] Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature, 1953;171(4356) 737-738.

[2] Kelman Z, White MF. Archaeal DNA replication and repair. Curr Opin Microbiol, 2005;(6) 669-676.

[3] Barry ER, Bell SD. DNA replication in the archaea. Microbiol Mol Biol Rev, 2006;70(4) 876-887.

[4] Wigley DB ORC proteins: marking the start. Curr Opin Struct Biol, 2009;19(1) 72-78.

[5] Leipe DD, Aravind L, Koonin EV. Did DNA replication evolve twice independently? Nucleic Acids Res, 1999;27(17) 3389-3401.

[6] Jacob F, Brenner SCR. Hebd Seances Acad Sci. 1963;256 298-300.

[7] Matsunaga F, Forterre P, Ishino Y, Myllykallio H. In vivo interactions of archaeal Cdc6/Orcl and minichromosome maintenance proteins with the replication origin. Proc. Natl. Acad. Sci. USA, 2001;98(20) 11152-11157.

[8] Uemori T, Sato Y, Kato I, Doi H, Ishino Y. A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. Genes Cells, 1997;2(8) 499-512.

[9] Lopez P, Philippe H, Myllykallio H, Forterre P. Identification of putative chromosomal origins of replication in Archaea. Mol Microbiol, 1999;32(4) 883-886.

[10] Matsunaga F, Norais C, Forterre P, Myllykallio H. Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. EMBO Rep, 2003;4(2) 154-158.

[11] Robinson NP, Dionne I, Lundgren M, Marsh VL. Bernander R, Bell SD. Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus* Cell, 2004;116(1) 25-38.

[12] Matsunaga F, Glatigny A, Mucchielli-Giorgi MH, Agier N, Delacroix H, Marisa L, et al. Genomewide and Biochemical Analyses of DNA-binding activity of Cdc6/Orc1 and Mcm proteins in *Pyrococcus* sp. Nucleic Acids Res, 2007;35(10)3214-3222.

[13] Lundgren M, Andersson A, Chen L, Nilsson P, Bernander R. Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. Proc Natl Acad Sci USA, 2004;101(18):7046-7051.

[14] Robinson NP, Bell SD. Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. Proc Natl Acad Sci USA, 2007;104(14) 5806-5811.

[15] Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, Allers T. Genetic and physical mapping of DNA replication origins in *Haloferax volcanii.* PLoS Genet, 2007;3(5) e77.

[16] Capes MD, Coker JA, Gessler R, Grinblat-Huse V, DasSarma SL, Jacob CG, et al. The information transfer system of halophilic archaea. Plasmid, 2011;65(2) 77-101.

[17] Liu J, Smith CL, DeRyckere D, DeAngelis K, Martin GS, Berger JM. Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. Mol Cell, 2000;6(3) 637-648.

[18] Singleton MR, Morales R, Grainge I., Cook N, Isupov MN, Wigley DB. Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. J Mol Biol, 2004;343(3) 547-557.

[19] Grainge I, Gaudier M, Schuwirth BS, Westcott SL, Sandall J, Atanassova N, et al. Biochemical analysis of a DNA replication origin in the archaeon *Aeropyrum pernix*. J Mol Biol, 2006;363(2) 355-369.

[20] Dueber EL, Corn JE, Bell, SD, Berger JM. Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. Science, 2007;317(5842) 1210-1213.

[21] Gaudier M, Schuwirth BS, Westcott SL, Wigley DB. Structural basis of DNA replication origin recognition by an ORC protein. Science, 2007;317(5842) 1213-1216.

[22] Majernik AI, Chong JP. A conserved mechanism for replication origin recognition and binding in archaea. Biochem J, 2008;409(2) 511-518.

[23] Matsunaga F, Takemura K, Akita M, Adachi A, Yamagami T, Ishino Y. Localized melting of duplex DNA by Cdc6/Orc1 at the DNA replication origin in the hyperthermophilic archaeon *Pyrococcus furiosus.* Extremophiles, 2010;14(1) 21-31.

[24] Akita M, Adachi A, Takemura K, Yamagami T, Matsunaga F, Ishino Y. Cdc6/Orc1 from *Pyrococcus furiosus* may act as the origin recognition protein and Mcm helicase recruiter. Gens Cells, 2010;15(5) 537-552.

[25] Masai H, You Z, Arai K. Control of DNA replication: regulation and activation of eukaryotic replicative helicase, MCM. IUBMB Life, 2005;57(4-5) 323-335.

[26] Ilves I, Petojevic T, Pesavento JJ, Botchan MR. Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. Mol Cell, 2010;37(2) 247-258,

[27] Chong JP, Hayashi MK, Simon MN, Xu RM, Stillman B. A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. Proc Natl Acad Sci USA, 2000;97(4) 1530-1535.

[28] Grainge I, Scaife S, Wigley DB. Biochemical analysis of components of the pre-replication complex of *Archaeoglobus fulgidus.* Nucleic Acids Res, 2003;31(16) 4888-4898.

[29] Kelman Z, Lee JK, Hurwitz J. The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* ΔH contains DNA helicase activity. Proc Natl Acad Sci USA, 1999;96(26) 14783-14788.

[30] Shechter DF, Ying CY, Gautier J, The intrinsic DNA helicase activity of *Methanobacterium thermoautotrophicum* ΔH minichromosome maintenance protein. J Biol Chem, 2000;275(20) 15049-15059.

[31] Yoshimochi T, Fujikane R, Kawanami M, Matsunaga F, Ishino Y. The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. J Biol Chem, 2008;283(3) 1601-1609.

[32] Sakakibara N, Kelman LM, Kelman Z. Unwinding the structure and function of the archaeal MCM helicase. Mol Microbiol, 2009;72(2) 286-896.

[33] Shin JH, Grabowski B, Kasiviswanathan R, Bell SD, Kelman Z. Regulation of minichromosome maintenance helicase activity by Cdc6. J Biol Chem, 2003;278(39) 38059-38067.

[34] Kasiviswanathan R, Shin JH, Kelman Z. Interactions between the archaeal Cdc6 and MCM proteins modulate their biochemical properties. Nucleic Acids Res, 2005;33(15) 4940-4950.

[35] De Felice M, Esposito L, Pucci B, Carpentieri F, De Falco M, Rossi M, et al. Biochemical characterization of a CDC6-like protein from the crenarchaeon *Sulfolobus solfataricus*. J Biol Chem, 2003;278(47) 46424-46431.

[36] Haugland GT, Shin JH, Birkeland NK, Kelman Z. Stimulation of MCM helicase activity by a Cdc6 protein in the archaeon *Thermoplasma acidophilum*. Nucleic Acids Res, 2006, 34(21) 6337-6344.

[37] Krupovic M, Gribaldo S, Bamford DH, Forterre P. The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. Mol Biol Evol, 2010;27(12) 2716-2732.

[38] Fukui T, Atomi H, Kanai T, Matsumi R, Fujiwara S, Imanaka T.. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakarensis* KOD1 and comparison with *Pyrococcus* genomes. Genome Res, 2005;15(3) 352-363.

[39] Sato T, Fukui T, Atomi H, Imanaka T. Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. J Bacteriol, 2003;185(1) 210-220.

[40] Sato T, Fukui T, Atomi H, Imanaka T. Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. Appl Environ Microbiol, 2005;71(7) 3889-3899.

[41] Pan M, Santangelo TJ, Li Z, Reeve JN, Kelman Z. *Thermococcus kodakarensis* encodes three MCM homologs but only one is essential. Nucleic Acids Res, 2011;39(22) 9671-9680.

[42] Ishino S, Fujino S, Tomita H, Ogino H, Takao K, Daiyasu H, et al. Biochemical and genetical analyses of the three mcm genes from the hyperthermophilic archaeon, *Thermococcus kodakarensis*. Genes Cells, 2011;16(12) 1176-1189.

[43] Fujikane R, Ishino S, Ishino Y. Forterre, P. Genetic analysis of DNA repair in the hyperthermophilic archaeon, *Thermococcus kodakarensis*. Genes Genet Syst, 2010; 85(4)243-257.

[44] Xia Q, Hendrickson E L, Zhang Y, Wang T, Taub F, Moore BC, et al. Quantitative proteomics of the archaeon *Methanococcus maripaludis* validated by microarray analysis and real time PCR. Mol Cell Proteomics, 2006;5(5) 868-881.

[45] Walters AD, Chong JP. An archaeal order with multiple minichromosome maintenance genes. Microbiology, 2010;156(Pt 5):1405-1414.

[46] Sakakibara N, Kelman LM, Kelman Z. How is the archaeal MCM helicase assembled at the origin? Possible mechanisms Biochem. Soc. Trans. 2007;37(Pt 1):7-11.

[47] Shin J-H, Heo GY, Kelman Z. The *Methanothermobacter thermautotrophicus* Cdc6-2 protein, the putative helicase loader, dissociates the minichromosome maintenance helicase. J Bacteriol, 2008;190(11) 4091-4094.

[48] Takayama Y, Kamimura Y, Okawa M, Muramatsu S, Sugino A, Araki H. et al. GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. Genes Dev, 2003;17(9)1153-1165.

[49] Makarova KS, Wolf YI, Mekhedov SL, Mirkin BG, Koonin EV. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. Nucleic Acids Res, 2005;33(14) 4626-4638.

[50] Marinsek N, Barry ER, Makarova KS, Dionne I, Koonin EV, Bell SD. GINS, a central nexus in the archaeal DNA replication fork. EMBO Rep, 2006;7(5) 539-545.

[51] Pacek M, Tutter AV, Kubota Y, Takisawa H, Walter JC. Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. Mol Cell, 2006;21(4) 581-587.

[52] Ogino H, Ishino S, Mayanagi K, Haugland GT, Birkeland NK, Yamagishi A, et al. The GINScomplex from the thermophilic archaeon, *Thermoplasma acidophilum* may function as a homotetramer in DNA replication. Extremophiles, 2011;15(4) 529-539.

[53]  Oyama T, Ishino S, Fujino S, Ogino H, Shirai T, Mayanagi K, et al. Architectures of archaeal GINS complexes, essential DNA replication initiation factors. BMC Biol, 2011;9 28.

[54]  Kamada K, Kubota Y, Arata T, Shindo Y, Hanaoka F. Structure of the human GINS complex and its assembly and functional interface in replication initiation. Nat Struct Mol Biol, 2007;14(5)388-396.

[55]  Choi JM, Lim HS, Kim JJ, Song OK, Cho Y. Crystal structure of the human GINS complex. Genes Dev, 2007;21(11) 1316-1321.

[56]  Chang, YP, Wang G, BermudezV, Hurwitz J, Chen XS, et al. Crystal structure of the GINS complex and functional insight into its role in DNA replication. Proc Natl Acad Sci USA, 2007;104(31) 12685-12690.

[57]  Sanchez-Pulido L, Ponting CP. Cdc45: the missing RecJ ortholog in eukaryotes? Bioinformatics, 2011;27(14) 1885-1888.

[58]  Li Z, Pan M, Santangelo T J, Chemnitz W, Yuan W, Edwards JL, et al. A novel DNA nuclease is stimulated by association with the GINS complex. Nucleic Acids Res, 2011;39(14) 6114-6123.

[59]  Krastanova I, Sannino V, Amenitsch H, Gileadi O, Pisani FM, Onesti S., et al. Structural and functional insights into the DNA replication factor Cdc45 reveal an evolutionary relationship to the DHH family of phosphoesterases. J Biol Chem, 2011;287(6) 4121-4128.

[60]  Kirk BW, Kuchta RD. Arg304 of human DNA primase is a key contributor to catalysis and NTP binding: primase and the family X polymerases share significant sequence homology. Biochemistry, 1999;38(31) 7727-7736

[61]  Desogus G, Onesti S, Brick P, Rossi M, Pisani FM. Identification and characterization of a DNA primase from the hyperthermophilic archaeon *Methanococcus jannaschii.* Nucleic Acids Res, 1999;27(22) 4444-4450.

[62]  Bocquier A, Liu L, Cann I, Komori K, Kohda D, Ishino Y. Archaeal primase: bridging the gap between RNA and DNA polymerase. Curr Biol, 2001;11(6) 452-456.

[63]  Liu L, Komori K, Ishino S, Bocquier AA, Cann IK, Kohda D, et al. The archaeal DNA primase: biochemical characterization of the p41-p46 complex from *Pyrococcus furiosus.* J Biol Chem, 2001;276(48) 45484-45490.

[64]  Lao-Sirieix SH, Bell SD. The heterodimeric primase of the hyperthermophilic archaeon *Sulfolobus solfataricus* possesses DNA and RNA primase, polymerase and 3'-terminal nucleotidyl transferase activities. J Mol Biol, 2004;344(5) 1251-1263.

[65]  Lao-Sirieix SH, Nookala RK, Roversi P, Bell SD, Pellegrini L.et al. Structure of the heterodimeric core primase. Nat Struct Mol Biol, 2005;12(12) 1137-1144.

[66] De Falco M, Fusco A, DeFelice M, Rossi M, Pisani FM. The DNA primase is activated by substrates containing a thymine-rich bubble and has a 3'-terminal nucleotidyl-transferase activity. Nucleic Acids Res, 2004;32(17) 5223-5230.

[67] Ito N, Nureki O, Shirouzu M, Yokoyama S, Hanaoka F. Crystal structure of the *Pyrococcus horikoshii* DNA primase-UTP complex: implications for the mechanism of primer synthesis. Genes Cells, 2003;8(12) 913-923.

[68] Matsui E, Nishio M, Yokoyama H, Harata K, Damas S, Matsui I. Distinct domain functions regulating de novo DNA synthesis of thermostable DNA primase from hyperthermophile *Pyrococcus horikoshii.* Biochemistry. 2003;42(50) 14968-14976.

[69] Ito N, Matsui I, Matsui E. Molecular basis for the subunit assembly of the primase from an archaeon *Pyrococcus horikoshii.* FEBS J, 2007;274(5):1340-1351.

[70] Le Breton M, Henneke G, Norais C, Flament D, Myllykallio H, Querellou J,et al. The heterodimeric primase from the euryarchaeon *Pyrococcus abyssi*: a multifunctional enzyme for initiation and repair? J Mol Biol, 2007;374(5) 1172-1185.

[71] Hu J, Guo L, Wu K, Lang S, Huang L. Template-dependent polymerization across discontinuous templates by the heterodimeric primase from the hyperthermophilic archaeon *Sulfolobus solfataricus.* Nucleic Acids Res, 2011;40(8) 3470-3483.

[72] Evguenieva-Hackenberg E, Walter P, Hochleitner E, Lottspeich F, Klug G. An exosome-like complex in *Sulfolobus solfataricus.* EMBO Rep, 2003;4(9) 889-934.

[73] Walter P, Klein F, Lorentzen E, Ilchmann A, Klug G, Evguenieva-Hackenberg E. et al. Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon *Sulfolobus solfataricus.* Mol Microbiol, 2006;62(4) 1076-1089.

[74] Zuo Z, Rodgers CJ, Mikheikin AL, Trakselis MA. Characterization of a functional DnaG-type primase in archaea: implications for a dual-primase system. J Mol Biol, 2010;397(3) 664-676.

[75] Wu K, Lai X, Guo X, Hu J, Xiang X, Huang L. Interplay between primase and replication factor C in the hyperthermophilic archaeon *Sulfolobus solfataricus*. Mol Microbiol, 2007;63(3) 826-837.

[76] Murzin AG. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J, 1993;12(3) 861-867.

[77] Braun KA, Lao Y, He Z, Ingles CJ, Wold MS. Wold MS Role of protein-protein interactions in the function of replication protein A (RPA): RPA modulates the activity of DNA polymerase alpha by multiple mechanisms. Biochemistry, 1997;36(28) 8443-8454.

[78] Lin YL, Chen C, Keshav KF, Winchester E, Dutta A. Dissection of functional domains of the human DNA replication protein complex replication protein A. J Biol Chem, 1996;271(29) 17190-17198.

[79] Bochkarev A, Bochkareva E, Frappier L, Edwards AM. The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. EMBO J, 1999;18(16) 4498-4504.

[80] Bochkareva E, Frappier L, Edwards A M, Bochkarev A.The RPA32 subunit of human replication protein A contains a single-stranded DNA-binding domain. J Biol Chem, 1998;273(7) 3932-3936.

[81] Philipova D, Mullen JR, Maniar HS, Lu J, Gu C, Brill SJ, et al. A hierarchy of SSB protomers in replication protein A. Genes Dev, 1996;10(17) 2222-2233.

[82] Gomes XV, Wold MS. Structural analysis of human replication protein A. Mapping functional domains of the 70-kDa subunit. J Biol Chem, 1995;270(9) 4534-4543.

[83] Mer G, Bochkarev A, Gupta R, Bochkareva E, Frappier L, Ingles CJ. Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. Cell, 2000;103(3) 449-456.

[84] Dornreiter I, Erdile LF, Gilbert IU, von Winkler D, Kelly TJ, Fanning E. Interaction of DNA polymerase alpha-primase with cellular replication protein A and SV40 T antigen. EMBO J, 1992;11(2) 769-776.

[85] Tsurimoto T, Stillman B. Multiple replication factors augment DNA synthesis by the two eukaryotic DNA polymerases, alpha and delta. EMBO J, 1989;8(12) 3883-3889.

[86] Kenny MK, Lee SH, Hurwitz J. Multiple functions of human single-stranded-DNA binding protein in simian virus 40 DNA replication: single-strand stabilization and stimulation of DNA polymerases alpha and delta. Proc Natl Acad Sci USA, 1989;86(24) 9757-9761.

[87] Chedin F, Seitz EM, Kowalczykowski SC. Novel homologs of replication protein A in archaea: implications for the evolution of ssDNA-binding proteins. Trends Biol Sci, 1998;23(8) 273-277.

[88] Kelly TJ, Simancek P, Brush GS. Identification and characterization of a single-stranded DNA-binding protein from the archaeon *Methanococcus jannaschii.* Proc Natl Acad Sci USA, 1998;95(25) 14634-14639.

[89] Kelman Z, Pietrokovski S, Hurwitz J. Isolation and characterization of a split B-type DNA polymerase from the archaeon *Methanobacterium thermoautotrophicum* ΔH. J Biol Chem, 1999;274(40) 28751-28761.

[90] Komori K, Ishino Y. Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. J Biol Chem, 2001;276(28) 25654-25660.

[91] Kerr ID, Wadsworth RI, Cubeddu L, Blankenfeldt W, Naismith JH, White MF Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein. EMBO J, 2003;22(11) 2561-270.

[92] Robbins JB, Murphy MC, White BA, Mackie RI, Ha T, Cann IK.et al. Functional analysis of multiple single-stranded DNA-binding proteins from *Methanosarcina acetivor-*

*ans* and their effects on DNA synthesis by DNA polymerase BI. J Biol Chem, 2004;279(8) 6315-6326.

[93] Lin Y, Lin LJ, Sriratana P, Coleman K, Ha T, Spies M, et al. Engineering of functional replication protein a homologs based on insights into the evolution of oligonucleotide/oligosaccharide-binding folds. J Bacteriol, 2008;190(17) 5766-5780.

[94] Braithwaite DK, Ito J. Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res, 1993, 1993;21(4) 787-802.

[95] Cann I, Ishino Y. Archaeal DNA replication: Identifying the pieces to solve a puzzle. Genetics, 1999;152(4) 1249-1267.

[96] Ishino, Y., Cann, I. The euryarchaeotes, a subdomain of archaea, survive on a single DNA polymerase: fact or farce? Genes Genet Syst, 1998;73(6) 323-336.

[97] Lipps G, Röther S, Hart C, Krauss G. A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. EMBO J, 2003;22(10) 2516-2525.

[98] Ohmori H, Friedberg EC, Fuchs RP, Goodman MF, Hanaoka F, Hinkle D, et al. The Y-family of DNA polymerases. Mol Cell;8(1) 7-8.

[99] Guo C, Kosarek-Stancel JN, Tang TS Friedberg EC. Y-family DNA polymerases in mammalian cells. Cell Mol Life Sci, 2009;66(14) 2363-2381.

[100] Pisani FM, De Martino C, Rossi M. A DNA polymerase from the archaeon *Sulfolobus solfataricus* shows sequence similarity to family B DNA polymerases. Nucleic Acids Res, 1992;20(11) 2711-2716.

[101] Perler, FB, Comb, DG, Jack, WE, Moran, LS, Qiang, B, Kucera, RB, et al. Intervening sequences in an Archaea DNA polymerase gene. Proc. Natl. Acad. Sci. USA 1992;89(12) 5577-5581.

[102] Uemori T, Ishino Y, Toh H, Asada K, Kato I. Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*. Nucleic Acids Res, 1992;21(2) 259-265.

[103] Uemori T, Ishino Y, Doi H, Kato I. The hyperthermophilic archaeon *Pyrodictium occultum* has two alpha-like DNA polymerases. J Bacteriol, 1995;177(8) 2164-2177.

[104] Cann I, Ishino S, Nomura N, Sako Y, Ishino Y. Two family B DNA polymerases in *Aeropyrum pernix*, an obligate aerobic hyperthermophilic crenarchaeote. J Bacteriol, 1999;181(19) 5984-5992.

[105] Cann I, Komori K, Toh H, Kanai S, Ishino Y. A heterodimeric DNA polymerase: evidence that members of the euryarchaeota possess a novel DNA polymerase. Proc. Natl. Acad. Sci. USA, 1998;95(24) 14250-14255.

[106] Ishino S, Ishino Y. Conprehensive search for DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*. Nucleosides, Nucleotides, and Nucleic Acids. 2006;25(4-6):681-691.

[107] Imamura M, Uemori T, Kato I, Ishino Y.A non-$\alpha$-like DNA polymerase from the hyperthermophilic archaeon *Pyrococcus furiosus.* Biol Pharm Bull, 1995;18(12) 1647-1652.

[108] Ishino Y, Ishino S. Novel DNA polymerases from Euryarchaeota. Meth Enzymol, 2001, 334 249-260

[109] Gueguen Y, Rolland JL, Lecompte O, Azam P, Le Romancer G, Flament D, et al. Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi.* Eur J Biochem, 2001;268(22) 5961-5969.

[110] Shen Y, Musti K, Hiramoto M, Kikuchi H, Kawarabayashi Y, Matsui I. Invariant Asp-1122 and Asp-1124 are essential residues for polymerization catalysis of family D DNA polymerase from *Pyrococcus horikoshii.* J Biol Chem, 2001;276(29) 27376-27383.

[111] Shen Y, Tang X, Matsui I. Subunit interaction and regulation of activity through terminal domains of the family D DNA polymerase from *Pyrococcus horikoshii.* J Biol Chem, 2003;278(23) 21247-21257.

[112] Tang XF, Shen Y, Matsui E, Matsui I. Domain topology of the DNA polymerase D complex from a hyperthermophilic archaeon *Pyrococcus horikoshii.* Biochemistry, 2004;43(37) 11818-11827.

[113] Jokela M, Eskelinen A, Pospiech H, Rouvinen J, Syväoja JE. Characterization of the 3' exonuclease subunit DP1 of *Methanococcus jannaschii* replicative DNA polymerase D. Nucleic Acids Res, 2004;32(8)2430-2440.

[114] Henneke G, Flament D, Hübscher U, Querellou J, Raffin JP. The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. J Mol Biol, 2005;350(1) 53-64.

[115] Castrec B, Laurent S, Henneke G, Flament D, Raffin JP. The glycine-rich motif of *Pyrococcus abyssi* DNA polymerase D is critical for protein stability. J Mol Biol, 2010;396(4) 840-848.

[116] Berquist BR, DasSarma P, DasSarma S. Essential and non-essential DNA replication genes in the model halophilic Archaeon, *Halobacterium* sp. NRC-1. BMC Genet, 2007;8 31.

[117] Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nat Rev Microbiol, 2008;6(3) 245-252.

[118] Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H, et al. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic Acids Res, 2011;39(8) 3204-3223.

[119] Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L, et al. A korarchaeal genome reveals insights into the evolution of the Archaea. Proc Natl Acad Sci USA, 2008;105(23) 8102-8107.

[120] Tahirov TH, Makarova KS, Rogozin IB, Pavlov YI, Koonin EV. 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, 2009;4 11.

[121] Zillig W, Kletzin A, Schleper C, et al. Screening for *Sulfolobales*, their plasmids and their viruses in Icelandic solfataras. Syst Appl Microbiol, 1993, 16: 609–628

[122] Lipps G, Röther S, Hart C, Krauss G. A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. EMBO J, 2003;22(10) 2516-2525.

[123] Lipps G. Molecular biology of the pRN1 plasmid from *Sulfolobus islandicus.* Biochem Soc Trans, 2009;37(Pt 1) 42-45.

[124] Lipps G. The replication protein of the *Sulfolobus islandicus* plasmid pRN1. Biochem Soc Trans, 2004;32(Pt 2) 240-244.

[125] Soler N, Marguet E, Cortez D, Desnoues N, Keller J, van Tilbeurgh H. Two novel families of plasmids from hyperthermophilic archaea encoding new families of replication proteins. Nucleic Acids Res, 2010;38(15) 5088-5104.

[126] Moldovan G-L, Pfander B, Jentsch S. PCNA, the maestro of the replication fork. Cell, 2007;129(4)665-679.

[127] Pan M, Kelman L, Kelman Z. The archaeal PCNA proteins. Biochem Soc Trans, 2011;39(1) 20-24.

[128] Cann I, Ishino S, Hayashi I, Komori K, Toh H, Morikawa K, et al. Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. J. Bacteriol, 1999, Nov;181(21):6591-6599.

[129] Ishino Y, Tsurimoto T, Ishino S, Cann IK. Functional interactions of an archaeal sliding clamp with mammalian clamp loader and DNA polymerase δ. Genes Cells, 2001;6(8) 699-706.

[130] Matsumiya S, Ishino S, Ishino Y, Morikawa K. Physical interaction between proliferating cell nuclear antigen and replication factor C from *Pyrococcus furiosus.* Genes Cells, 2002;7(9) 911-922.

[131] Matsumiya S, Ishino S, Ishino Y, Morikaw K. Intermolecular ion pairs maintain toroidal structure of *Pyrococcus furiosus* PCNA. Prot Sci, 2003;12(4) 823-831.

[132] Tori K, Kimizu M, Ishino S, Ishino Y. Both DNA polymerase BI and D from the hyperthermophilic archaeon, *Pyrococcus furiosus* bind to PCNA at the C-terminal PIP box motifs. J Bacteriol, 2007;189(15) 5652-5657.

[133] Cann I, Ishino S, Yuasa M, Daiyasu H, Toh H, Ishino Y. Biochemical analysis of the replication factor C from the hyperthermophilic archaeon *Pyrococcus furiosus.* J Bacteriol, 2001;183(8) 2614-2623.

[134] Mayanagi K, Miyata T, Oyama T, Ishino Y, Morikawa K. Three-dimensional electron microscopy of clamp loader small subunit from *Pyrococcus furiosus.* J Struct Biol, 200;134(1) 35-45.

[135] Ishino S, Oyama T, Yuasa M, Morikawa K, Ishino Y. Mutational Analysis of *Pyrococcus furiosus* Replication Factor C based on the Three-Dimensional Structure. Extremophiles, 2003;7(3) 169-175.

[136] Oyama T, Ishino Y, Cann I, Ishino S, Morikawa K. Atomic structure of the clamp loader small subunit from *Pyrococcus furiosus*. Mol Cell, 2001;8(2) 455-463.

[137] Indiani C, O'Donnell M. The replication clamp-loading machine at work in the three domains of life. Nat Rev Mol Cell Biol, 2006;7(10) 751-761.

[138] Miyata T, Suzuki H, Oyama T, Mayanagi K, Ishino Y, Morikawa K. Open clamp structure in the clamp-loading complex visualized by electron microscopic image analysis. Proc Natl Acad Sci USA, 2005;102(39) 13795-13800.

[139] Kelch BA, Makino DL, O'Donnell M, Kuriyan J. The replication clamp-loading machine at work in the three domains of life. Science, 2011;334(6063) 1675-1680.

[140] Warbrick EM. The puzzle of PCNA's many partners. BioEssays, 2000;22(11) 997-1006.

[141] Vivona JB, Kelman Z. The diverse spectrum of sliding clamp interacting proteins. FEBS Lett, 2003;546(2-3) 167-172.

[142] Nishida H, Mayanagi K, Kiyonari S, Sato Y, Oyama T, Ishino Y, et al. Structural determinant for switching between the polymerase and exonuclease modes in the PCNA-replicative DNA polymerase complex. Proc Natl Acad Sci USA, 2009;106(49) 20693-20698.

[143] Shamoo Y, Steitz TA. Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. Cell, 1999;99(2) 155-166.

[144] Franklin MC, Wang J. Steitz TA. Structure of the replicating complex of a pol α family DNA polymerase. Cell, 2001;105(5) 657-667.

[145] Mayanagi K, Kiyonari S, Nishida H, Saito M, Kohda D, Ishino Y, et al. Architecture of the DNA polymerase B-proliferating cell nuclear antigen (PCNA)-DNA ternary complex. Proc Natl Acad Sci USA, 2011;108(5):1845-1849

[146] Daimon K, Kawarabayasi Y, Kikuchi H, Sako Y, Ishino Y. Three proliferating cell nuclear antigen-like proteins found in the hyperthermophilic archaeon *Aeropyrum pernix*: interactions with the two DNA polymerases. J Bacteriol, 2002;184(3) 687-694.

[147] Dionne I, Nookala RK, Jackson SP, Doherty AJ, Bell SD. A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. Mol Cell, 2003;11(1) 275-282.

[148] Imamura K, Fukunaga K, Kawarabayasi Y, Ishino Y. Specific interactions of three proliferating cell nuclear antigens with replication-related proteins in *Aeropyrum pernix*. Mol Microbiol, 2007;64(2) 308-318.

[149] Williams GJ, Johnson K, Rudolf J, McMahon SA, et al. Structure of the heterotrimeric PCNA from *Sulfolobus solfataricus.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2006;62(Pt 10) 944-948.

[150] Doré AS, Kilkenny ML, Jones SA, Oliver AW, Roe SM, Bell SD, et al. Structure of an archaeal PCNA1-PCNA2-FEN1 complex: elucidating PCNA subunit and client enzyme specificity. Nucleic Acids Res, 2006;34(16) 4515-4526.

[151] Kuba Y., Ishino, S., Yamagami, T., Tokuhara, M., Kanai, T., Fujikane, R., et al. Comparative analyses of the two PCNAs from the hyperthermophilic archaeon, *Thermococcus lodakarensis.* Genes Cells, 2012, 17, in press

[152] Ladner JE, Pan M, Hurwitz J, Kelman Z. Crystal structures of two active proliferating cell nuclear antigens (PCNAs) encoded by *Thermococcus kodakaraensis.* Proc Natl Acad Sci USA, 2011;108(7) 2711-2716.

[153] Chen YH, Kocherginskaya SA, Lin Y, Sriratana B, Lagunas AM, Robbins JB, et al. Biochemical and mutational analyses of a unique clamp loader complex in the archaeon *Methanosarcina acetivorans.* J Biol Chem, 2005;280(51) 41852-41863.

[154] Chen YH, Lin Y, Yoshinaga A, Chhotani B, Lorenzini JL, Crofts AA, et al. Molecular analyses of a three-subunit euryarchaeal clamp loader complex from *Methanosarcina acetivorans.* J Bacteriol, 2009;191(21) 6539-6549.

[155] Wilkinson A, Day J, Bowater R. Bacterial DNA ligases. Mol Microbiol, 2001;40(6) 1241-1248.

[156] Tomkinson AE, Vijayakumar S, Pascal JM, Ellenberger T. DNA ligases: structure, reaction mechanism, and function. Chem Rev, 2006;106(2) 687-699.

[157] Poidevin L, MacNeill SA. Biochemical characterisation of LigN, an NAD+-dependent DNA ligase from the halophilic euryarchaeon *Haloferax volcanii* that displays maximal in vitro activity at high salt concentrations. BMC Mol Biol, 2006;7 44.

[158] Benarroch D, Shuman S. Characterization of mimivirus NAD$^+$-dependent DNA ligase. Virology, 2006;353(1) 133-143.

[159] Kletzin A. Molecular characterisation of a DNA ligase gene of the extremely thermophilic archaeon *Desulfurolobus ambivalens* shows close phylogenetic relationship to eukaryotic ligases. Nucleic Acids Res, 1992;20(20) 5389-5396.

[160] Nakatani M, Ezaki S, Atomi H, Imanaka T. A DNA ligase from a hyperthermophilic archaeon with unique cofactor specificity. J Bacteriol, 2000;182(22) 6424-6433.

[161] Rolland JL, Gueguen Y, Persillon C, Masson JM, Dietrich J. Characterization of a thermophilic DNA ligase from the archaeon *Thermococcus fumicolans.* FEMS Microbiol Lett, 2004;236(2)267-273.

[162] Kim YJ, Lee HS, Bae SS., Jeon JH, Yang SH, Lim JK, et al. Cloning, expression, and characterization of a DNA ligase fro3 a hyperthermophilic archaeon *Thermococcus* sp. Biotechnol Lett, 2006;28(6) 401-407.

[163] Ferrer M, Golyshina OV, Beloqui A, Böttger LH, Andreu JM, Polaina J, et al. A purple acidophilic di-ferric DNA ligase from *Ferroplasma*. Proc Natl Acad Sci USA, 2008;105(26) 8878-8883.

[164] Jeon SJ, Ishikawa K. A novel ADP-dependent DNA ligase from *Aeropyrum pernix* K1. FEBS Lett, 2003;550(1-3) 69-73.

[165] Seo MS, Kim YJ, Choi JJ, Lee MS, Kim JH, Lee JH, et al. Cloning and expression of a DNA ligase from the hyperthermophilic archaeon *Staphylothermus marinus* and properties of the enzyme. J Biotechnol, 2007;128(3) 519-530.

[166] Sun Y, Seo MS, Kim JH, Kim YJ, Kim GA, Lee JI, et al. Novel DNA ligase with broad nucleotide cofactor specificity from the hyperthermophilic crenarchaeon *Sulfophobococcus zilligii*: influence of ancestral DNA ligase on cofactor utilization. Environ Microbiol, 2008;10(12) 3212-3224.

[167] Keppetipola N, Shuman S. Characterization of a thermophilic ATP-dependent DNA ligase from the euryarchaeon *Pyrococcus horikoshii*. J Bacteriol, 2005;187(20) 6902-6908.

[168] Kiyonari S, Takayama K, Nishida H, Ishino Y. Identification of a novel binding motif in *Pyrococcus furiosus* DNA ligase for the functional interaction with proliferating cell nuclear antigen. J Biol Chem, 2006;281(38) 28023-28032.

[169] Nishida H, Kiyonari S, Ishino Y, Morikawa K. The closed structure of an archaeal DNA ligase from *Pyrococcus furiosus.* J Mol Biol, 2006;360(5) 956-967.

[170] Levin DS, Bai W, Yao N. An interaction between DNA ligase I and proliferating cell nuclear antigen: implications for Okazaki fragment synthesis and joining. Proc Natl Acad Sci USA, 1997;94(24) 12863-12868.

[171] Jónsson ZO, Hindges R, Hübscher U. Regulation of DNA replication and repair proteins through interaction with the front side of proliferating cell nuclear antigen. EMBO J, 1998;17(8) 2412-2425.

[172] Tom S, Henricksen LA, Park MS, Bambara RA. DNA ligase I and proliferating cell nuclear antigen form a functional complex. J Biol Chem, 2001;276(27) 24817-24825.

[173] Kiyonari S, Kamigochi T, Ishino Y. A single amino acid substitution in the DNA-binding domain of *Aeropyrum pernix* DNA ligase impairs its interaction with proliferating cell nuclear antigen, Extremophiles, 2007;11(5) 675-684.

[174] Pascal JM, O'Brien PJ, Tomkinson AE, Ellenberger T. Human DNA ligase I completely encircles and partially unwinds nicked DNA. Nature, 2004;432(7016) 473-478:

[175] Pascal JM, Tsodikov OV, Hura GL, Song W, Cotner EA, Classen S, et al. A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. Mol Cell, 2006;24(2) 279-291.

[176] Mayanagi K, Kiyonari S, Saito M, Shirai T, Ishino Y, Morikawa K. Mechanism of replication machinery assembly as revealed by the DNA ligase-PCNA-DNA complex architecture. Proc Natl Acad Sci USA, 2009;106(12) 4647-4652.

[177] Liu Y, Kao HI, Bambara RA. Flap endonuclease 1: a central component of DNA metabolism. Annu Rev Biochem, 2004, 2004;73 589-615.

[178] Zheng L, Jia J, Finger LD, Guo Z, Zer C, Shen B. Functional regulation of FEN1 nuclease and its link to cancer. Nucleic Acids Res, 2011;39(3) 781-794.

[179] Rossi ML, Bambara RA. Reconstituted Okazaki fragment processing indicates two pathways of primer removal. J Biol Chem, 2006;281(36):26051-26061.

[180] Tom S, Henricksen LA, Bambara RA. Mechanism whereby proliferating cell nuclear antigen stimulates flap endonuclease 1. J Biol Chem, 2000;275(14) 10498-10505.

[181] Sakurai S, Kitano K, Yamaguchi H, Hamada K, Okada K, Fukuda K, et al. Structural basis for recruitment of human flap endonuclease 1 to PCNA. EMBO J, 2005;24(4) 683-693.

[182] Shen B, Qiu J, Hosfield D, Tainer JA. Flap endonuclease homologs in archaebacteria exist as independent proteins. Trends Biochem Sci, 1998;23(5) 171-173.

[183] Hwang KY, Baek K, Kim HY, CHo Y. The crystal structure of flap endonuclease-1 from *Methanococcus jannaschii*. Nat Struct Biol, 1998;5(8) 707-713.

[184] Hosfield DJ, Mol CD, Shen B, Tainer JA. Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. Cell, 1998;95(1)135-146.

[185] Matsui E, Musti KV, Abe J, Yamasaki K, Matsui I, Harata K. Molecular structure and novel DNA binding sites located in loops of flap endonuclease-1 from *Pyrococcus horikoshii*. J Biol Chem, 2002;277(40) 37840-37847.

[186] Chapados BR, Hosfield DJ, Han S, Qiu J, Yelent B, Shen B, et al. Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. Cell, 2004;116(1) 39-50.

[187] Matsui E, Kawasaki S, Ishida H, Ishikawa K, Kosugi Y, Kikuchi H, et al. Thermostable flap endonuclease from the archaeon, *Pyrococcus horikoshii*, cleaves the replication fork-like structure endo/exonucleolytically. J Biol Chem, 1999;274(26) 18297-182309

[188] Matsui E, Abe J, Yokoyama H, Matsui I. Aromatic residues located close to the active center are essential for the catalytic reaction of flap endonuclease-1 from hyperthermophilic archaeon *Pyrococcus horikoshii*. J Biol Chem, 2004;279(16) 16687-16696.

[189] Beattie TR, Bell SD. Molecular machines in archaeal DNA replication. Curr Opin Chem Biol, 2011;15(5) 614-619.

[190] Ishino S and Ishino Y. Rapid progress of DNA replication studies in Archaea, the third domain of life. Sci China Life Sci, 2012;55(5) 386-403.

[191] Wang Y, Prosen DE, Mei L, Sullivan JC, Finney M, Vander Horn PB. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic Acids Res, 2004;32(3) 1197-1207

[192] Kawamura A, Ishino Y, Ishino S. Biophysical analysis of PCNA from *Pyrococcus furiosus.* J Jap Soc Extremophiles, 2012;11(1) 12-18

[193] Ishino S, Kawamura, A, Ishino Y. Application of PCNA to processive PCR by reducing the stability of its ring structure. J Jap Soc Extremophiles, 2012;11(1) 19-25

[194] Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol, 2008;26(10) 1135-1145.

[195] Ansorge WJ. Next-generation DNA sequencing techniques. Nat Biotechnol, 2009;25(4) 195-203

[196] Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet, 2010;11(1) 31-46.

# Proposal for a Minimal DNA Auto-Replicative System

Agustino Martinez-Antonio,
Laura Espindola-Serna and Cesar Quiñones-Valles

Additional information is available at the end of the chapter

## 1. Introduction

DNA replication allows cell division and population growth of living organisms. Here we will focus on DNA replication in prokaryotic single celled microorganisms. Several excellent reviews of the molecular processes that carry out DNA replication in bacteria already exist, *E. coli* being the model described in most detail (Langston LD et al., 2009; Quiñones-Valles et al., 2011). Briefly, the process begins when DnaA (DNA initiator replication protein) in its activated form (DnaA-ATP) recognizes and binds the *oriC* (origin of replication on the bacterial chromosome). In the following step, the replisome is assembled and binds to the complex of DnaA-ATP at the *oriC*. Next, the DNA strands are separated and synthesis of the complementary strands initiates followed by elongation steps. The molecular mechanisms of elongation differ depending on the strand used as a template; the leading strand is replicated continuously starting from a unique RNA primer, whereas on the lagging strand DNA polymerase III must recognize several RNA primers, previously synthesized by DnaG, and then replicate each DNA fragment (Okazaki fragments). This is followed by the replacement of RNA primers by DNA polymerase I, and removal of nicks by a DNA ligase. The whole process concludes when replisomes reach the *ter* site, almost opposite to *oriC* on the circular DNA molecule. Tus proteins are attached to the *ter* sites and when replisomes reach these complexes, they collide and finally are disassembled (see Figure 1 for an overview of the whole process).

From another aspect, one of the more challenging areas of Synthetic Biology is the design and construction of minimal cells. The accomplishment of this aim might contribute to answering basic questions about the minimal components necessary to sustain life systems, in addition to cell auto-organization, function and evolution. In a practical application, mini-

mal cells can be used as a background chassis for the generation of dedicated biological systems designed for the synthesis or degradation of diverse compounds of interest.
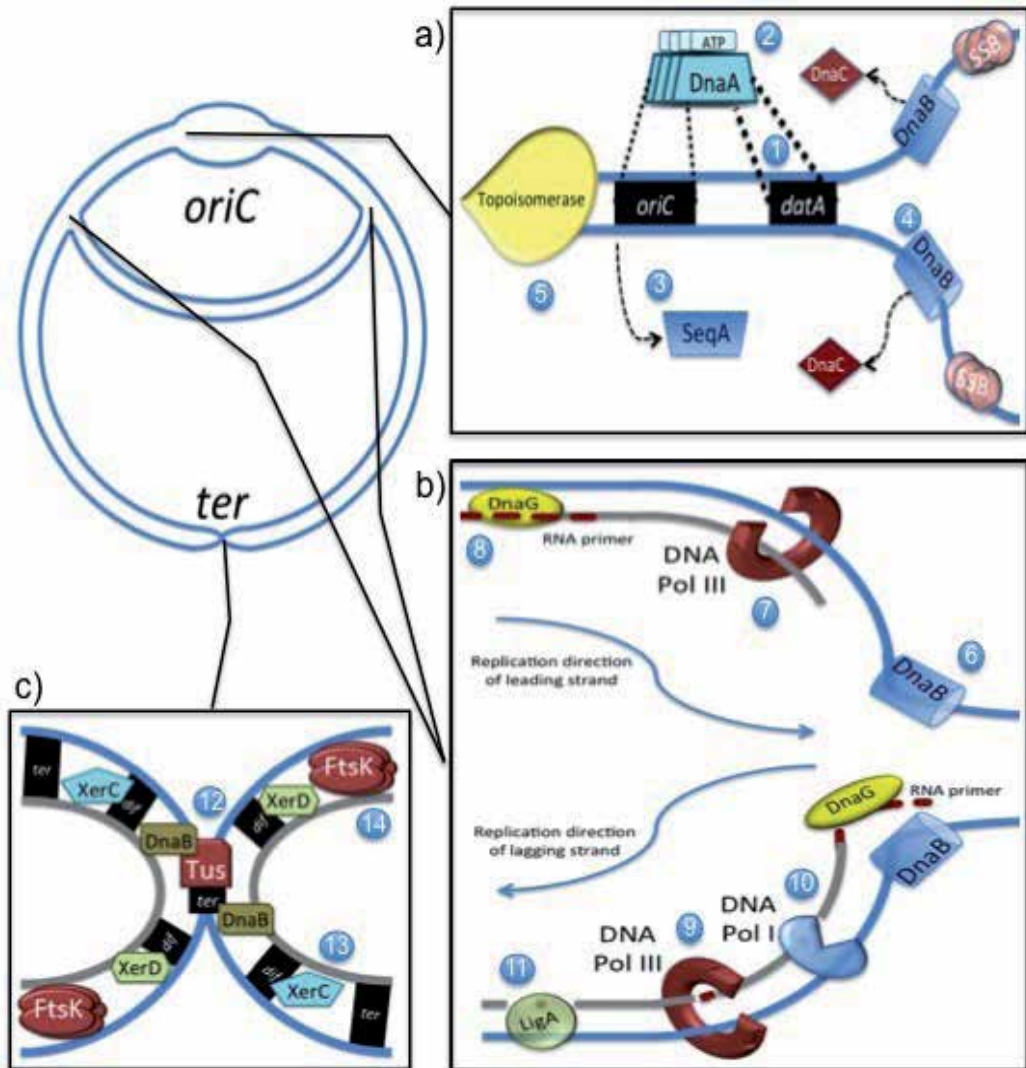


**Figure 1. Main steps of DNA replication in bacteria.** a) Initiation of DNA replication; the *datA* locus has a high affinity for binding DnaA (1). DnaA binds to ATP, homo-multimers of DnaA-ATP are formed (2). These homo-multimers bind to *oriC* and once replication is initiated SeqA binds this region and prevents initiation of a new replication event (3). The SSB (single strand binding) protein and DnaB assist the complex to open the DNA strands and release DnaC (4). A DNA topoisomerase helps to further unfold the DNA strands (5). b) The elongation phase; the replication fork is formed and the replisome is assembled (6). DNA polymerase III replicates the leading strand (7). DnaG incorporates RNA primers as primers for replication of the lagging strand (8). Polymerase III can now replicate Okazaki fragments on the lagging strand (9). DNA polymerase I replaces RNA nucleotides for DNA nucleotides (10). A DNA ligase (LigA) seals the nicks on contiguous DNA fragments (11). c) Termination of DNA replication; The protein Tus binds to the *ter* sites, when replisomes reach Tus, replication ceases (12). The recombinases XerC and XerD resolve the replicated DNA strands (13). Finally, FtsK translocates the DNA strands and each double-stranded DNA molecule can be liberated (14).

In recent years, the essential properties and capabilities necessary to develop minimum cells have been broadly speculated (MacDonald et al., 2011). Among these characteristics it is evident that DNA replication should be a fundamental property of these biosystems. Many genes for DNA replication are found to be conserved when comparative analysis of bacterial genomes is carried out. These types of genes are considered as informational genes, in charge of maintaining the genetic code, and are among the genes less frequently found be horizontally transferable (Jain et al, 1999). Therefore by genomic comparisons and functional analyses it is possible to propose a minimum core of genes capable of supporting the process of DNA replication.

From a genetic point of view, and for the purpose of this study it is important to state our definition of a minimal DNA auto-replicative system (MiDARS) as: *a genetic system comprising the minimum number of DNA components, including regulatory elements and gene products necessary for the auto-replication of the DNA molecule on which they are encoded, functioning in an in vitro condition.*

In this chapter we will develop a proposal for the construction of such an auto-replicative DNA system. This system is designed to serve as a scaffold for the incorporation of additional biological functions such as transcription and translation, etc. For the scaffold design we exploit information of genes necessary for replication in *E. coli* that are highly conserved in bacteria with extremely reduced genomes and analyze their functional role in DNA replication in order to finally propose a minimal genetic system with a DNA auto-replicatory function.

## 2. Minimal cells and minimal genetic systems

A minimal cell can be defined as a biological system that has the minimal number of genetic parts and molecular components for supporting life functions under defined growth conditions. In other words, it includes only the necessary number of genes and derived biomolecular machinery that are considered basic to support life functions (Jewett and Forster, 2010).

The concept of life is intrinsically complex; in biochemical terms it could be defined by three basic characteristics (Luisi et al., 2006):

1. auto-regulation of metabolism,

2. auto-replication of the genetic material and,

3. controlled evolution of their components and functions.

The design and synthesis of minimal cells depends on the environmental conditions the systems will be exposed to. Initially, we might consider that a minimal cell should be exposed to the most favorable conditions in order to facilitate its conception and function. These favorable conditions will require an environment where the cell is not suffering any kind of environmental stress. Nonetheless, even this ideal scenario is a challenging condition to direct the rational design of components of a minimal cellular system since the genes for many

cellular functions are not yet totally defined. What we could do is to start to reconstruct minimal biological functions that are more or less well defined. These might be the processes relating to the central dogma of molecular biology: DNA replication, DNA transcription and mRNA translation (Figure 2). Some of these functions have been the object of different studies; e. g. transcription and translation were successfully recreated in the experiment of Asahara (2010) by separately expressing the components of the *E. coli* RNA polymerase, including the sigma70 factor and reconstituting the function of the complete enzyme *in vitro*.

Since one of the fundamental characteristics of life systems is the replication of their own genetic material, we can consider the design of minimal genetic systems that sustain DNA auto-replication as an important to starting point.
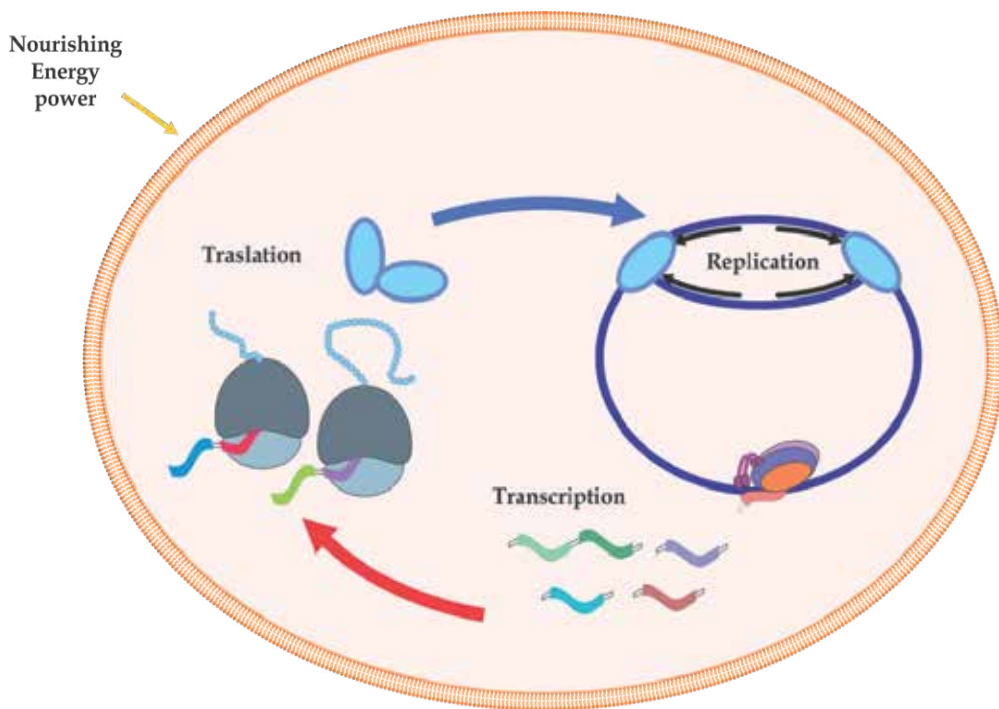
**Figure 2. Representation of a hypothetical minimal auto-replicative system.** One of the key features of the minimal cell is that it should perform basic functions such as transcription, translation and replication of the genetic information contained in its genome.

## 3. Approaches for the development of minimal genetic systems

Currently there are two approaches for the study of minimal biological systems. These are the *top down* and *bottom up* strategies (Delaye L & Moya A, 2009; Murtas, 2009). The *top down* approach considers the analysis of existing biological systems and, by following a reductionist approach, looks to minimize the number of components either by searching for con-

served genetic elements or by experimentally reducing the genome without losing functionality. This strategy was used to reduce the *E. coli* genome by 15% by deleting non-essential genes, recombinogenic and mobile DNA elements, and cryptic genes. The resulting cells had good growth profiles and showed improved performance for protein production (Pósfai et al., 2006). Another focus of this approach is to carry out comparative genomics and define a set of conserved genes such as those in charge of specific functions (Gil et al., 2004; Forster & Church, 2006).

On the other hand, the *bottom up* approach involves the construction of complex systems starting from relatively simple molecular precursors. A classical example is the experiment of Miller, who obtained amino acids from a mixture of simple organic and inorganic molecules (Miller, 1953).

Considering the design and construction of minimal genetic systems, benefits should be obtained by employing both complementary *top down* and *bottom up* approaches.

## 4. *Escherichia coli* as a model organism for the design of a minimal DNA auto-replicative system

*Escherichia coli* is a bacillary Gram-negative, aerobic, facultative and non-sporulating organism. It was discovered in 1885 by the physician Theodore von Escherich and is now classified as part of the Enterobacteriaceae family of the Gamma-proteobacterias (Blattner et al., 1997).

This bacterium lives in the intestine of mammals, and assists its hosts with assimilation of nutrients, providing some vitamins and preventing the establishment of bacterial pathogens. Since its discovery, *E. coli* has been widely used as a working model in the laboratory to study biochemistry and diverse molecular processes. In addition, it has been widely used in biotechnology as a vehicle for the expression of multiple recombinant proteins and whole metabolic pathways.

Arthur Kornberg was one of the most prominent investigators in molecular biology and a pioneer in the description of the replicative process using *E. coli* as a model. For his accomplishments in the field he was awarded the Nobel Prize in Physiology and Medicine in 1959. He discovered DNA polymerase I (Bessman et al., 1958; Lehman et al., 1958a), and describes the synthesis of DNA as a process based on the use of a single strand of DNA as a template (Lehman et al., 1958b). Later, Kornberg and his collaborators discovered additional enzymes involved in DNA replication: DNA primase, DNA helicase, DnaA, PriA among others. Nowadays, the replication process and the replicative enzymes of *E. coli* are the best understood and characterized of any organism.

From a biotechnological standpoint, *E. coli* shows three important characteristics that make it an ideal organism to serve as the platform for the design of a synthetic cellular program (Foley & Shuler, 2010):

1. functionally it is the organism best characterized at the molecular and biochemical levels in terms of components of metabolism,

2. it has proven to be a robust vehicle for the expression of multiple biotechnological processes,

3. it has a short growth cycle and is easy to manipulate genetically.

Additionally, the genome of *E. coli* serves as the principal source of standardized genetic parts for the construction of genetic circuits, the "BioBricks", in a project whose aim is to standardize genetic parts to facilitate biological engineering (http://partsregistry.org), (Smolke, 2009). Most biobricks are designed to function in *E. coli*, therefore, we think *E. coli* is the best organism of choice for the design of a DNA auto-replicative system.

## 5. Comparison of the DNA replicative machinery of *E. coli* with that of bacteria with reduced genomes

Comparative genomics is a powerful approach that allows the identification of genetic sequences sharing identity/similarity among different organisms. Through these comparisons it is possible to identify conserved genes and predict the components of the replicative machinery in several different organisms.

For our purpose, among the organisms of interest to consider in our design are those with extremely reduced genomes. A characteristic of these organisms is that they are incapable of growth in a free-living manner. The genomes of organisms with these characteristics correspond to those having the minimum number of genes possible in nature. From these we chose the 25 organisms with the most reduced genomes known to date (Table 1). All of these genomes contain less than 1,200 kbp of DNA and all are endosymbiotic bacteria, most of which are thought to survive at the expense of the host.

In these organisms, we searched for genes encoding enzymes involved in DNA replicative functions with orthology to the replicative machinery from *E. coli* (Table 2). To find orthologous genes we followed two complementary strategies: we looked for Clusters of Orthologous Groups (COGs, Tatusov et al., 2003) and also used bidirectional best blast hits (Moreno-Hagelsieb & Latimer, 2008). In Table 2 the blue cells indicate where genes orthologous to *E. coli* are present in the target organism. In the table we show orthologous genes to be present in at least fifteen of these bacteria. Remarkably, bacteria with the most reduced genomes; *Carsonella ruddii* PV (Nakabachi et al., 2006; Tamames et al., 2007), *Hodgkinia cicadicola Dsem* (McCutcheon et al., 2009) and *Tremblaya princeps* PCIT (López-Madrigal et al., 2011; McCutcheon & Moran, 2011) had only 5, 3 and 5 genes related to replication respectively which were orthologous to *E. coli*. These three organisms are strict endosymbionts of insects, with the smallest genomes known to date (Table 2). The fact that these bacteria showed fewer genes related to DNA replication in comparison to bacteria with larger genomes (Figure 3), indicates that the minimal replicative machinery in these organisms might be composed by a small number of constituents. This observation raises many open questions, for instance:

| Organism | Abrev | Total length | Total genes | Replic. genes | Classification | Symbiotic Relation/interaction | Host |
|---|---|---|---|---|---|---|---|
| Escherichia coli K-12 MG1655 | eco | 4,639,675 | 4,146 | 228 | Proteobacteria-gamma | Free living | Homo sapiens |
| Anaplasma marginale St. Maries | ama | 1,197,687 | 948 | 55 | Proteobacteria-alpha | Parasitic, Endosymbiotic | Homo sapiens |
| Midichloria mitochondrii IricVA | mmn | 1,183,732 | 1,211 | 65 | Proteobacteria-alpha | Pathogen | Ixodes ricinus |
| Chlamydophila caviae GPIC | cca | 1,181,356 | 1,005 | 85 | Chlamydiae | Parasitic, Endosymbiotic | Homo sapiens, swine |
| Ruthia magnifica Cm | rma | 1,160,782 | 976 | 61 | Proteobacteria-gamma | Endosymbiotic | Calyptogena magnifica |
| Treponema pallidum Nichols | tpa | 1,138,011 | 1,036 | 74 | Spirochaetes | Host-associated Free living | Homo sapiens |
| Rickettsia prowazekii Madrid E | rpr | 1,111,523 | 835 | 85 | Proteobacteria-alpha | Parasitic-endosymbiotic | Homo sapiens |
| Chlamydia trachomatis D/UW- | ctr | 1,042,519 | 895 | 83 | Chlamydiae | Parasitic, Endosymbiotic | Homo sapiens |
| Vesicomyosocius okutanii HA | vok | 1,022,154 | 937 | 49 | Proteobacteria-gamma | Endosymbiotic | Calyptogena okutanii |
| Tropheryma whipplei TW08/27 | tws | 925,938 | 783 | 68 | Actinobacteria | Pathogen | Homo sapiens |
| Neorickettsia sennetsu Miyayama | nse | 859,006 | 932 | 46 | Proteobacteria-alpha | Pathogen | Homo sapiens |
| Mesoplasma florum L1; ATCC 33453 | mfl | 793,224 | 682 | 85 | Firmicutes | Parasite, non-pathogen | Citrus limon |
| Ureaplasma parvum ATCC 700970 | uur | 751,719 | 614 | 52 | Tenericutes | Pathogen | Homo sapiens |
| Blochmannia floridanus | bfl | 705,557 | 583 | 37 | Proteobacteria-gamma | Endosymbiont | Camponotus floridanus |
| Wigglesworthia glossinidia | wbr | 703,004 | 617 | 41 | Proteobacteria-gamma | Endosymbiotic | Glossina brevipalpis |
| Baumannia cicadellinicola Hc | bci | 686,194 | 595 | 46 | Proteobacteria-gamma | Endosymbiotic | Insect endosymbiont |
| Buchnera aphidicola APS | buc | 655,725 | 574 | 48 | Proteobacteria-gamma | Mutualistic, Endosymbiotic | Acyrthosiphon pisum |
| Blattabacterium sp. BPLAN | bpi | 640,442 | 581 | 39 | Bacteroidetes | Endosymbiotic | Periplaneta americana |
| Riesia pediculicola USDA | rip | 582,127 | 556 | 31 | Proteobacteria-gamma | Endosymbiotic | Pediculus humanus |
| Mycoplasma genitalium G37 | mge | 580,076 | 475 | 40 | Tenericutes | Host-associated | Homo sapiens |
| Moranella endobia PCIT | men | 538,294 | 406 | 41 | Proteobacteria-gamma | Host-associated | Planococcus citri |
| Sulcia muelleri GWSS | smg | 245,530 | 227 | 11 | Bacteroidetes | Endosymbiotic | Glassy-winged sharpshooter |
| Zinderia insecticola CARI | zin | 208,564 | 202 | 15 | Proteobacteria beta | Endosymbiotic | Spittlebugs |
| Carsonella ruddii PV | cpr | 159,662 | 182 | 5 | Proteobacteria-gamma | Endosymbiotic intracellular | Pachypsylla venusta |
| Hodgkinia cicadicola Dsem | hci | 143,795 | 169 | 3 | Proteobacteria-alpha | Endosymbiotic | Diceroprocta semicincta |
| Tremblaya princeps PCIT | tpn | 138,927 | 121 | 5 | Proteobacteria-beta | Endosymbiotic | Mealybug Plano-coccus citri (Risso) |

**Table 1.** Bacteria with genome sizes less than 1200 kbp
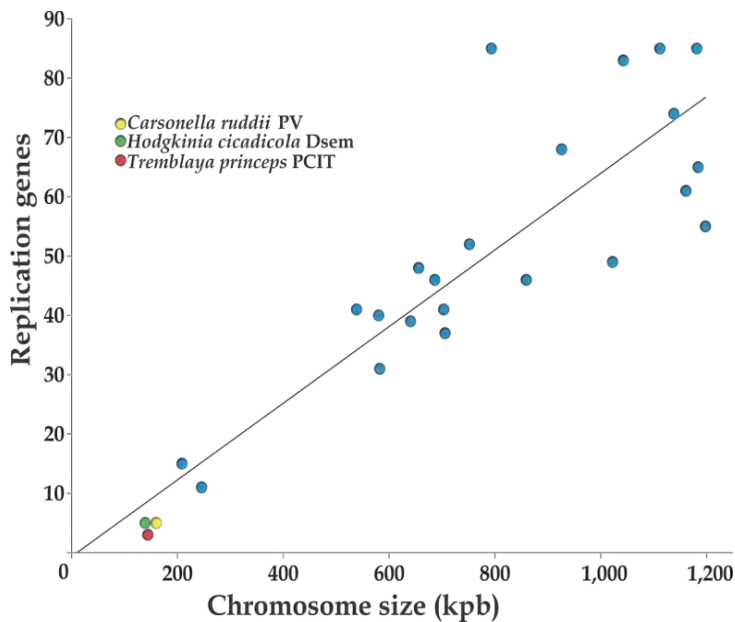
**Figure 3.** Conservation of DNA replicative machinery in bacteria with reduced genomes. Graph showing the relationship between number of genes annotated with DNA replicative functions versus genome sizes.

**i.**     Are these genes sufficient to sustain the process of replication of an entire chromosome?;

**ii.**    Does the host supply the missing elements for replication of the endosymbionts DNA? and,

**iii.**   Do these organisms use additional proteins in comparison to those currently described for the process of DNA replication?

The apparent requirement of only a handful of genes for DNA replication in extremely reduced genomes, compared with the 228 annotated in *E. coli,* might suggest a parsimonious mechanism of DNA replication in endosymbiont bacteria since they are always living in stable environments. The genes which are more highly conserved in both reduced genomes and *E. coli* are those whose products form the replisome (*dnaE, dnaB, dnaN, dnaG, dnaX, dnaQ, ssb, holA* and *holB*), the genes encoding for DNA topoisomerase type II (*gyrA* and *gyrB*), and the gene for the NAD(+)-dependent DNA-ligase, *ligA* (Figure 4).

## 6. Components of a Minimal DNA Auto-Replicative System (MiDARS)

Of the three organisms with the most reduced genomes in nature, *Carsonella ruddii* is the more closely related phylogenetically to *E. coli* (Nakabachi et al., 2006). For this reason in our design we used the information of the replicative machinery in *Carsonella ruddii* and the functions known in *E. coli.* For the physical construction of the systems, however, we will use genes from *E. coli* for two main reasons:

| Organims/ Gene | eco | rma | tpa | cca | ama | rpr | tws | ctr | vok | mmn | mfl | uur | wbr | bfl | bci | nse | bpi | rip | buc | mge | men | smg | zin | crp | hci | tpn |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *dnaA* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *priA* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *dnaB* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *dnaG* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *ssb* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *dnaE* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *dnaN* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *dnaQ* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *dnaX* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *holA* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *holB* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *polA* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *ligA* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *gyrA* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *gyrB* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *topA* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *recA* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *rnhA* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *ruvA* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *ruvB* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *tatD* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *ung* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *uvrD* | | | | | | | | | | | | | | | | | | | | | | | | | | |
| *yqgF* | | | | | | | | | | | | | | | | | | | | | | | | | | |

**Table 2.** Conservation of replicative genetic machinery in bacteria with less than 1200 kbp

1. it is difficult to obtain genomic DNA from *Carsonella ruddii* since it cannot be cultured *in vitro.*

2. *E. coli* is the best chassis for applications in synthetic biology as mentioned above and therefore adequate for the incorporation of additional functions.

For the design of the minimal DNA auto-replicative system we will attempt to include the minimal elements present in *Carsonella* and -in a conservative manner- those we presume as necessary to perform the process of DNA replication in *E. coli*. In addition to the coding sequences, it is also necessary to define the regulatory regions of the genes and we propose to conserve the operative regions as defined for *E. coli* with the future aim of expanding the minimal functions of *E. coli* including the regulatory functions. Other important regions to include in the design are: the DNA replication origin (*oriC*) and the signals for termination of replication. Below we propose the genetic components that would constitute a MiDARS.
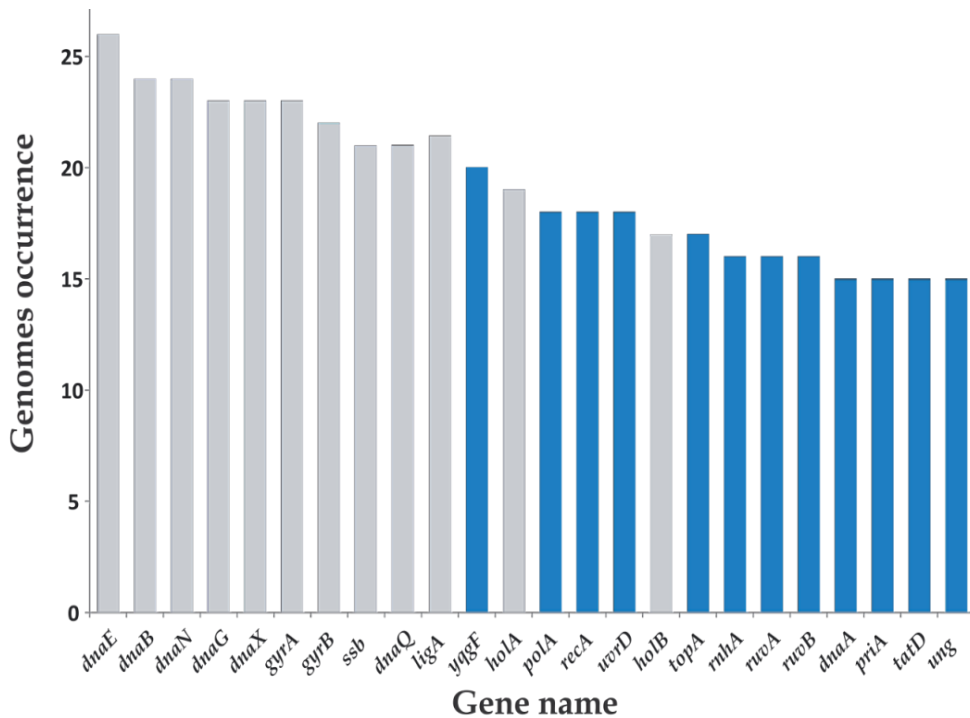
**Figure 4. Conservation of genes for DNA replication** in 25 reduced genomes and in E. *coli.* Grey bars represent the genes proposed to be essential for DNA auto-replication in a minimal genetic system.

### The DNA initiator protein (*dnaA)*

At the beginning of the replication process, check-point proteins have to recognize and unfold the initiation site for replication at *oriC*. In *E. coli* DnaA is the principal protein employed for this purpose and is highly conserved among bacteria with reduced genomes. Therefore, *dnaA* should be present in the MiDARS.

### The DNA helicase (*dnaB)*

The next candidate gene is *dnaB,* which encodes a DNA helicase. The role of the product of this gene is to unwind the DNA strands, a very important process during the elongation stage of replication.

### The DNA primase (*dnaG)*

The gene that encodes the primase (*dnaG*) should also be considered. It is important for the synthesis of the RNA primers that permit the elongation of new DNA strands.

### The single strand stabilization protein (*ssb)*

Another important function is the stabilization of single strands, carried out by the SSB protein, encoded by the *ssb* gene.

**The core components of DNA polymerase III (*dnaE* and *dnaQ*)**

The gene for the α subunit (*dnaE*) of DNA polymerase III is present in all 25 organisms with reduced genomes and the gene for the ε subunit (*dnaQ*) in twenty-one. These proteins form part of the core of DNA polymerase III, which carries out the essential polymerization and proofreading activities during DNA synthesis.

**The clamp components (*dnaX*, *holA*, *holB*, *dnaN*)**

During the elongation stage, two very important structures are formed; the leader and slider clamps. The first has the function of anchorage between DNA polymerase III and the DNA helicase; (Reyes-Lamothe R. et al., 2010) allowing the synthesis of the DNA in a synchronized manner between the leading and lagging strand. It is composed of the following subunits (genes): τ (dnaX), γ (*dnaX*), δ (*holA)* and δ'(*holB*). The circular slider clamp is constituted by two β-subunits (both products of *dnaN* gene), that recognize and bind to DNA-RNA hybrids (Georgescu R. et al., 2010). The slider clamp assists the core of DNA pol III to bind the lagging strand and allows the extension of the Okazaki fragments.

**The DNA ligase (*ligA*)**

The function of a ligase is needed for sealing nicks formed when the RNA primers are removed and replaced by DNA in the Okazaki fragments on the lagging strand.

**Type II DNA topoisomerase (*gyrA* and *gyrB*)**

We consider that a relaxing system produced by a DNA helicase may be necessary. This could be provided by the DNA gyrase complex (Type II Topoisomerase) composed of the A (*gyrA*) and B (*gyrB*) subunits.

**Protein for termination of replication (*tus*)**

Although there are several proteins that could contribute to termination of DNA replication we think that in a minimal system, the action of Tus could be enough to ensure this.

**Origin of DNA replication *(oriC)***

This DNA sequence of around 245 bp in *E. coli* (Tabata et al., 1983) is needed to enable the DnaA protein to initiate the process of DNA replication

**Termination of DNA replication (*terB* and *terC*)**

These sequences are used by the Tus proteins to form the trap which terminates DNA replication.

The proposed elements that constitute the auto-replicative system are also listed in Table 3. This proposal is somewhat similar to previous reports, where genes that could constitute a minimal cell based on a comparative genomics study among various endosymbionts are described (Gil et al., 2004). In the present study however we also considered the inclusion of the DNA regions for initiation and termination of replication, as well as the *dnaA*, *ssb* and *tus* genes.

| Gene/DNA element | Product | Size (bp. *E. coli*) |
|:---:|:---|:---:|
| *dnaA* | Chromosomal replication initiator protein DnaA | 1404 |
| *dnaB* | Replicative DNA helicase | 1416 |
| *dnaG* | DNA primase | 1746 |
| *ssb* | Single-stranded DNA-binding protein | 537 |
| *dnaE* | DNA polymerase III α subunit | 3483 |
| *dnaN* | DNA polymerase III, β subunit | 1101 |
| *dnaQ* | DNA polymerase III ε subunit | 732 |
| *dnaX* | DNA polymerase III, τ and γsubunits | 1932 |
| *holA* | DNA polymerase III,δ subunit | 1032 |
| *holB* | DNA polymerase III, δ' subunit | 1005 |
| *polA* | DNA pol I 5'-3' and 3'-5' exonuclease ; 3'-5' polymerase | 2787 |
| *ligA* | DNA ligase, NAD(+)-dependent | 2016 |
| *gyrA* | DNA gyrase (type II topoisomerase), subunit A | 2628 |
| *gyrB* | DNA gyrase, subunit B | 2415 |
| *tus* | Termination DNA replication protein | 930 |
| *oriC* | DNA region for initiation, origin of replication | 245 |
| *ter* | DNA region for termination of replication | 23 |

**Table 3.** Components of a minimal DNA auto-replicative system.

## 7. Expression of the replicative proteins of the MiDARS

A primary condition for the operation of an auto replicative system is that the protein-machinery encoded in it should be expressed. For transcription of the assembled group of genes, we propose use the *E. coli* RNA polymerase and its transcription factor sigma70 since all the genes of the system have a sigma70 factor promoter. The essential components of the RNA polymerase and their sigma70 factors have previously been successfully expressed separately and their activity reconstituted as mentioned previously (Asahara & Chong, 2010). We propose these components can be assembled as an additional functional module whose activity can be assayed separately and subsequently integrated into the system. The resulting mRNA (16) could be translated in an *in vitro* system such as the Pure System™ (Ueda et al., 1992; Shimizu & Ueda, 2010); containing ribosomes, aminoacyl-tRNAs, chaperones and initiation, elongation and termination factors among other elements essential for translation. Once protein synthesis is completed, the products could initiate replication of the DNA molecule for which the addition of deoxynucleotide triphosphates (dNTPs) and the appropriate buffers will be necessary. The source of energy for the

system will be creatine phosphate with the creatine kinase enzyme as the regenerator (Shimizu et al., 2006). An outline for the operation of the DNA auto-replicative system is shown in Figure 5.

## 8. Perspectives

Previous efforts have been made to propose the design of minimal cells however this objective is still far from being accomplished. From the standpoint of Synthetic Biology, biological systems that are robust, predictable in performance and highly efficient are desired (Jewett & Forster, 2010). In this work, we present a proposal to build an auto-replicative DNA system as the first step toward the development of synthetic biosystems. Additional cellular processes will need to be designed and constructed in a modular way including: transcriptional and translational functions and a minimal metabolism in order to maintain cell growth and produce energy.

Once this first prototype has been constructed and tested for performance, some further reduced combinations of the proposed number of genes could be tested to determine the absolute minimum set of genes sufficient to sustain DNA auto-replication; e.g. the few genes present in *Carsonella ruddii* PV.

The system proposed in this work can be assembled using methodologies such as that used when working with Biobricks (Smolke, 2009). Once the mini-chromosome is assembled it could function in cell-free systems, in anucleated mini-cells (Adler et al., 1967), in spores that lack DNA (Siccardi et al., 1975), in micelles or lipidic vesicles, and in some commercial systems. An important achievement in this sense has previously been reported by another research group, namely DNA replication achieved by using the Phi29 DNA polymerase, inside a lipidic vesicle. In this report only one strand was linearly replicated and circularized (Kurihara, 2011).

The successful development of a DNA auto-replicative system as proposed here could be a very important platform for the development of synthetic biology and the potential for such a system is great:

1.  in the refinement of biotechnological processes since cellular energy could be directed to the desired biosynthetic pathways;

2.  in the study of synthetic or natural circuits at a higher resolution and sharpness due to the minimalization of cellular noise; and

3.  to test some evolutionary hypotheses, such as the proposed components of last common ancestors and components of rudimentary first cells, among others.
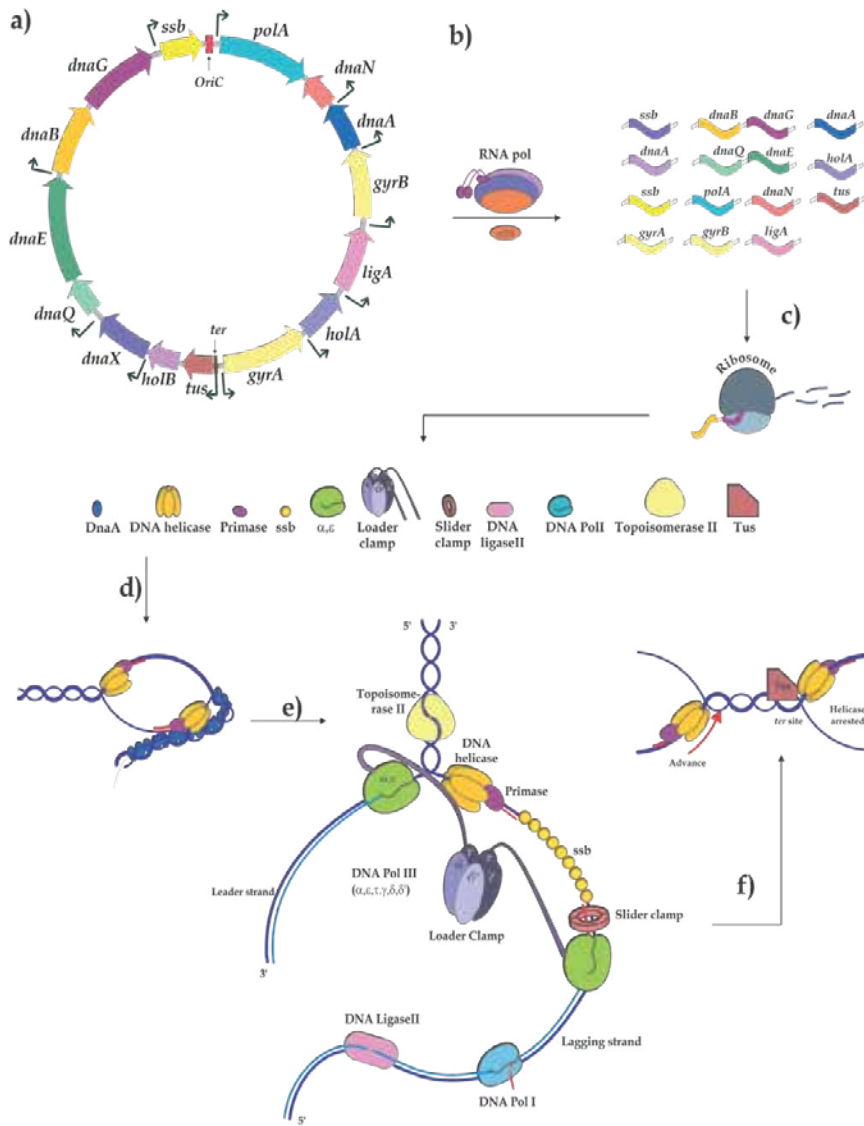
**Figure 5. Proposal for the minimal components of a MiDARS and their function.** a) The genetic system is a simplified version of a prokaryotic DNA mini-chromosome (25432 bp). The system contains the initiation region (*oriC*) and termination (*ter*) sites for DNA replication as well as a set of genes from *E. coli* (Table 4). The genes can be organized in the same order as in the native chromosome and contain their native operator regions to control expression. b) Transcription and, c) Translation can be carried out in solution using commercial kits (e.g. Pure System), RNA Polymerase and the *E. coli* sigma70 factor. d) The initiation of replication is regulated by DnaA-ATP and the helicase will join to the lagging strand in order to form the replication forks. The primase will bind to the helicase to carry out the synthesis of RNA primers that permit the activity of DNA pol III. The SSB protein stabilizes single strands of DNA. e) Two core subunits (α and ε) of the DNA Pol III, perform the elongation and proofreading of DNA. The DNA ligase and DNA Pol I replace the RNA primers, sealing the nicks between contiguous DNA fragments on the lagging strand. Topoisomerase II will relax the DNA template as the replication fork progresses. f) The Tus protein bound to the *ter* sites serves as a trap for the replicative machinery headed by the DNA helicase, stopping its movement and promoting the separation of the new MiDARS.

## 9. Conclusions

Here we propose a design for the construction of a minimal genetic system for DNA auto-replication. This proposal is based on the consideration of the latest knowledge of the details of the mechanisms and controls of DNA replication in *E. coli* and by taking into account the conservation of the replicative machinery in bacteria with extremely reduced genomes particularly those present in *Carsonella ruddii* PV.

The proposed auto-replicative device consists of 17 DNA elements (27822 bp including their operator regions) taken from the *E. coli* genome and incorporating the most conserved elements of the replicative machinery found in bacteria with extremely reduced genomes. These genetic elements will maintain their native operator and termination regions. Their products encode proteins encompassing the minimal number of predicted activities involved in DNA replication. Finally we propose some conditions in which the system might function.

## Acknowledgments

## Author details

Agustino Martinez-Antonio, Laura Espindola-Serna and Cesar Quiñones-Valles

Departamento de Ingeniería Genética, Cinvestav, Irapuato-León, Irapuato Gto, México

## References

[1] Adler, H., Fisher, W. D., Cohen, A., & Hardigree, A. (1967). Miniature Escherichia coli cells deficient in DNA. Proc. Natl. Acad. Sci. USA , 57, 321-326.

[2] Asahara, H., & Chong, S. (2010). In vitro genetic reconstruction of bacterial transcription initiation by coupled synthesis and detection of RNA polymerase holoenzyme. Nucleic Acids Res 3813:e141.

[3] Bessman, M. J., Lehman, I. R., Simms, E. S., & Kornberg, A. (1958). Enzymatic synthesis of deoxyribonucleic acid. II. General properties of the reaction. J Biol Chem ., 233, 171-177.

[4]  Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., & Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. Science , 277, 1453-1462.

[5]  Delaye, L., & Moya, A. (2009). Evolution of reduced prokaryotic genomes and the minimal cell concept: Variations on a theme. *BioEssays*, 32, 281-287.

[6]  Foley, P. L., & Shuler, M. L. (2010). Considerations for the design and construction of a synthetic platform cell for biotechnological applications. Biotechnol Bioeng , 105, 26-36.

[7]  Forster, A. C., & Church, G. M. (2006). Towards synthesis of a minimal cell. Mol Syst Biol 2:45.

[8]  Georgescu, R., Yao, N. Y., & O'Donnell, O. (2010). Single-molecule analysis of the *Escherichia coli* replisome and use of clamps to bypass replication barriers. *FEBS Letters*, 584, 2596-2605.

[9]  Gil, R., Silva, F. J., Peretó, J., & Moya, A. (2004). Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev , 68, 518-37.

[10]  Jain, R., Rivera, M. C., & Lake, J. (1999). Horizontal gene transfer among genomes: The complexity hypothesis. PNAS. , 96, 3801-3806.

[11]  Jewett, M. C., & Forster, A. C. (2010). Update on designing and building minimal cells. Curr Opin Biotechnol , 21, 697-703.

[12]  Keyamura, K., & Katayama, T. (2011). DnaA protein DNA-binding domain binds to Hda protein to promote inter-AAA+ domain interaction involved in regulatory inactivation of DnaA. J Biol Chem , 286, 29336-29346.

[13]  Kurihara, K., Tamura, M., Shohda, K., Toyota, T., Suzuki, K., & Sugawara, T. (2011). Self-reproduction of supramolecular giant vesicles combined with the amplification of encapsulated DNA. *Nature Chemistry*, 3, 775-781.

[14]  Langston, L. D., Indiani, C., & O'Donnell, M. (2009). Whither the replisome: emerging perspectives on the dynamic nature of the DNA replication machinery. *Cell Cycle*, 8, 2686-91.

[15]  Lehman, I. R., Bessman, M. J., Simms, E. S. ., & Kornberg, A. (1958a). Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*. J. Biol Chem , 233, 163-170.

[16]  Lehman, I. R., Zimmerman, S. B., Adler, J., Bessman, M. J., Simms, E. S., & Kornberg, A. (1958b). Enzymatic synthesis of deoxyribonucleic acid. V. Chemical composition of enzymatically synthesized deoxyribonucleic acid. Proc Natl Acad. Sci U S A , 44, 1191-1196.

[17] López-Madrigal, S., Latorre, A., Porcar, M., Moya, A., & Gil, R. (2011). Complete genome sequence of "Candidatus *Tremblaya princeps*" strain PCVAL, an intriguing translational machine below the living-cell status. J Bacteriol. , 193, 5587-8.

[18] Luisi, P. L., Ferri, F., & Stano, P. (2006). Approaches to semi-synthetic minimal cells: a review. *Naturwissenschaften*, 93, 1-13.

[19] Mac, Donald. J. T., Barnes, C., Kitney, R. I., Freemont, P. S., & Stan, G. B. (2011). Computational design approaches and tools for synthetic biology. Integr Biol (Camb). , 3, 97-108.

[20] Mc Cutcheon, J., & Moran, N. (2011). Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol , 10, 13-26.

[21] Mc Cutcheon, J. P., Mc Donald, B. R., & Moran, N. A. (2009). Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. PLoS Genet. 5:, e1000565 EOF.

[22] Miller, S. L. (1953). A Production of Amino Acids Under Possible Primitive Earth Conditions. Science. , 117, 528-529.

[23] Moreno-Hagelsieb, G., & Latimer, K. (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatic , 24, 319-234.

[24] Murtas, G. (2009). Artificial assembly of a minimal cell. Mol BioSyst , 5, 1292-1297.

[25] Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H. E., Moran, N. A., & Hattori, M. (2006). The 160 -Kilobase Genome of the Bacterial Endosymbiont Carsonella. Science. 314:267.

[26] Pósfai, G., Plunkett, G., 3rd , , Fehér, T., Frisch, D., Keil, G. M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S. S., de Arruda, M., Burland, V., Harcum, S. W., & Blattner, F. R. (2006). Emergent properties of reduced-genome *Escherichia coli.* Science , 312, 1044-1046.

[27] Quiñones-Valles, C., Espíndola-Serna, L., & Martínez-Antonio, A. (2011). Mechanisms and Controls of DNA Replication in Bacteria. Fundamental Aspects of DNA Replication. InTech , 219-244.

[28] Reyes-Lamothe, R., Sherratt, D. J., & Leake, M. C. (2010). Stoichiometry and Architecture of Active DNA Replication Machinery in *Escherichia coli*. *Science*, 328, 498-501.

[29] Shimizu, Y., Kuruma, Y., Ying, B. W., Umekage, S., & Ueda, T. (2006). Cell-free translation systems for protein engineering. FEBS Journal , 273, 4133-4140.

[30] Shimizu, Y., & Ueda, T. (2010). PURE technology. Methods Mol Biol. , 607, 11-21.

[31] Siccardi, A. G., Galizzi, A., Mazza, G., Clivio, A., & Albertini, A. M. (1975). Synchronous germination and outgrowth of fractionated *Bacillus subtilis* spores: tool for the analysis of differentiation and division of bacterial cells. J Bacteriol. , 121, 13-19.

[32] Smolke, C. D. (2009). Building outside of the box: iGEM and the BioBricks Foundation. Nat Biotechnol , 27, 1099-1102.

[33] Tabata, S., Oka, A., Sugimoto, K., Takanami, M., Yasuda, S., & Hirota, Y. (1983). The 245 base-pair oriC sequence of the *E. coli* chromosome directs bidirectional replication at an adjacent region. Nucleic Acids Res. , 11, 2617-26.

[34] Tamames, J., Gil, R., Latorre, A., Peretó, J., Silva, F. J., & Moya, A. (2007). The frontier between cell and organelle: genome analysis of Candidatus Carsonella ruddii. BMC Evolutionary Biology 7:181.

[35] Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., & Natale, D. A. (2003). The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41.

[36] Ueda, T., Tohda, H., Chikazumi, N., Eckstein, F., & Watanabe, K. (1991). Cell-free translation system using phosphorothioate-containing mRNA. Nucleic Acids Symp Ser. , 25, 151-2.

[37] Williams, K. P., Gillespie, J. J., Sobra, B. W., Nordberg, E. K., Snyder, E. E., Shallom, J. M., & Dickerman, A. W. (2010). Phylogeny of gammaproteobacteria. J Bacteriol , 192(9), 2305-1234.

# Extending the Interaction Repertoire of FHA and BRCT Domains

Lindsay A. Matthews and Alba Guarné

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/51670

## 1. Introduction

All living organisms are connected by the necessity to replicate their DNA. However, the process of unwinding the parental DNA to serve as a template at replication forks is danger‐ ous. The ssDNA generated by helicases is inherently cytotoxic; not only because it is more prone to damage, but it can also be an inappropriate target for nucleases and recombination proteins leading to loss of genetic material or gross chromosomal rearrangements [1]. Nor‐ mally, the replication machinery rapidly restores this single-stranded template DNA to its more stable double-stranded form. However, replication forks are prone to stalling if they encounter obstacles that the DNA polymerase is unable to bypass, such as sites of damage or DNA sequences with complex secondary structures [2], resulting in long stretches of ssDNA remaining exposed [3]. Replication stress, therefore, represents an important mecha‐ nism that erodes the genetic integrity of organisms. Not surprisingly, replication stress has been linked to aging in budding yeast, which can likely be extrapolated to higher eukaryotes as well [4]. Furthermore, inducing replication stress in normal human fibroblasts results in pathogenic changes in copy number due to duplication or deletion events [5]. Therefore, the ability of eukaryotes to detect, stabilize and resolve stalled replication forks using the repli‐ cation checkpoint represents an important safeguard for genomic stability.

The replication checkpoint response relies on a cascade of kinases that either remain local‐ ized to the stalled fork or disseminate the stress signal to distal sites resulting in the sup‐ pression of late origin firing, pausing of the cell cycle, and increasing the expression of DNA repair enzymes [6]. Overall, this checkpoint involves individual proteins and pro‐ tein complexes coming together to assemble intricate supramolecular complexes triggered by stalled replication forks. For simplicity's sake, this chapter will focus on the *Saccharo‐ myces cerevisiae* system and nomenclature. However, regardless of the organism being con‐

sidered, the general recurring theme is that most stress dependent protein interactions involve either BRCT or FHA domains. In fact, BRCT and FHA domains are rarely found in cytosolic proteins, but they are overrepresented in nuclear proteins involved in DNA replication, as well as the detection and response to DNA damage [7, 8]. Both domains share an ability to specifically recognize phosphorylated epitopes, although with different specificities. BRCT domains primarily recognize phospho-serine (pSer) containing epitopes, while FHA domains exclusively recognize phospho-threonine (pThr). However, phosphorylation-independent interactions have recently been described for both domains.
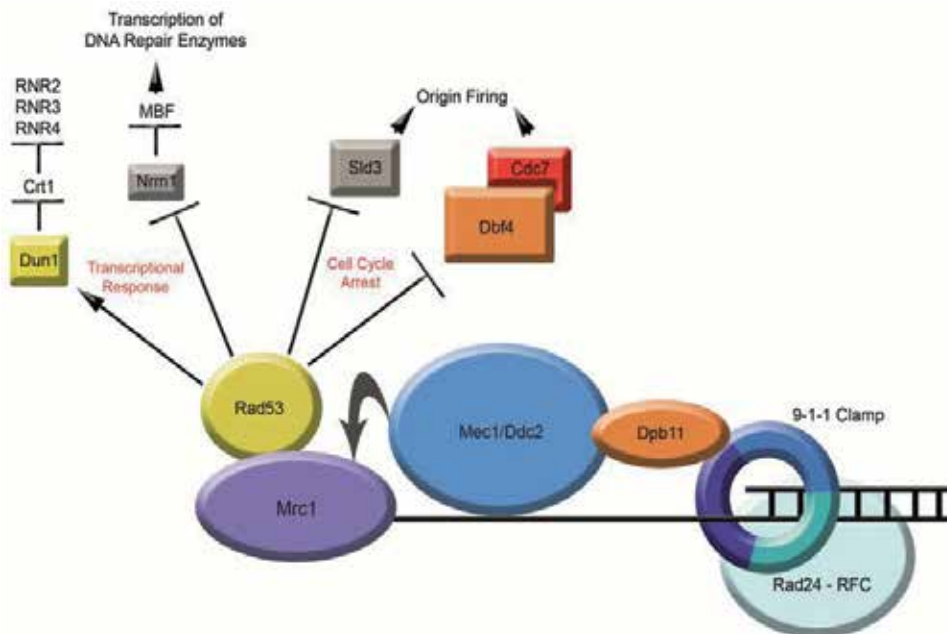


**Figure 1.** BRCT and FHA domains mediate protein interactions that relay the replication stress signal. A stalled replication fork is represented as a region of ssDNA recognized by a variety of checkpoint proteins colored in shades of blue. Proteins that mediate interactions involved in relaying the stress signal are colored orange if they contain BRCT domains or yellow if they contain FHA domains. The curved arrow indicates a phosphorylation event that takes place on Mrc1 to allow for Rad53 recruitment, while the straight arrows indicate interactions that take place at distal sites to the stalled fork.

Several BRCT-containing proteins are essential for the assembly of the pre-replication complex. For instance, the regulatory subunit of the Cdc7 kinase, Dbf4, contains a single BRCT domain [9, 10] and Dpb11, whose function is essential to activate pre-replication complexes, contains two BRCT pairs [8]. The Rfc1 subunit of the RFC complex, which functions in loading the sliding clamp onto ssDNA/dsDNA junctions during replication [11], also contains a BRCT domain required for DNA binding [12]. Most replication proteins containing BRCT domains are also involved in the replication checkpoint response, suggesting that BRCT domains may have fundamental roles in preserving DNA integrity. During the DNA damage response, BRCT domains are often used to recognize the site of damage, as they can bind to

DNA breaks directly [13, 14] or indirectly by recognizing phosphorylated histone H2A that marks areas of damage [15-20]. In the case of replication stress, at least two first responders are recruited to the ssDNA gap at stalled forks (Figure 1). The first is the protein kinase Mec1 and its targeting subunit Ddc2 that mediates the interaction with RPA-coated ssDNA [21]. The second is the 9-1-1 clamp that encircles the DNA at the ssDNA/dsDNA junction [22]. This requires the action of a clamp loader, which is composed of Rad24 and Rfc2-5. Interestingly, this complex is the alter ego of the clamp loader RFC, which differs only in having Rfc1 instead of Rad24 [23]. In contrast to the sliding clamp, the 9-1-1 complex is held statically by protein-protein interactions at the stalled fork [24].

After the recognition of DNA damage or stalled forks, both FHA- and BRCT-containing proteins feature prominently in bridging protein–protein interactions that disseminate the stress signal (Figure 1). For instance, Dpb11 bridges the interaction between the 9-1-1 clamp and Mec1, leading to the full activation of the Mec1 kinase [24, 25]. Mec1 has many roles at the stalled fork including facilitating the activation of the next downstream kinase in the pathway, Rad53 [26]. This is accomplished after Mec1 phosphorylates Mrc1—a protein naturally associated with the stalled replisome [27]. This creates phospho-epitopes that act as beacons for the FHA domain of Rad53. Multiple copies of Rad53 are thus recruited, increasing the local concentration of this kinase, which can then be autophosphorylated *in trans* or phosphorylated by additional kinases present at the stalled fork [28]. Hyperphosphorylation of Rad53 presents phospho-epitopes to other FHA-containing proteins such as Dun1, which leads to increased synthesis of nucleotides [28-30]. Additionally, Rad53 uses its own FHA domains to bind a variety of substrates, including the regulatory subunit of the Dbf4-dependent kinase (DDK) complex [31], consequently suppressing the firing of late origins [32-34]. Rad53 also modulates the activity of Nrm1 leading to a burst of expression of DNA repair enzymes contributing to recovery of the stalled fork [35]. BRCT domains also feature in this recovery process. For example, an important scaffolding protein involved in coordinating the recruitment of repair enzymes to the stalled fork, Rtt107, has six BRCT domains [20].

These themes are echoed in other DNA damage response pathways, where BRCT and FHA domains are known to mediate important interactions. Not surprisingly, mutations in the BRCT and FHA domains of critical damage repair proteins such as Chk2 [36], Nbs1 [37] and BRCA-1 [38, 39], lead to cancer predisposing syndromes. Why does nature rely so heavily on BRCT and FHA domains to respond to stress? The answer may seem to lie in the ability of BRCT and FHA domains to recognize phospho-epitopes, since they are an important cue during replication stress and the DNA damage response, when a number of kinases are awakened. However, their power could also lie in the plasticity of these two domains that can use multiple interaction surfaces to mediate additional interactions beyond phospho-epitope recognition. Such plasticity could, in turn, mediate the interaction network sustaining the formation of the large protein complexes required to promote genome stability in eukaryotes. Along with the well-characterized phosphate recognition ability of FHA and BRCT domains, these varied and unique alternative interaction surfaces will be considered in this chapter. Interactions occurring during the replication checkpoint will be discussed, but examples from other cellular pathways will also be included.

## 2. Interaction Modes of FHA Domains

An FHA domain consists of an 11-stranded β sandwich connected by loops that often contain short helical regions. The phospho-epitope binding groove is located at the apical surface of the β-sandwich, with the N- and C-termini at the opposite end of the domain. Unlike BRCT domains, that are often present in multiple copies in a single protein, FHA domains are almost always singular. Only two proteins are known to possess two FHA domains in the same polypeptide: Rad53 from *S. cerevisiae* and Rv1747 from *Mycobacterium tuberculosis* (*M. tuberculosis*). In the case of Rad53, these domains (FHA1 and FHA2) are found at opposite ends of the protein and have independent functions [40], reinforcing the idea that FHA domains function as single units. This, however, does not diminish the power of the FHA domain as a scaffold to build large protein complexes in response to stress. FHA domains can bind partners in a phosphorylation dependent or independent manner, the latter of which can utilize either the phospho-epitope binding pocket or alternative surfaces.

### 2.1. Phospho-epitope dependent interactions

FHA domains recognize phosphorylated proteins with a strict specificity for pThr-containing epitopes. The majority of interactions between FHA domains and their phosphorylated partners have been studied using short peptides including a central phosphorylated threonine [41-46]. These phospho-peptides bind to the apical surface of the FHA domain in an extended conformation using two pockets that determine their binding specificity. The β3–β4 and β4–β5 loops from the FHA domain primarily define the pThr-binding site, where a conserved arginine and serine (Arg70 and Ser85 in Rad53) provide critical contacts with the phosphate group. An extensive hydrogen-bond network mediated by non-conserved residues further stabilizes the interaction with the phosphorylated threonine. The second pocket recognizes the third residue C-terminal to the phospho-threonine (pThr+3), and is usually defined by the β6–β7 and β10–β11 loops of the FHA domain (Figure 2A and B).

A unique aspect of the pThr-binding pocket in FHA domains is its ability to distinguish between phospho-threonine and phospho-serine residues, a talent not shared by other phospho-epitope recognition modules. For instance, MH2 domains share a common ancestor with FHA domains and, yet, MH2 domains can also bind phospho-serine with high affinity [42, 47]. The difference is that the pThr-binding pocket in the FHA domain includes a well-defined hydrophobic nook that provides a docking site for the methyl moiety of the phosphorylated threonine (Figure 2C). While most residues in this nook are not conserved, the hydrophobic nature of this pocket is strictly maintained, thus providing a number of Van der Waals interactions that orient the phospho-threonine such that its phosphate moiety is locked in the most favorable geometry for the interaction with the domain [48]. Thus, the pThr pocket of an FHA domain can be thought of as a glove where only a phospho-threonine can fit (Figure 2C). Despite this rigid mode of binding, certain FHA domains can accommodate deviations. For instance, the FHA2 domain from Rad53 can bind pTyr-containing peptides with low affinity [47] and the FHA domain found in Dun1 has a unique preference for phosphorylated substrates including two pThr residues [49].
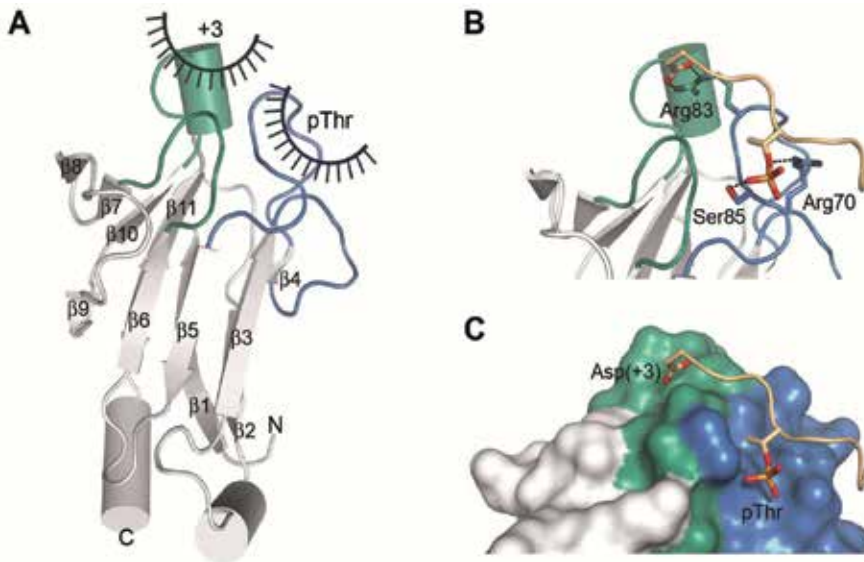
**Figure 2.** Phospho-peptides interact with FHA domains using at least two pockets. (A) Ribbon diagram of the FHA1 domain from Rad53 (PDB ID: 1G6G) with the loops defining the pThr (β3-β4 and β4-β5) and pThr+3 (β6-β7 and β10-β11) binding pockets highlighted in blue and green, respectively. The N- and C-terminal residues of the domain, which lie on the opposite side of the β-sandwich, are labeled for clarity. (B) Detailed view of a phospho-peptide bound to the pThr and pThr+3 pockets. The side chains of two conserved residues important for the recognition of the phosphate group (Ser85 and Arg70), as well as that of a non-conserved residue important for the specificity of the pThr+3 pocket are shown as sticks. Hydrogen bonds are indicated with dashed lines. (C) Surface representation, shown in the same orientation and color-coding as panel B, indicating the presence of defined pockets for both the phosphate and methyl moieties of the pThr.

In contrast to the pThr pocket, the residues defining the pThr+3 pocket are not conserved allowing for different domains to have different target-sequence specificities. This is called the "pThr+3 rule" wherein different FHA domains have different specificities for the pThr+3 residue. While this provides a convenient way to classify FHA domains, it should be noted that the specificity of the pThr+3 pocket is not fixed. For example, the FHA1 domain from Rad53 prefers aspartic acid as the pThr+3 residue using short peptides *in vitro*, but binds to a bulky, hydrophobic isoleucine in its partner Mdt1 *in vivo*, with the pThr+3 residue (Asp or Ile) occupying physically different pockets in each case [45]. Although FHA domain interactions have been disrupted *in vivo* by mutating the pThr residue, similar experiments with the pThr+3 residue are not available and, hence, the importance of the pThr+3 pocket is unclear. It has been proposed that this ancillary pocket may only have relevance for determining the specificity of the phospho-epitope in small peptides, whereas full-length partners may use different binding mechanisms [50]. Consequently, a detailed understanding of how FHA domains recognize phosphorylated binding partners will necessitate the structural analysis of FHA domains bound to full-length proteins rather than short phosphorylated peptides.

FHA domains often bind only weakly to pThr-containing peptides [43, 51], supporting the idea that additional contact points beyond the phospho-epitope binding site are necessary to form high avidity complexes with their partners. For example, the interaction between the FHA domain from Chk2 and the tandem BRCT repeat from BRCA-1 requires an addi-

tional hydrophobic patch on the surface of one of the FHA β-sheets. Mutating either the phosphate binding pocket or this hydrophobic patch destroys the interaction with BRCA-1 even though structurally these two sites are more than 20 Å apart [52]. Reinforcing the importance of this additional interaction surface *in vivo*, mutation of an isoleucine (Ile157Thr) within this hydrophobic patch results in the cancer predisposing Li-Fraumeni syndrome [52]. While this hydrophobic surface is not a common feature of all FHA domains—not even amongst Chk2 homologues [52], it is possible that unique patches exist within the surfaces of the β-sandwich of other FHA domains that provide auxiliary contacts to enhance binding to phosporylated target proteins. Due to the inherent difficulty in obtaining uniformly phosphorylated proteins, solving the structures of FHA domains interacting with their full-length phosphorylated partners is a lofty goal. This is further compounded by the fact that FHA domains cannot be fooled by phospho-mimetic mutations—a trick commonly used to study phosphorylation-dependent interactions—at least when using small peptides [43]. The structure of the Ki67 FHA domain interacting with NIFK1 has recently shed light onto this problem. This interaction was recapitulated with a very long phospho-peptide (consisting of 44 amino acids), which, in addition to occupying the phospho-epitope binding site, also wraps around and extends one of the β-sheets in the Ki67 FHA domain by providing an additional β-strand [50] (Figure 3).
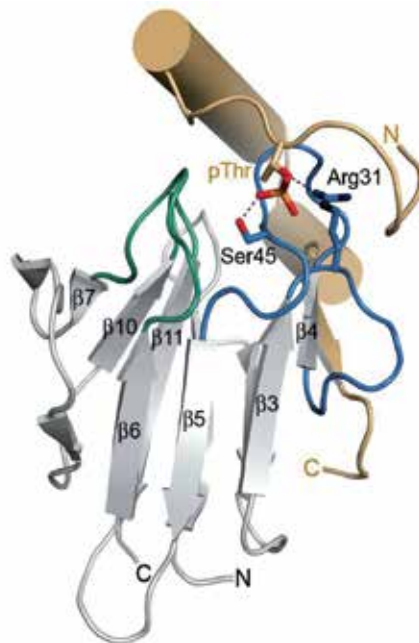


**Figure 3.** Phosphorylated binding partners of FHA domains can occupy extensive interaction surfaces. Ribbon diagram of the FHA domain of Ki67 bound to a phospho-peptide encompassing residues 226-269 of human NIFK (PDB ID: 2AFF). This long phospho-peptide interacts with three distinct surfaces on the FHA domain, but does not occupy the pThr+3 pocket identified in the structures of FHA domains bound to short phospho-peptides. The phospho-threonine occupies the canonical pThr-binding pocket defined by the β3–β4 and β4–β5 loops (blue), the α helix following the pThr covers a hydrophobic surface partially defined by the β4–β5 and β10–β11 loops (green), and the β-strand at the C-terminus of the peptide extends the β-sheet defined by β7–β10–β11–β1–β2–β4.

This long phospho-peptide does not conform to the "pThr+3" rule. Instead, binding to the FHA domain induces the formation of an $\alpha$ helix that nestles in a hydrophobic pocket formed by the β4–β5 and β10–β11 loops [50], underscoring the need for additional structural information using full-length phosphorylated binding partners.

## 2.2. Phospho-epitope independent interactions

Although FHA domains were initially identified as pThr binding domains, it has been predicted that they can also mediate phosphorylation-independent interactions. Members of the kinesin-3 family, a class of motor proteins that transport vesicles to the tips of axons in neural cells [53], contain an N-terminal FHA domain in addition to their motor domain and coiled-coil regions [53]. One member of this family, KIF13B, uses its FHA domain to transport PIP3-rich vesicles in order to facilitate axon development. This involves the formation of a tetrameric complex with CENTA1, which has been studied through X-ray crystallography [54]. This complex has two CENTA1 molecules and two kinesin molecules, with the FHA domain of each kinesin involved in two simultaneous interactions (Figure 4). The first is with the ArfGAP domain of one of the CENTA1 molecules, which contacts the FHA loops that normally recognize a pThr. However, this interaction is phosphorylation-independent because the FHA domain of KIF13B lacks the conserved residues for phospho-threonine recognition [54]. The second CENTA1 molecule in the tetramer uses its Pleckstrin Homology 1 (PH1) domain to contact a surface on the β-sandwich of the same KIF13B FHA domain. This situation is reminiscent of the interaction between Chk2 and BRCA-1, and suggests that auxiliary contacts mediated by the β-sandwich may enhance both phosphorylation dependent and independent interactions. Phosphorylation independent interactions are not exclusive to FHA domains that lack the pThr-recognition residues. Another member of the kinesin-3 family, KIF1A, has a canonical phospho-epitope binding site and, yet, is also suspected of using this pocket for a phosphorylation-independent interaction [55]. Similarly, the FHA domain of *S. cerevisiae* Rad53 has a canonical phosphoepitope binding site, but is presumed to interact with the BRCT domain of Dbf4 in a phosphorylation-independent manner, though the molecular determinants of this interaction are unclear [10].

Some bacterial proteins can also interact with the phospho-epitope binding site of an FHA domain in a phosphorylation-independent manner. Although bacteria primarily rely on histidine kinases and their associated regulatory responders for phosphorylation-dependent signaling, some also utilize eukaryotic-like Ser/Thr Protein Kinases (STPKs) [56]. Like eukaryotes, bacteria can use STPKs to respond to stress, but they also participate in other processes such as pathogenicity, thereby providing important drug targets [56]. The best characterized bacterium in this regard is *M. tuberculosis*, the causative agent of tuberculosis, which tops the charts in the prokaryotic kingdom with eleven STPKs [57]. Proteins that work downstream of bacterial STPKs often contain FHA domains, with *M. tuberculosis* having five such proteins [58]. One of them, Rv1827, is of special interest because it engages a phospho-epitope present in its N-terminal tail intramolecularly [59]. This effectively occludes the phospho-epitope binding site preventing the interaction of other binding partners [59]. Intriguingly, at least three different binding partners can compete with this intramolec-

ular phospho-epitope even though none of them includes a phosphorylation site [59]. Therefore, the interactions mediated by the FHA domain of *M. tuberculosis* Rv1827 reveal two recurrent features: the ability of an intramolecular interaction to negatively regulate the interactions of an FHA domain, and the use of an FHA phospho-epitope binding site to engage in phosphorylation-independent interactions. These features are also reminiscent of the interactions mediated by the FHA domains found in KIF1A [55], reinforcing the idea that FHA domains may use competing interactions to fine-tune cellular processes.
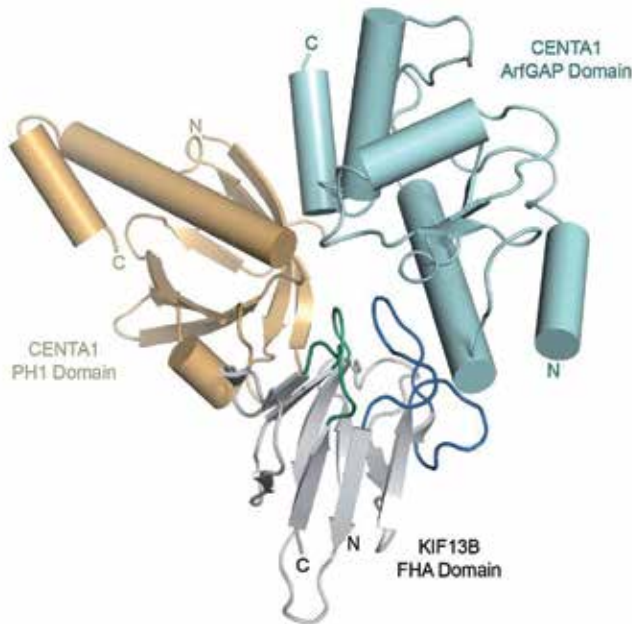


**Figure 4.** The FHA domain of kinesin KIF13B mediates two phosphorylation independent interactions simultaneously. Ribbon diagram of the FHA domain of KIF13B bound to CENTA1 (PDB ID: 3MDB). In the crystal structure, the FHA domain of KIF13B (white) contacts the PH1 (tan) and the ArfGAP (cyan) domains of two adjacent CENTA1 molecules simultaneously. The interaction between the FHA and PH1 domains is mediated by one of the faces of the β-sandwich, while the loops that normally define the pThr-binding pocket are involved on the recognition of the ArfGAP domain.

## 3. Interaction Modes of BRCT Domains

BRCT domains are named after the breast cancer associated protein 1 (BRCA-1) C-terminus because they were originally identified at this end of BRCA-1. However, the BRCT is an ancient domain that originates in prokaryotic NAD+ ligases where it is used to bind to DNA. Eukaryotes obtained the BRCT domain through horizontal gene transfer, and while some eukaryotic BRCT domains still retain DNA binding function, the vast majority have evolved to recognize protein partners instead [60]. BRCT domains are defined by a central four-stranded parallel β-sheet surrounded by three helices: $\alpha1$ and $\alpha3$ on one side, and $\alpha2$ on the other of the β-sheet [61]. However, additional secondary structure elements have been de-

scribed in the loop regions. There are also BRCT domains that lack elements, notably helix $\alpha2$ [61]. Rap1 is the most extreme example of this, having only three strands in its central β-sheet and all helices packed against the same side of the sheet, leaving the other side exposed [62]. This unique structure is highly flexible and relatively unstable [62].

BRCT domains can occur both as single or multiple units, which usually consist of two BRCT domains and are referred to as tandem BRCT repeats. The structural diversity of BRCT domains is perhaps best illustrated by the human homologue of Dpb11 (human TopBP1), which contains eight BRCT domains that function as single, double or triple BRCT units [63-65]. Similarly to FHA domains, BRCT domains are overrepresented in DNA damage response proteins where they recognize phosphorylated targets generated during damage recognition and repair [7]. While the molecular intricacies of phospho-epitope recognition by tandem BRCT repeats have been extensively studied, phosphorylation-independent interactions mediated by BRCT tandems or the functions of single BRCT domains remain poorly characterized. Elegant proteolysis studies have shown that mutations in the BRCT tandem repeat of BRCA-1, with a causal link to early onset breast and ovarian cancer, destabilize the BRCT fold [39], suggesting that BRCT domains may work as protein scaffolds. These studies also revealed the hypersensitivity of BRCT domains to mutations, an effect that was attributed to its minimal size (95-100 amino acids) [66] and, consequently, the fact that every residue contributes to either maintaining the domain fold or mediating interactions with BRCT-binding partners [67].

### 3.1. Tandem BRCT Domains: The Two-Knob Hypothesis

The tandem BRCT repeat was formed through a gene duplication event, in which the binding pocket originally used to bind to the phosphate backbone of DNA evolved to recognize a phospho-epitope in a target protein [60]. Being an $\alpha/\beta$ fold, the BRCT domain has a topological switch point; a region along the C-terminal edge of the β-sheet whereupon a groove is formed in the connecting loops. Tandem BRCT repeats use the topological switch point—termed the P1 pocket—of their first BRCT domain to bind phospho-serine (pSer) residues in their interaction partners (Figure 5). Similar to FHA domains, this interaction involves the side chain of a conserved serine residue, but in BRCT domains the phosphate moiety of the pSer residue is further stabilized by the interaction with the side chain of a conserved lysine, as well as the backbone atoms from the glycine immediately following the conserved serine [68].

Similar to FHA domains, tandem BRCT repeats also subscribe to a "pSer+3 rule" to enhance phospho-epitope binding specificity [68]. However, the pSer+3-binding pocket—termed the P2 pocket—only forms in tandem BRCT repeats as it is defined by residues at the interface between the first and second BRCT domains. When two BRCT domains coalesce to define a tandem repeat, the central β-sheet of both domains adopt a parallel arrangement that defines an intervening hydrophobic three-helix bundle form by $\alpha2$ from the N-terminal and $\alpha1$ and $\alpha3$ from the C-terminal BRCT domains (Figure 5) [69]. Helical bundles are known to facilitate molecular interactions [70] and, in the case of tandem BRCT domains, it allows for the recognition of a bulky hydrophobic residue at the pSer+3 position [68]. Indeed, the high specificity of tandem BRCT repeats for their phospho-epito-

pes is primarily due to the presence of the P2 pocket that imposes the need for a second knob in the phospho-epitope, thus precluding non-specific binding.
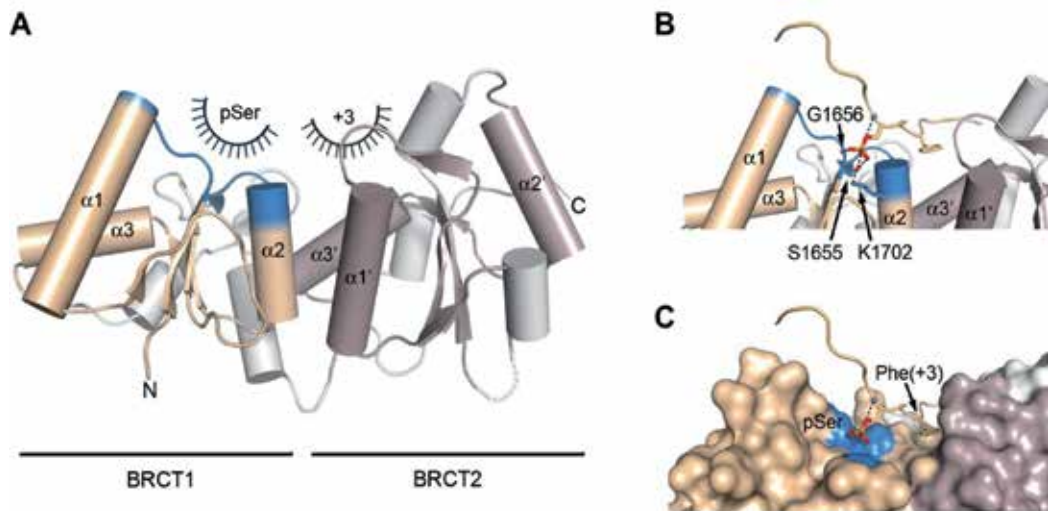


**Figure 5.** Phospho-epitope recognition by tandem BRCT repeats. (A) Ribbon diagram of the tandem BRCT domain from BRCA-1 (PDB ID: 1T2V) with secondary structure elements defining the BRCT fold shown in tan (BRCT1) and brown (BRCT2), and those not common to the BRCT fold shown in white. The pSer-binding site is located entirely within BRCT1, whereas the pSer+3 pocket is located in a three-helical bundle created at the interface between the two domains. (B) The side chains of Ser1655, Lys1702, as well as the main chain nitrogen of Gly1656 stabilize the phosphate moiety of the pSer. Additionally, the phosphate group engages in hydrogen bonds with the backbone amino groups represented as blue spheres. (C) The surface representation, shown in the same orientation as in panel B, reveals that the pSer binds to a shallow pocket that could not accommodate any other phosphorylated residue, while the conserved aromatic residue at the pSer+3 position is nestled into a well-defined hydrophobic pocket.

In certain tandem BRCT repeats, the P1 pocket is found in the C-terminal rather than the N-terminal BRCT, however due to the absence of the P2 pocket it is presumed that this mode of interaction is weaker than the canonical binding mode [71]. In fact, BRCT repeats containing a P1 pocket on the C-terminal BRCT are known to mediate phosphorylation-independent protein–protein interactions [72], suggesting that the binding specificity for a phospho-epitope may not be as critical. In the structure of the tandem BRCT repeats found in 53BP1 bound to the DNA-binding domain of p53 [73, 74], the inter-domain linker is critical to mediate the interaction between 53BP1 and p53. Similarly, in the structure of the *Schizosaccharomyces pombe* Crb2 homodimer [75], the linker connecting the two BRCT domains mediates protein dimerization. Collectively, these structures underscore the fact that tandem BRCT repeats define single functional units with multiple interaction surfaces.

### 3.2. BRCT "Super-domains": Expanding the two-knob model

The individual units in a tandem BRCT repeat are dependent on each other for structural stability due to the hydrophobicity of the $\alpha$ helices that define the intervening helix bundle [69]. Thus, a tandem BRCT repeat can actually be considered one single domain module, distinct from single BRCT domains. Beyond single and tandem BRCT arrangements,

a number of BRCT structures over the past decade have revealed many unexpected terti-
ary structures formed by the combination with other functional domains. This is not only
in the number of BRCT domains involved, such as the structure of the triple BRCT repeat
from the human TopBP1 [76], but also in the diversity of domains that can be ensnared
by a BRCT neighbor. This includes the Fibronectin Type III domain (FN3) found at the N-
terminus of a tandem BRCT domain in the *S. cerevisiae* protein Chs5 [77], or the FHA do-
main that does likewise in *S. pombe* Nbs1 [78, 79].
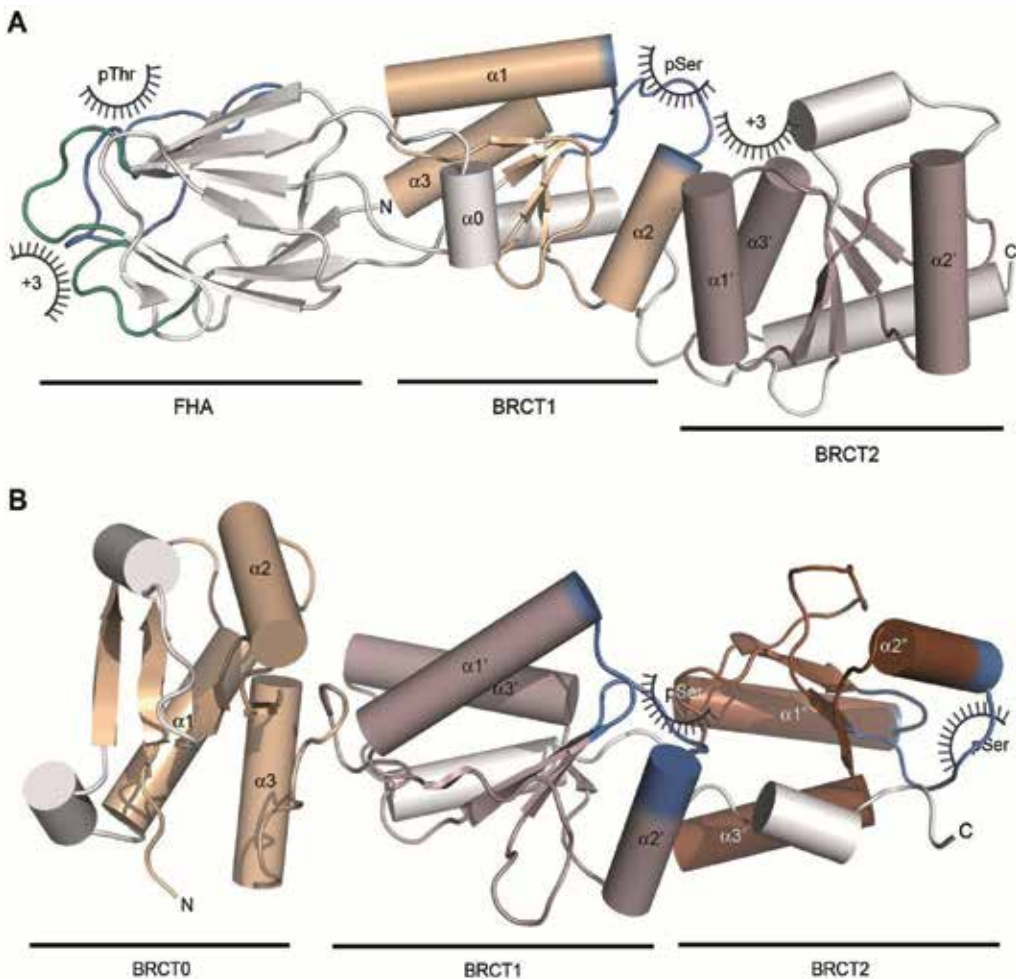


**Figure 6.** BRCT super-domains. (A) Ribbon diagram of the FHA–BRCT–BRCT super-domain found in Nbs1 (PDB ID:
3ION). Four knobs are present: pThr and pThr+3 pockets in the phospho-epitope binding site in the FHA domain
and pSer and pSer+3 pockets in the phospho-epitope binding site defined by the tandem BRCT repeat. Helix α0 in
the FHA domain interacts with helices α1 and α3 in the BRCT1 domain, however the relative orientation of these
helices does not resemble that of the characteristic three-helix bundle found at the interface of tandem BRCT re-
peats. (B) Ribbon diagram of the triple BRCT repeat found in human TopBP1 (PDB ID: 2XNH) with the structural el-
ements defining each BRCT domain shown in different shades of brown. The pSer binding sites present in the
second and third BRCT domains are highlighted in blue, while additional structural elements not common to the
BRCT fold are shown in white.

The crystal structure of the FHA–BRCT–BRCT super-domain of Nbs1 emphasizes the ability of BRCT domains to build scaffolds capable of multiple interaction modes. While the two BRCT domains in Nbs1 associate to form a canonical tandem BRCT repeat with a phospho-epitope binding site, the FHA domain interacts with the hydrophobic core of the first BRCT domain leaving helices $\alpha1$ and $\alpha3$ exposed to the solvent (Figure 6) [78, 79]. In contrast to other BRCT domains, these two helices are amphipathic and, hence, break the theme of BRCT domains using hydrophobic three-helix bundles to build super-domains. Surprisingly, the FHA and the first BRCT domain (FHA–BRCT1) form the most stable unit of Nbs1, whereas the second BRCT (BRCT2) is quite flexible despite forming a canonical tandem repeat with the first BRCT domain [78, 79]. Given its tertiary structure, the FHA–BRCT–BRCT super-domain could bind two phospho-epitopes simultaneously, suggesting that the interactions of Nbs1 with its binding partners may be highly regulated.

The recent structure of the triple BRCT repeat in TopBP1 (BRCT0/1/2) provides another example of a BRCT super-domain deviating from the canonical three-helix bundle interface (Figure 6). In this case, neither BRCT0/1 nor BRCT1/2 associate to form canonical tandem BRCT repeats, primarily due to the unusually short inter-domain linkers that connect adjacent BRCTs [76]. Beyond connecting adjacent BRCT domains, the inter-domain linkers in some tandem BRCT repeats actively mediate protein-protein interactions [73, 74] and, in extreme cases, it is the linker rather than the BRCT domains that mediates the interaction. For example, the damage response protein XRCC4 interacts exclusively with the inter-domain linker connecting the two BRCT domains of ligase IV [80]. Collectively, these structures demonstrate that not only the BRCT repeat, but also the length and composition of the inter-domain linker joining the two domains affect the binding plasticity of BRCT repeats and their ability to form higher order structures with diverse binding specificities.

### 3.3. Single BRCT Domains: Is One the Loneliest Number?

The association of multiple BRCT domains with other functional domains within a single polypeptide chain is becoming a common theme found in many DNA damage response proteins. This poses the question as to whether the increased binding specificity, and hence the underlying ability to fine tune interactions during the checkpoint response, is the driving force for BRCT domains to build such complex structures. Surprisingly, the majority of eukaryotic proteins that possess BRCT domains have at least one that functions solo. The exposed $\alpha$ helices ($\alpha1$, $\alpha2$ and $\alpha3$) in single BRCT domains are amphipathic, with a hydrophobic face interacting with the central $\beta$-sheet and a polar face exposed to the solvent, unlike their tandem counterparts where both faces are chiefly hydrophobic. This enhances the stability of single BRCT domains but does not shed light on their functions.

There is mounting evidence indicating that both single and tandem BRCT domains may require additional secondary structural elements to form functional units [10, 12, 81, 82]. Some of these structural elements are required for structural stability, such as the additional C-terminal $\alpha$-helix in the PARP-1 BRCT [81], while others enhance function (Figure 7). An example of the latter comes from the largest subunit of the budding yeast clamp loader, Rfc1. Rfc1 has a single BRCT domain that is required for DNA-binding [12]. However, as an isolated unit, this do-

main is unable to recognize DNA despite having a positively charged patch positioned at the conserved P1 pocket [12]. The inclusion of an N-terminal extension recapitulates DNA binding with a $K_D$ in the nM range. The extension encompasses an additional $\alpha$ helix that directly contacts DNA [12]. Rather than being a structural element integrated into the BRCT fold, however, this helix is predicted to act as an auxiliary element to enhance function [12]. Based on this model, both the BRCT domain and its auxiliary helix likely bind synergistically to the DNA leading to a robust interaction (Figure 7). Similarly, the presence of an additional N-terminal helix necessary to bind DNA is also predicted in the single BRCT domain from the translesion polymerase Rev1 [12, 82]. It is conceivable that these additional structural elements play the dual role of stabilizing the BRCT fold and enhancing its function. In fact, this is the case of *S. cerevisiae* Dbf4, the regulatory subunit of the Cdc7 kinase, where an $\alpha$-helix immediately precedes the β1 strand and its presence is important for the stability of the domain, as well as the interaction with Rad53 during the checkpoint response [10, 83].
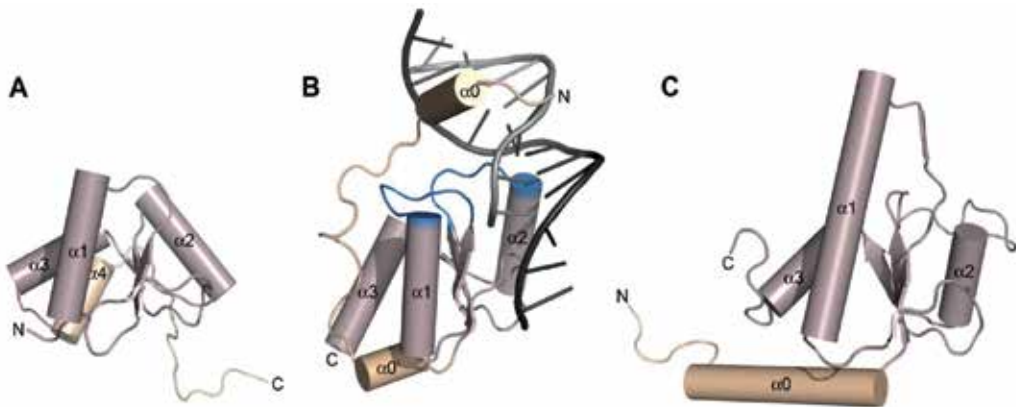


**Figure 7.** Single BRCT domains including additional structural elements. (A) Ribbon diagram of the BRCT domain of PARP-1 (PDB ID: 2LE0), highlighting an additional C-terminal α helix (α4) necessary for the structural integrity of the domain. (B) Ribbon diagram of the molecular model of Rfc1 bound to DNA (PDB ID: 2K7F). Similar to Dbf4, this model predicts two additional helices at the N-terminus of the BRCT domain of Rfc1. Helix α0 is not part of the BRCT fold, but it is essential for DNA binding by Rfc1. Conversely, the predicted helix α0' does not appear to be involved in DNA binding but its relative orientation and interaction with the BRCT core resembles that of helix α0 in Dbf4. (C) Ribbon diagram of the HBRCT domain of Dbf4 (PDB ID: 3QBZ), highlighting an additional N-terminal α helix (α0) necessary to mediate the interaction between this domain of Dbf4 and the FHA1 domain of Rad53. Helix α0 is an integral part of the fold as it anchors itself to the BRCT core through hydrophobic interactions, thereby introducing the idea of BRCT domains being building blocks that can be decorated to form super-structures with broader binding specificities.

## 4. Dbf4/Rad53: A Case Study for phosphorylation-independent BRCT and FHA Interactions

The DDK complex, formed by the association of the Cdc7 kinase and its regulatory subunit Dbf4, is required for initiating DNA replication and, hence, it is essential for the life of all eukaryotes [84]. Like many other replication proteins, DDK is also involved in the replica-

tion stress checkpoint [31-34, 85, 86]. The ability of Dbf4 to crossover into the stress response pathway is partly due to a conserved motif at the N-terminus of the protein (motif N). Based on sequence alignments, it was a matter of debate whether motif N was a bona fide BRCT domain [9, 87], however concerns were laid to rest when the crystal structure of the N-terminal region of *S. cerevisiae* Dbf4 was determined [10]. The structure revealed that this region of Dbf4 folds as a modified BRCT domain that requires an additional N-terminal $\alpha$-helix to form a stable unit (Figure 7). A fragment of Dbf4 consisting of the canonical BRCT domain but missing the additional helix did not support binding to the FHA1 domain of the checkpoint effector kinase, Rad53 [9, 10, 31, 88]. It was proposed that this additional helix ($\alpha$0) defines, at least in part, the interaction interface [10]. Due to the functional and structural relevance of the $\alpha$0 helix, this domain of Dbf4 is referred to as H–BRCT to signify the location of the additional helix. In contrast to the additional N-terminal helix identified in the BRCT domain of Rfc1 that is completely independent from the BRCT domain, helix $\alpha$0 in Dbf4 is physically latched onto the BRCT domain in the crystal structure of Dbf4 [10] (Figure 7). This is through hydrophobic residues from $\alpha$0 interacting with a hydrophobic pocket between the central $\beta$-sheet and $\alpha$1 of the BRCT domain [10]. Therefore, helix $\alpha$0 is in a sense decorating the surface of the BRCT and maintained as an integral part of the domain [10, 83]. This serves as a note of caution when studying BRCT domains in general, as the functional and the structurally stable forms of the domain may not necessarily coincide. Thus, functional BRCT units can only be reliably ascertained through empirical assays with different sized protein fragments and not through structure-guided sequence alignments.

The interaction between the H–BRCT domain of Dbf4 and the first FHA domain of Rad53 also poses an interesting paradigm during the checkpoint response because Dbf4 and Rad53 interact in a phosphorylation independent manner using domains notable for recognizing phosphorylated epitopes. This interaction was initially proposed to depend on the recognition of a phospho-epitope in Dbf4 because a point mutation in the conserved arginine (R70A) involved in phospho-threonine recognition by FHA1 effectively abolished the interaction between the two proteins in a yeast two-hybrid experiment [31]. Yeast two-hybrid experiments conducted in the past have demonstrated that phosphorylation-dependent interactions with Rad53 can indeed be detected, likely due to the activity of endogenous kinases [89]. However, all the threonine residues located within or surrounding the H–BRCT domain of Dbf4 can be mutated without abrogating the interaction with Rad53 [10]. Several possible scenarios can explain the critical role of Arg70 in mediating the interaction with Dbf4. First, Arg70 may be part of the interaction interface despite the lack of a phospho-epitope in Dbf4. This scenario would be reminiscent of *M. tuberculosis* Rv1827 that can bind several binding partners using the apical surface of the FHA domain containing the conserved arginine in a phosphorylation-independent manner [59]. Alternatively, the R70A mutation could destabilize the FHA1 fold, in which case the Dbf4-binding defect associated with this mutation would be indirect. This scenario seems unlikely, given that FHA are stable domains and that Arg70 is solvent exposed (Figure 2). Lastly, Dbf4 could have a dual interaction with Rad53, where two independent interactions would need to occur simultaneously to form a high-avidity complex. In this case, the H–BRCT domain of Dbf4 could interact with the FHA1 domain of Rad53 in a phos-

phorylation-independent manner, while FHA1 recognizes a phospho-epitope located in another region of the DDK complex. While this idea awaits validation, other dual interactions have been previously observed in the structures of other FHA domains, including that of KIF13B and Chk2 [52, 54]. This seems to suggest that simultaneous phosphorylation dependent and independent interactions may be a broader mechanism to regulate interactions mediated by FHA domains than previously anticipated.

## 5. Conclusion

Through the use of short phospho-peptides, the basic mechanism underlying pThr recognition by FHA domains has been elucidated. However, it is clear that full-length binding partners harboring pThr-epitopes will likely contact the FHA domain at multiple sites in addition to the pThr pocket, many of which will be unique to that particular interaction. Additionally, proteins are capable of interacting with FHA domains in a phosphorylation independent manner, using the phospho-epitope binding site as well as alternative interfaces, most often the surfaces of the β-sandwich. While phosphorylation-dependent interactions have a clear mode for turning the interaction on and off, regulation of phosphorylation-independent interactions remains unexplored. If these interactions are mediated by the phospho-epitope binding site it may simply be a matter of availability of phosphorylated binding partners that can compete with the unmodified protein. However, if the canonical phosphate binding residues are absent or the interaction takes place on an alternative surface, the mechanism for control is less clear.

BRCT domains have previously been divided in tandem repeats, which can interact with phosphorylated partners in a well-defined manner, and the enigmatic single domains. The plethora of interaction mechanisms used by single BRCT domains can seem overwhelming, but like FHA domains a common theme is now emerging. The BRCT fold may serve as a structural core upon which more complex and unique assemblies can be built. In this way, specific interaction surfaces can be created allowing for the BRCT domain to gain function. Therefore, BRCT domains can use extra secondary structural elements, either integrated into the fold as in Dbf4 or as an auxiliary element as in Rfc1, or entire domains such as in the case of Chs5 and Nbs1—or even in tandem BRCT repeats—to modulate their functions.

Functional and structural analyses of FHA and BRCT domains during the last decade have unveiled a complex repertoire of interactions mediated by these two domains. Once regarded as mere phospho-epitope binding units, we now know that they can mediate very sophisticated interactions regulated by multiple binding knobs. Further structural and functional analysis of protein complexes mediated by these two domains will delineate the common mechanisms that regulate the DNA damage response, and will extend the lessons learned from studying the replication stress pathway in yeast to a variety of stress response networks that rely on BRCT and FHA domains across all kingdoms of life.

## Author details

Lindsay A. Matthews and Alba Guarné[*]

*Address all correspondence to: guarnea@mcmaster.ca

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada

## References

[1] Longhese, M. P., Clerici, M., & Lucchini, G. (2003). The S-phase checkpoint and its regulation in Saccharomyces cerevisiae. *Mutat Res.*, 532, 41-58.

[2] Branzei, D., & Foiani, M. (2010). Maintaining genome stability at the replication fork. *Nat Rev Mol Cell Biol.*, 11, 208-19.

[3] Sogo, J. M., Lopes, M., & Foiani, M. (2002). Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science.*, 297, 599-602.

[4] Burhans, W. C., & Weinberger, M. (2007). DNA replication stress, genome instability and aging. *Nucleic Acids Res.*, 35, 7545-56.

[5] Arlt, M. F., Mulle, J. G., Schaibley, V. M., Ragland, R. L., Durkin, S. G., Warren, S. T., & Glover, T. W. (2009). Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. *Am J Hum Genet.*, 84, 339-50.

[6] Branzei, D., & Foiani, M. (2009). The checkpoint response to replication stress. *DNA Repair Amst*, 8, 1038-46.

[7] Mohammad, D. H., & Yaffe, M. B. (2009). 14-3-3 proteins, FHA domains and BRCT domains in the DNA damage response. *DNA repair*, 8, 1009-17.

[8] Pospiech, H., Grosse, F., & Pisani, F. M. (2010). The initiation step of eukaryotic DNA replication. *Subcell Biochem.*, 50, 79-104.

[9] Gabrielse, C., Miller, C. T., Mc Connell, K. H., De Ward, A., Fox, C. A., & Weinreich, M. (2006). A Dbf4p BRCA1 C-terminal-like domain required for the response to replication fork arrest in budding yeast. *Genetics.*

[10] Matthews, L. A., Jones, D. R., Prasad, A. A., Duncker, B. P., & Guarne, A. (2012). Saccharomyces cerevisiae Dbf4 has unique fold necessary for interaction with Rad53 kinase. *J Biol Chem.*, 287, 2378-87.

[11] Majka, J., & Burgers, P. M. (2004). The PCNA-RFC families of DNA clamps and clamp loaders. *Prog Nucleic Acid Res Mol Biol.*, 78, 227-60.

[12] Kobayashi, M. A. B. E., Bonvin, A. M. J. J., & Siegal, G. (2009). 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, 10087-97.

[13] Yamane, K., Katayama, E., & Tsuruo, T. (2000). The BRCT regions of tumor suppressor BRCA1 and of XRCC1 show DNA end binding activity with a multimerizing feature. *Biochem Biophys Res Commun.*, 279, 678-84.

[14] Yamane, K., & Tsuruo, T. (1999). Conserved BRCT regions of TopBP1 and of the tumor suppressor BRCA1 bind strand breaks and termini of DNA. *Oncogene*, 18, 5194-203.

[15] Sofueva, S., Du, L. L., Limbo, O., Williams, J. S., & Russell, P. (2010). BRCT domain interactions with phospho-histone H2A target Crb2 to chromatin at double-strand breaks and maintain the DNA damage checkpoint. *Mol Cell Biol.*, 30, 4732-43.

[16] Hammet, A., Magill, C., Heierhorst, J., & Jackson, S. P. (2007). Rad9 BRCT domain interaction with phosphorylated H2AX regulates the G1 checkpoint in budding yeast. *EMBO Rep.*, 8, 851-7.

[17] Kobayashi, J, Tauchi, H, Sakamoto, S, Nakamura, A, Morishima, K, Matsuura, S, Kobayashi, T, Tamai, K, Tanimoto, K, & Komatsu, K. (2002). NBS1 localizes to gamma-H2AX foci through interaction with the FHA/BRCT domain. *Curr Biol.*, 12, 1846-51.

[18] Williams, J. S., Williams, R. S., Dovey, C. L., Guenther, G., Tainer, J. A., & Russell, P. (2010). gammaH2A binds Brc1 to maintain genome integrity during S-phase. *EMBO J.*, 29, 1136-48.

[19] Stucki, M., Clapperton, J. A., Mohammad, D., Yaffe, M. B., Smerdon, S. J., & Jackson, S. P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell.*, 123, 1213-26.

[20] Li, X., Liu, K., Li, F., Wang, J., Huang, H., Wu, J., & Shi, Y. (2012). Structure of C-terminal tandem BRCT repeats of Rtt107 protein reveals critical role in interaction with phosphorylated histone H2A during DNA damage repair. *J Biol Chem.*, 287, 9137-46.

[21] Zou, L., & Elledge, S. J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science.*, 300, 1542-8.

[22] Majka, J., Binz, S. K., Wold, MS, & Burgers, P. M. (2006). Replication protein A directs loading of the DNA damage checkpoint clamp to 5'-DNA junctions. *J Biol Chem.*, 281, 27855-61.

[23] Green, C. M., Erdjument-Bromage, H., Tempst, P., & Lowndes, N. F. (2000). A novel Rad24 checkpoint protein complex closely related to replication factor C. *Curr Biol.*, 10, 39-42.

[24] Navadgi-Patil, V. M., & Burgers, P. M. (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.

[25] Mordes, D. A., Nam, E. A., & Cortez, D. (2008). Dpb11 activates the Mec1-Ddc2 complex. *Proc Natl Acad Sci*, U S A, 105, 18730-4.

[26] Chen, S. H., & Zhou, H. (2009). Reconstitution of Rad53 activation by Mec1 through adaptor protein Mrc1. *J Biol Chem.*, 284, 18593-604.

[27] Osborn, A. J., & Elledge, S. J. (2003). Mrc1 is a replication fork component whose phosphorylation in response to DNA replication stress activates Rad53. *Genes Dev.*, 17, 1755-67.

[28] Lee, S. J., Schwartz, M. F., Duong, J. K., & Stern, D. F. (2003). Rad53 phosphorylation site clusters are important for Rad53 regulation and signaling. *Mol Cell Biol.*, 23, 6300-14.

[29] Zhao, X., & Rothstein, R. (2002). The Dun1 checkpoint kinase phosphorylates and regulates the ribonucleotide reductase inhibitor Sml1. *Proc Natl Acad Sci*, U S A, 99, 3746-51.

[30] Huang, M., Zhou, Z., & Elledge, S. J. (1998). The DNA replication and damage checkpoint pathways induce transcription by inhibition of the Crt1 repressor. *Cell*, 94, 595-605.

[31] Duncker, B. P., Shimada, K., Tsai-Pflugfelder, M., Pasero, P., & Gasser, S. M. (2002). An N-terminal domain of Dbf4p mediates interaction with both origin recognition complex (ORC) and Rad53p and can deregulate late origin firing. *Proc Natl Acad Sci.*, U S A, 99, 16087-92.

[32] Duch, A., Palou, G., Jonsson, Z. O., Palou, R., Calvo, E., Wohlschlegel, J., & Quintana, D. G. (2011). A Dbf4 mutant contributes to bypassing the Rad53-mediated block of origins of replication in response to genotoxic stress. *J Biol Chem.*, 286, 2486-91.

[33] Lopez-Mosqueda, J., Maas, N. L., Jonsson, Z. O., Defazio-Eli, L. G., Wohlschlegel, J., & Toczyski, D. P. (2010). Damage-induced phosphorylation of Sld3 is important to block late origin firing. *Nature*, 467, 479-83.

[34] Zegerman, P., & Diffley, J. F. (2010). Checkpoint-dependent inhibition of DNA replication initiation by Sld3 and Dbf4 phosphorylation. *Nature*, 467, 474-8.

[35] Travesa, A., Kuo, D., de Bruin, R. A., Kalashnikova, T. I., Guaderrama, M., Thai, K., Aslanian, A., Smolka, M. B., Yates, J. R., 3rd, Ideker, T., & Wittenberg, C. (2012). DNA replication stress differentially regulates G1/S genes via Rad53-dependent inactivation of Nrm1. *EMBO J.*, 31, 1811-22.

[36] Wu, X., Webster, S. R., & Chen, J. (2001). Characterization of tumor-associated Chk2 mutations. *J Biol Chem.*, 276, 2971-4.

[37] Varon, R., Reis, A., Henze, G., von Einsiedel, H. G., Sperling, K., & Seeger, K. (2001). Mutations in the Nijmegen Breakage Syndrome gene (NBS1) in childhood acute lymphoblastic leukemia (ALL). *Cancer Res.*, 61, 3570-2.

[38] Drikos, I., Nounesis, G., & Vorgias, C. E. (2009). Characterization of cancer-linked BRCA1-BRCT missense variants and their interaction with phosphoprotein targets. *Proteins.*, 77, 464-76.

[39] Williams, R. S., Chasman, D. I., Hau, D. D., Hui, B., Lau, A. Y., & Glover, J. N. (2003). Detection of protein folding defects caused by BRCA1-BRCT truncation and missense mutations. *J Biol Chem.*, 278, 53007-16.

[40] Tam, A. T., Pike, B. L., & Heierhorst, J. (2008). Location-specific functions of the two forkhead-associated domains in Rad53 checkpoint kinase signaling. *Biochemistry*, 47, 3912-6.

[41] Yaffe, M. B., & Smerdon, S. J. (2004). The use of in vitro peptide-library screens in the analysis of phosphoserine/threonine-binding domain structure and function. *Annu Rev Biophys Biomol Struct.*, 33, 225-44.

[42] Hammet, A., Pike, B. L., Mc Nees, C. J., Conlan, L. A., Tenis, N., & Heierhorst, J. (2003). FHA domains as phospho-threonine binding modules in cell signaling. *IUBMB Life.*, 55, 23-7.

[43] Durocher, D., Henckel, J., Fersht, A. R., & Jackson, S. P. (1999). The FHA domain is a modular phosphopeptide recognition motif. *Mol Cell.*, 4, 387-94.

[44] Durocher, D., & Jackson, S. P. (2002). The FHA domain. *FEBS Lett.*, 513, 58-66.

[45] Mahajan, A., Yuan, C., Pike, B. L., Heierhorst, J., Chang, C. F., & Tsai, M. D. (2005). FHA domain-ligand interactions: importance of integrating chemical and biological approaches. *J Am Chem Soc.*, 127, 14572-3.

[46] Liang, X., & Van Doren, S. R. (2008). Mechanistic insights into phosphoprotein-binding FHA domains. *Acc Chem Res.*, 41, 991-9.

[47] Wang, P., Byeon, I. J., Liao, H., Beebe, K. D., Yongkiettrakul, S., Pei, D., & Tsai, M. D. (2000). II. Structure and specificity of the interaction between the FHA2 domain of Rad53 and phosphotyrosyl peptides. *J Mol Biol.*, 302, 927-40.

[48] Pennell, S., Westcott, S., Ortiz-Lombardia, M., Patel, D., Li, J., Nott, T. J., Mohammed, D., Buxton, R. S., Yaffe, M. B., Verma, C., & Smerdon, S. J. (2010). Structural and functional analysis of phosphothreonine-dependent FHA domain interactions. *Structure.*, 18, 1587-95.

[49] Lee, H., Yuan, C., Hammet, A., Mahajan, A., Chen, E. S. W., Wu, M. R., Su, M. I., Heierhorst, J., & Tsai, M. D. (2008). Diphosphothreonine-specific interaction between an SQ/TQ cluster and an FHA domain in the Rad53-Dun1 kinase cascade. *Mol Cell.*, 30, 767-78.

[50] Byeon, I. J., Li, H., Song, H., Gronenborn, A. M., & Tsai, M. D. (2005). Sequential phosphorylation and multisite interactions characterize specific target recognition by the FHA domain of Ki67. *Nat Struct Mol Biol.*, 12, 987-93.

[51]   Durocher, D., Taylor, I. A., Sarbassova, D., Haire, L. F., Westcott, S. L., Jackson, S. P., Smerdon, S. J., & Yaffe, M. B. (2000). The molecular basis of FHA domain:phospho-peptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol Cell.*, 6, 1169-82.

[52]   Li, J., Williams, B. L., Haire, L. F., Goldberg, M., Wilker, E., Durocher, D., Yaffe, M. B., Jackson, S. P., & Smerdon, S. J. (2002). Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. *Mol Cell.*, 9, 1045-54.

[53]   Hirokawa, N., & Noda, Y. (2008). Intracellular transport and kinesin superfamily proteins, KIFs: structure, function, and dynamics. *Physiol Rev.*, 88, 1089-118.

[54]   Tong, Y., Tempel, W., Wang, H., Yamada, K., Shen, L., Senisterra, G. A., Mac Kenzie, F., Chishti, A. H., & Park, H. W. (2010). Phosphorylation-independent dual-site binding of the FHA domain of KIF13 mediates phosphoinositide transport via centaurin alpha1. *Proc Natl Acad Sci*, U S A, 107, 20346-51.

[55]   Lee, J. R., Shin, H., Choi, J., Ko, J., Kim, S., Lee, H. W., Kim, K., Rho, S. H., Lee, J. H., Song, H. E., Eom, S. H., & Kim, E. (2004). An intramolecular interaction between the FHA domain and a coiled coil negatively regulates the kinesin motor KIF1A. *EMBO J.*, 23, 1506-15.

[56]   Danilenko, V. N., Osolodkin, D. I., Lakatosh, S. A., Preobrazhenskaya, M. N., & Shtil, A. A. (2011). Bacterial eukaryotic type serine-threonine protein kinases: from structural biology to targeted anti-infective drug design. *Curr Top Med Chem.*, 11, 1352-69.

[57]   Av-Gay, Y., & Everett, M. (2000). The eukaryotic-like Ser/Thr kinases of Mycobacterium tuberculosis. *Trends Microbiol.*, 8, 238-44.

[58]   Pallen, M., Chaudhuri, R., & Khan, A. (2002). Bacterial FHA domains: neglected players in the phospho-threonine signalling game? *Trends Microbiol.*, 10, 556-63.

[59]   Nott, T. J., Kelly, G., Stach, L., Li, J., Westcott, S., Patel, D., Hunt, D. M., Howell, S., Buxton, R. S., O'Hare, H. M., & Smerdon, S. J. (2009). An intramolecular switch regulates phosphoindependent FHA domain interactions in Mycobacterium tuberculosis. *Sci Signal.*, 2, ra12.

[60]   Sheng, Z. Z., Zhao, Y. Q., & Huang, J. F. (2011). Functional Evolution of BRCT Domains from Binding DNA to Protein. *Evol Bioinform Online.*, 7, 87-97.

[61]   Glover, J. N., Williams, R. S., & Lee, M. S. (2004). Interactions between BRCT repeats and phosphoproteins: tangled up in two. *Trends Biochem Sci.*, 29, 579-85.

[62]   Zhang, W., Zhang, J., Zhang, X., Xu, C., & Tu, X. (2011). Solution structure of Rap1 BRCT domain from Saccharomyces cerevisiae reveals a novel fold. *Biochem Biophys Res Commun.*, 404, 1055-9.

[63] Gong, Z., Kim, J. E., Leung, C. C., Glover, J. N., & Chen, J. (2010). BACH1/FANCJ acts with TopBP1 and participates early in DNA replication checkpoint control. *Mol Cell.*, 37, 438-46.

[64] Leung, C. C., Gong, Z., Chen, J., & Glover, J. N. (2011). Molecular basis of BACH1/ FANCJ recognition by TopBP1 in DNA replication checkpoint control. *J Biol Chem.*, 286, 4292-301.

[65] Leung, C. C., Kellogg, E., Kuhnert, A., Hanel, F., Baker, D., & Glover, J. N. (2010). Insights from the crystal structure of the sixth BRCT domain of topoisomerase IIbeta binding protein 1. *Protein Sci.*, 19, 162-7.

[66] Koonin, E. V., Altschul, S. F., & Bork, P. (1996). BRCA1 protein products... Functional motifs. *Nat Genet.*, 13, 266-8.

[67] Glover, J. N. (2006). Insights into the molecular basis of human hereditary breast cancer from studies of the BRCA1 BRCT domain. *Fam Cancer.*, 5, 89-93.

[68] Williams, R. S., Lee, M. S., Hau, D. D., & Glover, J. N. (2004). Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. *Nat Struct Mol Biol.*, 11, 519-25.

[69] Williams, R. S., Green, R., & Glover, J. N. (2001). Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1. *Nature Structural Biology*, 8, 838-42.

[70] Lofblom, J., Feldwisch, J., Tolmachev, V., Carlsson, J., Stahl, S., & Frejd, F. Y. (2010). Affibody molecules: engineered proteins for therapeutic, diagnostic and biotechnological applications. *FEBS Lett.*, 584, 2670-80.

[71] Leung, C. C., & Glover, J. N. (2011). BRCT domains: easy as one, two, three. *Cell Cycle.*, 10, 2461-70.

[72] Cescutti, R., Negrini, S., Kohzaki, M., & Halazonetis, T. D. (2010). TopBP1 functions with 53BP1 in the G1 DNA damage checkpoint. *EMBO J.*, 29, 3723-32.

[73] Derbyshire, D. J., Basu, B. P., Serpell, L. C., Joo, W. S., Date, T., Iwabuchi, K., & Doherty, A. J. (2002). Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor. *EMBO J.*, 21, 3863-72.

[74] Joo, W. S., Jeffrey, P. D., Cantor, S. B., Finnin, M. S., Livingston, D. M., & Pavletich, N. P. (2002). Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brca1 BRCT structure. *Genes Dev.*, 16, 583-93.

[75] Du, L. L., Moser, B. A., & Russell, P. (2004). Homo-oligomerization is the essential function of the tandem BRCT domains in the checkpoint protein Crb2. *J Biol Chem.*, 279, 38409-14.

[76] Rappas, M., Oliver, A. W., & Pearl, L. H. (2011). Structure and function of the Rad9-binding region of the DNA-damage checkpoint adaptor TopBP1. *Nucleic Acids Res.*, 39, 313-24.

[77] Martin-Garcia, R., de Leon, N., Sharifmoghadam, M. R., Curto, M. A., Hoya, M., Bustos-Sanmamed, P., & Valdivieso, M. H. (2011). The FN3 and BRCT motifs in the exomer component Chs5p define a conserved module that is necessary and sufficient for its function. *Cell Mol Life Sci.*, 68, 2907-17.

[78] Lloyd, J., Chapman, R. J., Clapperton, J. A., Haire, L. F., Hartsuiker, E., Li, J. J., Carr, A. M., Jackson, S. P., & Smerdon, S. J. (2009). A supramodular FHA/BRCT-repeat architecture mediates Nbs1 adaptor function in response to DNA damage. *Cell.*, 139, 100-11.

[79] Williams, R. S., Dodson, G. E., Limbo, O., Yamada, Y., Williams, J. S., Guenther, G., Classen, S., Glover, J. N., Iwasaki, H., Russell, P., & Tainer, J. A. (2009). Nbs1 flexibly tethers Ctp1 and Mre11-Rad50 to coordinate DNA double-strand break processing and repair. *Cell.*, 139, 87-99.

[80] Wu, P. Y., Frit, P., Meesala, S., Dauvillier, S., Modesti, M., Andres, S. N., Huang, Y., Sekiguchi, J., Calsou, P., Salles, B., & Junop, MS. (2009). Structural and functional interaction between the human DNA repair proteins DNA ligase IV and XRCC4. *Mol Cell Biol.*, 29, 3163-72.

[81] Loeffler, P. A., Cuneo, M. J., Mueller, G. A., DeRose, E. F., Gabel, S. A., & London, R. E. (2011). Structural studies of the PARP-1 BRCT domain. *BMC Struct Biol.*, 11, 37.

[82] de Groote, F. H., Jansen, J. G., Masuda, Y., Shah, D. M., Kamiya, K., de Wind, N., & Siegal, G. (2011). The Rev1 translesion synthesis polymerase has multiple distinct DNA binding modes. *DNA Repair (Amst).*, 10, 915-25.

[83] Matthews, L. A., Duong, A., Prasad, A. A., Duncker, B. P., & Guarne, A. (2009). Crystallization and preliminary X-ray diffraction analysis of motif N from Saccharomyces cerevisiae Dbf4. *Acta Crystallogr Sect F Struct Biol Cryst Commun.*, 65, 890-4.

[84] Johnston, L. H., Masai, H., & Sugino, A. (1999). First the CDKs, now the DDKs. *Trends Cell Biol.*, 9, 249-52.

[85] Ogi, H., Wang, C. Z., Nakai, W., Kawasaki, Y., & Masumoto, H. (2008). The role of the Saccharomyces cerevisiae Cdc7-Dbf4 complex in the replication checkpoint. *Gene*, 414, 32-40.

[86] Stead, B. E., Brandl, C. J., Sandre, M. K., & Davey, M. J. (2012). Mcm2 phosphorylation and the response to replicative stress. *BMC Genet.*, 13, 36.

[87] Masai, H., & Arai, K. (2000). Dbf4 motifs: conserved motifs in activation subunits for Cdc7 kinases essential for S-phase. *Biochem Biophys Res Commun.*, 275, 228-32.

[88] Varrin, A. E., Prasad, A. A., Scholz, R. P., Ramer, M. D., & Duncker, B. P. (2005). A mutation in Dbf4 motif M impairs interactions with DNA replication factors and confers increased resistance to genotoxic agents. *Mol Cell Biol.*, 25, 7494-504.

[89] Pike, B. L., Yongkiettrakul, S., Tsai, M. D., & Heierhorst, J. (2004). Mdt1, a novel Rad53 FHA1 domain-interacting protein, modulates DNA damage tolerance and G(2)/M cell cycle progression in Saccharomyces cerevisiae. *Mol Cell Biol.*, 24, 2779-88.

# Intrinsically Disordered Proteins in Replication Process

Apolonija Bedina Zavec

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/51673

## 1. Introduction

Intrinsically disordered proteins (IDPs) are proteins that lack stable tertiary conformation (3D structure) under physiological conditions and are biologically active in their unstructured form. IDPs are disordered either along their entire lengths, but more often they are disordered only in localized regions, intrinsically disordered regions (IDRs).

IDRs often undergo transitions to more ordered states after binding to their targets and adopt a fixed three dimensional structures. Folding transition enables specificity without excessive binding strength. Important characteristic of IDRs is multispecificity. One IDR is able to bind multiple targets (multispecific recognition) because it can adopt different conformations upon interaction with different binding partners [1]. IDPs are able to simultaneously bind their partners, which enable the assembly of large complexes. An additional functional advantage of IDPs is increased speed of the interaction due to greater capture radius and larger interaction surfaces.

The level of IDPs is tightly regulated in a cell and diverse post-translational modifications facilitate regulation of their function [2].

IDRs with multispecific recognition capabilities are especially important for the complex recognition processes. Therefore, IDRs are particularly enriched in proteins implicated in cell signalling. It is known that the majority of transcription factors and proteins involved in signal transduction contain long disordered segments [3]. How about IDPs in replication process? The analysis of the yeast proteome showed that IDPs are often located in the cell nucleus [4]. In addition, IDRs are abundant in DNA-binding proteins and many replication and recombination proteins are DNA-binding proteins. Many IDPs are involved in recognition and regulation pathways, because interactions with multiple partner molecules and high-specificity/low-affinity interactions are extremely important in these pathways. Additional interesting feature of IDRs is that they are very sensitive to the environment (Subchapter 4.2.). Summarizing these

findings, a high level of protein disorder is to be expected in processes that take place in the cell nucleus and the highest level of disorder is expected in processes involved in responses to environmental changes. Therefore it is expected that in the nucleus, transcription is a process with the highest level of IDPs. Recombination and repair processes are also expected to have many IDPs; however, these processes are tightly linked to DNA replication and many proteins are used by all three processes. DNA replication is a process that proceeds by a precise program with a defined temporal order. The structural and functional properties of IDPs indicate that a disordered structure is likely present to a lesser extent in DNA replication process. Because of the need for responsiveness to the environment, the initiation of DNA replication should engage more IDPs than the elongation of DNA replication. It is expected that the majority of IDPs in these processes are regulatory proteins.

In this chapter, the binding mechanism of IDRs, the level of IDRs in replication and recombination proteins, and the role of IDPs in replication and recombination processes are discussed.

## 2. Intrinsically disordered proteins

IDPs contain one or more long intrinsically disordered regions (IDRs) or they are disordered along their entire lengths. Structural disorder can span from short stretches, through long regions, to entire proteins [5]. The majority of IDRs is not fully disordered, but contains some secondary structure and sometimes even partial tertiary structure. IDRs are dynamic fluctuating systems that exist as structural ensembles of rapidly interconverting alternative conformations and perform their biological functions in a highly dynamic disordered state; however, they often have more compact configurations than simply a random coil and contain sites of molecular recognition [6]. The structure of IDPs is similar to a molten globule or pre-molten globule, which preserve the main elements of the native secondary structure and the approximate positions of the folded state, while the loops and ends are flexible. Structural flexibility is a major feature and a major functional advantage of these proteins. IDRs are rich in binding sites for various partners and these binding sites mean that many IDPs with flexible structure are polyfunctional proteins.

The disordered structure gives IDPs specific properties. They need no stable conformation to remain functional; therefore, they are more robust to different changes. Contrary to globular proteins, IDPs are stable at extreme temperatures and extreme pHs [7]. Increases and decreases in temperature or pH can even induce partial folding of IDPs. It has been shown that IDPs partially fold at extreme pH due to minimization of their large net charge present at neutral pH. An increase in temperature can also induce the partial folding of IDPs; in addition, they are resistant to freeze-thaw treatment [8].

### 2.1. Amino acid (AA) composition of IDPs

IDPs have a specific AA composition that differs from the AA composition of ordered proteins. In particular, IDPs are depleted in hydrophobic (Ile, Leu, Val) and aromatic AA (Trp, Tyr, Phe) that stabilize the structure of folded proteins, while they are enriched in

hydrophilic and charged AAs. The charge/hydropathy (C/H) ratio has been suggested to govern the degree of compaction in IDPs [9]. The combination of low hydrophobicity and high net charge represents an important prerequisite for the disordered structure under physiological conditions.

### 2.2. Evolution

IDPs are more abundant in eukaryotes than in archaea and prokaryotes, while multicellular eukaryotes have much more predicted disorder than unicellular eukaryotes [4,10]. IDPs play an important role in complex organisms by participating in recognition and in various signalling and regulatory pathways.

IDRs show higher robustness against mutations [11], presumably because changes in protein sequence do not affect protein stability and function as severely. IDRs are more tolerant of mutations than structured proteins. It was found that flexible proteins exhibiting functional promiscuity are the foundation stones of protein evolvability [12]. They are able to accumulate a large number of mutations and thereby facilitate adaptation. Structural disorder seems to enable the rapid appearance of novel, 'less-evolved' proteins [13]. It has been shown that in alternative splicing both alternative proteins have high disorders, because the chance is very low that dual coding would result in two sequences that are both capable of folding into well-defined, functional, 3D structures [14].

## 3. The binding mechanism of IDRs

IDPs bind to their molecular partners and perform their biological functions by regulation of the function of their binding partners or by promotion the assembly of multi-molecular complexes. One IDR is able to bind many different partners because of its flexible structure; on the other hand, some IDRs do not bind to any partner, but they provide flexible linkers between domains that maintain constant motion during functioning or they provide flexible tails that regulate the structured domains [7,15].

IDPs have functionally relevant characteristics:

- They frequently fold up upon binding to their biological targets [16]. The interaction of a disordered protein with a structured partner, very often induces a disorder-to-order transition thereby forming stable structures, enabling high-specificity-low-affinity interactions [17,18].

- They have possibility of overlapping binding sites (binding diversity) due to extended linear conformation [19]. Structural flexibility of IDPs enables their interactions with numerous biological targets.

- IDRs enable a very large accessible surface area [20]. Greater capture radius and larger interaction surfaces enable increased speed of interactions [15].

- IDPs undergo tighter regulation by post-translational modification as compared to structured proteins [2].

### 3.1. Complexes with IDPs

Molecular complexes with IDPs are diverse: the IDR may bind on the surface of the binding partner (Figure 1), by wrapping around the binding partner, or by penetrating deep inside the binding partner [21].
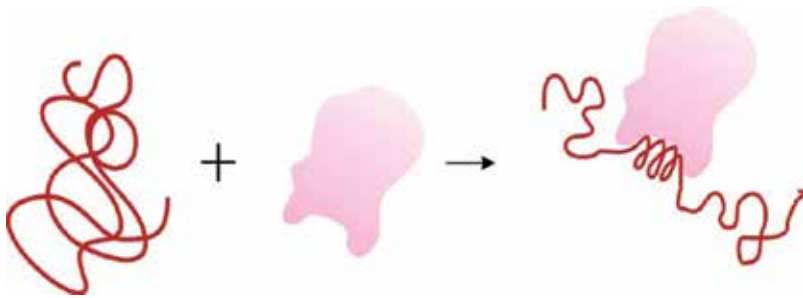


**Figure 1.** Intrinsically disordered protein forms complex with structured protein.

IDRs in complexes may control the degree of motion between domains, mask binding sites, enable transient binding of different binding partners, and be targets of post-translational modifications. IDPs are often involved in the binding of large partners or they are proteins involved in the binding of large number of small partners. In the latter case, they often function as scaffold proteins that enable the assembly of the relevant proteins into specific multi-molecular complexes and increase the efficiency of the interaction between partner molecules (Figure 2).



**Figure 2.** Intrinsically disordered proteins often function as scaffold proteins that enable assembling the relevant proteins into multi-molecular complexes.

The majority of intrinsically disorder-based complexes are ordered and relatively static due to disorder-to-order transitions; however, there are also dynamic complexes where IDRs go through an ensemble of rapidly interconverting conformations. Dynamic complexes do not involve significant ordering of the interacting protein segments but rely exclusively on transient contacts.

### 3.2. Mechanisms of formation of the complexes with IDPs

The primary mechanism by which disorder is utilized in molecular interactions is that the same IDR may fold differently and bind to several structurally diverse partners. On the other hand, different IDRs with different AA sequences may use their flexibility to bind to the same protein partner [6,22]. Their associations are dynamic. The lack of structure of highly flexible IDRs enables more diverse functionality [23]. IDRs are ensembles of conformations and each individual conformation has a dynamic structure. The binding partner selects the most binding-compatible conformation from this ensemble to form a complex [21,15]. The equilibrium is thus shifted towards this interaction-prone member of the conformational ensemble.

Models of IDRs interaction processes:

- The 'binding and folding' mechanism with disorder-to-order transition is the most accepted model for the binding of IDR, where a highly structured conformation is formed by binding to the partner molecule. A structured conformation is formed on binding IDR (the local disorder-to-order transition) or on the entire molecule of IDP (the global disorder-to-order transition) [25,26]. An IDR binds weakly at a relatively large distance followed by folding when the protein comes close to the binding site. One model utilizes a prediction that an IDR with an open structure has a larger binding surface and a greater capture radius for a specific binding site than the ordered protein and therefore the binding rate is significantly enhanced over the binding rate of the ordered proteins [21]. The binding induced disorder-to-order transition is accompanied by a dramatic decrease in accessible surface area and by the release of a large number of water molecules [6]. A large decrease in conformation entropy during this process enables highly specific but easily reversible interactions.

- The 'polyelectrostatic' model describes the interaction of highly charged IDR with several similar binding motifs and a folded partner with one binding site [27,28]. Multiple disordered binding motifs interact with the partner's folded binding site in a dynamic equilibrium. The flexibility of the IDR makes all binding motifs equally accessible. Weak affinities of the individual interactions permit their efficient exchange. In this model, the IDR generates an electrostatic field representing the cumulative electrostatic interaction of all charges in the IDR.

- The 'multi-step interaction' model describes the binding of an IDR that depends on the conformational selection of the structural ensemble via the pre-formed elements that dominate the ensemble [29]. When the IDP comes close to the binding site of the partner molecule, an encounter complex is formed that either proceeds towards the final complex or dissociates again. Electrostatic forces are the most important for encounter complex formation [30]. Interacting partners in the encounter complex affect the conformational landscapes of each other. Consecutive steps depend on the preceding steps and cooperation between protein partners. This process is called an interdependent protein dance [31,32]. The structural variability of complexes with IDPs can be considered a reflection of interdependent protein dynamics, where the structure of the complex is a result of coordinated mutual co-folding [21]. In such encounters 'pre-

organized' complexes, mainly non-specific electrostatic interactions are involved and multiple conformations and orientations are employed. In the 'multi-step interaction' model, IDPs interacts with their partners by a biphasic process with a fast Phase I leading to the formation of disordered complexes and slower Phase II leading to the formation of ordered complexes. Phase II includes the 'Binding and folding' model that may or may not (binding without folding) follow a Phase I [33]. 'Polyelectrostatic' complexes are probably the stopped stages of encounter complexes [21].

It is the most likely that the IDR contains a conformational preference for the structure it will take upon binding.

### 3.3. Levels of IDPs in the cell and modulation of their activity

The level of IDPs inside the cell is precisely controlled. IDPs are more tightly regulated as compared to structured proteins. Obviously it is very important that they are available at the appropriate time and in the appropriate amount. The level of IDPs is controlled at the synthesis and clearance levels and their activity is further modulated via interaction with specific binding partners and post-translational modifications. IDRs are more solvent-accessible then folded regions and therefore suitable for diverse post-translational modifications, such as phosphorylation, sumoylation, ubiquitination, acetylation, etc. Such modification can change the electrostatic properties of IDRs and affect their affinity for charged molecules like DNA.

The predicted intrinsic disorder is the strongest determinant of dosage sensitivity - proteins become harmful when they are overexpressed [34]. The likely cause of dosage sensitivity is the binding promiscuity of IDPs [11]. IDRs are prone to make promiscuous interactions when their concentration is increased; it has been demonstrated that this is a likely cause of pathology when genes are overexpressed [34].

## 4. The role of IDPs in replication processes

This chapter refers to the proteins of budding yeast *Saccharomyces cerevisiae*, because *S.cerevisiae* is the best studied eukaryotic model organism that providing the most integrated view of replication and recombination processes.

### 4.1. DNA replication process is tightly linked to recombination process

DNA replication and DNA recombination are central characteristics of life that cooperate to maintain biological inheritance and genomic integrity. Replication enables the formation of two identical DNA molecules from a single double-stranded DNA, while recombination enables accurate repair of errors that occur on both strands of DNA, as well as the formation of new combinations of genes. Both processes are tightly intertwined [35]. The recombination system plays a crucial role in DNA replication ensuring that the replication machines can complete their task of genome duplication. DNA replication forks stall or collapse at DNA

lesions or problematic genomic regions. When replication forks collapse, recombination is the most important rescue mechanism. The recombination mechanism forms substrates for the assembly of a new replication fork thus allowing continued DNA replication. On the other hand, DNA synthesis is a crucial step during the recombination process. After Rad51-mediated DNA strand invasion, DNA synthesis is the next step in recombination to restore the integrity of the chromosome. Repair DNA synthesis during the recombination process is similar to normal S-phase replication, but has specific properties. Thus recombination is part of DNA replication and, vice versa, DNA synthesis is part of the recombination process.

Clearly then, the replication process requires both, replication and recombination proteins, but then again so does the recombination process. This is why replication and recombination proteins are discussed within the same functional group.

## 4.2. Predicted level of IDPs in replication and recombination processes

There are some facts to consider when predicting the level of IDPs in replication and recombination processes:

- An analysis of the yeast proteome showed that IDPs are often located in the cell nucleus [4]. IDRs are abundant in DNA-binding proteins, while many replication and recombination proteins are DNA-binding proteins. IDRs play a crucial role in DNA-binding proteins by increasing the affinity and specificity of DNA binding [36]. The ability of IDRs to interact with DNA is tightly linked to the high content of charged residues in IDRs; IDRs that bind to DNA are rich in positively charged residues and their positive charges are highly clustered.

- Many IDPs are involved in recognition and regulation pathways, because interactions with multiple partner molecules and high-specificity/low-affinity interactions are extremely important in these pathways [2].

- Interesting feature of IDRs is that they are very sensitive to the environment. Flexible IDPs more readily undergo conformational change in response to environmental perturbations than rigid proteins [37,38]. Due to flexible structure, their local and global structures can easily be shaped by their environment. High-specificity/low-affinity interactions with their partners enable extremely sensitive functioning of IDPs, which is favourable for responses to the environmental changes. In addition, the level of IDPs inside the cell is precisely controlled (Subchapter 3.4.) allowing rapid and accurate responses of the cell to changing environmental conditions. Higher and more regulated synthesis, higher degradation rates, and tightly regulated activity make the levels of IDPs very sensitive to the environment.

Summarizing these findings, a high level of protein disorder is to be expected in processes that take place in the cell nucleus, especially within regulatory proteins. The highest level of disorder is expected in processes involved in responses to environmental changes. According to those findings, in the nucleus, transcription should be the process with the highest level of IDPs. Recombination and repair processes are also expected to have many IDPs; however, these processes are tightly linked to DNA replication and many proteins are used by all these processes. DNA replication is a process that proceeds by a precise program with

a defined temporal order. The structural and functional properties of IDPs indicate that a disordered structure is likely present to a lesser extent in DNA replication process. However, the initiation of DNA replication would be expected to engage more IDPs than the elongation of DNA replication due to the need for responsiveness to the environment (Figure 3). It is expected that the majority of IDPs in these processes are regulatory proteins.
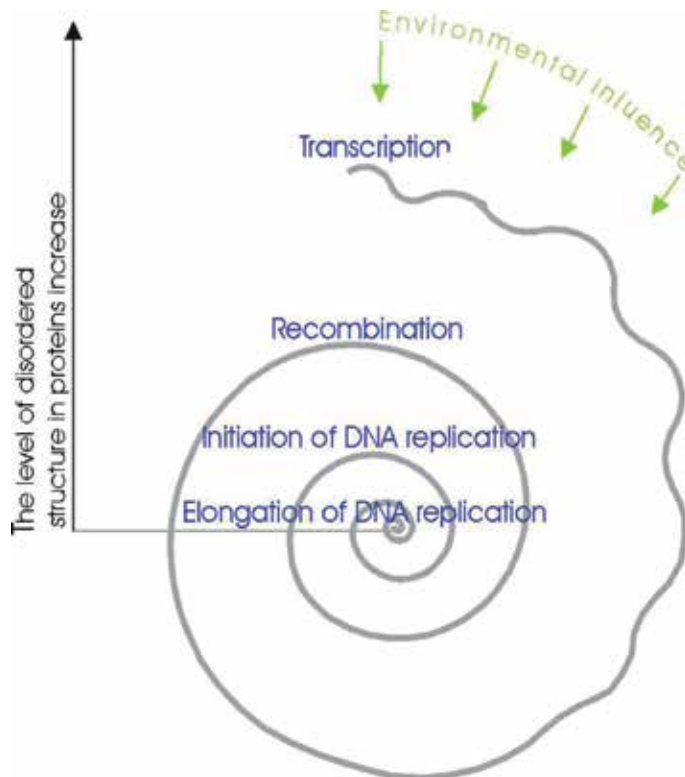


**Figure 3.** Processes in the nucleus: global prediction of protein disordered structure in the processes linked to DNA, considering responsiveness to changes in the environment.

### 4.3. IDRs in replication proteins

Analysis of predicted IDRs within proteins that have a role in DNA replication was done (Table 1). The majority of them functions also in recombination and repair processes.

It was found that proteins with the role in initiation of DNA replication have more predicted disordered structure (26%) than proteins with the role in elongation of DNA replication (20%). Difference is significant among proteins with very short IDRs; there is 35% proteins with the role in initiation of DNA replication that contain less than 10% disordered structure, while there is as much as 60% such proteins with the role in elongation of DNA replication. However, difference is tiny among proteins with very large IDRs; 22% proteins with the role in initiation of DNA replication contain more than 50% disordered structure and there is 20% such proteins with the role in elongation of DNA replication.

| Protein | Protein Length (AA) | *Location of IDR (AA) | *IDR Length (AA) | % of disorder | Recombination and Repair |
|---|---|---|---|---|---|
| | | Elongation of DNA replication | | | |
| Pol2 | 2222 | 0-44, 1186-1263 | 44, 77 | 5.4 | R |
| Dpb2 | 689 | 92-159 | 67 | 9.7 | R |
| Dpb3 | 201 | 96-201 | 105 | **52.2** | R |
| Dpb4 | 196 | 0-26, 119-196 | 26, 77 | **52.6** | R |
| Pol3 | 1097 | 0-100 | 100 | 9.1 | R |
| Pol31 | 487 | 0-32 | 32 | 6.6 | R |
| Pol32 | 350 | 118-350 | 232 | **66.3** | R |
| PCNA | 258 | - | 0 | 0 | R |
| Cdc9 | 755 | 0-144 | 144 | 19.1 | R |
| Rfa1 | 621 | 126-183 | 57 | 9.2 | R |
| Rfa2 | 273 | 0-39, 177-235 | 39, 58 | 35.5 | R |
| Rfa3 | 122 | - | 0 | 0 | R |
| Rfc1 | 861 | 0-155, 230-296, 780-861 | 155, 66, 81 | 35.1 | R |
| Rfc2 | 353 | 0-25 | 25 | 7.1 | |
| Rfc3 | 340 | - | 0 | 0 | |
| | | Initiation of DNA replication | | | |
| Pol1 | 1468 | 82-341 | 259 | 17.6 | R |
| Pol12 | 705 | 70-203 | 133 | 18.8 | R |
| Pri1 | 409 | - | 0 | 0 | R |
| Pri2 | 528 | 0-43 | 43 | 8.1 | R |
| Orc1 | 914 | 0-42, 195-428 | 42, 233 | 30.1 | |
| Orc2 | 620 | 0-267 | 267 | 43.1 | |
| Orc3 | 616 | 0-40 | 40 | 6.5 | |
| Mcm2 | 868 | 0-68, 112-189 | 68, 77 | 16.7 | R |
| Mcm3 | 971 | 742-897 | 155 | 16.0 | R |
| Mcm4 | 933 | 0-177 | 177 | 19.0 | R |
| Mcm10 | 571 | 41-154, 338-571 | 113, 233 | **60.6** | |
| Sld2 | 453 | 0-453 | 453 | **100** | |
| Sld3 | 668 | 87-155, 292-335, 417-668 | 68, 43, 271 | **57.2** | |
| Sld5 | 294 | 17-45 | 28 | 9.5 | |
| Psf1 | 208 | - | 0 | 0 | |
| Psf2 | 213 | 195-213 | 18 | 7.8 | |
| Psf3 | 194 | - | 0 | 0 | |
| Dpb11 | 764 | 226-321, 567-764 | 95, 197 | 38.2 | R |
| Cdt1 | 604 | 427-507 | 80 | 13.2 | R |
| Cdc6 | 513 | 0-61 | 61 | 11.9 | |
| Cdc7 | 507 | - | 0 | 0 | R |
| Dbf4 | 704 | 0-110, 317-658 | 110, 341 | **64.1** | R |
| Ecm11 | 302 | 0-183 | 183 | **60.6** | R |

Data about proteins were obtained from *Saccharomyces genome* database [39] and references therein.
Server Disopred2 [40] was used for protein disorder prediction.
*Predicted disorder regions more than 30 AA long (more than 20 AA long for very small proteins).
R - proteins involved in recombination and repair processes;
Disordered regions at N- or C- terminus are underlined.

**Table 1.** Predicted disorder regions of replication proteins.

The majority (63%) of IDRs are at N- or C- terminus in proteins that have a role in DNA replication. Most often IDRs are at the N-terminus, in 44% of replication proteins; 19% of replication proteins have IDRs at the C-terminus.

## 4.4. IDRs in DNA polymerases

### 4.4.1. DNA polymerase α

Polymerase α is the only enzyme that can synthesize DNA de novo. It is required for initiation of chromosomal DNA replication during mitosis and meiosis, intragenic recombination, and repair of double stranded DNA breaks. Pol1, the catalytic subunit of polymerase α complex, has a lot of disordered structure in comparison to Pol3 and Pol2 (Table 1). This fact is consistent with the hypothesis concerning the expected average level of disordered structures, since Pol1 is required for the initiation of DNA replication and Pol3 and Pol2 are required for the elongation of DNA replication.

It was shown that IDR of Pol1 interacts with Cdc13: Pol1 residues 13-392 [41] or Pol1 residues 47-560 [42]. Actually, a short fragment of Pol1 consisting only of residues 215-250 is necessary and sufficient for binding with Cdc13. This disordered region of Pol1 becomes well ordered, folded into a single amphipathic α-helix, when it is in complex with Cdc13, as evidenced by good electron density in the crystals [43]. The interaction between IDR of Pol1 and Cdc13 is primarily mediated by a highly positively charged groove of Cdc13 and a negatively charged acidic convex surface of Pol1. These two surfaces are not only opposite in charge distribution but also complementary in shape.

### 4.4.2. DNA polymerase δ

DNA polymerase δ is a major replicative DNA polymerase and is primarily required for the lagging strand synthesis. It is a heterotrimeric complex composed of the catalytic subunit Pol3, the structural subunit Pol31, and an additional auxiliary subunit Pol32. Pol32 is highly disordered protein (Figure 4). While structured Pol3 and Pol31 are essential for viability, the disordered Pol32 is not essential. Pol3 and Pol31 are highly conserved in eukaryotes; on the other hand, the disordered Pol32 shows an extreme divergence in its AA sequence [44]. Hydrodynamic studies of polymerase δ have shown an unusually high Stokes radius [45]. This deviation from globularity may be due to the disordered structure of Pol32.

Pol32 is bound to Pol3 through Pol31. The C-terminus of Pol3 interacts with the conserved region of Pol31 [46]. Deletion of the last four C-terminal AAs of Pol3, which are required for the interaction between the Pol3 and Pol31, does not affect DNA replication but leads to defects in homologous recombination and in break-induced replication repair pathways. Deletion of Pol32 leads to signs of DNA replication defects and DNA repair defects, with increased sensitivity to ultraviolet (UV) radiation and methylation damage [47].

Pol32 binds to Pol31 by the N-terminus (92 AA) and to PCNA by the C-terminus [48]. The structured N-terminus of Pol32, which enables binding to Pol3 through Pol31, is essential for damage-induced mutagenesis. Highly disordered C-terminus of Pol32 interacts with the

C-terminus of PCNA during DNA synthesis. Although the C-terminus of Pol32 is highly disordered, there is one motif that is highly conserved in this region: the consensus PCNA-binding motif 338-QGTLESFFKRKAK-350 (conserved amino acids in bold).
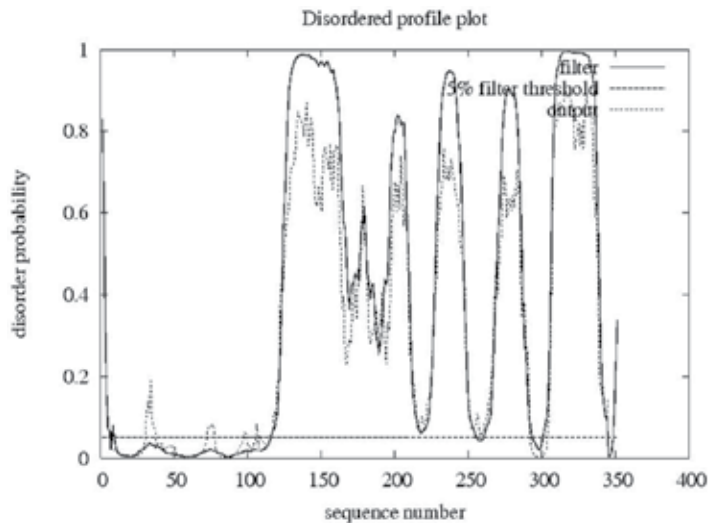


**Figure 4.** Predicted disordered regions of Pol32 (Disopred2).

It has also been shown that Pol32 interacts also with Pol1 that is a part of polymerase $\alpha$, suggesting that Pol $\delta$ and Pol $\alpha$ interact *via* the Pol32 subunit [48]. These findings show diverse role of Pol32 as typical IDP.

For the replication of the lagging strand where the polymerase must dissociate from the DNA after extension of each Okazaki fragment, Pol$\delta$ utilizes a collision-release mechanism where the Pol$\delta$ is released from PCNA. Pol$\delta$ exhibits a very high processivity in synthesizing DNA with the PCNA sliding clamp. It has been shown that the N-terminal region of Pol3 interacts with PCNA, and that this interaction increases Pol3 processivity [49]. The N-terminal of Pol3 is predicted to be highly disordered (Table 1). Pol31 and Pol32 also have binding sites for PCNA and all three subunits contribute to PCNA-stimulated DNA synthesis by Pol$\delta$ [50].

### 4.4.3. DNA polymerase ε

DNA polymerase ε is primarily required for the leading strand synthesizes. Pol2 is the catalytic subunit of the polymerase ε complex. It has been shown that the highly structured C-terminus of Pol2 is essential for DNA replication [51].

Dpb3 and Dpb4 are nonessential small subunits of the DNA polymerase ε complex, which have a histone fold. Both Dpb3 and Dpb4 are highly disordered. They form a sub-assembly that interacts with histones and functions in transcriptional silencing caused by chromatin structures [52].

### 4.5. Dbf4, IDP with the role in replication and recombination

The complex Cdc7–Dbf4, also known as Dbf4-dependent kinase (DDK), has a role at eukaryotic origins of replication. DDK is required for origin firing and replication fork progression in mitotic S phase, for pre-meiotic DNA replication, meiotic double strand break formation, recruitment of the monopolin complex to kinetochores during meiosis I and as a gene-specific regulator of the meiosis-specific transcription factor Ndt80p. DDK is a Ser/Thr kinase whose activity depends on the association of the Cdc7 catalytic subunit with a regulatory subunit Dbf4. The level of Dbf4 is changes during the cell cycle and is the highest during metaphase I [53]. Both subunits Cdc7 and Dbf4 are essential for growth.
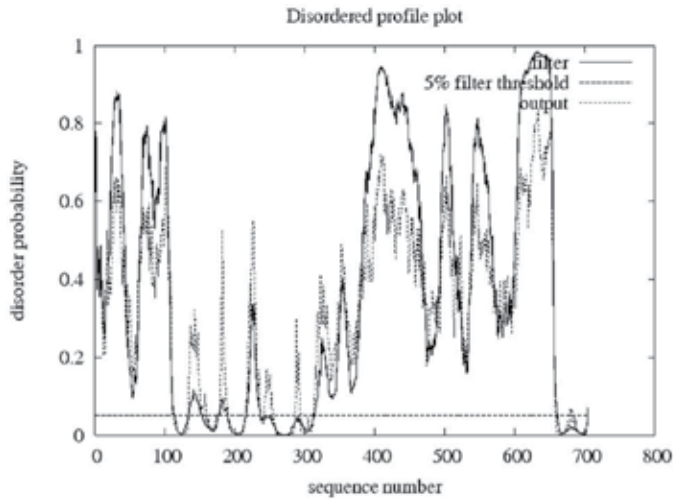


**Figure 5.** Predicted disordered regions of Dbf4 (Disopred2).
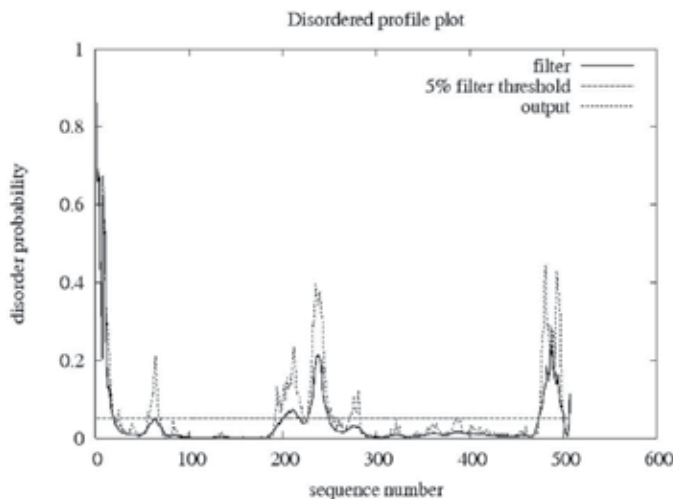


**Figure 6.** Predicted disordered regions of Cdc7 (Disopred2).

As expected, the regulatory subunit Dbf4 is highly disordered (Figure 5), while catalytic subunit Cdc7 is a highly structured protein (Figure 6).

Dbf4 is a highly disordered protein with a disordered N-terminus (110 AA) and an IDR that is half the length of the protein at the C-terminus; it has only a 200 AA long structural region between both IDRs (Figure 5). It was shown that highly disordered C-terminus of Dbf4 has a role in response to mutation by HU and that it is required in meiosis [54]. Superfamily assignments [55] show no confident structure prediction for Dbf4, while Cdc7 has a predicted cyclin-dependent protein kinase function at 1-304 AA with protein, ATP, and DNA binding activity. Cdc7 has well conserved subdomains (30-195, 275-348, 438-465) found in the eukaryotic protein kinase superfamily, while Dbf4 contains only three short conserved regions, termed N (135-179), M (260-309), and C (659-696) [56]. Two of the three conserved regions (N and M) are found in the the structural region of Dbf4.

### 4.5.1. Dbf4 in initiation of replication during mitosis

DDK phosphorylates the Mcm2-7 helicase, and is probably required for helicase activation or for recruitment of pre-IC factors. DDK preferentially phosphorylates the MCM complexes that are most tightly linked to the DNA [57]. Dbf4 associates with origins in an ORC-dependent manner [58]. The pre-RC components Mcm2, Mcm4, Orc2, and Orc3 have each been identified as binding partners for Dbf4 [59,60,61]. The N-terminal half of Dbf4 is critical for recruitment of DDK to the origin. The highly disordered C-terminal half of Dbf4 is required to bind the Cdc7 kinase [58]; more precisely, region 573-695 is required for interaction with Cdc7, while the structured region of Dbf4 (110-296) is required for binding the Mcm2–7 complex [60].

### 4.5.2. Dbf4 in checkpoint control

During the replication checkpoint response, Dbf4 is phosphorylated by checkpoint kinase Rad53 allowing inhibition of initiation of replication at late origins. Checkpoint control during S-phase slows the rate of DNA replication in response to DNA damage and blocks the replication fork. This regulation is achieved through the Rad53 kinase-dependent block of late origins of replication [62]. Dbf4 has been shown to be phosphorylated in a Rad53-dependent manner in response to replication stress, which correlates with a reduced DDK activity [63]. It was shown that mutations at predicted Rad53 phosphorylation sites (Ser84, Ser235, Ser377, Thr467, Thr506, Ser507, and Thr551) contribute to bypassing such control [64].

The conserved region N of Dbf4 (66-221) is necessary for the interaction of Cdc7-Dbf4 with the checkpoint kinase Rad53. The core of this binding region folds as a BRCT domain; in addition, it includes an additional N-terminal helix unique to Dbf4 that is essential for the interaction with Rad53 [65]. This unique N-terminal part of the conserved region N is predicted to be an IDR (Figure 6) and probably becomes helix-structured after binding with Rad53.

### 4.5.3. Dbf4 in meiosis

DDK is required for replication, recombination and segregation events during meiosis in yeast. It has been shown that in addition to the initiation of DNA replication, DDK has an

important role in the initiation of meiotic recombination [66]. DDK phosphorylates the double strand break protein Mer2 and facilitates meiotic recombination [67]. CDK-S and DDK function sequentially phosphorylate Mer2 on adjacent serines, Ser30 and Ser29, allowing formation of meiotic double strand breaks.

DDK plays a role in meiotic segregation. DDK allows expression of *NDT80*, a global transcription factor in meiosis, required for the induction of genes required for meiotic progression and spore formation. DDK promotes *NDT80* transcription by relieving repression mediated by a complex of Sum1, Rfm1, and histone deacetylase Hst1. Sum1 exhibits meiosis-specific Cdc7-dependent phosphorylation. By this function, DDK links DNA replication to the segregation of homologous chromosomes in meiosis I [68,69].

DDK is also necessary for recruitment of monopolin Mam1 to sister kinetochores, which is required for mono-orientation of sister kinetochores in the reductional segregation occurring during meiosis I.

The use of the same Cdc7-Dbf4 complex to regulate many distinct meiosis-specific processes could be important for the coordination of these processes during meiosis [68]. DDK is a link between DNA replication, recombination and mono-orientation during meiosis I in budding yeast [70]. In addition to the unifying role in meiosis, DDK has a role in initiation of replication during mitosis and in checkpoint control. Highly flexible structure of Dbf4 is very likely crucial for such a complex role of DDK.

### 4.6. Ecm11, IDP with the role in replication and recombination

Ecm11 is a protein with a strong meiotic phenotype; it affects meiotic DNA synthesis and recombination [71]. Homozygous deletion of the *ECM11* gene causes delay in a process of meiosis, lower efficiency of ascii formation and lower spore viability.
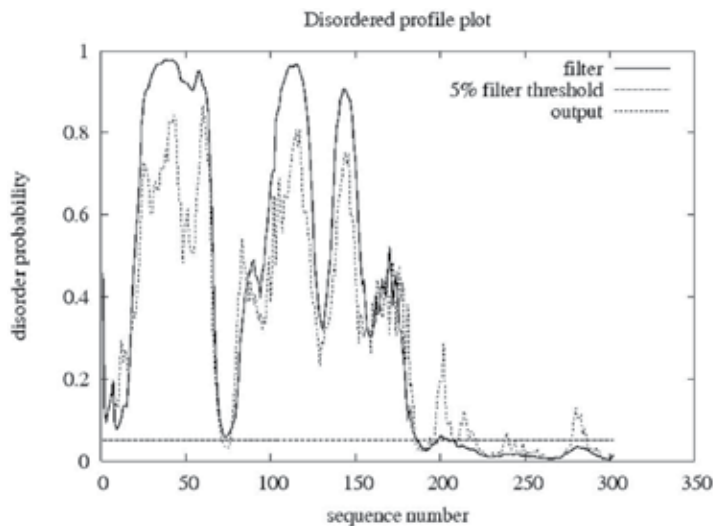


**Figure 7.** Predicted disordered regions of Ecm11 (Disopred2).

Ecm11 is highly disordered protein; 2/3 of Ecm11 is unstructured (Table 1). Ecm11 has 302 mostly hydrophilic AA as expected for IDP (Subchapter 2.1.). IDR of 183 AA is located at the N-terminal end of Ecm11 (Figure 7). The C-terminal end is predicted to be mostly helical and contain coiled-coil motif at the very C-terminus. Superfamiliy assignments [55] show no confident structure prediction for Ecm11.

### 4.6.1. The ecm11 mutation affects sporulation efficiency

It was showed that *ecm11* homozygous diploid strains sporulate more slowly and less efficiently than the wild type strains [71]. Wild type strains carrying additional *ECM11* on the centromeric plasmid also showed reduced sporulation efficiency comparing to wild types. Obviously, sporulation efficiency depends on the copy number of Ecm11 protein in the cell during meiosis. As more Ecm11 than usual in the cell make lower sporulation efficiency, Ecm11 is probably a part of heterologous protein complex, demanding exactly correct balance among those proteins.

### 4.6.2. Ecm11 has a role in meiotic recombination

It was showed that *ecm11* homozygous spores have reduced viability for 50% [71]. The majority of *ecm11* ascii (56%) produced only two viable spores, while only 1% of such ascii were observed in the parental strain. This result shows non-disjunction of homologous chromosomes at the first meiotic division. By recombination tests was demonstrated that *ECM11* is required for crossing over, but not for gene conversion. This result raises the possibility that *ecm11* mutation impairs the crossover process at an early step of recombination, at the differentiation of intermediates into crossovers or non-crossovers.

### 4.6.3. Ecm11 is required for meiotic DNA replication

Deletion of the *ECM11* gene cause diminished DNA replication in meiosis [71].

In the two-hybrid screen it was found out that Ecm11 strongly interacts with Cdc6 that has a pivotal role in the initiation of DNA replication [72]. Genetic interactions between Cdc6 and Ecm11 were also observed. Moderate supression of *cdc6-1* mutation by overexpression of *ECM11* was detected [72] and deletion of *ECM11* in *cdc6-1* genetic background enhances thermo-sensitivity of *cdc6-1* mutation (Zavec AB, unpublished result). These data suggest direct involving of Ecm11 in initiation of DNA replication process.

### 4.6.4. Ecm11 is modified by SUMO during meiosis

IDPs are tightly regulated in a cell and diverse post-translational modifications (such as ubiquitination, sumoylation, and phosphorylation) facilitate regulation of their function (Subchapter 3.4.). It was shown that the majority of Ecm11 protein in the cell is sumoylated during meiosis [73]. Lys5 at the highly disordered N-terminus of Ecm11 is modified by SUMO. It was shown that sumoylation is essential for biological role of Ecm11 in meiosis and that sumoylation directly regulates Ecm11 function in meiosis.

## 5. Conclusion

Cell nuclei contain high levels of IDPs. In this work, a hypothesis has been made, that in the nucleus, transcription is a process with the highest level of IDPs and that a disordered structure is likely present to a lesser extent in DNA replication process. However, the initiation of DNA replication would be expected to engage more IDPs than the elongation of DNA replication due to the need for responsiveness to the environment. By analysis of predicted disordered structure in replication proteins, it was confirmed that proteins with the role in initiation of DNA replication have more disordered structure than proteins with the role in elongation of DNA replication. The majority of IDRs in these proteins are at N- or C- terminus, most often IDRs are at the N-terminus of IDPs.

## Acknowledgements

## Author details

Apolonija Bedina Zavec[*]

Address all correspondence to: polona.bedina@ki.si

National Institute of Chemistry, Department for Molecular biology and Nanobiotechnology, Slovenia

## References

[1]  Uversky, N., Oldfield, C. J., & Dunker, A. K. (2005). Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit*, 18, 343-384.

[2]  Babu, M. M., Lee, R., Groot, S. N., & Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. *Curr Opin Struct Bio*, 21(3), 432-40.

[3]  Lakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z., & Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol.*, 323(3), 573-84.

[4]  Ward, J. J., Sodhi, J. S., Mc Guffin, L. J., Buxton, B. F., & Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*, 337(3), 635-45.

[5] Tompa, P., Fuxreiter, M., Oldfield, C. J., Simon, I., Dunker, A. K., & Uversky, V. N. (2009). Close encounters of the third kind: disordered domains and the interactions of proteins. *Bioessays*, 31(3), 328-35.

[6] Uversky, V. N., & Dunker, A. K. (2010). Understanding protein non-folding. *Biochim Biophys Acta*, 1804(6), 1231-64.

[7] Uversky, V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*, 11(4), 739-56.

[8] Tantos, P., Friedrich, P., & Tompa, P. ( 2009). Cold stability of intrinsically disordered proteins. *FEBS Lett*, 583(2), 465-9.

[9] Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L., & Pappu, R. V. (2010). Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci*, USA, 107(18), 8183-8.

[10] Oldfield, C. J., Cheng, Y., Cortese, M. S., Brown, C. J., Uversky, V. N., & Dunker, A. K. (2005). Comparing and combining predictors of mostly disordered proteins. *Biochemistry*, 44(6), 1989-2000.

[11] Bellay, J., Han, S., Michaut, M., Kim, T., Constanzo, M., Andrews, B. J., Boone, C., Bader, G. D., Myers, C. L., & Kim, P. M. (2011). Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol*, 12(2), R14.

[12] Tokuriki, N., & Tawfik, D. S. (2009). Protein Dynamism and Evolvability. *Science*, 324, 203-207.

[13] Tompa, P. (2011). Unstructural biology coming of age. *Current Opinionin Structural Biology*, 21, 1-7.

[14] Kovacs, E., Tompa, P., Liliom, K., & Kalmar, L. (2010). Dual coding in alternative reading frames correlates with intrinsic protein disorder. *Proc Natl Acad Sci*, USA, 107(12), 5429-5434.

[15] Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem Sci*, 27(10), 527-33.

[16] Dyson, H. J., & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol*, 12, 54-60.

[17] Demchenko, A. P. (2001). Recognition between flexible protein molecules: induced and assisted folding. *J Mol Recognit*, 14, 42-61.

[18] Tompa, P., Szasz, C., & Buday, L. (2005). Structural disorder throws new light on moonlighting. Trends Biochem Sci. , 30, 484-489.

[19] Dyson, H. J., & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol.*, 6(3), 197-208.

[20] Marsh, J. A., & Teichmann, S. A. (2011). Relative Solvent Accessible Surface Area Predicts Protein Conformational Changes upon Binding Structure. 19(6), 859-867.

[21] Uversky, V. N. (2011). Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chem Soc Rev.*, 40(3), 1623-34.

[22] Oldfield, C. J., Meng, J., Yang, J. Y., Yang, M. Q., Uversky, V. N., & Dunker, A. K. (2008). Flexible nets: disorder and induced fit in the associations of 53 and 14-3-3 with their partners. *BMC Genomics*, 9(1), S1.

[23] Cortese, M. S., Uversky, V. N., & Dunker, A. K. (2008). Intrinsic disorder in scaffold proteins: getting more from less. *Prog Biophys Mol Biol*, 98(1), 85-106.

[24] Tobi, D., & Bahar, I. (2005). Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci*, USA, 102(52), 18908-13.

[25] Shoemaker, B. A., Portman, J. J., & Wolynes, P. G. (2000). Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci*, USA, 97(16), 8868-73.

[26] Pontius, B. W. (1993). Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem Sci*, 18(5), 181-6.

[27] Mittag, T., Orlicky, S., Choy, W. Y., Tang, X., Lin, H., Sicheri, F., Kay, L. E., Tyers, M., & Forman-Kay, J. D. (2008). Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc Natl Acad Sci*, USA, 105(46), 17772-7.

[28] Mittag, T., Marsh, J., Grishaev, A., Orlicky, S., Lin, H., Sicheri, F., Tyers, M., & Forman-Kay, J. D. (2010). Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure*, 18(4), 494-506.

[29] Fuxreiter, M., Simon, I., Friedrich, P., & Tompa, P. (2004). Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol*, 338(5), 1015-26.

[30] Ubbink, M. (2009). The courtship of proteins: understanding the encounter complex. *FEBS Lett*, 583(7), 1060-6.

[31] Kovács, I.A. , Szalay, M. S., & Csermely, P. (2005). Water and molecular chaperones act as weak links of protein folding networks: energy landscape and punctuated equilibrium changes point towards a game theory of proteins. *FEBS Lett*, 579(11), 2254-60.

[32] Antal, M. A., Böde, C., & Csermely, P. (2009). Perturbation waves in proteins and protein networks: applications of percolation and game theories in signaling and drug design. *Curr Protein Pept Sci*, 10(2), 161-72.

[33] Sigalov , A. B. (2010). Protein intrinsic disorder and oligomericity in cell signaling. *Mol Biosyst*, 6(3), 451-61.

[34] Vavouri, T., Semple, J. I., Garcia-Verdugo, R., & Lehner, B. (2009). Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell*, 138(1), 198-208.

[35] Kusic-Tisma, J. (2011). *DNA Replication and Related Cellular Processes*, Rijeka, InTech.

[36] Vuzman, D., & Levy, Y. (2012). Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol Biosyst*, 8(1), 47-57.

[37] Wright, P. E., & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J of Mol Bio.*, 293(2), 321-331.

[38] Romero, P., Obradovic, Z., & Dunker, A. K. (2004). Natively disordered proteins: functions and predictions. *Appl Bioinformatics*, 3(2-3), 105-13.

[39] Saccharomyces genome database. (2012). http://www.yeastgenome.org/, accessed 20 June.

[40] The DISOPRED2 Prediction of Protein Disorder Server. (2012). http://bioinf.cs.ucl.ac.uk/disopred/, accessed 20 June.

[41] Qi, H., & Zakian, V. A. (2000). The Saccharomyces telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomer-ase-associated est1 protein. *Genes Dev*, 14(14), 1777-88.

[42] Hsu, C. L., Chen, Y. S., Tsai, S. Y., Tu, P. J., Wang, M. J., & Lin, J. J. (2004). Interaction of Saccharomyces Cdc13p with Pol1Imp4pSir4p and Zds2p is involved in telomere replication, telomere maintenance and cell growth control. *Nucleic Acids Res*, 32(2), 511-21.

[43] Sun, J., Yang, Y., Wan, K., Mao, N., Yu, T. Y., Lin, Y. C., De Zwaan, D. C., Freeman, B. C., Lin, J. J., Lue, N. F., & Lei, M. (2011). Structural bases of dimerization of yeast telo-mere protein Cdc13 and its interaction with the catalytic subunit of DNA polymerase α. *Cell Res*, 21(2), 258-74.

[44] Cliften, P. F., Hillier, L. W., Fulton, L., Graves, T., Miner, T., Gish, W. R., Waterston, R. H., & Johnston, M. (2001). Surveying Saccharomyces genomes to identify function-al elements by comparative DNA sequence analysis. *Genome Res*, 11(7), 1175-86.

[45] Johansson, E., Majka, J., & Burgers, P. M. (2001). Structure of DNA polymerase delta from Saccharomyces cerevisiae. *J Biol Chem*, 276(47), 43824-8.

[46] Brocas, C., Charbonnier, J. B., Dhérin, C., Gangloff, S., & Maloisel, L. (2010). Stable interactions between DNA polymerase δ catalytic and structural subunits are essen-tial for efficient DNA repair. *DNA Repair*, 9(10), 1098-111.

[47] Haracska, L., Prakash, S., & Prakash, L. (2000). Replication past O(6)-methylguanine by yeast and human DNA polymerase eta. *Mol Cell Biol*, 20(21), 8001-7.

[48] Johansson, E., Garg, P., & Burgers, P. M. (2004). . The Pol32 subunit of DNA polymerase delta contains separable domains for processive replication and proliferating cell nuclear antigen (PCNA) binding. J Biol Chem , 279(3), 1907-15.

[49] Brown, W. C., & Campbell, J. L. (1993). Interaction of proliferating cell nuclear antigen with yeast DNA polymerase delta. *J Biol Chem*, 268(29), 21706-10.

[50] Acharya, N., Klassen, R., Johnson, R. E., 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(44), 17927-32.

[51] Dua, R., Levy, D. L., & Campbell, J. L. (1999). Analysis of the essential functions of the C-terminal protein/protein interaction domain of Saccharomyces cerevisiae pol epsilon and its unexpected ability to support growth in the absence of the DNA polymerase domain. *J Biol Chem*, 274(32), 22283-8.

[52] Tsubota, T., Tajima, R., Ode, K., Kubota, H., Fukuhara, N., Kawabata, T., Maki, S., & Maki, H. (2006). Double-stranded DNA binding, an unusual property of DNA polymerase epsilon, promotes epigenetic silencing in Saccharomyces cerevisiae. *J Biol Chem*, 281(43), 32898-908.

[53] Matos, J., Lipp, J. J., Bogdanova, A., Guillot, S., Okaz, E., Junqueira, M., Shevchenko, A., & Zachariae, W. (2008). Dbf4-dependent CDC7 kinase links DNA replication to the segregation of homologous chromosomes in meiosis I. *Cell*, 135(4), 662-78.

[54] Davey, M. J., Andrighetti, H. J., Ma, C. J., & Brandl, X. (2011). A synthetic human kinase can control cell cycle progression in budding yeast. *G3*, 1(4), 317-25.

[55] Malmström, L., Riffle, M., Strauss, C. E., Chivian, D., Davis, T. N., Bonneau, R., & Baker, D. (2007). Superfamily assignments for the yeast proteome through integration of structure prediction with the gene ontology. *PLoS Biol*, e76.

[56] Masai, H., & Arai, K. (2000). Dbf4 motifs: conserved motifs in activation subunits for Cdc7 kinases essential for S-phase. *Biochem Biophys Res Commun.*, 275(1), 228-32.

[57] Francis, L. I., Randell, J. C., Takara, T. J., Uchima, L., & Bell, S. P. (2009). Incorporation into the prereplicative complex activates the Mcm2-7 helicase for Cdc7-Dbf4 phosphorylation. *Genes Dev.*, 23(5), 643-54.

[58] Dowell, S. J., Romanowski, P., & Diffley, J. F. (1994). Interaction of Dbf4, the Cdc7 protein kinase regulatory subunit, with yeast replication origins in vivo. *Science*, 265(5176), 1243-6.

[59] Duncker, B. P., Shimada, K., Tsai-Pflugfelder, M., Pasero, P., & Gasser, S. M. (2002). An N-terminal domain of Dbf4p mediates interaction with both origin recognition complex (ORC) and Rad53p and can deregulate late origin firing. *Proc Natl Acad Sci*, USA, 99(25), 16087-92.

[60] Varrin, A. E., Prasad, A. A., Scholz, R. P., Ramer, M. D., & Duncker, B. P. (2005). A mutation in Dbf4 motif M impairs interactions with DNA replication factors and confers increased resistance to genotoxic agents. *Mol Cell Biol.*, 25(17), 7494-504.

[61] Sheu, Y.J. , & Stillman, B. (1999). Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. *Mol Cell 2006*, 24(1), 101-13.

[62] Zegerman, P. , & Diffley, J. F. (2010). Checkpoint-dependent inhibition of DNA replication initiation by Sld3 and Dbf4 phosphorylation. *Nature*, 467(7314), 474-8.

[63] Weinreich, M. , & Stillman, B. Cdc7p-Dbf4p kinase binds to chromatin during S phase and is regulated by both the APC and the RAD53 checkpoint pathway. *EMBO J.*, 18(19), 5334-46.

[64] Duch, A. , Palou, G., Jonsson, Z. O., Palou, R., Calvo, E. , Wohlschlegel, J. , & Quintana, D. G. (2011). A Dbf4 mutant contributes to bypassing the Rad53-mediated block of origins of replication in response to genotoxic stress. *J Biol Chem*, 286(4), 2486-91.

[65] Matthews, L. A., Jones, D. R., Prasad, A. A., Duncker, B. P., & Guarné, A. (2012). Saccharomyces cerevisiae Dbf4 has unique fold necessary for interaction with Rad53 kinase. *J Biol Chem*, 287(4), 2378-87.

[66] Wan, L., Niu, H., Futcher, B., Zhang, C., Shokat, K. M., Boulton, S. J., & Hollingsworth, N. M. (2008). Cdc28-Clb5 (CDK-S) and Cdc7-Dbf4 (DDK) collaborate to initiate meiotic recombination in yeast. *Genes Dev.*, 22(3), 386-97.

[67] Sasanuma, H., Hirota, K., Fukuda, T., Kakusho, N., Kugou, K., Kawasaki, Y., Shibata, T., Masai, H., & Ohta, K. (2008). Cdc7-dependent phosphorylation of Mer2 facilitates initiation of yeast meiotic recombination. *Genes Dev.*, 22(3), 398-410.

[68] Lo, H. C., Wan, L., Rosebrock, A., Futcher, B., & Hollingsworth, N. M. (2008). Cdc7-Dbf4 regulates NDT80 transcription as well as reductional segregation during budding yeast meiosis. *Mol Biol Cell*, 19(11), 4956-67.

[69] Matos, J., Lipp, J. J., Bogdanova, A., Guillot, S., Okaz, E., Junqueira, M., Shevchenko, A., & Zachariae, W. (2008). Dbf4-dependent CDC7 kinase links DNA replication to the segregation of homologous chromosomes in meiosis I. *Cell*, 135(4), 662-78.

[70] Marston, A. L. (2009). Meiosis: DDK is not just for replication. *Curr Biol*, 19(2), 74-6.

[71] Zavec, A. B., Lesnik, U., Komel, R., & Comino, A. (2004). The Saccharomyces cerevisiae gene ECM11 is a positive effector of meiosis. *FEMS Microbiol Lett*, 241(2), 193-9.

[72] Zavec, P. B., Comino, A., Watt, P., & Komel, R. (2000). Interaction trap experiment with CDC6. *Pflugers Arch*, 439(3), R 94-6.

[73] Zavec, A. B., Comino, A., Lenassi, M., & Komel, R. (2008). Ecm11 protein of yeast Saccharomyces cerevisiae is regulated by sumoylation during meiosis. *FEMS Yeast Res*, 8(1), 64-70.

# Mechanisms that Protect Chromosome Integrity During DNA Replication

# Preserving the Replication Fork in Response to Nucleotide Starvation: Evading the Replication Fork Collapse Point

Sarah A. Sabatinos and Susan L. Forsburg

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/51393

## 1. Introduction

Replication fork progression is blocked by a variety impediments including DNA damage, aberrant DNA structures, or nucleotide depletion [1-3]. The response to replication fork stalling varies according the type of replication inhibition, the number of stalled forks and the duration of the treatment [3-7]. Stalled replication forks are at increased risk for DNA damage, which can lead to mutation or cell death [7-13]. The cell relies on the Intra-S phase checkpoint and DNA damage response proteins to preserve fork structure to allow recovery and resumption of the cell cycle [5, 10, 14-19]. Thus, the mechanisms that maintain replication fork structure are crucial for genome maintenance, and form a primary barrier to malignant transformation [20, 21].

The drug hydroxyurea (HU) induces a reversible early S-phase arrest by causing deoxynucleotide triphosphate (dNTP) depletion [22-24]. HU is a venerable chemotherapeutic, used for its ability to inhibit cell proliferation, but also because it predisposes proliferating cells to genome instability. The loss of replication fork stability and its associated DNA damage following HU treatment is loosely termed "replication fork collapse". Changes in dNTP pool levels through other mechanisms (*e.g.* exogenous thymidine or 5-bromo-2'deoxyuridine treatment) are known to cause point mutations [25-27], plasmid instability [28] and polyploidization [29]. Further, dNTP pool changes in human cells may cause hypersensitivity to secondary treatment with alkylating agents [30, 31].

Wild type cells recover from HU arrest and complete S-phase once drug is removed from the culture medium. Alternatively, some cultures may recover from HU arrest prior to its remov-

al by up-regulating nucleotide synthesis and overcoming HU replication inhibition to slow-ly complete S-phase [17, 32, 33]. The ability to recover stalled replication out of an HU arrest requires restoration of replication forks, restart of DNA synthesis and completion of S-phase.

Whether a replication fork successfully recovers, or collapses with DNA damage, de-pends in part on the Intra-S phase checkpoint pathway. Cells lacking the checkpoint suf-fer fork collapse and death. Notably, cells that do not trigger the Intra-S phase checkpoint continue to synthesize DNA despite the presence of HU. Continued synthesis in the pres-ence of low dNTP pools leads to reduced replication rates and increased single stranded DNA (ssDNA) [33-37]. This is a fragile state of "open" DNA that is prone to double strand breaks (DSBs) [38-40]. Further, altered dNTP levels during DNA replication enhance point mutations, in which the base inserted shifts towards that of the dominant pool or away from the lowest pool [25, 41-44]. This explains why the replication checkpoint is a crucial barrier to genome instability.

Thus, replication fork collapse in checkpoint mutants does not occur immediately after HU treatment, detection of decreased dNTP levels, or failure to mount a checkpoint response. Instead, replication fork collapse across a population of forks, within a culture of cells, is a consequence of continued fork activity. The signs and symptoms of replication fork collapse represent a new execution point, the Replication Fork Collapse Point. This metric describes the time at which the majority of replication forks in a cell population become non-function-al. In this review, we describe the causes and symptoms of the Replication Fork Collapse Point, with particular regard to the Intra S-phase checkpoint.

## 2. Replication Fork Structure is Maintained During Stalling

The replication fork describes a region of denatured DNA where DNA synthesis is actively occurring, resembling a two-tined fork. The replisome encompasses the forked DNA, and the entire complex is large and dynamic, coupling DNA unwinding and polymerization [45-47]. Unwinding is performed by a conserved hexameric helicase (MCM) and its associat-ed proteins Cdc45 and GINS. The processive helicase produces single strand DNA (ssDNA) which becomes transiently coated with replication protein A (RPA), a ssDNA binding pro-tein homologue. ssDNA is the substrate for leading- (polε) and lagging-strand (polδ and polα-primase) polymerases.

These functions must be linked to facilitate DNA synthesis. Coupling generation of ssDNA with its use in replication is particularly important, because ssDNA is vulnerable to form-ing secondary structures, which leads to DNA damage [40, 48, 49], and recombination [50-52]. Thus, fork proteins limit the amount of DNA unwinding and ssDNA [39, 53]. In normal conditions, synthesis may occur rapidly and the goal of minimizing ssDNA production (<200bp) is easily accomplished [54]. However, if either the leading or lagging strand poly-merases become stalled or arrested in a slow zone, the helicase must also be slowed down

to prevent it from generating excessive ssDNA and potentially dissociating entirely from the replisome.

Helicase and polymerases are linked by the replication Fork Protection Complex (FPC), which contributes to replication fidelity and later chromosome segregation. Tim1 (*S. pombe* Swi1) and Tipin (*Sp*Swi3) are evolutionarily conserved core components of the FPC that are essential for fork stability [11, 55-58](Table 1). This core is joined by AND1 (*Sp*Mcl1) and CLASPIN (*Sp*Mrc1), two proteins that bridge the helicase and polymerases. AND1 links the lagging strand primase (pol$\alpha$) [59-62], while CLASPIN connects the leading strand polymerase (pol$\varepsilon$) [58, 63].

Because of its role maintaining replisome structure, the FPC promotes replication fork efficiency and speed, particularly during fork stalling or pausing. While not essential for DNA replication [58, 64-66], the FPC contributes to processivity [67-70], and has additional roles in response to replication stalling [55, 71, 72], and facilitating sister chromatid cohesion, which is essential for faithful chromosome segregation [73-75].

| Human FPC component | Homologues | References |
|---|---|---|
| TIMELESS (TIM) | Tim1 (*M. musculus, X. laevis*) | [56, 57, 63, 65, 72] |
| | Tof1 (*S. cerevisiae*) | |
| | Swi1 (*S. pombe*) | |
| TIPIN (TIP) | Tipin (*M. musculus, X. laevis*) | [11, 56-58, 63, 75, |
| | Csm3 (*S. cerevisiae*) | 75-77] |
| | Swi3 (*S. pombe*) | |
| CLASPIN | Claspin (*M. musculus, X. laevis*) | [14, 63, 78-82] |
| | Mrc1 (*S. cerevisiae, S. pombe*) | |
| AND1 | And1 (*M. musculus, X. laevis*) | [59, 61, 73, 83] |
| | Ctf4 (*S. cerevisiae*) | |
| | Mcl1 (*S. pombe*) | |

**Table 1.** Fork Protection Complex Proteins in Various Species.

## 3. Causes of Replication Fork Stalling

DNA replication occurs in a short period during the cell cycle. In yeasts, replication of the ~12 Mb genome occurs within 20 to 30 minutes out of a 2.5 to 3h cell cycle. Human cells require several hours, a fraction of a full cell cycle, to replicate a substantially larger genome. The rate-limiting factor is replication fork velocity at 1–2 kb/min. This is an astonishing rate, considering secondary and tertiary structure of the genome packaged into higher order chromatin domains. The tight links between helicase, polymerase and FPC promote highly processive replication. Importantly, they also contribute to replication fidelity. Disruption of

any one component (if not already lethal) leads to significant disruptions in processivity and/or fidelity. This is particularly true when impediments to replication are encountered.

Replication pausing and stalling is caused by both natural barriers and external factors [3, 84, 85]. Some regions of DNA cause replication fork stalling through sequence elements (*e.g.* DNA secondary structure in repetitive elements), or protein interference (*i.e.* transcription). Replication "slow zones" have been described in many model organisms, and these may contribute to genome instability and chromosome fragility. One characteristic shared by many "difficult templates" is the presence of repetitive sequence elements that cause fork stalling [86-89].

A replication termination sequence (RTS1) at the mating locus of fission yeast also promotes unidirectional fork progression by binding the replication termination factor 1 (Rtf1) [88, 90-95]. Unidirectional DNA replication is required to establish an imprint that directs mating type switching. RTS1 replication fork pausing is polar, meaning that forks approaching the barrier from one direction will be affected; forks from the opposite direction continue replication [93, 96].

Similarly, ribosomal DNA (rDNA) arrays are an example of a natural, repetitive element that is at risk for fork pausing. Each of the rDNA repeats contains a polar replication terminator, which ensures that forks proceed unidirectionally through each element [86, 97-100]. This occurs as a response to the binding of a fork arrest protein. For example, in fission yeast the Reb1 protein binds the replication termination element Ter3, which promotes long-range DNA interactions with other chromosomal Ter sequences [101, 102]. Localized to the nucleolus, this may nucleate a zone for replication termination [103]. Based on similarity to prokaryotic replication terminators, Reb1-Ter binding may stop the MCM helicase from creating more ssDNA leading to fork pausing and stalling. Pausing of the fork at this site also depends on FPC proteins Swi1 and Swi3.

Replication termination at rDNA is also seen in budding yeast and mammals. In *S.cerevisiae*, the FOB1 protein binds to ribosomal fork barrier elements and arrests progression of the replication fork so that replication is in concert with rDNA transcription [104, 105]. In mice, transcription termination factor 1 (TTF1) binds to termination sites in the rDNA and causes fork arrest [106-108]. It is suggested that Reb1/FOB1/TTF1 binding to their specific rDNA elements blocks the replicative MCM helicase and arrests forks.

The rDNA elements define one type of genomic sequence that causes replication slowing or pausing sites. Other regions of the genome may also cause fragile sites, which are broadly characterized as replication slow zones that are prone to forming DNA breaks [38, 40, 109, 110]. These may be dependent upon the chromatin context, transcriptional activity, or impairment of the fork by external agents, such as HU [111].

HU inhibits the activity of ribonucleotide reductase, which causes a reduction of dNTP pools [112]. HU is frequently used as a reversible early-S phase block reagent in cultured cells. In this sense, HU response is similar to excess thymidine treatment, which changes dNTP pools and induces an early S-phase arrest in metazoan cells [35]. The size of dNTP pools is intimately linked to cell cycle and checkpoint responses [24, 32, 113-115]. Critically,

checkpoint proficiency allows cells to survive HU arrest, hold forks stable, and efficiently re-start during release.

## 4. Intra-S Checkpoint: keeping things connected

The Intra-S phase checkpoint is a kinase cascade that responds to HU treatment. It serves to stabilize replication forks and arrest replication until dNTP pools recover. The checkpoint also prevents DNA damage from forming, particularly DNA double strand breaks, by re-stricting endonucleases such as Mus81 that can act on stalled fork structures [9, 10]. In addi-tion, the Intra-S checkpoint regulates recombination enzymes (*e.g. Sp*Rqh1/*Sc*Sgs1/*Hs*BLM, Rad60), to preserve stalled forks in a state competent for restart without loss of genetic infor-mation [18, 116].

The remainder of this review will focus on the effects of the checkpoint on the replisome it-self. During checkpoint activation, the helicase is restrained and stabilized, to prevent exces-sive unwinding and allow the fork to restart when HU is removed or bypassed. DNA synthesis is also restrained, preventing mutations that may occur during replication in the presence of altered dNTP pools. Late replication origins are prohibited from firing, conserv-ing these "second-chance" origins for later replication restart. These activities help to stabi-lize established forks after HU treatment, later allowing them to restart. Alternatively, new forks may be established from the late origins in restart to rescue collapsed forks and com-plete DNA synthesis.

Wild type cells are actively inhibited from DNA synthesis during HU block [10, 17, 36, 58, 65, 117, 118]. That is, the forks do not cease synthesis because they run out of nucleotides. Rather, the checkpoint ensures that the forks are slowed or stopped before such starvation occurs, saving them from the mutagenic effects of dNTP imbalance [34, 42, 119]. These ob-servations are consistent with depletion, rather than exhaustion of specific dNTP pools [22], and extremely slow residual synthesis [33]. This fork arrest is accompanied by inhibition of the helicase [15, 53, 54, 65], which reduces ssDNA accumulation and concomitant RPA bind-ing until HU is removed.

The Intra-S phase checkpoint is a key component of the response to HU and actively re-strains forks during replication stress. The initial signal to activate the checkpoint is provid-ed by increased ssDNA created as replication forks stall [39, 53, 54]. ssDNA is bound by RPA and recruits the Rad9-Rad1-Hus1 (9-1-1) complex and ATR family kinases to stalled forks [120, 121]. Thus, the symptom of slow or stalled forks (generation of ssDNA) initiates the checkpoint [4, 120-125]. However, the FPC and checkpoint together ensure that the heli-case cannot generate too much ssDNA, which provides one defense against replication fork collapse during HU stalling.
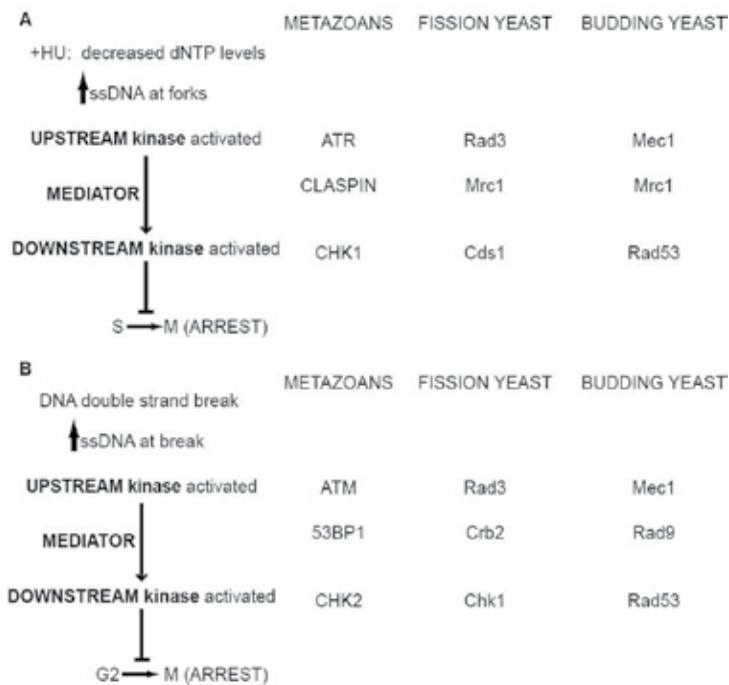
**Figure 1.** The Intra-S phase checkpoint across species. Key components of the checkpoint and their names are described for metazoan (Human, mouse, *Xenopus*), budding yeast (*S. cerevisiae*) and fission yeast (*S. pombe*). A) The replication checkpoint signals the presence of replication fork stalling and ssDNA accumulation through the upstream kinase (ATR/Rad3/Mec1), which activates the downstream effector kinase (CHK1/Cds1/Rad53). CLASPIN/Mrc1 is the mediator of the replication checkpoint and is responsible for efficient dimerization and activation of the downstream kinase. B) Similarities to the DNA damage response checkpoint. When DSBs are generated in G2, ssDNA is created around the break, which recruits the upstream kinase (ATM/Rad3/Mec1). Through the 53BP1/Crb2/Rad9 mediator, the upstream kinase phosphorylates and activates the downstream kinase (CHK2/Chk1/Rad53), which arrests the cell cycle and allows time for repair.

Checkpoint activation is also coupled to the FPC proteins, particularly CLASPIN and its yeast equivalent, Mrc1 [118, 126, 127]. In fission yeast, Mrc1 is phosphorylated by the upstream Rad3/ATR kinase to a checkpoint-active form [128]. This activation recruits the downstream Cds1 kinase to the stalled replication fork and is essential to signal amplification and transmission by activated Cds1. This pathway is conserved: in humans and budding yeast, respectively, Chk1/Rad53 is recruited to stalled forks by CLASPIN/Mrc1 and ATR/Mec1 kinase [6, 14, 16, 124, 129-131].

This S phase checkpoint has a parallel structure to the DNA double strand break (DSB) response: Mrc1 is a replication-specific version of the *Hs*53BP1/*Sc*RAD9/*Sp*Crb2 mediator, which brings together master kinases (ATM/ATR, Mec1 and Rad3) with an effector kinase (CHK2, Rad53 or Chk1) for DSB response [14, 129] (Figure 1).

Fission yeast *cds1Δ* and *mrc1Δ* cells rapidly die in HU [117, 118, 127]. These cells lack the Intra-S phase checkpoint and cannot restrain late origin firing or nuclease activity at stalled forks [34, 36, 51, 58, 132]. In contrast to wild type cells, however, these mutants continue

DNA synthesis during HU block [53, 133]. The replication forks develop extensive ssDNA which can be observed by RPA binding. However, the fork proteins do not remain together, suggesting that the link between helicase and polymerase components is lost. Upon release from HU, *cds1Δ* or *mrc1Δ* cells manage a limited amount of further DNA synthesis. Their DNA synthesis proceeds slowly, but cells never achieve a fully replicated amount of DNA. Thus, the forks are failing as they reach the Replication Fork Collapse point, which results in S-phase failure and ultimately, cell death.

## 5. The Rules of Replisome Restraint and Restart,1: Fork Movement

Considering the phenotype of checkpoint mutants, we infer that an active mechanism restrains the helicase during HU treatment. Genome-wide studies in budding yeast show accumulation of single stranded DNA occurs in checkpoint mutants, adjacent to replication forks upon treatment with HU [39, 58, 133]. Similarly, in fission yeast checkpoint mutants, large masses of RPA can be visualized in whole cells treated with HU, which depend upon the MCM helicase [53].

A simple interpretation is that the helicase becomes uncoupled from the stalling polymerase and unwinds DNA ahead of it. This excessive unwinding generates ssDNA that is prone to breakage, which generates a characteristic DSB marker, phosphorylated histone H2A(X) [15, 39]. In many cases, the RPA signal is associated with markers of DNA synthesis, such as incorporation of the nucleotide analogue BrdU [53], or proximity to replication fork proteins [133]. Importantly, this uncoupling and unwinding occurs at the same time as DNA synthesis during both HU block and release. This suggests a more subtle effect in which leading and lagging strand synthesis is uncoupled, which leads to simultaneous accumulation of ssDNA and markers of synthesis, either because they are in the same region or because the ssDNA is a functional template.

## 6. Rules of Restraint and Restart, 2: Synthesis

The second key to restraint and successful restart is modulating the DNA polymerases. Wild type cells incorporate minimal amounts of nucleoside analogue in the presence of HU. Forks slow but remain stable [7, 17, 34, 54]. The rate of nucleotide analogue incorporation decreases, and DNA content does not increase significantly [36, 117, 134, 135]. In the yeast system, studies suggest that early replication forks extend about 5kb from the origin in the presence of HU before stopping [134, 136]. Decreased dNTP pools slow replication elongation during HU arrest. However, ectopic expansion of dNTPs by expressing ribonucleotide reductase from a plasmid can increase fork velocity even in HU [34]. Upon release from HU, replication rapidly restarts, whether from new origins or reactivation of existing forks, which results in rapid completion of DNA synthesis before cell division.

Budding yeast dNTP metabolism is quite robust and resistant to challenge, sensitive only to high levels of HU or significant NTP imbalance. In contrast, fission yeast [137] and metazo-

an cells are sensitive to low levels of HU, or modest dNTP imbalance, both which are sufficient to provoke replication arrest [7]. In all systems, there is an intimate connection to the Intra-S phase checkpoint.

Surprisingly, checkpoint mutants do not block DNA synthesis in HU, indicating that they are not actually starved for nucleotides, but rather lacking the ability to monitor pool levels [53]. Fission yeast *cds1Δ* mutants continue to synthesize DNA and incorporate analogue. In analysis of chromatin fibers, these can be visualized as extended tracts of newly synthesized DNA despite the presence of HU. Upon release from the drug, *cds1Δ* mutants continue to incorporate some analogue before reaching a plateau, by which time they have accumulated approximately 66% the total amount of DNA incorporated in wild type (continuous labeling, block and release) [53]. These differences can be measured by detection of analogue incorporation, but are obscured by total DNA content analysis, which is prone to artifact [22, 36, 134].

The difference between the two situations is that much of this synthesis occurs *during* HU treatment in the mutant, and only *after release* in wild type [53]. Thus, it is not until the recovery period that the majority of *cds1Δ* forks break down and can no longer synthesize DNA. We define the point at which synthesis ceases as the Replication Fork Collapse Point (see discussion of the Collapse Point in section "The Collapse Point: A Metric for Fork Stability"). Importantly, this is an extended window of time where there is a stochastic probability of forks arresting and suffering collapse. Our data show that regions of DNA synthesis upon HU release have high levels of RPA [53], which indicates that fork collapse is accompanied by accumulation of ssDNA. This may reflect a burden of damage, incurred in HU, which is remembered during release at the Collapse Point.

Polymerase ε is coupled to the helicase by Mrc1 and the FPC proteins Tof1 (Swi1) and Csm1 (Swi3) [63, 65, 132, 138]. This is thought to stabilize leading strand components at stalled forks in HU. Asynchronously growing *mrc1Δ* cells lack this connection, which leads to intrinsic damage and a higher level of basal Cds1 phosphorylation even without added replication stress [128]. This essentially uncouples leading and lagging strands. *mrc1Δ* cells treated with HU incorporate more nucleotide analogue but in shorter DNA fiber tracts. This is consistent with a role for Mrc1 in modulating origin firing, as well as rate [139-141]. The increase in DNA synthesis is only slightly higher after release, which could be attributed to slower forks or polε uncoupling as in *S. cerevisiae* [81, 141]. These data suggest that in the absence of Mrc1, forks continue to synthesize a low, steady level of DNA and this is independent of Cds1.

Mrc1 brings Cds1 and Rad3 together to phosphorylate Cds1 on threonine 11 [128, 142]. Subsequently, Cds1 activation is amplified by dimerization and autophosphorylation, setting in motion the full Intra-S phase checkpoint [128]. HU treatment induces little Cds1-T11 phosphorylation in *mrc1Δ*; instead, the damage response kinase Chk1 is activated. This suggests that there is conversion of stalled synthesis into DNA damage. Consistent with this, phospho-H2A accumulates in *mrc1Δ* nuclei and replicated fiber tracts after HU release [53]. The damage signal is frequently coincident with areas of synthesis, but often distinct from RPA-

heavy areas. Thus, unwinding and synthesis are likely uncoupled and distinct in *mrc1Δ* with HU treatment.

## 7. Rules of Restraint and Restart, 3: the late origins

An additional function of the Intra-S phase checkpoint is to restrain late origins from firing. Upon release from HU, these origins become competent for replication, and establish "rescue forks" that ensure completion of DNA replication [33, 36, 143-145]. Could these origins explain the post-release DNA synthesis observed in the checkpoint mutants? While late origin firing must contribute to some of the synthesis after release, we suggest that much of the post-release DNA synthesis does not occur from late origin firing, for the following reasons.

First, origin firing is de-regulated in HU blocked checkpoint mutants, which suggests that many late origins have already fired at the time of release, and are not available for this further synthesis. Recent work on dNTP pools in budding yeast suggests that >200 additional origins are fired in a *rad53-11* mutant compared to wild type in HU [34]. Deleting the ribonucleotide reductase inhibitor Sml1 results in activation of late origins in HU, while a *rad53-11sml1Δ* double mutant shows increased replication tracts in HU [34]. Sml1 is regulated by the replication checkpoint [114], and *sml1Δ* cells have increased dNTP pools [113, 146]. *sml1Δ* mutant backgrounds are frequently used in replication checkpoint studies because they overcome the lethal effects of *rad53Δ* or *mec1Δ*, but this makes direct comparison of these double mutants with other organisms, which retain controls of NTP pools, difficult. However, HU has also been shown to arrest replication without completely exhausting dNTP pools [22, 34], which suggests that cells sense small dNTP changes. Perhaps the Intra-S phase checkpoint also contributes to fork slowing and stalling, and is not limited to signal transduction at stalled forks.

Second, it is likely that late origins that fire in checkpoint mutants after HU release are incapable of synthesizing more than a short tract length, due to lack of nucleotides [137]. More analogue is incorporated in *cds1Δ* compared to wild type for the first 30 minutes after HU release, suggesting that start-up replication is both different and faster in the mutant cells. Additionally, more origins fire in mutants during HU block [140, 143, 147], suggesting that fewer origins remain to be activated after release. These observations imply that for forks established during HU arrest in *cds1Δ* and *mrc1Δ*, synthesis cannot proceed past a point of increased fork collapse and template damage [9, 15, 116, 148-150]. The role of origin repression, using mutant cells that impede origin firing, is required to confirm the degree to which DNA synthesis occurring in checkpoint mutants after HU release is dependent on late origins.

Together, these observations from multiple systems suggest that wild type cells survive HU block and release through coordination of several mechanisms: control of late origin firing, maintenance of existing replication forks, and later restart of the stabilized forks. Wild type

cells do not encounter the Replication Fork Collapse Point because forks are maintained, replication is successfully restarted, and DNA synthesis completed.

## 8. Converting Stalled Forks to Restart

After HU is removed from culture medium, stabilized replication forks are returned to competence for DNA synthesis. In theory, immediate restart from a stabilized fork may be possible if all components are in place, having been protected from disassembly during HU arrest. In many cases this is likely to involve recombination pathways and the Rad51 recombinase. Rad51 binds to replisome components in HU, and around damaged replication forks [7, 15, 151]. Rad51 binds to ssDNA and promotes homologous recombination by allowing broken DNA to invade a homologous region for repair [52, 152, 153]. Checkpoint mutants have additional ssDNA, and experience "branch migration" of the fork structure [7, 52, 94, 154]. The resulting "chicken foot" structure is at risk for becoming a break or collapsed fork. Alternatively, the cruciform structure can be resolved by exonuclease Exo1, but leads to a partially replicated structure that cannot be replicated without *de novo* polymerase recruitment or break-induced replication [154-158].

The amount of time in HU until release has different effects in yeast and metazoan cultures. Both budding and fission yeast begin to arrest in HU within the first hour of HU exposure (e.g. [53, 134, 144, 159]). After a few hours at normal growth temperature, adaptation occurs, probably through changes in ribonucleotide reductase activity. In budding yeast, long-term HU exposure causes normal replication profiles to proceed at a glacial pace [33]. Similarly, human cells show increased sensitivity to HU over time, where fewer forks are observed with extended HU dose [7]. Peterman *et.al.* (2010) demonstrated that restarting long-HU treatment forks depends on both Rad51 and the repair protein XRCC3. Thus, protracted HU exposure may change that repair pathway used in fork repair, slowing the entire replication program [7, 153]. Slightly later than ssDNA accumulation, we detect Rad52 foci in live cells, increasing during HU block and release but lagging behind RPA accumulation [53]. Intriguingly, MRN/MRX- components co-localize at the replication fork and are important for fork stability (e.g. [35, 149, 160-162]), pointing at the essential role of recombination repair in restarting stalled replication forks.

## 9. The Collapse Point: A Metric for Fork Stability

The concept of replication fork "collapse" encompasses the observations that DNA damage and broken forks lead to loss of replication. ssDNA accumulates at susceptible forks and is a marker of increased risk of collapse [39, 54, 133, 137]. The DNA damage created at a stalled fork at or before collapse may not simply be DSBs. In fact, single strand breaks may form an

important part in the damage process, converted to DSBs either during fork regression or in a second S-phase [163-165].

We propose the Collapse Point as the time when the balance between replication fork processivity and instability tips toward disaster. The time when the majority of forks in a cell have irreversibly, irrecoverably failed and replication will not be completed. Ongoing synthesis in checkpoint mutants during HU treatment sets the forks on a course to destruction, but actual collapse does not occur until the attempt to recover. We suggest some replication forks retain activity and undergo a shortened replication restart after HU release. This is consistent with data in fission and budding yeasts that fork components are retained and move during HU arrest in checkpoint deficient cells [53, 133].
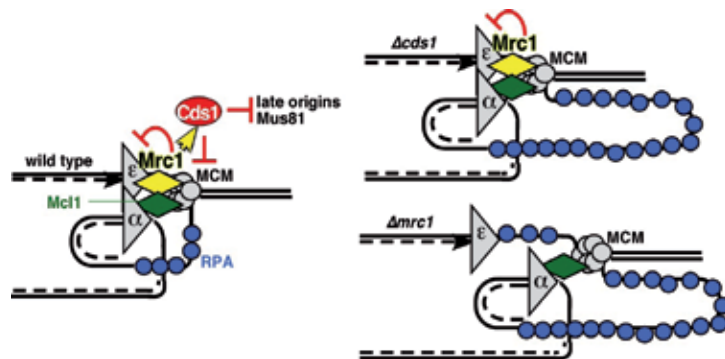


**Figure 2.** Model of the Replication Fork in HU before the Collapse Point.

Model of how replication fork architecture is affected by HU treatment in wild type, *cds1*Δ or *mrc1*Δ fission yeast cells. Left, wild type forks are stabilized through the Fork Protection Complex of Mrc1 (yellow diamond) and Mcl1 (green diamond), and the Intra-S phase checkpoint. As part of the FPC, Mrc1 forms a link between polymerase epsilon (ε), while Mcl1 links the helicase to polymerase alpha (α). A small loop of ssDNA forms at the fork in response to HU. This signals through the mediator function of Mrc1, activating Cds1 kinase. Cds1 kinase stabilizes and stalls the helicase activity of the replisome. Cds1 also inhibits Mus81 endonucleolytic activity and late origin firing, maintaining wild type cells in a state where replication can be restarted and/or finished from late origins after HU is removed. Right, top; when Cds1 is absent, a larger ssDNA loop forms from failing to slow the helicase. ssDNA is coated with replication protein A (RPA, blue circles) which form large foci in nuclei during HU block and release. Mrc1 is present, likely stabilizing polε, but the ssDNA may serve as template for lagging strand synthesis by pol α. Note that a ssDNA loop is presumed on the *cds1*Δ leading strand, but is omitted for clarity. The Intra-S phase checkpoint is not activated, late origins fire, and nucleases are not regulated.

Right, bottom; in the absence of Mrc1 the helicase is deregulated and potentially detached from polymerases. Large amounts of ssDNA form at individual forks and are coated by RPA. Although there is a great deal of ssDNA template, the length of *mrc1*Δ tracts in HU suggests that replication is slow, and that ssDNA areas are used as template during block and release up to a point where damage is encountered and forks collapse. In both *cds1*Δ and *mrc1*Δ cases, loss of replicative activity correlates with increased ssDNA foci, which build during HU block and release, suggesting that both mutants reach the Replication Fork Collapse Point during release.

The Replication Fork Collapse Point has no meaning for an individual fork; instead, it is the emergent property of the sum behavior of forks in a cell. The Collapse Point will generally be *estimated* by ensemble averaging across all cells in a culture.

While forks in *cds1Δ* and *mrc1Δ* retain synthesis activity, they are not necessarily the same as wild type (Figure 2). The amount of ssDNA and DNA damage signal (phospho-H2A(X)) is increased proximal to *cds1Δ* and *mrc1Δ* replication forks [39, 53], which could represent uncoupling of leading and lagging strand synthesis in advance of replication fork collapse. Together, these observations suggest that replication fork activity in checkpoint mutants shapes their stability. The inability of the mutant cells to restrain replication during HU and throughout release contributes to extensive ssDNA, DNA damage, and eventual collapse.

These results point to the *cds1Δ* and *mrc1Δ* Replication Fork Collapse Points occurring later than previously expected, and largely during HU release as cells attempt to resume the cell cycle. Thus, forks do not immediately collapse, but instead retain synthesis activity. The ends of new synthesis in release bear the marks of ssDNA and DNA damage. We conclude that fork collapse for these mutants is delayed, but its seeds are sown during HU block, only coming into full effect during release. The Replication Fork Collapse Point may be used as a descriptor for other genotypes to describe both how and when the majority of replication forks are destroyed in a population of cells. This is an execution point: while the cells are not viable by this time, the Collapse Point signals the time at which decay leading to death is fully established. We anticipate that the Collapse Point will be much later for *rad51Δ* or other repair-deficient mutants, which collapse by failing to properly restart. In contrast, Fork Protection Complex mutants may show an intermediate timing, or an incomplete Fork Collapse Point.

In turn, these studies prompt further questions. Do dNTP pools recover after release in *cds1Δ* and *mrc1Δ* cells? If this is the case, as suggested by the increase in replication after release in *cds1Δ* and *mrc1Δ* cells, what are the additional defining features of replication fork collapse?

## 10. Conclusions

Monitoring replication competency, accumulation of ssDNA and DNA damage signals around replication forks permits modeling to determine how replication forks respond to HU arrest and recovery. This, in turn, indicates what role checkpoint proteins Cds1 and Mrc1 play in fork stability and effective restart. The Replication Fork Collapse Point incorporates the signs and symptoms of fork collapse and attempts to put a time to when the majority of replication forks undergo collapse. This is likely different for different genetic backgrounds missing key components of checkpoint signal, fork stabilization and replication restart. Future work will dissect replication fork proteins in HU and release, and take the genome-wide data from microarray and sequencing, moving into monitoring patterns at individual replication forks. Since replication stability and fidelity is a key barrier to malignancy, defining when and how replication forks collapse in the absence of checkpoint will allow insights into the development and prevention of cancer.

## Acknowledgements

## Author details

Sarah A. Sabatinos[1] and Susan L. Forsburg[1*]

*Address all correspondence to: forsburg@usc.edu

1 Department of Molecular and Computational Biology, University of Southern California, USA

## References

[1] Torres-Rosell, J., De Piccoli, G., Cordon-Preciado, V., Farmer, S., Jarmuz, A., Machin, F., et al. (2007). Anaphase onset before complete DNA replication with intact checkpoint responses. *Science*, 315(5817), 1411-1415.

[2] Rothstein, R., Michel, B., & Gangloff, S. (2000). Replication fork pausing and recombination or "gimme a break". *Genes Dev*, 14(1), 1-10.

[3] Labib, K., & Hodgson, B. (2007). Replication fork barriers: pausing for a break or stalling for time? *EMBO Rep*, 8(4), 346-353.

[4] Lucca, C., Vanoli, F., Cotta-Ramusino, C., Pellicioli, A., Liberi, G., Haber, J., et al. (2004). Checkpoint-mediated control of replisome-fork association and signalling in response to replication pausing. *Oncogene*, 23(6), 1206-1213.

[5] Meister, P., Taddei, A., Vernis, L., Poidevin, M., Gasser, S. M., & Baldacci, G. (2005). Temporal separation of replication and recombination requires the intra-S checkpoint. *J Cell Biol*, 168(4), 537-544.

[6] Tourriere, H., & Pasero, P. (2007). Maintenance of fork integrity at damaged DNA and natural pause sites. *DNA Repair (Amst)*, 6(7), 900-913.

[7] Petermann, E., Orta, M. L., Issaeva, N., Schultz, N., & Helleday, T. (2010). Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. *Mol Cell*, 37(4), 492-502.

[8] Mao, N., Kojic, M., & Holloman, W. K. (2009). Role of Blm and collaborating factors in recombination and survival following replication stress in Ustilago maydis. *DNA Repair (Amst)*, 8(6), 752-9.

[9] Froget, B., Blaisonneau, J., Lambert, S., & Baldacci, G. (2008). Cleavage of stalled forks by fission yeast Mus81/Eme1 in absence of DNA replication checkpoint. *Mol Biol Cell*, 19(2), 445-456.

[10] Kai, M., Boddy, M. N., Russell, P., & Wang, T. S. (2005). Replication checkpoint kinase Cds1 regulates Mus81 to preserve genome integrity during replication stress. *Genes Dev*, 19(8), 919-932.

[11] Noguchi, E., Noguchi, C., Du, L. L., & Russell, P. (2003). Swi1 prevents replication fork collapse and controls checkpoint kinase Cds1. *Mol Cell Biol*, 23(21), 7861-7874.

[12] Bryant, H. E., Petermann, E., Schultz, N., Jemth, A. S., Loseva, O., & Issaeva, N. (2009). PARP is activated at stalled forks to mediate Mre11-dependent replication restart and recombination. *Embo J*, 28(17), 2601-2615.

[13] Bernstein, K. A., Shor, E., Sunjevaric, I., Fumasoni, M., Burgess, R. C., Foiani, M., et al. (2009). Sgs1 function in the repair of DNA replication intermediates is separable from its role in homologous recombinational repair. *Embo J*, 28(7), 915-925.

[14] Alcasabas, A. A., Osborn, A. J., Bachant, J., Hu, F., Werler, P. J., Bousset, K., et al. (2001). Mrc1 transduces signals of DNA replication stress to activate Rad53. *Nat Cell Biol*, 3(11), 958-965.

[15] Bailis, J. M., Luche, D. D., Hunter, T., & Forsburg, S. L. (2008). Minichromosome maintenance proteins interact with checkpoint and recombination proteins to promote s-phase genome stability. *Molecular and cellular biology*, 28(5), 1724-1738.

[16] Branzei, D., & Foiani, M. (2007). Interplay of replication checkpoints and repair proteins at stalled replication forks. *DNA Repair (Amst)*, 6(7), 994-1003.

[17] Lopes, M., Cotta-Ramusino, C., Pellicioli, A., Liberi, G., Plevani, P., Muzi-Falconi, M., et al. (2001). The DNA replication checkpoint response stabilizes stalled replication forks. *Nature*, 412(6846), 557-561.

[18] Marchetti, M. A., Kumar, S., Hartsuiker, E., Maftahi, M., Carr, A. M., & Freyer, G. A. (2002). A single unbranched S-phase DNA damage and replication fork blockage checkpoint pathway. *Proc Natl Acad Sci U S A*, 99(11), 7472-7477.

[19] Tsang, E., & Carr, A. M. (2008). Replication fork arrest, recombination and the maintenance of ribosomal DNA stability. *DNA Repair (Amst)*, 7(10), 1613-1623.

[20] Bartkova, J., Horejsi, Z., Koed, K., Kramer, A., Tort, F., & Zieger, K. (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature*, 434(7035), 864-870.

[21] Bartkova, J., Rezaei, N., Liontos, M., Karakaidos, P., Kletsas, D., Issaeva, N., et al. (2006). Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature*, 444(7119), 633-637.

[22] Koc, A., Wheeler, L. J., Mathews, C. K., & Merrill, G. F. (2004). Hydroxyurea arrests DNA replication by a mechanism that preserves basal dNTP pools. *J Biol Chem*, 279(1), 223-30.

[23] Matsumoto, M., Rey, D. A., & Cory, J. G. (1990). Effects of cytosine arabinoside and hydroxyurea on the synthesis of deoxyribonucleotides and DNA replication in L1210 cells. *Adv Enzyme Regul*, 30-47.

[24] Bianchi, V., Pontis, E., & Reichard, P. (1986). Changes of deoxyribonucleoside triphosphate pools induced by hydroxyurea and their relation to DNA synthesis. *J Biol Chem*, 261(34), 16037-16042.

[25] Goodman, M. F., Hopkins, R. L., Lasken, R., & Mhaskar, D. N. (1985). The biochemical basis of 5-bromouracil- and 2-aminopurine-induced mutagenesis. *Basic Life Sci*, 31, 409-423.

[26] Hakansson, P., Dahl, L., Chilkova, O., Domkin, V., & Thelander, L. (2006). Thelander L. The Schizosaccharomyces pombe replication inhibitor Spd1 regulates ribonucleotide reductase activity and dNTPs by binding to the large Cdc22 subunit. *The Journal of biological chemistry*, 281(3), 1778-1783.

[27] Kunz, BA, Kang, X. L., & Kohalmi, L. (1991). The yeast rad18 mutator specifically increases G.C----T.A transversions without reducing correction of G-A or C-T mismatches to G.C pairs. *Molecular and cellular biology*, 11(1), 218-225.

[28] Kohalmi, S. E., Haynes, R. H., & Kunz, B. A. (1988). Instability of a yeast centromere plasmid under conditions of thymine nucleotide stress. *Mutat Res*, 207(1), 13-16.

[29] Potter, C. G. (1971). Induction of polyploidy by concentrated thymidine. *Exp Cell Res*, 68(2), 442-448.

[30] Meuth, M. (1981). Role of deoxynucleoside triphosphate pools in the cytotoxic and mutagenic effects of DNA alkylating agents. *Somatic Cell Genet*, 7(1), 89-102.

[31] Meuth, M. (1983). Deoxycytidine kinase-deficient mutants of Chinese hamster ovary cells are hypersensitive to DNA alkylating agents. *Mutat Res*, 110(2), 383-391.

[32] Mulder, K. W., Winkler, G. S., & Timmers, H. T. (2005). DNA damage and replication stress induced transcription of RNR genes is dependent on the Ccr4-Not complex. *Nucleic Acids Res*, 33(19), 6384-6392.

[33] Alvino, G. M., Collingwood, D., Murphy, J. M., Delrow, J., Brewer, B. J., & Raghuraman, M. K. (2007). Replication in hydroxyurea: it's a matter of time. *Molecular and cellular biology*, 27(18), 6396-6406.

[34] Poli, J., Tsaponina, O., Crabbe, L., Keszthelyi, A., Pantesco, V., & Chabes, A. (2012). dNTP pools determine fork progression and origin usage under replication stress. *The EMBO journal*, 31(4), 883-894.

[35]  Bolderson, E., Scorah, J., Helleday, T., Smythe, C., & Meuth, M. (2004). ATM is required for the cellular response to thymidine induced replication fork stress. *Hum Mol Genet*, 13(23), 2937-2945.

[36]  Feng, W., Collingwood, D., Boeck, M. E., Fox, L. A., Alvino, G. M., Fangman, W. L., et al. (2006). Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nature cell biology*, 8(2), 148-55.

[37]  Vassin, V. M., Anantha, R. W., Sokolova, E., Kanner, S., & Borowiec, J. A. (2009). Human RPA phosphorylation by ATR stimulates DNA synthesis and prevents ssDNA accumulation during DNA-replication stress. *J Cell Sci*, Pt 22, 4070-4080.

[38]  Letessier, A., Millot, G. A., Koundrioukoff, S., Lachages, A. M., Vogt, N., Hansen, R. S., et al. (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature*, 470(7332), 120-123.

[39]  Feng, W., Di Rienzi, S. C., Raghuraman, M. K., & Brewer, B. J. (2011). Replication stress-induced chromosome breakage is correlated with replication fork progression and is preceded by single-stranded DNA formation. *G3 (Bethesda)*, 1(5), 327-35.

[40]  Durkin, S. G., & Glover, T. W. (2007). Chromosome fragile sites. *Annu Rev Genet*, 41-169.

[41]  Chabes, A., Georgieva, B., Domkin, V., Zhao, X., Rothstein, R., & Thelander, L. (2003). Survival of DNA damage in yeast directly depends on increased dNTP levels allowed by relaxed feedback inhibition of ribonucleotide reductase. *Cell*, 112(3), 391-401.

[42]  Davidson, M. B., Katou, Y., Keszthelyi, A., Sing, T. L., Xia, T., Ou, J., et al. (2012). Endogenous DNA replication stress results in expansion of dNTP pools and a mutator phenotype. *The EMBO journal*, 31(4), 895-907.

[43]  Fasullo, M., Tsaponina, O., Sun, M., & Chabes, A. (2010). Elevated dNTP levels suppress hyper-recombination in Saccharomyces cerevisiae S-phase checkpoint mutants. *Nucleic acids research*, 38(4), 1195-1203.

[44]  Kumar, D., Abdulovic, A. L., Viberg, J., Nilsson, A. K., Kunkel, T. A., & Chabes, A. (2011). Mechanisms of mutagenesis in vivo due to imbalanced dNTP pools. *Nucleic acids research*, 39(4), 1360-1371.

[45]  Muzi-Falconi, M., Giannattasio, M., Foiani, M., & Plevani, P. (2003). The DNA polymerase alpha-primase complex: multiple functions and interactions. *ScientificWorldJournal*, 3-21.

[46]  Langston, L. D., Indiani, C., & O'Donnell, M. (2009). Whither the replisome: emerging perspectives on the dynamic nature of the DNA replication machinery. *Cell Cycle*, 8(17), 2686-2691.

[47]  Hubscher, U. (2009). DNA replication fork proteins. *Methods Mol Biol*, 521-19.

[48] Lopez-Contreras, A. J., & Fernandez-Capetillo, O. (2010). The ATR barrier to replication-born DNA damage. *DNA repair*, 9(12), 1249-1255.

[49] Glover, T. W., Arlt, M. F., Casper, A. M., & Durkin, S. G. (2005). Mechanisms of common fragile site instability. *Hum Mol Genet* [2], R197-R205.

[50] Wang, X., & Haber, J. E. (2004). Role of Saccharomyces single-stranded DNA-binding protein RPA in the strand invasion step of double-strand break repair. *PLoS Biol*, E21.

[51] Alabert, C., Bianco, J. N., & Pasero, P. (2009). Differential regulation of homologous recombination at DNA breaks and replication forks by the Mrc1 branch of the S-phase checkpoint. *Embo J*, 28(8), 1131-1141.

[52] Sugiyama, T., & Kantake, N. (2009). Dynamic regulatory interactions of rad51, rad52, and replication protein-a in recombination intermediates. *J Mol Biol*, 390(1), 45-55.

[53] Sabatinos, S. A., Green, M. D., & Forsburg, S. L. (2012). Continued DNA synthesis in replication checkpoint mutants leads to fork collapse. *submitted.*

[54] Sogo, J. M., Lopes, M., & Foiani, M. (2002). Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science*, 297(5581), 599-602.

[55] Mc Farlane, R. J., Mian, S., & Dalgaard, J. Z. (2010). The many facets of the Tim-Tipin protein families' roles in chromosome biology. *Cell Cycle*, 9(4), 700-705.

[56] Gotter, A. L., Suppa, C., & Emanuel, BS. (2007). Mammalian TIMELESS and Tipin are evolutionarily conserved replication fork-associated factors. *J Mol Biol*, 366(1), 36-52.

[57] Noguchi, E., Noguchi, C., Mc Donald, W. H., Yates, J. R. 3rd, & Russell, P. (2004). Swi1 and Swi3 are components of a replication fork protection complex in fission yeast. *Mol Cell Biol*, 24(19), 8342-8355.

[58] Katou, Y., Kanoh, Y., Bando, M., Noguchi, H., Tanaka, H., Ashikari, T., et al. (2003). S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature*, 424(6952), 1078-1083.

[59] Williams, D. R., & McIntosh, J. R. (2002). mcl1+, the Schizosaccharomyces pombe homologue of CTF4, is important for chromosome replication, cohesion, and segregation. *Eukaryot Cell*, 1(5), 758-773.

[60] 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(12), 5724-6735.

[61] Gambus, A., van Deursen, F., Polychronopoulos, D., Foltman, M., Jones, R. C., & Edmondson, R. D. (2009). A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome. *Embo J*, 28(19), 2992-3004.

[62] Im, J. S., Ki, S. H., Farina, A., Jung, Hurwitz. J., & Lee, J. K. (2009). Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins. *Proc Natl Acad Sci U S A*, 106(37), 15628-15632.

[63] Bando, M., Katou, Y., Komata, M., Tanaka, H., Itoh, T., Sutani, T., et al. (2009). Csm3, Tof1, and Mrc1 form a heterotrimeric mediator complex that associates with DNA replication forks. *J Biol Chem*, 284(49), 34355-34365.

[64] Gambus, A., Jones, R. C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., & Edmondson, R. D. (2006). GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol*, 8(4), 358-366.

[65] Nedelcheva, M. N., Roguev, A., Dolapchiev, L. B., Shevchenko, A., Taskov, H. B., Shevchenko, A., et al. (2005). Uncoupling of unwinding from DNA synthesis implies regulation of MCM helicase by Tof1/Mrc1/Csm3 checkpoint complex. *J Mol Biol*, 347(3), 509-21.

[66] Calzada, A, Hodgson, B, Kanemaki, M, Bueno, A, & Labib, K. (2005). Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. *Genes Dev*, 19(16), 1905-1919.

[67] Hamdan, S. M., Johnson, D. E., Tanner, N. A., Lee, J. B., Qimron, U., Tabor, S., et al. (2007). Dynamic DNA helicase-DNA polymerase interactions assure processive replication fork movement. *Molecular cell*, 27(4), 539-549.

[68] Kim, S., Dallmann, H. G., Mc Henry, C. S., & Marians, K. J. (1996). Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement. *Cell*, 84(4), 643-650.

[69] Stano, N. M., Jeong, Y. J., Donmez, I., Tummalapalli, P., Levin, M. K., & Patel, S. S. (2005). DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase. *Nature*, 435(7040), 370-373.

[70] Tougu, K., & Marians, K. J. (1996). The interaction between helicase and primase sets the replication fork clock. *The Journal of biological chemistry*, 271(35), 21398-21405.

[71] Unsal-Kacmaz, K., Chastain, P. D., Qu, P. P., Minoo, P., Cordeiro-Stone, M., Sancar, A., et al. (2007). The human Tim/Tipin complex coordinates an Intra-S checkpoint response to UV that slows replication fork displacement. *Mol Cell Biol*, 27(8), 3131-3142.

[72] Yoshizawa-Sugata, N., & Masai, H. (2007). Human Tim/Timeless-interacting protein, Tipin, is required for efficient progression of S phase and DNA replication checkpoint. *J Biol Chem*, 282(4), 2729-2740.

[73] Errico, A., Cosentino, C., Rivera, T., Losada, A., Schwob, E., Hunt, T., et al. (2009). Tipin/Tim1/And1 protein complex promotes Pol alpha chromatin binding and sister chromatid cohesion. *Embo J*, 28(23), 3681-3692.

[74] Tanaka, H., Kubota, Y., Tsujimura, T., Kumano, M., Masai, H., & Takisawa, H. (2009). Replisome progression complex links DNA replication to sister chromatid cohesion in Xenopus egg extracts. *Genes Cells*, 14(8), 949-963.

[75] Leman, A. R., Noguchi, C., Lee, C. Y., & Noguchi, E. (2010). Human Timeless and Tipin stabilize replication forks and facilitate sister-chromatid cohesion. *J Cell Sci*, Pt 5, 660-670.

[76] Errico, A., Costanzo, V., & Hunt, T. (2007). Tipin is required for stalled replication forks to resume DNA replication after removal of aphidicolin in Xenopus egg extracts. *Proc Natl Acad Sci U S A*, 104(38), 14929-34.

[77] Tanaka, T., Yokoyama, M., Matsumoto, S., Fukatsu, R., You, Z., & Masai, H. (2010). Fission yeast Swi1-Swi3 complex facilitates DNA binding of Mrc1. *The Journal of biological chemistry*, 285(51), 39609-22.

[78] Kumagai, A., & Dunphy, W. G. (2000). Claspin, a novel protein required for the activation of Chk1 during a DNA replication checkpoint response in Xenopus egg extracts. *Mol Cell*, 6(4), 839-849.

[79] Kumagai, A., Kim, S. M., & Dunphy, W. G. (2004). Claspin and the activated form of ATR-ATRIP collaborate in the activation of Chk1. *J Biol Chem*, 279(48), 49599-45608.

[80] Lee, J., Gold, D. A., Shevchenko, A., Shevchenko, A., & Dunphy, W. G. (2005). Roles of replication fork-interacting and Chk1-activating domains from Claspin in a DNA replication checkpoint response. *Mol Biol Cell*, 16(11), 5269-5282.

[81] Lou, H., Komata, M., Katou, Y., Guan, Z., Reis, C. C., Budd, M., et al. (2008). Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint. *Molecular cell*, 32(1), 106-117.

[82] Osborn, A. J., & Elledge, S. J. (2003). Mrc1 is a replication fork component whose phosphorylation in response to DNA replication stress activates Rad53. *Genes & development*, 17(14), 1755-1767.

[83] Williams, D. R., & McIntosh, J. R. (2005). Mcl1p is a polymerase alpha replication accessory factor important for S-phase DNA damage survival. *Eukaryotic cell*, 4(1), 166-177.

[84] Arlt, M. F., Mulle, J. G., Schaibley, V. M., Ragland, R. L., Durkin, S. G., & Warren, S. T. (2009). Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. *Am J Hum Genet*, 84(3), 339-350.

[85] Mirkin, E. V., & Mirkin, S. M. (2007). Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev*, 71(1), 13-35.

[86] Krings, G., & Bastia, D. (2004). swi1- and swi3-dependent and independent replication fork arrest at the ribosomal DNA of Schizosaccharomyces pombe. *Proc Natl Acad Sci U S A*, 101(39), 14085-90.

[87] Noguchi, C., & Noguchi, E. (2007). Sap1 promotes the association of the replication fork protection complex with chromatin and is involved in the replication checkpoint in Schizosaccharomyces pombe. *Genetics*, 175(2), 553-566.

[88] Dalgaard, J. Z., & Klar, A. J. (2000). swi1 and swi3 perform imprinting, pausing, and termination of DNA replication in S. pombe. *Cell*, 102(6), 745-751.

[89] Razidlo, D. F., & Lahue, R. S. (2008). Mrc1, Tof1 and Csm3 inhibit CAG.CTG repeat instability by at least two mechanisms. *DNA Repair (Amst)*, 7(4), 633-640.

[90] Ahn, J. S., Osman, F., & Whitby, M. C. (2005). Replication fork blockage by RTS1 at an ectopic site promotes recombination in fission yeast. *The EMBO journal*, 24(11), 2011-2023.

[91] Eydmann, T., Sommariva, E., Inagawa, T., Mian, S., Klar, A. J., & Dalgaard, J. Z. (2008). Rtf1-mediated eukaryotic site-specific replication termination. *Genetics*, 180(1), 27-39.

[92] Codlin, S., & Dalgaard, J. Z. (2003). Complex mechanism of site-specific DNA replication termination in fission yeast. *The EMBO journal*, 22(13), 3431-3440.

[93] Dalgaard, J. Z., & Klar, A. J. (2001). A DNA replication-arrest site RTS1 regulates imprinting by determining the direction of replication at mat1 in S. pombe. *Genes & development*, 15(16), 2060-2068.

[94] Lambert, S., Mizuno, K., Blaisonneau, J., Martineau, S., Chanet, R., Freon, K., et al. (2010). Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange. *Molecular cell*, 39(3), 346-359.

[95] Vengrova, S., Codlin, S., & Dalgaard, J. Z. (2002). RTS1-an eukaryotic terminator of replication. *Int J Biochem Cell Biol*, 34(9), 1031-1034.

[96] Lee, B. S., Grewal, S. I., & Klar, A. J. (2004). Biochemical interactions between proteins and mat1 cis-acting sequences required for imprinting in fission yeast. *Molecular and cellular biology*, 24(22), 9813-9822.

[97] Coulon, S, Noguchi, E, Noguchi, C, Du, LL, Nakamura, TM, & Russell, P. (2006). Rad22Rad52-dependent repair of ribosomal DNA repeats cleaved by Slx1-Slx4 endonuclease. *Mol Biol Cell*, 17(4), 2081-2090.

[98] Kaplan, D. L., & Bastia, D. (2009). Mechanisms of polar arrest of a replication fork. *Mol Microbiol*, 72(2), 279-285.

[99] Krings, G., & Bastia, D. (2005). Sap1p binds to Ter1 at the ribosomal DNA of Schizosaccharomyces pombe and causes polar replication fork arrest. *J Biol Chem*, 280(47), 39135-39142.

[100] Maric, C., Levacher, B., & Hyrien, O. (1999). Developmental regulation of replication fork pausing in Xenopus laevis ribosomal RNA genes. *J Mol Biol*, 291(4), 775-788.

[101] Biswas, S., & Bastia, D. (2008). Mechanistic insights into replication termination as revealed by investigations of the Reb1-Ter3 complex of Schizosaccharomyces pombe. *Mol Cell Biol*, 28(22), 6844-6857.

[102] Zhao, A., Guo, A., Liu, Z., & Pape, L. (1997). Molecular cloning and analysis of Schiz-osaccharomyces pombe Reb1p: sequence-specific recognition of two sites in the far upstream rDNA intergenic spacer. *Nucleic acids research*, 25(4), 904-910.

[103] Singh, S. K., Sabatinos, S., Forsburg, S., & Bastia, D. (2010). Regulation of replication termination by Reb1 protein-mediated action at a distance. *Cell*, 142(6), 868-78.

[104] Bochman, M. L., Sabouri, N., & Zakian, V. A. (2010). Unwinding the functions of the Pif1 family helicases. *DNA Repair (Amst)*, 9(3), 237-49.

[105] Bairwa, N. K., Zzaman, S., Mohanty, B. K., & Bastia, D. (2010). Replication fork arrest and rDNA silencing are two independent and separable functions of the replication terminator protein Fob1 of Saccharomyces cerevisiae. *The Journal of biological chemistry*, 285(17), 12612-9.

[106] Evers, R., & Grummt, I. (1995). Molecular coevolution of mammalian ribosomal gene terminator sequences and the transcription termination factor TTF-I. *Proceedings of the National Academy of Sciences of the United States of America*, 92(13), 5827-31.

[107] Gerber, J. K., Gogel, E., Berger, C., Wallisch, M., Muller, F., Grummt, I., et al. (1997). Termination of mammalian rDNA replication: polar arrest of replication fork move-ment by transcription termination factor TTF-I. *Cell*, 90(3), 559-567.

[108] Langst, G., Becker, P. B., & Grummt-I, I. (1998). TTF-I determines the chromatin ar-chitecture of the active rDNA promoter. *The EMBO journal*, 17(11), 3135-45.

[109] Arlt, M. F., Durkin, S. G., Ragland, R. L., & Glover, T. W. (2006). Common fragile sites as targets for chromosome rearrangements. *DNA repair*, 5(9-10), 1126-1135.

[110] Howlett, N. G., Taniguchi, T., Durkin, S. G., D'Andrea, A. D., & Glover, T. W. (2005). The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. *Hum Mol Genet*, 14(5), 693-701.

[111] Bermejo, R., Capra, T., Gonzalez-Huici, V., Fachinetti, D., Cocito, A., Natoli, G., et al. (2009). Genome-organizing factors Top2 and Hmo1 prevent chromosome fragility at sites of S phase transcription. *Cell*, 138(5), 870-884.

[112] Yarbro, J. W. (1992). Mechanism of action of hydroxyurea. *Semin Oncol*, (9), 1-10.

[113] Tsaponina, O., Barsoum, E., Astrom, S. U., & Chabes, A. (2011). Ixr1 is required for the expression of the ribonucleotide reductase Rnr1 and maintenance of dNTP pools. *PLoS Genet*, 7(5), e1002061.

[114] Zhao, X., Chabes, A., Domkin, V., Thelander, L., & Rothstein, R. (2001). Thelander L, and Rothstein R. The ribonucleotide reductase inhibitor Sml1 is a new target of the Mec1/Rad53 kinase cascade during growth and in response to DNA damage. The EMBO journal , 20(13), 3544-3553.

[115] Huang, A., Fan, H., Taylor, W. R., & Wright, J. A. (1997). Ribonucleotide reductase R2 gene expression and changes in drug sensitivity and genome stability. *Cancer Res*, 57(21), 4876-4881.

[116] Miyabe, I., Morishita, T., Shinagawa, H., & Carr, A. M. (2009). Schizosaccharomyces pombe Cds1Chk2 regulates homologous recombination at stalled replication forks through the phosphorylation of recombination protein Rad60. *J Cell Sci*, Pt 20, 3638-3643.

[117] Lindsay, H. D., Griffiths, D. J., Edwards, R. J., Christensen, P. U., Murray, J. M., Osman, F., et al. (1998). S-phase-specific activation of Cds1 kinase defines a subpathway of the checkpoint response in Schizosaccharomyces pombe. *Genes & development*, 12(3), 382-395.

[118] Zhao, H., Tanaka, K., Nogochi, E., Nogochi, C., & Russell, P. (2003). Replication checkpoint protein Mrc1 is regulated by Rad3 and Tel1 in fission yeast. *Mol Cell Biol*, 23(22), 8395-8403.

[119] Kumar, D., Viberg, J., Nilsson, A. K., & Chabes, A. (2010). Highly mutagenic and severely imbalanced dNTP pools can escape detection by the S-phase checkpoint. *Nucleic acids research*, 38(12), 3975-3983.

[120] Zou, L., Liu, D., & Elledge, S. J. (2003). Replication protein A-mediated recruitment and activation of Rad17 complexes. *Proc Natl Acad Sci U S A*, 100(24), 13827-13832.

[121] Kanoh, Y., Tamai, K., & Shirahige, K. (2006). Different requirements for the association of ATR-ATRIP and 9-1-1 to the stalled replication forks. *Gene*, 377-388.

[122] Kemp, M., & Sancar, A. (2009). DNA distress: just ring 9-1-1. *Curr Biol*, 19(17), R733-R734.

[123] Yan, S., & Michael, W. M. (2009). TopBP1 and DNA polymerase-alpha directly recruit the 9-1-1 complex to stalled DNA replication forks. *J Cell Biol*, 184(6), 793-804.

[124] Branzei, D., & Foiani, M. (2006). The Rad53 signal transduction pathway: Replication fork stabilization, DNA repair, and adaptation. *Exp Cell Res*, 312(14), 2654-2659.

[125] Yan, S., & Michael, W. M. (2009). TopBP1 and DNA polymerase alpha-mediated recruitment of the 9-1-1 complex to stalled replication forks: implications for a replication restart-based mechanism for ATR checkpoint activation. *Cell Cycle*, 8(18), 2877-2884.

[126] Tanaka, K., & Russell, P. (2001). Mrc1 channels the DNA replication arrest signal to checkpoint kinase Cds1. *Nat Cell Biol*, 3(11), 966-972.

[127] Tanaka, K., & Russell, P. (2004). Cds1 phosphorylation by Rad3-Rad26 kinase is mediated by forkhead-associated domain interaction with Mrc1. *J Biol Chem*, 279(31), 32079-32086.

[128] Xu, Y. J., Davenport, M., & Kelly, T. J. (2006). Two-stage mechanism for activation of the DNA replication checkpoint kinase Cds1 in fission yeast. *Genes Dev*, 20(8), 990-1003.

[129] Osborn, A. J., Elledge, S. J., & Zou, L. (2002). Checking on the fork: the DNA-replication stress-response pathway. *Trends Cell Biol*, 12(11), 509-516.

[130] Naylor, M. L., Li, J. M., Osborn, A. J., & Elledge, S. J. (2009). Mrc1 phosphorylation in response to DNA replication stress is required for Mec1 accumulation at the stalled fork. *Proc Natl Acad Sci U S A*, 106(31), 12765-12770.

[131] Schleker, T., Nagai, S., & Gasser, S. M. (2009). Posttranslational modifications of repair factors and histones in the cellular response to stalled replication forks. *DNA Repair (Amst)*, 8(9), 1089-100.

[132] Tourriere, H., Versini, G., Cordon-Preciado, V., Alabert, C., & Pasero, P. (2005). Mrc1 and Tof1 promote replication fork progression and recovery independently of Rad53. *Mol Cell*, 19(5), 699-706.

[133] De Piccoli, G., Katou, Y., Itoh, T., Nakato, R., Shirahige, K., & Labib, K. (2012). Replisome stability at defective DNA replication forks is independent of S phase checkpoint kinases. *Molecular cell*, 45(5), 696-704.

[134] Kim, S. M., & Huberman, J. A. (2001). Regulation of replication timing in fission yeast. *The EMBO journal*, 20(21), 6115-6126.

[135] Feng, W., Bachant, J., Collingwood, D., Raghuraman, M. K., & Brewer, B. J. (2009). Centromere replication timing determines different forms of genomic instability in Saccharomyces cerevisiae checkpoint mutants during replication stress. *Genetics*, 183(4), 1249-1260.

[136] Lengronne, A., Pasero, P., Bensimon, A., & Schwob, E. (2001). Monitoring S phase progression globally and locally using BrdU incorporation in TK(+) yeast strains. *Nucleic acids research*, 29(7), 1433-1442.

[137] Sabatinos, S. A., Mastro, T. L., & Forsburg, S. L. (2012). Nucleoside analogues create DNA damage and sensitivity in fission yeast. *Eukaryotic Cell*, submitted.

[138] Chin, J. K., Bashkirov, V. I., Heyer, WD, & Romesberg, FE. (2006). Esc4/Rtt107 and the control of recombination during replication. *DNA repair*, 5(5), 618-628.

[139] Koren, A., Soifer, I., & Barkai, N. (2010). MRC1-dependent scaling of the budding yeast DNA replication timing program. *Genome Res*, 20(6), 781-790.

[140] Hayano, M, Kanoh, Y, Matsumoto, S, & Masai, H. (2011). Mrc1 marks early-firing origins and coordinates timing and efficiency of initiation in fission yeast. *Molecular and cellular biology*, 31(12), 2380-2391.

[141] Szyjka, S. J., Viggiani, C. J., & Aparicio, O. M. (2005). Mrc1 is required for normal progression of replication forks throughout chromatin in S. cerevisiae. *Molecular cell*, 19(5), 691-697.

[142] Xu, Y. J., & Kelly, T. J. (2009). Autoinhibition and autoactivation of the DNA replication checkpoint kinase Cds1. *The Journal of biological chemistry*, 284(23), 16016-16027.

[143] Hayashi, M., Katou, Y., Itoh, T., Tazumi, A., Yamada, Y., Takahashi, T., et al. (2007). Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. *The EMBO journal*, 26(5), 1327-1339.

[144] Santocanale, C., & Diffley, J. F. (1998). A Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication. *Nature*, 395(6702), 615-618.

[145] Heichinger, C., Penkett, C. J., Bahler, J., & Nurse, P. (2006). Genome-wide characterization of fission yeast DNA replication origins. *The EMBO journal*, 25(21), 5171-5179.

[146] Chabes, A., & Stillman, B. (2007). Constitutively high dNTP concentration inhibits cell cycle progression and the DNA damage checkpoint in yeast Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, 104(4), 1183-8.

[147] Mickle, K. L., Ramanathan, S., Rosebrock, A., Oliva, A., Chaudari, A., Yompakdee, C., et al. (2007). Checkpoint independence of most DNA replication origins in fission yeast. *BMC Mol Biol*, 8, 112.

[148] Hanada, K., Budzowska, M., Davies, S. L., van Drunen, E., Onizawa, H., Beverloo, H. B., et al. (2007). The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. *Nat Struct Mol Biol*, 14(11), 1096-1104.

[149] Robison, J. G., Elliott, J., Dixon, K., & Oakley, G. G. (2004). Replication protein A and the Mre11.Rad50.Nbs1 complex co-localize and interact at sites of stalled replication forks. *J Biol Chem*, 279(33), 34802-34810.

[150] Schlacher, K., Christ, N., Siaud, N., Egashira, A., Wu, H., & Jasin, M. (2011). Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell*, 145(4), 529-42.

[151] Ouyang, K. J., Woo, L. L., Zhu, J., Huo, D., Matunis, M. J., & Ellis, N. A. (2009). SU-MO modification regulates BLM and RAD51 interaction at damaged replication forks. *PLoS Biol*, 7(12), e1000252.

[152] Kurokawa, Y., Murayama, Y., Haruta-Takahashi, N., Urabe, I., & Iwasaki, H. (2008). Reconstitution of DNA strand exchange mediated by Rhp51 recombinase and two mediators. *PLoS Biol*, 6(4), e88.

[153] Wray, J., Liu, J., Nickoloff, J. A., & Shen, Z. (2008). Distinct RAD51 associations with RAD52 and BCCIP in response to DNA damage and replication stress. *Cancer Res*, 68(8), 2699-2707.

[154] Lambert, S., Froget, B., & Carr, A. M. (2007). Arrested replication fork processing: interplay between checkpoints and recombination. *DNA Repair (Amst)*, 6(7), 1042-1061.

[155] Aggarwal, M., Sommers, J. A., Morris, C., & Brosh, R. M. , Jr. (2010). Delineation of WRN helicase function with EXO1 in the replicational stress response. *DNA Repair (Amst)*, 9(7), 765-776.

[156] Tinline-Purvis, H., Savory, A. P., Cullen, J. K., Dave, A., Moss, J., Bridge, W. L., et al. (2009). Failed gene conversion leads to extensive end processing and chromosomal rearrangements in fission yeast. *The EMBO journal*, 28(21), 3400-3412.

[157] Tran, P. T., Fey, J. P., Erdeniz, N., Gellon, L., Boiteux, S., & Liskay, R. M. (2007). A mutation in EXO1 defines separable roles in DNA mismatch repair and post-replication repair. *DNA Repair (Amst)*, 6(11), 1572-1583.

[158] Lambert, S., Watson, A., Sheedy, D. M., Martin, B., & Carr, A. M. (2005). Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell*, 121(5), 689-702.

[159] Meister, P., Taddei, A., Ponti, A., Baldacci, G., & Gasser, S. M. (2007). Replication foci dynamics: replication patterns are modulated by S-phase checkpoint kinases in fission yeast. *The EMBO journal*, 26(5), 1315-1326.

[160] Willis, N., & Rhind, N. (2010). The fission yeast Rad32(Mre11)-Rad50-Nbs1 complex acts both upstream and downstream of checkpoint signaling in the S-phase DNA damage checkpoint. *Genetics*, 184(4), 887-897.

[161] Hashimoto, Y., Puddu, F., & Costanzo, V. (2012). RAD51- and MRE11-dependent reassembly of uncoupled CMG helicase complex at collapsed replication forks. *Nature structural & molecular biology*, 19(1), 17-24.

[162] Brugmans, L., Verkaik, N. S., Kunen, M., van Drunen, E., Williams, B. R., & Petrini, J. H. (2009). NBS1 cooperates with homologous recombination to counteract chromosome breakage during replication. *DNA repair*, 8(12), 1363-1370.

[163] Kuzminov, A. (2001). Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15), 8241-8246.

[164] Caldecott, K. W. (2007). Mammalian single-strand break repair: mechanisms and links with chromatin. *DNA repair*, 6(4), 443-453.

[165] Hutchinson, F. (1993). Induction of large DNA deletions by persistent nicks: a new hypothesis. *Mutat Res*, 299(3-4), 211-218.

# The Role of WRN Helicase/Exonuclease in DNA Replication

Lynne S. Cox and Penelope A. Mason

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/51520

## 1. Introduction

### 1.1. WRN is a RecQ helicase/exonuclease required for genome stability and to prevent premature ageing

#### 1.1.1. Clinical phenotype of Werner's syndrome

Humans possess five distinct RecQ helicases (see Figure 1), all of which possess a hallmark RecQ helicase domain. Mutation or loss in any one of three human RecQ helicases give rise to genetic instability syndromes: WRN mutation gives Werner's syndrome (WS), BLM loss results in Bloom syndrome (BS), and Rothmund-Thomson syndrome (RTS) is caused by mutation of RECQL4[1]. WRN has come to prominence because its loss of function results in human Werner's syndrome, a segmental progeria (premature ageing) characterised by many signs and symptoms of normal ageing at both the organismal and cellular levels, with shortened lifespan (median age of death 47 years [2]). In particular WS patients suffer from osteoporosis, athero-and arterio-sclerosis and a high cancer incidence (particularly sarcoma) together with metabolic disorders normally associated with increased age, especially type II diabetes and lipodystrophy. Furthermore, patients show outwardly recognisable signs of ageing such as cataracts, greying hair and skin wrinkling, while female WS patients suffer premature menopause and both sexes show hypogonadism, with decreased fertility (reviewed in ref. [2]).

---

1 RTS is found in a subset of patients with RECQL4 mutation; different mutations in the same gene give rise to RAPA-DILLINO syndrome [1]

---

**1.2. Cellular phenotype on WRN loss**

This premature ageing phenotype is also observed at the cellular level: fibroblasts from WS patients undergo highly premature replicative senescence in culture, failing to proliferate after only 9-11 population doublings, compared with the 50-60 doublings characteristic of wild type fibroblasts [3]. Transcriptomic studies have demonstrated that >90% gene expression changes associated with normal ageing are seen in young WS cells [4], while glycosylation of blood albumin (a biomarker of ageing) in young WS patients is equivalent to levels detected in normal centenarians [5]. Importantly, loss of function of WRN is associated with significant genome instability with a high frequency of chromosomal translocations and deletions [6, 7], which is thought to contribute to the increased cancer risk. Genome instability is a hallmark of defective S phase checkpoint proteins (reviewed in ref. [8]), suggesting either than WRN is directly involved in the checkpoint, or that it normally serves downstream of the checkpoint such that its loss prevents correct execution of the arrest and recovery pathways. Notably, it is not only WS patients who are more susceptible to cancer on WRN loss: epigenetic inactivation by methylation of CpG islands in the WRN gene promoter has been reported in epithelial and mesenchymal cancers with value in prognosis in colorectal cancer [9], while specific WRN SNPs have been correlated with breast cancer incidence [10], even though such genetic changes do not alter the helicase or exonuclease activities of the protein or modulate the levels expressed. WRN is therefore of interest not only to those attempting to understand the molecular basis of human ageing, but also to cancer biologists – indeed WRN knockdown is likely to promote cancer cell death and hypersensitise cells to current chemotherapeutic agents such as camptothecin that impact on DNA replication [9, 11, 12]. Small molecules that specifically inhibit WRN but not other RecQ helicases are therefore likely to have therapeutic potential [13].

**1.3. WRN protein**

The wide range of ageing-associated phenotypes in WS patients and their cells indicates a fundamental role for WRN in preventing premature ageing, but how can loss of one protein lead to the pleiotropic outcomes of human ageing? The most important clue came from cloning the WRN gene [14], which showed for the first time that the human WRN gene encodes a large protein of 1432 amino acid (~162kDa) with an amino terminal exonuclease domain conserved with proteins of the DnaQ family, and a central helicase domain of the RecQ family. In addition, DNA binding (RQC) and protein interaction (HRDC) domains exist distal to the helicase domain (Figure 1A). Immunofluorescence and mutational studies have demonstrated that WRN is a nuclear protein with both NLS and NoLS sequences situated at the C terminus [15]), that appears to be sequestered in the nucleolus [16] except during S phase or upon DNA damage, when it is redistributed to sites of DNA replication or repair ([17-19].

Of the five human RecQ proteins, WRN is the only one to possess exonuclease activity [20]. Acting in a 3'-5' direction (as shown using 3'- or 5'-end labelled substrates), WRN exonuclease has been demonstrated to bind onto overhanging 5' ends of the guide strand of duplex DNA and cleave the target strand sequentially, though with relatively low processivity [21]. While it cannot cleave blunt ended substrates, nor those where ends are blocked by

bulky lesions [22], WRN exonuclease degrades substrates that are likely to be found both during DNA repair and as intermediates in DNA replication, including forks and bubble substrates [23] (see Table 1). Despite early reports of lack of activity on short single-stranded DNA (e.g. [21]), WRN exonuclease can digest single stranded oligonucleotides over 50 bases in length [24, 25].
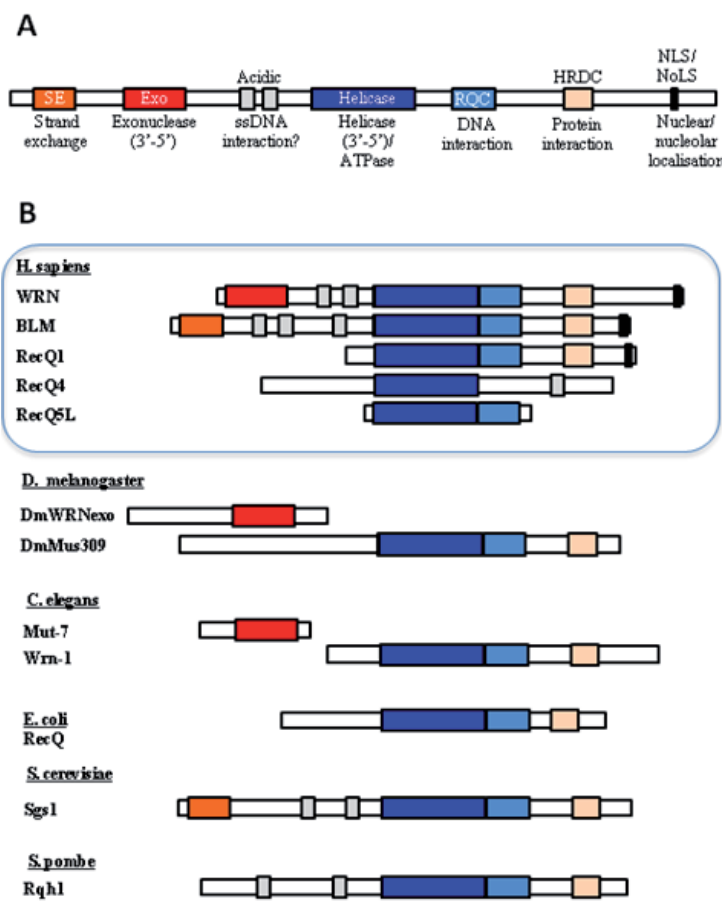


**Figure 1.** The RecQ helicase family. (A) Domain organization of human WRN. Note that for human WRN, the RQC serves in DNA binding and the HRDC is probably involved in protein-protein interaction, though these roles may be reversed in other RecQs. (B) Humans have 5 RecQ helicases (boxed), named after the archetypal RecQ of *E. coli*. Human WRN is unique in the family in possessing an exonuclease domain. In invertebrates such as *Drosophila* and *C. elegans*, the exonuclease (red) and helicase (blue) activities are encoded by separate genes.

| Helicase substrates | Exonuclease substrates |
|---|---|
| Holliday junction | Holliday junction |
| Bubble duplex | Bubble duplex |
| 3'-recessed duplex | 3'-recessed duplex |

| Helicase substrates | Exonuclease substrates |
|---|---|
| D-loop duplex | D-loop duplex |
| Duplex with 5'-flaps | Looped duplex |
| G-quadruplex | Nicked looped duplex |
| Partial duplex | |
| Forked duplex | |
| Triplex | |
| NOT blunt duplex | |
| NOT 5'-recessed duplex | |

**Table 1.** Substrates unwound by WRN helicase or degraded by WRN exonuclease. (Note that BLM also unwinds the same substrates as WRN helicase)

The helicase activity of WRN is highly conserved with other RecQ helicase family members, acting 3'-5'to unwind duplex DNA in an ATP-dependent manner [26]. Within the helicase domain are seven conserved motifs characteristic of the RecQ family. In general, RecQ helicases are adept at unwinding unusual DNA structures that can inhibit the course of normal DNA replication. Examples are tailed and forked duplexes, small gaps and flaps (commonly found as DNA repair and recombination intermediates), bubble substrates and displacement-loop triplex and Holliday junctions (common at telomeres and during recombinational repair and sister chromatid exchange), and G-quadruplexes which are often found at tracts rich in guanine such as at the telomere (e.g. [27], reviewed in ref. [28], see Table 1). It is important to note that the helicase and exonuclease activities do not simply act as independent entities in cells, but that their actions are almost certainly co-ordinated and interlinked. For example, co-operation between them is required during telomere maintenance ([29]; see section 4 below for more detail).

WRN helicase template specificity requires DNA binding that is probably mediated through the conserved RQC domain. X-ray crystallographic analysis has shown some unusual features, in that binding of WRN to DNA does not occur through a standard 'recognition helix', but instead through a beta wing of the RQC domain that inserts like a wedge between the terminal bases of blunt duplex DNA to unwind one base even in the absence of ATP [30]. How this binding correlates with WRN's lack of unwinding of blunt ended substrates remains to be determined. In addition to binding to DNA, WRN binds to many different proteins at the replication fork, the telomere and during fork recovery after stalling. Protein interaction with WRN may occur through the helicase-and-ribonuclease D/C-terminal (HRDC) domain; while this region is through to be important for DNA binding in *E. coli* RecQ and yeast Sgs1, the conserved for DNA interaction surface is lacking in human WRN, and the domain is unable to bind DNA *in vitro*, but that reveals many exposed alpha helicases that are likely to bind to protein partners [31].

## 1.4. WRN orthologues

While the exonuclease and helicase activities are both encoded by the same gene in verte-
brates, giving rise to one multifunctional protein, the enzymes are encoded by separate ge-
netic loci in plants, invertebrates and prokaryotes (Figure 1B, reviewed in ref. [32]), with
physical and/or functional interaction between the helicase and nuclease proposed *in vivo*.
(Figure 1, reviewed in ref. [32, 33] For example, in the fruit fly *Drosophila melanogaster*, we
have cloned and characterised the orthologue of human WRN exonuclease encoded by the
fly locus *CG7670* [34]. *Drosophila* WRN exonuclease (DmWRNexo) is a 3'-5' exonuclease [35]
that shows remarkable substrate conservation with human WRN exonuclease and utilises
conserved residues at the active site for nucleic acid cleavage [36]. Flies homozygous for a
strong hypomorphic mutation in *CG7670* have greatly elevated levels of recombination that
appears to occur through reciprocal exchange, and are hypersensitive to the topoisomerase
poison camptothecin, that leads to replication fork collapse [37]. Hence loss of only the
WRN exonuclease activity in flies results in many features characteristic of human WS, sug-
gesting a key role for the exonuclease in preventing premature ageing. We consider the pos-
sible role(s) of WRN exonuclease in replication fidelity, restart of stalled forks and telomere
maintenance in more detail below (see sections 2.2, 3.4 and 4 below).

A limitation to studying WRN in flies is the lack of a fully characterised WRN helicase or-
thologue. However, the nematode worm *C. elegans* has a highly conserved WRN-like heli-
case, encoded by the *wrn-1* gene, and two candidate exonucleases, at loci *ZK1098.8* (*mut-7*
[38]) and adjacent *ZK1098.3*. RNAi knockdown of *wrn-1* results in shortened lifespan [39]
and perturbation of the S phase checkpoint via ATM/R kinases [40], suggesting both that
WRN is important during DNA replication, and that its role is critical in maintaining normal
longevity of the organism. These outcomes are of particular interest since they so closely
echo the findings in humans, but in a genetically tractable and short-lived lower eukaryotic
model organism. In plants, WRN has been most studied in *Arabidopsis*, where physical and
function interaction has been described between the exonuclease (AtWEX) and helicase
(AtWRN) orthologues [41]. In budding yeast and fission yeast, there is only one RecQ heli-
case (Sgs1 and Rqh1, respectively); whether these proteins interacts directly with an exonu-
clease to reconstitute human WRN-like activity is yet to be determined, though genetic
interaction between Rqh1 and Mus81/Eme1 has been reported [42].

Because of the phenotypes resulting from WRN loss or mutation, it has been implicated in
many aspects of DNA metabolism, including transcription, DNA repair, recombination and
telomere maintenance. Its role in DNA replication will be discussed in this chapter, includ-
ing not only a direct role in normal processive DNA replication, and replication of the telo-
meres, but also in preventing replication fork stalling or assisting fork recovery after arrest.

## 1.5. S phase defects in WS cells

Fibroblasts and lymphoblastoid cells from Werner's syndrome patients show a defect in pro-
gression through S phase [17, 43]. FACS analysis demonstrates both a longer duration of S
phase and an overall significant increase in cell cycle time in primary fibroblasts from WS pa-
tients ([17] and in normal primary fibroblasts in which WRN was depleted by shRNAi by

80-90% [44]. Early studies on replication rates in WS fibroblasts used alkaline sucrose gradients to detect the size of nascent DNA, demonstrating slower replication in WS cells compared with normal controls [45]. The ability of WS cells to incorporate Texas-red-dUTP into nascent DNA is also significantly impaired [46]. Interestingly, while acute shRNAi-mediated WRN depletion in SV40 T antigen-transformed cells had no impact on cell cycle progression in the absence of imposed replication stress, primary fibroblasts depleted of WRN did show an S phase delay [44]. Hence it appears that loss of WRN protein results in an S phase phenotype.

WRN has been isolated within a large multi-protein replication complex [47] and found to interact *in vitro* with purified PCNA. The binding region has been localised to a PIP-like motif on WRN towards the amino terminus [18], which is likely to bind within the hydrophobic pocket of PCNA, as described for other PIP-containing proteins (see section 2.1). Studies on *Xenopus* egg cell-free extracts depleted of the frog orthologue of WRN, called FFA-1 (focus-forming activity-1) initially suggested that the protein was required for establishment of replication foci and thus served a central role in DNA synthesis [48]. (Note however that immunoprecipitation from *Xenopus* egg extracts is fraught with difficulties and accidental removal of other components such as membranes may inadvertently lead to loss of replication capacity). Subsequently, FFA-1 was shown to localise to sites of DNA synthesis coincident with RPA, and expression of a dominant negative GST-FFA-1 fusion protein blocked replication activity [49]. Similar immunofluorescence studies in both HeLa cells and primary human fibroblasts, supported by high-resolution immuno-electron microscopy, also showed WRN present at a subset (~60%) of replication foci, colocalising with PCNA [18]. This localisation is in the absence of replication stress, while on HU arrest, the majority of WRN relocates from the nucleolus to RPA-containing foci that are suggested to represent stalled forks [19]. Hence WRN is present at replication sites, and in its absence, cell cycle and DNA synthesis phenotypes are consistent with a replication defect.

## 2. WRN at the replication fork

In order to appreciate where WRN acts during DNA replication, it is necessary to understand the core structure of the DNA replication fork during the elongation stage of DNA replication. During elongation, processive polymerisation of the leading strand is carried out by DNA polymerase epsilon (pol ε) and the leading strand by DNA polymerase delta (pol δ) (based on mutational studies of the proof-reading domains of each in yeast) [50-52]. The replicative polymerases are tethered to the template by association with the homotrimeric sliding clamp protein PCNA (proliferating cell nuclear antigen) [53]. Co-ordination between leading and lagging strands may be achieved through the action of the GINS/Cdc45 complex that has been proposed to act as a replisome progression complex (RPC) [54]. On the lagging strand, repeated cycles of priming by DNA pol α-primase results in synthesis of 7-10 nucleotide of RNA primer followed by ~20 nucleotides of initiator DNA (with error rates of $10^{-2}$ and $10^{-4}$ respectively), followed by switching to the higher fidelity and more processive DNA pol δ on the lagging strand and pol ε on the leading strand. This switch occurs through a multistep loading process essentially requiring recognition of the

primer-template junction (where RPA is bound to the unwound single-stranded parental DNA) by RFC, an AAA+ ATPase that serves to load the sliding clamp PCNA. Pol δ is then recruited to PCNA through its p66 subunit to synthesise approximately 200 nucleotides of the Okazaki fragment. (For a more detailed discussion of fork establishment, see ref. [55]).

## 2.1. Okazaki fragment processing

Because of the low fidelity of pol $\alpha$-primase, it is essential to remove both the RNA primer and iDNA during Okazaki fragment processing (OFP) This is coincident with continued synthesis of nascent DNA on the lagging strand; processive replication by pol δ results in displacement of the RNA-iDNA primer as a 5′ flap and its removal by one of a range of postulated pathways involving RNase H1, FEN1, Dna2 (on long RPA-coated flaps) and other helicases/nucleases including Pif1 and possibly a RecQ helicase (Sgs1 in yeast, WRN in humans) (reviewed in ref. [56]). Pol δ synthesises DNA to fill the gap and DNA ligase seals the nick in the phosphodiester backbone. These steps in Okazaki fragment processing (OFP) may be co-ordinated through differential binding of the separate enzymes to PCNA, which has been suggested to act as a molecular 'toolbelt' in OFP [57]. Association of the OFP proteins[2] with PCNA occurs through a conserved PCNA-interacting peptide (PIP) of the general motif QxxL/M/IxxFF to the hydrophobic pocket of PCNA formed at the interdomain connector loop (e.g. [58, 60], reviewed in ref. [61]). Each PIP is likely to bind by an induced fit mechanism, since the crystal structures of PCNA bound by its various partners shows variation in this loop region [62]. Notably, WRN has a conserved PIP, and peptide ELISA studies showed that this region is sufficient for PCNA binding *in vitro* [18]. Additionally, WRN binds to and stimulates the nuclease activity of Fen1, which may contribute to efficiency of Okazaki fragment processing [63]; as WRN binds to Fen1 immediately adjacent to its PCNA binding site, it is likely that there is some interplay between the three proteins [64] that may be important in Okazaki fragment processing, though this has not been fully explored.

## 2.2. Proof-reading during processive DNA synthesis

DNA replication overall has an extremely low error rate of $10^{-9}$, achieved in part by the very high fidelity of the processive replicative polymerase ε and δ, and also by additional 'extrinsic' proofreading activities together with mismatch repair (MMR) to remove incorrectly incorporated bases. The high fidelity DNA polymerases ε and δ achieve an error rate of ~2 $x10^{-5}$ (reviewed in ref. [65]) through two key structural features. Firstly, the active site is only fully formed upon acceptance of the correct incoming dNTP to create a solvent–inaccessible site that is partially specified by correct helical geometry of duplex DNA, thus increasing enthalpy and decreasing entropy for correct nucleotides and allowing high discrimination over incorrect nucleotides. Secondly, these polymerases each possesses a 3′-5′ exonuclease active site whereby the nascent DNA swings through ~40° to present to this site [66], and where incorrect nucleotides are removed by hydrolysis of the phosphodiester backbone just created. X-ray crystal structures of the isolated WRN exonuclease domain have shown that

---

2 Many other proteins also bind to PCNA in this manner – some regulate PCNA's activity (e.g. p21) [58] while others are regulated by such binding (e.g. Cdt1 degradation is PIP-dependent) [59].

WRN shares structural homology with exonuclease domains of the high fidelity DnaQ family of replication polymerases, suggesting a possible role for WRN in editing DNA, either during DNA synthesis or in processing free ends, in collaboration with and stimulated by the end-binding protein Ku [67]. Very recently, it has been shown that WRN assists pol δ (possibly on the lagging strand during Okazaki fragment synthesis) by removing 3′ mismatches, thus allowing the polymerase to extend primers [68]. This supports a direct role for WRN in Okazaki fragment synthesis.

## 3. Replication fork stalling – the role of WRN

### 3.1. High rates of replication fork stalling in WS

Early electron microscopy studies of $^3$H-T labelled DNA in fibre autoradiographs suggested a problem with replication origin spacing in WS [69, 70], though subsequent higher resolution studies using fluorescent antibodies to halogenated nucleotides suggest rather that it is replication fork rate, not inter-origin distance, which is abnormal in WS cells [17, 44]. Indeed, these DNA combing studies, that analyse individual DNA molecules labelled during replication, have demonstrated a problem with replication fork progression in WS cells, resulting in a high degree of replication fork asymmetry from what should be bidirectional origins [17]. Such studies led to the proposal that replication forks stall at high frequency in cells lacking WRN protein. Why should WS cells be particularly prone to fork stalling?

### 3.2. Causes of fork stalling

The replication fork encounters barriers during normal replication, such as unusual DNA structures arising at G-rich regions (G4-quadruplex) or fragile sites. These structures must be unwound to present a single stranded template suitable for copying; a high incidence of replication fork stalling is likely if the normal mechanisms for tackling the unusual structures is lacking. Alteration in nucleotide pools through treatment with hydroxyurea (HU), or polymerase inhibition with the dCTP mimic aphidicolin results in replication fork arrest in the absence of template abnormalities or lesions. In addition, exogenous agents can cause formation of lesions in the DNA that the replication fork cannot easily pass over – for example, methylated or oxidized bases.

Replication fork pausing or stalling is therefore likely to be a common occurrence, and the cell has mechanisms to stabilise the fork, deal with the unusual structure or repair the damaged region, and allow fork restart. Where DNA synthesis pauses but the MCM replicative helicases proceed to unwind the duplex template, regions of single stranded parental DNA arise, that are rapidly coated with RPA. This forms a signal to the S phase checkpoint machinery, particularly the kinase ATR, that, together with other checkpoint kinases such as Mec1, Chk1 and Chk2 (Rad53) and mediator Mrc1, leads both to recruitment of proteins to deal with the particular fork progression barrier, and to stabilisation of the replisome at the stalled fork, reviewed in ref. [8]. Indeed, DNA pol ε has been shown to stay associated with stalled forks in yeast [71]

under the influence of Rad53 signalling. Replication fork restart then occurs once the damage has been resolved and the checkpoint lifted. More serious to the cell is the collapse of replication forks as they traverse regions of the template containing single strand breaks – single-stranded breaks are converted to double-strand breaks (DSBs) by the passage of the replication fork, forming highly cytotoxic and potentially recombinogenic lesions. Hence surveillance and rescue mechanisms must exist in the cell to deal both with stalled and collapsed forks. The RecQ helicase family has been implicated as key in this mechanism.

## 3.3. Dealing with unusual structures before they arrest the fork

The most efficient mode of replication involves the removal of barriers to fork progression before they lead to fork stalling. Importantly, WRN has been shown to be required by DNA pol $\delta$ (but not $\alpha$ or $\varepsilon$) to unwind G4 DNA [72], bubbles and D loops [68] to allow pol $\delta$-mediated synthesis over such template sequences without leading to fork stalling. In addition, the helicase activity of WRN is also required to limit the formation of single stranded DNA regions and gaps during replication of common fragile sites (CFS) [73, 74] and enhances processivity of DNA pol $\delta$ on fragile site FRA16D over hairpins and microsatellite regions, requiring either the helicase or DNA binding activities of WRN [75]. Hence one important role of WRN in DNA replication is to present the replisome with a template that is easy to replicate, but does it act at any other point to ensure efficient replication?

## 3.4. Is WRN involved in fork restart or progression following restart?

Where replication forks have stalled, replication restart can occur in one of a number of ways: (i) the block may be repaired (or removed); (ii) it may be bypassed using error-prone translesional synthesis (TLS), or (iii) it may be avoided by using an alternative template (e.g. the newly synthesised region on the opposite strand, resulting from fork regression or generated by recombination). The first option is usually the easiest and the least likely to have mutational consequences; translesional synthesis is inherently more likely to cause mutation (pol iota ($\iota$), for example, has an error rate of 0.72 i.e. it incorporates nucleotides almost at random, irrespective of the template sequence [76, 77]), whilst recombination requires a suitable donor template that is not always available. The type of lesion, whether it is on the leading or lagging strand, and the surrounding environment all contribute to how the replication block is dealt with. For example, nucleotide depletion following HU treatment imposes replication stress and can lead to fork stalling, but such stalling may be 'seen' differently by the checkpoint and restart machinery to forks that stall at physical barriers caused by damaging agents such as MMS.

It appears that RecQ helicases may aid in pathway 'choice', although the mechanisms that dictate which pathway is utilised are not fully understood. For instance, yeast complementation studies in *rad50* mutants have demonstrated that BLM is important in resistance to ionising radiation that causes double-strand breaks [78], while WRN confers resistance to drugs such as MMS that lead to replication fork stalling [79, 80]. In human cells, dual labelling of DNA before and after either HU or MMS treatment and analysis by fibre spreading (DNA

combing) has shown that cells acutely depleted of WRN using shRNAi were still able to preserve replisome integrity upon HU- or MMS-induced fork stalling, though following recovery, replication fork rates were slower in WRN-depleted cells than controls, as evidenced by much shorter tracts of labelled DNA post-treatment compared with those synthesised before treatment [44]. It has been proposed [44] that WRN leads to rapid elimination of single-stranded DNA tracts by promoting recombination (using the sister chromatid as template), by enhancing translesion polymerase-mediated gap filling, or by removing DNA immediately after fork passage. It has therefore been suggested that the genome instability in WS results from a defective response to stalled replication forks.

### 3.5. Error-prone translesional synthesis to relieve the replication block

Some lesions such as those caused by MMS or 4NQO present an insurmountable barrier to templating for the high fidelity B family DNA polymerases, but error-prone replication through these small lesions is often less costly for the cell than replication pausing and recruitment of repair complexes. Such error-prone synthesis is conducted by the Y family translesion DNA polymerases (TLS pols). These can pair nucleotides opposite modified and unusual bases, but at the cost of fidelity (ranging from error rates of $\sim 6 \times 10^{-3}$ for pol kappa ($\kappa$), through $3.5 \times 10^{-2}$ for pol eta ($\eta$) to the essentially random 0.72 error rate for pol $\iota$ [76, 77, 81, 82]). The active site of such polymerasis is much larger than that of the proofreading polymerasis, allowing for unusual base pairing geometry, helical distortion of the template DNA, and solvent access [83]. Consistent with an important role for WRN in replication fork progression after pausing, WRN has been found to promote the processivity of Y-family TLS pols on a wide range of substrates including oxidized bases, abasic sites, and thymine dimers [84]. This activity is specific to WRN, and appears to increase the apparent $V_{max}$ of polymerisation. This does not require either catalytic activity of WRN, as proteins with point mutations that ablate both helicase and exonuclease activities can still promote pol $\eta$ polymerisation, although neither catalytically-active BLM nor RecQ5 can substitute [84].

### 3.6. WRN suppresses illegitimate recombination at stalled forks

Whilst the experiments described above strongly support the assertion that WRN is required for fork progression after restart, others have suggested that WRN is itself required to promote restart, possibly through preventing either the accumulation of recombinogenic substrates or in suppressing recombination itself. High levels of spontaneous Rad51 foci in WS cells indicate the presence of an increased number of DNA double-strand breaks (DSBs) and elevated recombination when WRN is absent, supporting the assertion that WRN blocks excessive and illegitimate recombination. Indeed, stalled forks are thought to regress to 'chicken foot; structures with 4-way Holiday junctions that can either be removed by exonuclease degradation of the free ends, by branch migration to a point at which replication can simply restart, or by recombination at the junction (see Figure 2). WRN is likely to suppress the recombinational route, as shown by partial complementation of yeast cells defective in Sgs1 by expression of human WRN. Accumulation and persistence of Holliday junctions is likely, since ectopic expression of the bacterial RusA resolvase allows WS cells to

proliferate as rapidly as control cells, and to resist treatment with CPT or 4NQO (fork collapse and fork stalling agents) to which WS are normally hypersensitive [46].

WRN helicase may branch migrate the chicken foot to 'fold back' the regressed form and thus re-establish a normal fork structure (Figure 2). Indeed, fork regression by WRN on RPA-coated DNA has recently been reported [85]. Alternatively, WRN exonuclease may degrade regions of the chicken foot and allow reformation of a normal replication fork. In addition to its own exonuclease activity, WRN associates with human Exonuclease 1 (Exo1), stimulating its activity [86]. It may therefore be the case that the two nuclease activities combine to remove regressed forks. It has been suggested that in the absence of WRN, the recombinational route is used to process the accumulated HJs, and that this requires the action of the nuclease Mus81; fission yeast Rqh1 suppresses Mus81 mutation [42] and human WRN suppresses Mus81-mediated recombination [87].
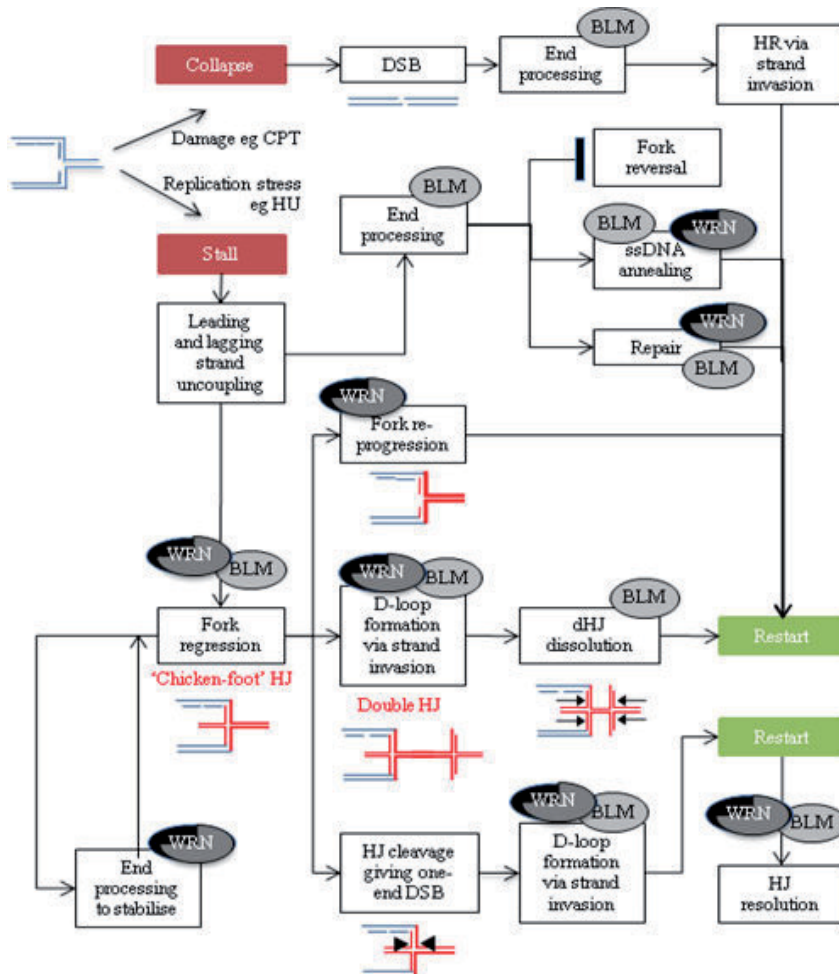


**Figure 2.** Possible roles of WRN in replication restart after fork stalling (see text for details)

### 3.7. Template switching at stalled forks

Leading strand blockage often uncouples the replicative helicases from the rest of the replisome, allowing significant unwinding to form long tracts of single-stranded DNA, with lagging strand synthesis continuing for a distance [88]. The accumulated long single stranded loop of leading strand DNA is highly susceptible to damage. Replication fork restart on the leading strand might simply utilise new priming by RPA-mediated recruitment of pol $\alpha$-primase to the region of transition between singe stranded and duplex DNA (i.e. where the previous polymerase ceased synthesis), in much the same way that it normally reassociates with the primer-template junction in Okazaki fragment synthesis. Alternatively, regression of the replication fork may permit annealing to the new lagging strand using 'template switching' to give a Holliday junction that can then be reversed past the lesion [89, 90]. In bacteria, this can be done by RecQ helicase, with RecJ exonuclease to remove the protruding lagging strand flap [91].

In mammals this is likely to require WRN and the flap endonuclease activity of FEN-1 [92]. WRN (and BLM) can induce fork regression over the lesion by local unwinding, and can lead to the formation of the chicken-foot. WRN can also reverse a regressed fork. Both BLM and WRN helicase activities can also catalyse branch migration of the DNA leading to recovery of the template daughter strand annealing via Rad51 [93], formation of a double Holliday junction and strand exchange. If the product here is a hemicatenenes, it can be resolved into either a chicken foot or a HJ and processed the same way. Ultimately, functional replication forks may be reformed [94]. Alternatively, the Holliday junction can then be cleaved by a resolvase and DSB repair as before. See Figure 2 (above) for a schematic of replication fork restart.

### 3.8. How is WRN recruited to stalled forks?

Stalling of replication forks initiates the caffeine-sensitive S phase checkpoint, mediated by RPA, ATR and Rad53. WRN recruitment to, or retention at, stalled forks may be direct through binding to RPA [85], but it also appears to require phosphorylation by the checkpoint kinase ATR [95]. When such phosphorylation is prevented, WRN cannot accumulate at repair sites and DNA strand breaks are detected [73]. That WRN is an *in vivo* as well as *in vitro* target of ATR has been confirmed by phosphoproteomic studies [96]. However, it is still the subject of research and debate as to whether WRN is an upstream sensor or downstream effector in the S phase checkpoint that responds to replication stress or stalled forks. For example, shRNAi-mediated WRN knockdown abrogated the S phase checkpoint on CPT treatment but did not affect checkpoint induction on HU exposure [97], suggesting that WRN may be an important 'sensor' of collapsed but not stalled forks, although the mechanism has yet to been defined. Perhaps fork collapse (e.g. upon CPT treatment) requires ATM, with its double-strand break sensing activity through recruitment by Ku and activation by DNA-PKcs (DNA-dependent protein kinase catalytic subunit), while fork stalling (e.g. on HU) uses the ATR pathway. This is consistent with differential regulation of WRN by the two kinases [73], and with a requirement for WRN not only in replication fork progression after stalling (see above) but also in directing recombination in concert with RAD51

and RAD54 [93]. Recently, it has been shown that WRN also interacts with the repair sliding clamp 9-1-1 (homologous structurally and functionally to PCNA, though acting in repair rather than replication), and that upon fork arrest, the 9-1-1 complex recruits TopBP1 that in turn recruits ATR which phosphorylates WRN [98]. Perhaps the initial type of damage that leads to fork arrest is therefore a deciding factor in the pathways of WRN recruitment and post-translational modification.

### 3.9. Role of WRN at stalled forks on the lagging strand

Lagging strand blocks do not uncouple the replication fork; rather, lagging strand polymerase merely stutters to the next primer to restart synthesis of the next Okazaki fragment [99, 100]. The resulting single-stranded gap is repaired by translesional synthesis as above (which may be error-prone) or by homologous recombination with the sister chromatid (which is more likely to retain fidelity). In *E. coli*, this requires formation of a double Holliday junction and resolution via non-crossover [101]. In mammals, BLM has the ability to mobilise double Holliday junctions and the resulting catenated DNA is resolved by topoisomerase III without crossover [102]. WRN does not interact with TopoIII and cannot migrate a double Holliday junction [103], although structures involved in intermediate formation (D-loops, G-quadruplex) might require either WRN or BLM. WRN can process a mobile D-loop [104] using the co-ordinate action of both helicase and exonuclease.

However, WRN is also linked to the functionality of the lagging strand polymerase, pol δ. WRN stimulates the base incorporation of pol δ (but not α or ε) even in the absence of PCNA [105]. Pol δ is slowed at fragile sites and repetitive runs likely to cause hairpin or bubble structures, but this can be alleviated by the helicase functionality of WRN [75]. Like WRN, pol δ has a 3′-5′ exonuclease capability which it can use to proofread bases after insertion [106]. WRN can substitute at this proofreader, and cells with low levels of WRN show increased mutation of the lagging strand [107]. Interestingly, the exonuclease activity of pol δ is active on WRN-preferred DNA substrates such as Holliday junctions, D-loops and bubble duplex, and can form a complex with WRN [107] that increases the degradation of these substrates. WRN exonuclease is blocked by many common lesions [22, 108]; it will be interesting to find out whether the nuclease activity of pol δ is complementary to this, and might suggest why the two would functionally substitute within the lagging strand complex.

Ultimately, fork restart requires proximal repositioning of the replication complex; this remodelling may make use of WRN nuclease activities to further process DNA ends and allow removal of damage. Interactions with PCNA and either strict (pol δ, pol α) or promiscuous (TLS pathway) repair polymerases and FEN1 flap removal activity can allow bypass of nicks and modified DNA bases at the same time as restart positioning, allowing many lesions to be handled.

# 4. Involvement of WRN in telomere maintenance

## 4.1. Telomere structure and replication

Mammalian telomeres consist of a few kilobases of repetitive non-coding G-rich sequence (the human sequence is (TTAGGG)n) which must be 'capped' rather like a bootlace in order to stop the DNA end being recognised as a DSB via p53/p21 signalling [109] and instigating profligate double-strand break (DSB) repair [110]. Functional capping forms a lasso-like structure [111] called the telomere-loop (T-loop) where the repetitive telomere sequence folds back upon itself to displace a short segment of proximal sequence with a 3' single stranded end to give a displacement-loop (the D-loop)[112]. The proteins that make up the telosome (or core shelterin complex [112]) include TRF1 and TRF2 [113], which bind and stabilise telomeric duplex DNA at the T-loop [114], and POT1 [115], a DNA-binding protein which coats and protects the tracts of single stranded telomeric sequence that occur at the telomeric D-loop and during telomeric replication and processing. Figure 3 shows the T and D loop structure with associated proteins.

Telomeres are replicated by passage of a replication fork that initiated upstream of the chromosome end: obviously it is not possible to load the replisome or prime DNA synthesis beyond the end of the chromosome. At each round of replication, the telomeric sequence is unwound from the D (and possibly also the T) loops, and passively replicated by an incoming fork. While early reports suggested that priming on the lagging strand was defective at the very end of the chromosome, it has become apparent that both leading and lagging strands are normally replicated but that regeneration of a 3' overhang for strand invasion to form the D loop involves end resection of the leading strand, thus removing sequence information and shortening the telomere at each round of replication.

## 4.2. Telomere shortening leads to replicative senescence and genome instability

Telomere shortening acts as a counting mechanism to indicate the number of cell divisions a somatic cell has passed through, and normal fibroblasts generally arrest at the Hayflick limit of 55-60 population doublings [3] under the influence of this telomere attrition. Hence cellular ageing is in a large part caused by progressive telomere loss – cells that lose telomeres more rapidly senesce more quickly that those with long telomeres, and people with prematurely short telomeres (e.g. mothers of chronically sick children[116], carers of partners with dementia and low paid workers experiencing
work-related stress) age prematurely [117, 118]. (Note that this is not the case in mice, where lab strains have extremely long telomeres and cells senesce prior to telomeres reaching a critical length).

To overcome this cellular ageing, it is vital that immortal cells such as those of the germline have a mechanism to restore telomeric DNA at every round of replication. Such cells express active telomerase, a reverse transcriptase which utilises its endogenous RNA template to regenerate telomeric sequence [119], but telomerase levels are extremely low or absent in most somatic cells [120]. Notably, immortalisation of cancer cells is accompanied by re-expression

of telomerase [121] in about 85% of all human cancers, while the remaining 15% are able to maintain their telomere lengths in the absence of telomerase, by alternatives mechanisms, reviewed in ref. [122] (see section 4.6).
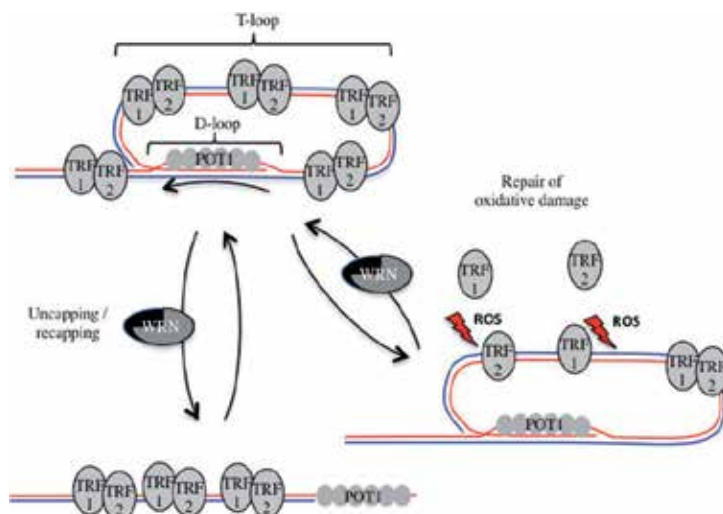


**Figure 3.** The structure of the telomere, showing the large telomere (T) loop and the smaller displacement (D) loop. Proteins TRF1, TRF2 and POT1 are critically important in stabilising the telomeric structure. WRN binds to all of these proteins

Dysfunctional telomeres that become uncapped are liable to degradation or immediate repair by homologous recombination (HR) or non-homologous end-joining (NHEJ), the latter causing chromosome fusions that are usually catastrophic for the cell. However, the tightly capped telomere cannot serve as a template during replication, so regulated disassembly of the shelterin complex and unwinding of the D (and possibly T) loop is necessary for efficient copying of telomeric regions. The transient uncapping that occurs during replication is recognised by repair proteins as DNA damage [123], and the correct reformation of the T-loop requires correct handling and processing by repair enzymes. Uncontrolled uncapping is therefore a powerful cause of genomic instability, and loss of telomeres shortens replicative lifespan; both are hallmarks of WS.

### 4.3. Are telomeres defective in WS?

The major clinical characteristics of WS are premature ageing, presumably resulting from the highly premature replicative senescence, and elevated cancer risk, which is caused by excess genome instability. Since replicative senescence is caused, at least in part, by telomere shortening, and chromosome fusions result from telomere loss, it has been of major importance to determine whether telomeres are indeed defective in WS cells, and whether WRN plays any role in telomere maintenance. Human WS cells in culture show elevated rates of telomere loss [124]. Contradictory to this, however, are data from single telomere length analysis (STELA) that suggest WS cells do not experience exceptional rates of telomere

shortening, at least in clonal populations, though in bulk cultures of WS fibroblast, telomere loss ranges from a normal 99bp/PD to a four fold increase at 355 bp/PD [125].

Support for the importance of telomeric dysfunction in WS replicative senescence comes from studies of mouse models that are null for WRN. However, mice lacking WRN do not exhibit the premature ageing symptoms seen in humans [126] because laboratory mouse strains possess much longer telomeres than humans (40-80kb compared to 2-10kb) and detectable levels of telomerase even in somatic cells [127]. When mice deficient in telomerase are bred for several generations to reduce their telomere lengths to that approaching the normal human mean, removal of WRN gives similar premature ageing characteristics to those seen in human WS [128, 129]. Crucially, later-generation telomerase-null mice that still retain longer telomeres do not show this phenotype even though premature senescence is seen in their littermates that have short telomeres. Hence short telomeres combined with lack of WRN results in premature ageing.
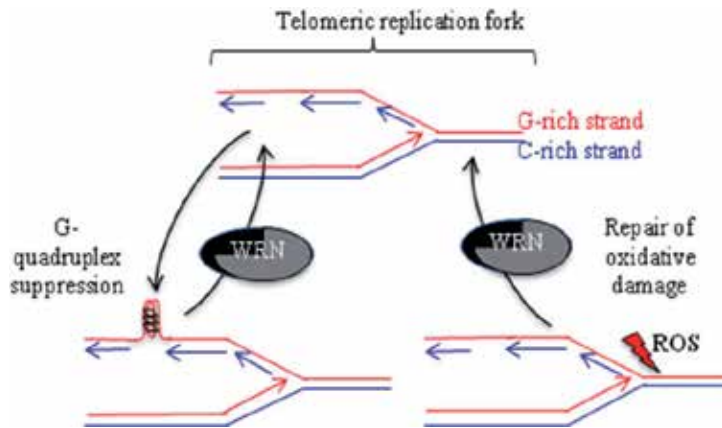


**Figure 4.** Roles of WRN at the telomere include unwinding of G4 DNA, that would otherwise lead to replication fork stalling, and repair of oxidative damage to which the telomeric DNA is exquisitely sensitive.

### 4.4. WRN helicase and exonuclease co-operate at the telomere

The repetitive nature of telomeric DNA arises as a consequence of the short RNA template within telomerase; this, combined with the G-rich nature leads to these sequences forming secondary structures called G-quadruplexes, which stall replication machinery much as any bulky lesion or DNA gap or break will. It is therefore essential for cells to unwind telomeric DNA ahead of the replication fork to prevent stalling, or worse, collapse. As discussed above, D-loops, recombination intermediates and G-quadruplexes may all require WRN and other RecQ helicases to remove these blockages (Figure 4). Under experimental conditions *in vitro*, WRN localises to a sub-set of telomeres during S-phase without the induction of stress [29], and is enriched when cells are subjected to damaging agents that cause replication stress such as CPT. Thus WRN catalysis is needed to police both endogenous replication fork blocks and induced damage.

WRN interacts with many of the proteins making up the shelterin complex or telosome [130-133] (see also Figure 3, above). Such interactions are likely to have functional consequences: for example, POT1 stimulates WRN helicase activity on linear and D-loop structures *in vitro* [134], whilst the presence of TRF1 and TRF2 can modulate their activity. TRF2 recruits WRN to D-loops and therefore stimulates unwinding [134], but it inhibits the helicase activity of WRN if binding to telomeric HJ substrates [135].

There are fewer pathways for replication fork recovery at telomeric ends because of the lack of downstream origins [136]. This obviously increases the need for proteins such as WRN that can dissolve or resolve replication blocks and promote fork progression before irreversible fork collapse occurs. One such block is G-quadruplex DNA: it has been shown to stall the major replicative polymerase δ [72]. G-quadruplex structures can arise spontaneously in single-stranded telomeric sequence [137] and can be suppressed by the binding of POT1 to release single stranded telomeric sequence during uncapping [138]. WRN preferentially unwinds G-quadruplex DNA [139] and its presence will suppress polymerase δ stalling [72], suggesting it is a good candidate for this role.

Interestingly, WRN – the only human RecQ helicase to also have exonuclease activity – unwinds D-loops *in vitro* in the absence of other proteins, using co-ordinate activity of both its helicase and exonuclease functions (RecQ helicase activity on these substrates is not particularly processive without stimulation for example by RPA [140]). The catalytic subunit of DNA-PK has also been shown to interact with WRN at telomeres [141], acting to suppress its exonuclease function and allow longer tracts to be unwound by the helicase activity. Therefore in the presence of DNA-PKcs, WRN processing of telomeric DNA does not shorten telomeric ends.

## 4.5. WRN acts on the lagging strand during telomere replication

Despite these detailed studies, the exact catalytic role(s) of WRN in telomere maintenance are still not fully defined. There is good evidence that cells lacking WRN have defective lagging strand synthesis at the telomere [142], as metaphase chromosomes in WRN helicase-deficient cells show a characteristic (if low-level) loss of telomeric sequence on one but not both sister chromatids. This is called sister telomere loss (STL), and suggests dysfunctional processing of one strand of the telomere during replication. The sister telomere lost is always the one resulting from lagging strand synthesis [142] This phenotype is thought to arise because in the absence of WRN activity, G-quadruplexes accumulate in the G-rich template strand and cause failure of lagging-telomere replication. Expression of active (but not inactive) telomerase suppresses STLs in cells lacking WRN [142], suggesting that sister telomere loss occurs during WRN-dependent processing of telomeres at times other than normal S phase when telomeres are uncapped for replication elongation.

The low levels of STL that occur (if the experimental data from chromosomal FISH reflect the underlying levels) suggest that the events that cause the telomere loss might be difficult to process or close to irreparable, or are merely rare; they might be alternatively-processed in a pathway that does not induce loss. Conversely, the catalytic activity supplied by WRN might be substituted by other enzymes – its helicase role by one of the other RecQs, or its

exonuclease function by another appropriate 3′-5′ exonuclease such as ExoI [143]. Whilst this has not yet been determined, however it is notable that cells deficient in BLM also show telomeric defects, although these are not the end-fusions arising from DSB repair as seen with WRN, but seem to be catenated associations possibly from aberrant HR [131]. BLM may thus have a role in resolving late-replicating DNA intermediates at telomeres distinct from WRN, as the rate of telomere dysfunction seen in cells with either single-null genotype is exacerbated in a double null [144].

Since the processing of Okazaki fragments during lagging strand synthesis gives rise to regions of ssDNA, and G-rich sequences have a tendency to form G-quadruplex structures spontaneously, at the telomere there is increased likelihood of G-quadruplex formation in the single-stranded tracts. POT1 cannot actively dissociate the structure by binding, strongly suggesting that the G-quadruplex must first be dissociated before POT1 can bind and protect the telomeric sequence, and implicating a role for WRN in removing the replication block before problems arise (see review [145]). Interestingly, the available levels of POT1 may modulate the coupling of the leading and lagging strands at telomeres in the absence of WRN, allowing uncoupled synthesis of leading strand without processing of the lagging strand block [146].

Supporting this hypothesis, recent research in yeast suggests an alternative protein that may function to suppress G-quadruplex formation at telomeres, but this time on the leading strand. Pif1 is a 5′-3′ helicase that negatively regulates telomere length [147]. Loss of Pif1 leads to slow replication fork progression, and *in vitro* Pif1 can unwind replication substrates [148]. Recently it was shown that cells without Pif1 have chromosome breakage at sites of G-tracts, and Pif1 can unwind G-quadruplex DNA that forms in the leading strand [149]. The higher eukaryote *C. elegans* also possesses a helicase (DOG-1) that is able to inhibit loss of guanine tracts, presumably by suppression of G-quadruplex structures [150]. It is tempting to speculate that genome surveillance utilises Pif1 on the leading strand and WRN on the lagging strand to suppress G-quadruplex formation and subsequent replication fork blockage at sites of high guanine content such as fragile sites and telomeric sequence.

The loss of WRN in this putative mechanism inherently implies loss specifically of lagging-strand DNA at the telomeres. In this model, the rarity of STL may be explained by a low rate of G-quadruplex formation at single-stranded telomeric tracts during Okazaki fragment replication, the ability of BLM (or another RecQ) to substitute for WRN, or the specific need for WRN in a small subset of these events – perhaps because exonuclease processing is also required. The WRN exonuclease activity is itself specifically implicated in processing of the 3′-end of the telomere, although other nucleases such as ExoI or perhaps FEN-1 [151] might possess the capability to substitute for WRN. Addition of exogenous DNA oligonucleotides homologous to the 3′-overhang structure of an uncapped telomeric end to cells lacking WRN results in an increase in DNA damage responses and ultimately cell senescence [152].

The loss of WRN in telomerase-positive cells *in vivo* causes the generation of extrachromosomal telomeric structures [153], [154] and this requires both helicase and exonuclease activities. WRN has exonuclease activity here that requires telomeric sequence in both double-stranded and single-stranded portions, and shows a characteristic limited degradation

pattern [155]. TRF2 recruits WRN to telomeric sequence and *in vitro* it synergistically enhances the ability of WRN to degrade the G-rich 3'-overhangs of telomeric D-loops substrates [132, 156]. POT1 inhibits WRN exonuclease activity here [155]. TRF2 or WRN alone exhibit little or no stimulation on these substrates. Non-telomeric substrates show similarly little WRN-dependent degradation, presumably because TRF2 does not bind/recruit and stimulate WRN exonuclease, whilst TRF2 bound to telomeric sequence completely inhibits the activity of other nucleases such as ExoIII [132]. WRN helicase and exonuclease, together with TRF2, POT1, and Ku therefore probably act together to prevent telomeric free ends from becoming substrates for HR or other aberrant pathways. Taken together, these results support the specificity of WRN exonuclease in reducing the length of the telomeric 3'-end to the optimal length for regeneration of the T-loop after replication, and suppression of extrachromosomal telomeric circles.

## 4.6. WRN may be important in ALT

Telomeres can be lengthened without the use of telomerase using recombination to generate the template DNA needed. In yeast, this ALT[3] pathway requires Sgs1 (the RecQ homologue in *S. cerevisiae*), for which WRN and BLM may both partially substitute [131, 157, 158]. Both WRN and BLM have also been seen to interact with telomeric DNA in human cells that utilize the ALT pathway [29, 130, 131], albeit only a small proportion. Although the ALT pathways are not yet elucidated, most models suggest recombinational mechanisms where strand invasion into telomeric DNA of the same (or different) chromosome or chromatid is utilized as template for resynthesis (e.g. see [159, 160]). BLM-deficient cells show elevated rates of sister chromatid exchange [161] that were not detected in cells lacking WRN, however finer resolution experiments suggest that WS cells do show elevated SCE, but only at telomeres [162, 163]. The WS mouse models with shortened telomeres (described in section 4.3 above) show elevated levels of this telomere-specific SCE [164], as do cells deficient in POT1, or Ku and TRF2 together [165]. Ku stimulates both helicase and exonuclease activities of WRN [166, 167], and suppresses telomeric recombination brought on by the absence of TRF2 and consequent telomeric uncapping [110]. Taken together, these data suggest that WRN is prominent in a pathway that specifically suppresses telomeric recombination or dissolves junctions, and it is at least partially distinct from the role of BLM.

## 4.7. Telomeric DNA is hypersensitive to oxidative lesions – a further role for WRN

The G-rich nature of telomeric sequence means it is a rich target for oxidative damage[4] [168], and oxidative stress and mitochondrial dysfunction often give rise to concomitant telomeric dysfunction [169], which can be reduced using antioxidants. Notably, artificial replicative senescence can be induced with a burst of oxidative damage [170, 171]. Oxidation of telomeric bases can disrupt DNA binding of TRF1 and TRF2, and presumably therefore telo-

---

3 ALT = alternative lengthening of telomeres

4 8-oxoG is a common product of oxidative attack of DNA

some and T-loop assembly [166], whilst over-expression of TRF2 protect cells with shortened telomeres from early senescence [172].

WRN is a central component of base excision repair (BER) of oxidative lesions, interacting with most of the key proteins in the pathway such as pol beta (β) and FEN1 [173, 174]. Consistent with an important role for WRN in removing oxidative lesions, WS cells show increased oxidative damage [175, 176].

It has been shown that D-loops containing oxidised bases can be bound by POT1 and are a preferred substrate for WRN [177]. The strand-displacement activity of pol β, the repair polymerase in BER, is also stimulated by TRF2 [178], and TRF1, TRF2 and POT1 can enhance all the constituent steps of long patch BER [179]. As previously mentioned, WRN itself can also stimulate TLS pols to replicate past an oxidative block [84]. This suggests active recruitment and stimulation of anti-oxidative damage processes at telomeres involving RecQ helicases. These findings partly illustrate how the activities of RecQ helicases are tightly controlled by the surrounding milieu in order to differentiate their roles in replication and repair.

Although the wider significance of all these data is yet to be determined, it is obvious that WRN is active at multiple points in telomere replication and repair.

## 5. Conclusions

The helicase/exonuclease WRN has been shown to be critically important in DNA replication, acting to enhance fidelity, regulate template unwinding to prevent fork stalling at unusual structures, assist with replication fork restart and/or enhance processivity post-restart, aid translesion synthesis over otherwise unreplicatable lesions, promote regression of stalled replication forks to allow error-free restart, modulate recombination at collapsed replication forks, and aid telomere replication. It is recruited to sites of DNA synthesis, possibly through association with the sliding clamp PCNA, and to sites of stalled/collapsed forks probably by RPA in concert with the S phase checkpoint kinase ATR and its downstream effectors and mediators Chk1, Rad53, Mec1 and Mrc1. Loss of WRN results in high levels of chromosomal instability and elevated cancer risk, and the defects in DNA replication on WRN loss also results in premature onset of replicative senescence with concomitant organismal ageing, manifest as progeroid Werner's syndrome. While much has been discovered as to WRN's mode of action, there is still an enormous amount to learn as to how its activities are co-ordinated with the cell during DNA replication.

## Acknowledgements

## Author details

Lynne S. Cox[*] and Penelope A. Mason

*Address all correspondence to: Lynne.cox@bioch.ox.ac.uk

Department of Biochemistry, University of Oxford, South Parks Road, Oxford

## References

[1] Larizza, L., Roversi, G., & Volpi, L. Rothmund-Thomson syndrome. *Orphanet J Rare Dis*, vol. 5, 2.

[2] Goto, M. (2001). Clinical characteristics of Werner syndrome and other premature aging syndromes: pattern of aging in progeroid syndromes *From premature gray hair to helicase- Werner syndrome: implications for aging and cancer, M. Goto and R. W. Miller eds., Tokyo: Japan Scientific Societies Press*, 27-39.

[3] Hayflick, L. (1979). The cell biology of aging. *J Invest Dermatol*, 73(1), 8-14.

[4] Kyng, K. J., May, A., Kolvraa, S., & Bohr, V. A. (2003). Gene expression profiling in Werner syndrome closely resembles that of normal aging. *Proc Natl Acad Sci U S A*, 100(21), 12259-12264.

[5] Vanhooren, V., Desmyter, L., Liu, X. E., Cardelli, M., Franceschi, C., Federico, A., Libert, C., Laroy, W., Dewaele, S., Contreras, R., & Chen, C. (2007). N-glycomic changes in serum proteins during human aging. *Rejuvenation Res*, 10(4), 521-531a.

[6] Fukuchi, K., Martin, G. M., & Monnat, R. J. (1989). Mutator phenotype of Werner syndrome is characterized by extensive deletions. *Proc Natl Acad Sci U S A*, 86(15), 5893-5897.

[7] Scappaticci, S., Cerimele, D., & Fraccaro, M. (1982). Clonal structural chromosomal rearrangements in primary fibroblast cultures and in lymphocytes of patients with Werner's Syndrome,. *Hum Genet*, 62(1), 16-24.

[8] Segurado, M., & Tercero, J. A. (2009). The S-phase checkpoint: targeting the replication fork. *Biol Cell*, 101(11), 617-627.

[9] Agrelo, R., Cheng, W. H., Setien, F., Ropero, S., Espada, J., Fraga, M. F., Herranz, M., Paz, M. F., Sanchez-Cespedes, M., Artiga, M. J., Guerrero, D., Castells, A., von, C., Kobbe, V. A., & Esteller, M. (2006). Epigenetic inactivation of the premature aging Werner syndrome gene in human cancer. *Proc Natl Acad Sci U S A*, 103(23), 8822-8827.

[10] Ding, S. L., Yu, J. C., Chen, S. T., Hsu, G. C., & Shen, C. Y. (2007). Genetic variation in the premature aging gene WRN: a case-control study on breast cancer susceptibility. *Cancer Epidemiol Biomarkers Prev*, 16(2), 263-269.

[11] Bird, J. L., Jennert-Burston, K. C., Bachler, M. A., Mason, P. A., Lowe, J. E., Heo, S. J., Campisi, J., Faragher, R. G., & Cox, L. S. (2012). Recapitulation of Werner syndrome sensitivity to camptothecin by limited knockdown of the WRN helicase/exonuclease. *Biogerontology*, 13(1), 49-62.

[12] Futami, K., Takagi, M., Shimamoto, A., Sugimoto, M., & Furuichi, Y. (2007). Increased chemotherapeutic activity of camptothecin in cancer cells by siRNA-induced silencing of WRN helicase. *Biol Pharm Bull*, 30(10), 1958-1961.

[13] Aggarwal, M., Sommers, J. A., Shoemaker, R. H., & Brosh, R. M. (2011). Inhibition of helicase activity by a small molecule impairs Werner syndrome helicase (WRN) function in the cellular response to DNA damage or replication stress. *Proc Natl Acad Sci U S A*, 108(4), 1525-1530.

[14] Yu, C. E., Oshima, J., Fu, Y. H., Wijsman, E. M., Hisama, F., Alisch, R., Matthews, S., Nakura, J., Miki, T., Ouais, S., Martin, G. M., Mulligan, J., & Schellenberg, G. D. (1996). Positional cloning of the Werner's syndrome gene,. *Science*, 272(5259), 258-262.

[15] von Kobbe, C., & Bohr, V.A. (2002). A nucleolar targeting sequence in the Werner syndrome protein resides within residues 949-1092. *J Cell Sci*, 115, 3901-3907.

[16] Marciniak, R. A., Lombard, D. B., Johnson, F. B., & Guarente, L. (1998). Nucleolar localization of the Werner syndrome protein in human cells. *Proc Natl Acad Sci U S A*, 95(12), 6887-6892.

[17] Rodriguez-Lopez, A. M., Jackson, D. A., Iborra, F., & Cox, L. S. (2002). Asymmetry of DNA replication fork progression in Werner's syndrome. *Aging Cell*, 1(1), 30-39.

[18] Rodriguez-Lopez, A. M., Jackson, D. A., Nehlin, J. O., Iborra, F., Warren, A. V., & Cox, L. S. (2003). Characterisation of the interaction between WRN, the helicase/exonuclease defective in progeroid Werner's syndrome, and an essential replication factor, PCNA. *Mech Ageing Dev*, 124(2), 167-174.

[19] Constantinou, A., Tarsounas, M., Karow, J. K., Brosh, R. M., Bohr, V. A., Hickson, I. D., & West, S. C. (2000). Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest. *EMBO Rep*, 1(1), 80-84.

[20] Huang, S., Li, B., Gray, M. D., Oshima, J., Mian, I. S., & Campisi, J. (1998). The premature ageing syndrome protein, WRN, is a 3'-->5' exonuclease. *Nat Genet*, 20(2), 114-116.

[21] Huang, S., Beresten, S., Li, B., Oshima, J., Ellis, N. A., & Campisi, J. (2000). Characterization of the human and mouse WRN 3'-->5' exonuclease. *Nucleic Acids Res*, 28(12), 2396-2405.

[22] Harrigan, J. A., Fan, J., Momand, J., Perrino, F. W., Bohr, V. A., & Wilson, D. M. (2007). WRN exonuclease activity is blocked by DNA termini harboring 3′ obstructive groups. *Mech Ageing Dev*, 128(3), 259-266.

[23] Ozgenc, A., & Loeb, L. A. (2005). Current advances in unraveling the function of the Werner syndrome protein. *Mutat Res*, 577(1-2), 237-251.

[24] Xue, Y., Ratcliff, G. C., Wang, H., Davis-Searles, P. R., Gray, M. D., Erie, D. A., & Redinbo, M. R. (2002). A minimal exonuclease domain of WRN forms a hexamer on DNA and possesses both 3′- 5′ exonuclease and 5′-protruding strand endonuclease activities. *Biochemistry*, 41(9), 2901-2912.

[25] Machwe, A., Xiao, L., & Orren, D. K. (2006). Length-dependent degradation of single-stranded 3′ ends by the Werner syndrome protein (WRN): implications for spatial orientation and coordinated 3′ to 5′ movement of its ATPase/helicase and exonuclease domains. *BMC Mol Biol*, 7, 6.

[26] Gray, M. D., Shen, J. C., Kamath-Loeb, A. S., Blank, A., Sopher, B. L., Martin, G. M., Oshima, J., & Loeb, L. A. (1997). The Werner syndrome protein is a DNA helicase. *Nat Genet*, 17(1), 100-103.

[27] Brosh, R. M., Waheed, J., & Sommers, J. A. (2002). Biochemical characterization of the DNA substrate specificity of Werner syndrome helicase. *J Biol Chem*, 277(26), 23236-23245.

[28] Shen, J., & Loeb, L. A. (2001). Unwinding the molecular basis of the Werner syndrome. *Mech Ageing Dev*, 122(9), 921-944.

[29] Opresko, P. L., Otterlei, M., Graakjaer, J., Bruheim, P., Dawut, L., Kolvraa, S., May, A., Seidman, M. M., & Bohr, V. A. (2004). The Werner syndrome helicase and exonuclease cooperate to resolve telomeric D loops in a manner regulated by TRF1 and TRF2. *Mol Cell*, 14(6), 763-774.

[30] Kitano, K., Kim, S. Y., & Hakoshima, T. (2010). Structural basis for DNA strand separation by the unconventional winged-helix domain of RecQ helicase WRN. *Structure*, 18(2), 177-187.

[31] Kitano, K., Yoshihara, N., & Hakoshima, T. (2007). Crystal structure of the HRDC domain of human Werner syndrome protein, WRN. *J Biol Chem*, 282(4), 2717-2728.

[32] Hartung, F., & Puchta, H. (2006). The RecQ gene family in plants. *J Plant Physiol*, 163(3), 287-296.

[33] Opresko, P. L., Cheng, W. H., & Bohr, V. A. (2004). Junction of RecQ helicase biochemistry and human disease. *J Biol Chem*, 279(18), 18099-18102.

[34] Cox, L. S., Clancy, D. J., Boubriak, I., & Saunders, R. D. (2007). Modeling Werner Syndrome in Drosophila melanogaster: hyper-recombination in flies lacking WRN-like exonuclease. *Ann N Y Acad Sci*, 1119, 274-288.

[35]  Boubriak, I., Mason, P. A., Clancy, D. J., Dockray, J., Saunders, R. D., & Cox, L. S. (2009). DmWRNexo is a 3'-5' exonuclease: phenotypic and biochemical characterization of mutants of the Drosophila orthologue of human WRN exonuclease. *Biogerontology*, 10(3), 267-277.

[36]  Mason, P. A., Boubriak, I., Robbins, T., Lasala, R., Saunders, R., & Cox, L. S. (2012). The Drosophila orthologue of progeroid human WRN exonuclease, DmWRNexo, cleaves replication substrates but is inhibited by uracil or abasic sites : Analysis of DmWRNexo activity in vitro. *Age (Dordr)*.

[37]  Saunders, R. D., Boubriak, I., Clancy, D. J., & Cox, L. S. (2008). Identification and characterization of a Drosophila ortholog of WRN exonuclease that is required to maintain genome integrity. *Aging Cell*, 7(3), 418-425.

[38]  Ketting, R. F., Haverkamp, T. H., van Luenen, H. G., & Plasterk, R. H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell*, 99(2), 133-141.

[39]  Lee, S. J., Yook, J. S., Han, S. M., & Koo, H. S. (2004). A Werner syndrome protein homolog affects C. elegans development, growth rate, life span and sensitivity to DNA damage by acting at a DNA damage checkpoint. *Development*, 131(11), 2565-2575.

[40]  Lee, S. J., Gartner, A., Hyun, M., Ahn, B., & Koo, H. S. (2010). The Caenorhabditis elegans Werner syndrome protein functions upstream of ATR and ATM in response to DNA replication inhibition and double-strand DNA breaks. *PLoS Genet*, 6(1), e1000801.

[41]  Hartung, F., Plchova, H., & Puchta, H. (2000). Molecular characterisation of RecQ homologues in Arabidopsis thaliana. *Nucleic Acids Res*, 28(21), 4275-4282.

[42]  Doe, C. L., Ahn, J. S., Dixon, J., & Whitby, M. C. (2002). Mus81-Eme1 and Rqh1 involvement in processing stalled and collapsed replication forks. *J Biol Chem*, 277(36), 32753-32759.

[43]  Poot, M., Hoehn, H., Runger, T. M., & Martin, G. M. (1992). Impaired S-phase transit of Werner syndrome cells expressed in lymphoblastoid cell lines. *Exp Cell Res*, 202(2), 267-273.

[44]  Sidorova, J. M., Li, N., Folch, A., & Monnat, R. J. (2008). The RecQ helicase WRN is required for normal replication fork progression after DNA damage or replication fork arrest. *Cell Cycle*, 7(6), 796-807.

[45]  Fujiwara, Y., Higashikawa, T., & Tatsumi, M. (1977). A retarded rate of DNA replication and normal level of DNA repair in Werner's syndrome fibroblasts in culture. *J Cell Physiol*, 92(3), 365-374.

[46]  Rodriguez-Lopez, A. M., Whitby, M. C., Borer, C. M., Bachler, M. A., & Cox, L. S. (2007). Correction of proliferation and drug sensitivity defects in the progeroid Werner's Syndrome by Holliday junction resolution. *Rejuvenation Res*, 10(1), 27-40.

[47] Lebel, M., Spillare, E. A., Harris, C. C., & Leder, P. (1999). The Werner syndrome gene product co-purifies with the DNA replication complex and interacts with PCNA and topoisomerase I. *J Biol Chem*, 274(53), 37795-37799.

[48] Yan, H., & Newport, J. (1995). FFA-1, a protein that promotes the formation of replication centers within nuclei. *Science*, 269(5232), 1883-1885.

[49] Chen, C. Y., Graham, J., & Yan, H. (2001). Evidence for a replication function of FFA-1, the Xenopus orthologue of Werner syndrome protein. *J Cell Biol*, 152(5), 985-996.

[50] Garg, P., & Burgers, P. M. (2005). DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit Rev Biochem Mol Biol*, 40(2), 115-128.

[51] Pursell, Z. F., Isoz, I., Lundstrom, E. B., Johansson, E., & Kunkel, T. A. (2007). Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science*, 317(5834), 127-130.

[52] McCulloch, S.D., & Kunkel, T.A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res*, 18(1), 148-161.

[53] Podust, V. N., Tiwari, N., Stephan, S., & Fanning, E. (1998). Replication factor C disengages from proliferating cell nuclear antigen (PCNA) upon sliding clamp formation, and PCNA itself tethers DNA polymerase delta to DNA. *J Biol Chem*, 273(48), 31992-31999.

[54] Gambus, A., Jones, R. C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R. D., & Labib, K. (2006). GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol*, 8(4), 358-366.

[55] Cox, L.S. (2009). Molecular Themes in DNA Replication in Book Molecular Themes in DNA Replication. *Series Molecular Themes in DNA Replication*, Editor ed.^eds., City: Royal Society of Chemistry, 2009, 443.

[56] Budd, M. E., Cox, L. S., & Campbell, J. L. (2009). Coordination of nucleases and helicases during DNA replication and double-strand break repair. *Molecular Themes in DNA Replication, L. S. Cox ed., Cambridge: Royal Society of Chemistry*, 112-155.

[57] Beattie, T.R., & Bell, S.D. (2012). Coordination of multiple enzyme activities by a single PCNA in archaeal Okazaki fragment maturation. *EMBO J*, 31(6), 1556-1567.

[58] Warbrick, E., Lane, D. P., Glover, D. M., & Cox, L. S. (1995). A small peptide inhibitor of DNA replication defines the site of interaction between the cyclin-dependent kinase inhibitor p21WAF1 and proliferating cell nuclear antigen. *Curr Biol*, 5(3), 275-282.

[59] Guarino, E., Shepherd, M. E., Salguero, I., Hua, H., Deegan, R. S., & Kearsey, S. E. (2011). Cdt1 proteolysis is promoted by dual PIP degrons and is modulated by PCNA ubiquitylation. *Nucleic Acids Res*, 39(14), 5978-5990.

[60] Warbrick, E., Lane, D. P., Glover, D. M., & Cox, L. S. (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(19), 2313-2321.

[61] Cox, L. S. (1997). Who binds wins: Competition for PCNA rings out cell-cycle changes. *Trends Cell Biol*, 7(12), 493-498.

[62] Cox, L. S., & Kearsey, S. (2009). Ring structures and six-fold symmetry in DNA replication. *Molecular Themes in DNA Replication, L. S. Cox ed., Cambridge: Royal Society of Chemistry*, 47-85.

[63] Brosh, R. M., von Kobbe, C., Sommers, J. A., Karmakar, P., Opresko, P. L., Piotrowski, J., Dianova, I., Dianov, G. L., & Bohr, V. A. (2001). Werner syndrome protein interacts with human flap endonuclease 1 and stimulates its cleavage activity. *EMBO J*, 20(20), 5791-5801.

[64] Sharma, S., Sommers, J. A., Gary, R. K., Friedrich-Heineken, E., Hubscher, U., & Brosh, R. M. (2005). The interaction site of Flap Endonuclease-1 with WRN helicase suggests a coordination of WRN and PCNA. *Nucleic Acids Res*, 33(21), 6769-6781.

[65] Nick, S. A., Mc Elhinny, Z. F., Pursell, S., & Kunkel, T. A. (2009). Mechanisms for high fidelity DNA replication. *Molecular Themes in DNA Replication, L. S. Cox ed., Cambridge, UK: Royal Society of Chemistry*, 86-111.

[66] Swan, M. K., Johnson, R. E., Prakash, L., Prakash, S., & Aggarwal, A. K. (2009). Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase delta. *Nat Struct Mol Biol*, 16(9), 979-986.

[67] Perry, J. J., Yannone, S. M., Holden, L. G., Hitomi, C., Asaithamby, A., Han, S., Cooper, P. K., Chen, D. J., & Tainer, J. A. (2006). WRN exonuclease structure and molecular mechanism imply an editing role in DNA end processing. *Nat Struct Mol Biol*, 13(5), 414-422.

[68] Kamath-Loeb, A. S., Shen, J. C., Schmitt, M. W., & Loeb, L. A. (2012). The Werner syndrome exonuclease facilitates DNA degradation and high fidelity DNA polymerization by human DNA polymerase delta. *J Biol Chem*, 287(15), 12480-12490.

[69] Takeuchi, F., Hanaoka, F., Goto, M., Akaoka, I., Hori, T., Yamada, M., & Miyamoto, T. (1982). Altered frequency of initiation sites of DNA replication in Werner's syndrome cells. *Hum Genet*, 60(4), 365-368.

[70] Hanaoka, F., Yamada, M., Takeuchi, F., Goto, M., Miyamoto, T., & Hori, T. (1985). Autoradiographic studies of DNA replication in Werner's syndrome cells. *Adv Exp Med Biol*, 190, 439-457.

[71] Foiani, M., Ferrari, M., Liberi, G., Lopes, M., Lucca, C., Marini, F., Pellicioli, A., Muzi, M., Falconi, , & Plevani, P. (1998). S-phase DNA damage checkpoint in budding yeast. *Biol Chem*, 379(8-9), 1019-1023.

[72] Kamath-Loeb, A. S., Loeb, L. A., Johansson, E., Burgers, P. M., & Fry, M. (2001). Interactions between the Werner syndrome helicase and DNA polymerase delta specifi-

cally facilitate copying of tetraplex and hairpin structures of the d(CGG)n trinucleotide repeat sequence. *J Biol Chem*, 276(19), 16439-16446.

[73] Ammazzalorso, F., Pirzio, L. M., Bignami, M., Franchitto, A., & Pichierri, P. (2010). ATR and ATM differently regulate WRN to prevent DSBs at stalled replication forks and promote replication fork recovery. *EMBO J*, 29(18), 3156-3169.

[74] Murfuni, I., De Santis, A., Federico, M., Bignami, M., Pichierri, P., & Franchitto, A. (2012). Perturbed replication induced genome-wide or at common fragile sites is differently managed in the absence of WRN. *Carcinogenesis*, 33(9), 1655-63.

[75] Shah, S. N., Opresko, P. L., Meng, X., Lee, M. Y., & Eckert, K. A. (2010). DNA structure and the Werner protein modulate human DNA polymerase delta-dependent replication dynamics within the common fragile site FRA16D,. *Nucleic Acids Res*, 38(4), 1149-1162.

[76] Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S., & Prakash, L. (2000). Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. *Nature*, 406(6799), 1015-1019.

[77] Choi, J.Y., & Guengerich, F.P. (2006). Kinetic evidence for inefficient and error-prone bypass across bulky N2-guanine DNA adducts by human DNA polymerase iota. *J Biol Chem*, 281(18), 12315-12324.

[78] Nimonkar, A. V., Ozsoy, A. Z., Genschel, J., Modrich, P., & Kowalczykowski, S. C. (2008). Human exonuclease 1 and BLM helicase interact to resect DNA and initiate DNA repair. *Proc Natl Acad Sci U S A*, 105(44), 16906-16911.

[79] Aggarwal, M., Sommers, J. A., Morris, C., & Brosh, R. M. (2010). Delineation of WRN helicase function with EXO1 in the replicational stress response. *DNA Repair (Amst)*, 9(7), 765-776.

[80] Aggarwal, M., & Brosh, R. M. (2010). Genetic mutants illuminate the roles of RecQ helicases in recombinational repair or response to replicational stress. *Cell Cycle*, 9(16), 3139-3141.

[81] Matsuda, T., Bebenek, K., Masutani, C., Rogozin, I. B., Hanaoka, F., & Kunkel, T. A. (2001). Error rate and specificity of human and murine DNA polymerase eta. *J Mol Biol*, 312(2), 335-346.

[82] Zhang, Y., Yuan, F., Xin, H., Wu, X., Rajpal, D. K., Yang, D., & Wang, Z. (2000). Human DNA polymerase kappa synthesizes DNA with extraordinarily low fidelity. *Nucleic Acids Res*, 28(21), 4147-4156.

[83] Vasquez Del, R., Carpio, T. D., Silverstein, S., Lone, M. K., Swan, J. R., Choudhury, R. E., Johnson, S., Prakash, L., & Aggarwal, A. K. (2009). Structure of human DNA polymerase kappa inserting dATP opposite an 8-OxoG DNA lesion. *PLoS One*, 4(6), e5766.

[84]  Kamath-Loeb, A. S., Lan, L., Nakajima, S., Yasui, A., & Loeb, L. A. (2007). Werner syndrome protein interacts functionally with translesion DNA polymerases. *Proc Natl Acad Sci U S A*, 104(25), 10394-10399.

[85]  Machwe, A., Lozada, E., Wold, M. S., Li, G. M., & Orren, D. K. (2011). Molecular co-operation between the Werner syndrome protein and replication protein A in relation to replication fork blockage. *J Biol Chem*, 286(5), 3497-3508.

[86]  Sharma, S., Sommers, J. A., Driscoll, H. C., Uzdilla, L., Wilson, T. M., & Brosh, R. M. (2003). The exonucleolytic and endonucleolytic cleavage activities of human exonuclease 1 are stimulated by an interaction with the carboxyl-terminal region of the Werner syndrome protein. *J Biol Chem*, 278(26), 23487-23496.

[87]  Franchitto, A., Pirzio, L. M., Prosperi, E., Sapora, O., Bignami, M., & Pichierri, P. (2008). Replication fork stalling in WRN-deficient cells is overcome by prompt activation of a MUS81-dependent pathway. *J Cell Biol*, 183(2), 241-252.

[88]  Byun, T. S., Pacek, M., Yee, M. C., Walter, J. C., & Cimprich, K. A. (2005). Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev*, 19(9), 1040-1052.

[89]  Singleton, M. R., Scaife, S., Raven, N. D., & Wigley, D. B. (2001). Crystallization and preliminary X-ray analysis of RecG, a replication-fork reversal helicase from Thermotoga maritima complexed with a three-way DNA junction. *Acta Crystallogr D Biol Crystallogr*, 57(11), 1695-1696.

[90]  Singleton, M. R., Scaife, S., & Wigley, D. B. (2001). Structural analysis of DNA replication fork reversal by RecG. *Cell*, 107(1), 79-89.

[91]  Courcelle, J., Donaldson, J. R., Chow, K. H., & Courcelle, C. T. (2003). DNA damage-induced replication fork regression and processing in Escherichia coli. *Science*, 299(5609), 1064-1067.

[92]  Sharma, S., Otterlei, M., Sommers, J. A., Driscoll, H. C., Dianov, G. L., Kao, H. I., Bambara, R. A., & Brosh, R. M. (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(2), 734-750.

[93]  Otterlei, M., Bruheim, P., Ahn, B., Bussen, W., Karmakar, P., Baynton, K., & Bohr, V. A. (2006). Werner syndrome protein participates in a complex with RAD51, RAD54, RAD54B and ATR in response to ICL-induced replication arrest. *J Cell Sci*, 119, 5137-5146.

[94]  Machwe, A., Karale, R., Xu, X., Liu, Y., & Orren, D. K. (2011). The Werner and Bloom syndrome proteins help resolve replication blockage by converting (regressed) holliday junctions to functional replication forks. *Biochemistry*, 50(32), 6774-6788.

[95]  Pichierri, P., & Rosselli, F. (2004). The DNA crosslink-induced S-phase checkpoint depends on ATR-CHK1 and ATR-NBS1-FANCD2 pathways. *EMBO J*, 23(5), 1178-1187.

[96] Bennetzen, M. V., Marino, G., Pultz, D., Morselli, E., Faergeman, N. J., Kroemer, G., & Andersen, J. S. (2012). Phosphoproteomic analysis of cells treated with longevity-related autophagy inducers. *Cell Cycle*, 11(9), 1827-1840.

[97] Patro, B. S., Frohlich, R., Bohr, V. A., & Stevnsner, T. (2011). WRN helicase regulates the ATR-CHK1-induced S-phase checkpoint pathway in response to topoisomerase-I-DNA covalent complexes. *J Cell Sci*, 124, 3967-3979.

[98] Pichierri, P., Nicolai, S., Cignolo, L., Bignami, M., & Franchitto, A. (2011). The RAD9-RAD1-HUS1 (9.1.1) complex interacts with WRN and is crucial to regulate its response to replication fork stalling. *Oncogene*, 31(23), 2809-2823.

[99] Svoboda, D.L., & Vos, J.M. (1995). Differential replication of a single, UV-induced lesion in the leading or lagging strand by a human cell extract: fork uncoupling or gap formation. *Proc Natl Acad Sci U S A*, 92(26), 11975-11979.

[100] Mc Inerney, P., & O'Donnell, M. (2004). Functional uncoupling of twin polymerases: mechanism of polymerase dissociation from a lagging-strand block. *J Biol Chem*, 279(20), 21543-21551.

[101] Heller, R.C., & Marians, K.J. (2006). Replisome assembly and the direct restart of stalled replication forks. *Nat Rev Mol Cell Biol*, 7(12), 932-943.

[102] Wu, L., & Hickson, I. D. (2003). The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature*, 426(6968), 870-874.

[103] Wu, L., Chan, K. L., Ralf, C., Bernstein, D. A., Garcia, P. L., Bohr, V. A., Vindigni, A., Janscak, P., Keck, J. L., & Hickson, I. D. (2005). The HRDC domain of BLM is required for the dissolution of double Holliday junctions. *EMBO J*, 24(14), 2679-2687.

[104] Opresko, P. L., Sowd, G., & Wang, H. (2009). The Werner syndrome helicase/exonuclease processes mobile D-loops through branch migration and degradation. *PLoS One*, 4(3), e4825.

[105] Kamath-Loeb, A. S., Johansson, E., Burgers, P. M., & Loeb, L. A. (2000). Functional interaction between the Werner Syndrome protein and DNA polymerase delta. *Proc Natl Acad Sci U S A*, 97(9), 4603-4608.

[106] Kunkel, T. A., Sabatino, R. D., & Bambara, R. A. (1987). Exonucleolytic proofreading by calf thymus DNA polymerase delta. *Proc Natl Acad Sci U S A*, 84(14), 4865-4869.

[107] Kamath-Loeb, A., Loeb, L. A., & Fry, M. (2012). The Werner syndrome protein is distinguished from the Bloom syndrome protein by its capacity to tightly bind diverse DNA structures. *PLoS One*, 7(1), e30189.

[108] Bukowy, Z., Harrigan, J. A., Ramsden, D. A., Tudek, B., Bohr, V. A., & Stevnsner, T. (2008). WRN Exonuclease activity is blocked by specific oxidatively induced base lesions positioned in either DNA strand. *Nucleic Acids Res*, 36(15), 4975-4987.

[109] Maser, R.S., & DePinho, R.A. (2004). Telomeres and the DNA damage response: why the fox is guarding the henhouse,. *DNA Repair (Amst)*, 3(8-9), 979-988.

[110]  d'Adda, F., di Fagagna, S. H., & Jackson, S. P. (2004). Functional links between telomeres and proteins of the DNA-damage response. *Genes Dev*, 18(15), 1781-1799.

[111]  Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., & de Lange, T. (1999). Mammalian telomeres end in a large duplex loop. *Cell*, 97(4), 503-514.

[112]  de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev*, 19(18), 2100-2110.

[113]  Broccoli, D., Smogorzewska, A., Chong, L., & de Lange, T. (1997). Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat Genet*, 17(2), 231-235.

[114]  Liu, D., O'Connor, M. S., Qin, J., & Songyang, Z. (2004). Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins. *J Biol Chem*, 279(49), 51338-51342.

[115]  Baumann, P., & Cech, T. R. (2001). Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science*, 292(5519), 1171-1175.

[116]  O'Donovan, A., Tomiyama, A. J., Lin, J., Puterman, E., Adler, N. E., Kemeny, M., Wolkowitz, O. M., Blackburn, E. H., & Epel, E. S. (2012). Stress appraisals and cellular aging: a key role for anticipatory threat in the relationship between psychological stress and telomere length. *Brain Behav Immun*, 26(4), 573-579.

[117]  Epel, E. S., Lin, J., Dhabhar, F. S., Wolkowitz, O. M., Puterman, E., Karan, L., & Blackburn, E. H. (2010). Dynamics of telomerase activity in response to acute psychological stress. *Brain Behav Immun*, 24(4), 531-539.

[118]  Damjanovic, A. K., Yang, Y., Glaser, R., Kiecolt-Glaser, J. K., Nguyen, H., Laskowski, B., Zou, Y., Beversdorf, D. Q., & Weng, N. P. (2007). Accelerated telomere erosion is associated with a declining immune function of caregivers of Alzheimer's disease patients. *J Immunol*, 179(6), 4249-4254.

[119]  Smogorzewska, A., & de Lange, T. (2004). Regulation of telomerase by telomeric proteins. *Annu Rev Biochem*, 73, 177-208.

[120]  Harley, C. B., Futcher, A. B., & Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature*, 345(6274), 458-460.

[121]  Harley, C. B., Kim, N. W., Prowse, K. R., Weinrich, S. L., Hirsch, K. S., West, M. D., Bacchetti, S., Hirte, H. W., Counter, C. M., Greider, C. W., et al. (1994). Telomerase, cell immortality, and cancer,. *Cold Spring Harb Symp Quant Biol*, 59, 307-315.

[122]  Cesare, A.J., & Reddel, R.R. (2010). Alternative lengthening of telomeres: models, mechanisms and implications. *Nat Rev Genet*, 11(5), 319-330.

[123]  Verdun, R. E., & Karlseder, J. (2007). Replication and protection of telomeres. *Nature*, 447(7147), 924-931.

[124] Schulz, V. P., Zakian, V. A., Ogburn, C. E., Mc Kay, J., Jarzebowicz, A. A., Edland, S. D., & Martin, G. M. (1996). Accelerated loss of telomeric repeats may not explain accelerated replicative decline of Werner syndrome cells. *Hum Genet*, 97(6), 750-754.

[125] Baird, D. M., Davis, T., Rowson, J., Jones, C. J., & Kipling, D. (2004). Normal telomere erosion rates at the single cell level in Werner syndrome fibroblast cells. *Hum Mol Genet*, 13(14), 1515-1524.

[126] Lombard, D. B., Beard, C., Johnson, B., Marciniak, R. A., Dausman, J., Bronson, R., Buhlmann, J. E., Lipman, R., Curry, R., Sharpe, A., Jaenisch, R., & Guarente, L. (2000). Mutations in the WRN gene in mice accelerate mortality in a p53-null background. *Mol Cell Biol*, 20(9), 3286-3291.

[127] Chang, S. (2005). A mouse model of Werner Syndrome: what can it tell us about aging and cancer ? *Int J Biochem Cell Biol*, 37(5), 991-999.

[128] Du, X., Shen, J., Kugan, N., Furth, E. E., Lombard, D. B., Cheung, C., Pak, S., Luo, G., Pignolo, R. J., De Pinho, R. A., Guarente, L., & Johnson, F. B. (2004). Telomere shortening exposes functions for the mouse Werner and Bloom syndrome genes. *Mol Cell Biol*, 24(19), 8437-8446.

[129] Chang, S., Multani, A. S., Cabrera, N. G., Naylor, M. L., Laud, P., Lombard, D., Pathak, S., Guarente, L., & De Pinho, R. A. (2004). Essential role of limiting telomeres in the pathogenesis of Werner syndrome. *Nat Genet*, 36(8), 877-882.

[130] Stavropoulos, D. J., Bradshaw, P. S., Li, X., Pasic, I., Truong, K., Ikura, M., Ungrin, M., & Meyn, M. S. (2002). The Bloom syndrome helicase BLM interacts with TRF2 in ALT cells and promotes telomeric DNA synthesis. *Hum Mol Genet*, 11(25), 3135-3144.

[131] Lillard-Wetherell, K., Machwe, A., Langland, G. T., Combs, K. A., Behbehani, G. K., Schonberg, S. A., German, J., Turchi, J. J., Orren, D. K., & Groden, J. (2004). Association and regulation of the BLM helicase by the telomere proteins TRF1 and TRF2. *Hum Mol Genet*, 13(17), 1919-1932.

[132] Machwe, A., Xiao, L., & Orren, D. K. (2004). TRF2 recruits the Werner syndrome (WRN) exonuclease for processing of telomeric DNA. *Oncogene*, 23(1), 149-156.

[133] Opresko, P. L., von, C., Kobbe, J. P., Laine, J., Harrigan, I. D., & Hickson, V. A. (2002). Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases. *J Biol Chem*, 277(43), 41110-41119.

[134] Opresko, P. L., Mason, P. A., Podell, E. R., Lei, M., Hickson, I. D., Cech, T. R., & Bohr, V. A. (2005). POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. *J Biol Chem*, 280(37), 32069-32080.

[135] Nora, G. J., Buncher, N. A., & Opresko, P. L. (2010). Telomeric protein TRF2 protects Holliday junctions with telomeric arms from displacement by the Werner syndrome helicase. *Nucleic Acids Res*, 38(12), 3984-3998.

[136]  Lee, J. Y., Kozak, M., Martin, J. D., Pennock, E., & Johnson, F. B. (2007). Evidence that a RecQ helicase slows senescence by resolving recombining telomeres. *PLoS Biol*, 5(6), e160.

[137]  Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., & Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev*, 18(13), 1618-1629.

[138]  Zaug, A. J., Podell, E. R., & Cech, T. R. (2005). Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc Natl Acad Sci U S A*, 102(31), 10864-10869.

[139]  Mohaghegh, P., & Hickson, I. D. (2001). DNA helicase deficiencies associated with cancer predisposition and premature ageing disorders. *Hum Mol Genet*, 10(7), 741-746.

[140]  Opresko, P. L., Laine, J. P., Brosh, R. M., Seidman, M. M., & Bohr, V. A. (2001). Coordinate action of the helicase and 3′ to 5′ exonuclease of Werner syndrome protein. *J Biol Chem*, 276(48), 44677-44687.

[141]  Kusumoto-Matsuo, R., Opresko, P. L., Ramsden, D., Tahara, H., & Bohr, V. A. (2010). Cooperation of DNA-PKcs and WRN helicase in the maintenance of telomeric D-loops. *Aging (Albany NY)*, 2(5), 274-284.

[142]  Crabbe, L., Verdun, R. E., Haggblom, C. I., & Karlseder, J. (2004). Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. *Science*, 306(5703), 1951-1953.

[143]  Dewar, J. M., & Lydall, D. (2010). Pif1- and Exo1-dependent nucleases coordinate checkpoint activation following telomere uncapping. *EMBO J*, 29(23), 4020-4034.

[144]  Barefield, C., & Karlseder, J. (2012). The BLM helicase contributes to telomere maintenance through processing of late-replicating intermediate structures. *Nucleic Acids Res*, 40(15), 7358-7367.

[145]  Opresko, P. L. (2008). Telomere ResQue and preservation--roles for the Werner syndrome protein and other RecQ helicases. *Mech Ageing Dev*, 129(1-2), 79-90.

[146]  Arnoult, N., Saintome, C., Ourliac-Garnier, I., Riou, J. F., & Londono-Vallejo, A. (2009). Human POT1 is required for efficient telomere C-rich strand replication in the absence of WRN. *Genes Dev*, 23(24), 2915-2924.

[147]  Schulz, V.P., & Zakian, V.A. (1994). The saccharomyces PIF1 DNA helicase inhibits telomere elongation and de novo telomere formation. *Cell*, 76(1), 145-155.

[148]  George, T., Wen, Q., Griffiths, R., Ganesh, A., Meuth, M., & Sanders, C. M. (2009). Human Pif1 helicase unwinds synthetic DNA structures resembling stalled DNA replication forks. *Nucleic Acids Res*, 37(19), 6491-6502.

[149] Paeschke, K., Capra, J. A., & Zakian, V. A. (2011). DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell*, 145(5), 678-691.

[150] Cheung, I., Schertzer, M., Rose, A., & Lansdorp, P. M. (2002). Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA. *Nat Genet*, 31(4), 405-409.

[151] Vallur, A. C., & Maizels, N. (2010). Distinct activities of exonuclease 1 and flap endonuclease 1 at telomeric g4 DNA. *PLoS One*, 5(1), e8908.

[152] Eller, M. S., Liao, X., Liu, S., Hanna, K., Backvall, H., Opresko, P. L., Bohr, V. A., & Gilchrest, B. A. (2006). A role for WRN in telomere-based DNA damage responses. *Proc Natl Acad Sci U S A*, 103(41), 15073-15078.

[153] Li, B., Jog, S. P., Reddy, S., & Comai, L. (2008). WRN controls formation of extrachromosomal telomeric circles and is required for TRF2DeltaB-mediated telomere shortening. *Mol Cell Biol*, 28(6), 1892-1904.

[154] Reddy, S., Li, B., & Comai, L. (2010). Processing of human telomeres by the Werner syndrome protein. *Cell Cycle*, 9(16), 3137-3138.

[155] Li, B., Reddy, S., & Comai, L. (2009). Sequence-specific processing of telomeric 3′ overhangs by the Werner syndrome protein exonuclease activity. *Aging (Albany NY)*, 1(3), 289-302.

[156] Orren, D. K., Theodore, S., & Machwe, A. (2002). The Werner syndrome helicase/exonuclease (WRN) disrupts and degrades D-loops in vitro. *Biochemistry*, 41(46), 13483-13488.

[157] Johnson, F. B., Marciniak, R. A., Mc Vey, M., Stewart, S. A., Hahn, W. C., & Guarente, L. (2001). The Saccharomyces cerevisiae WRN homolog Sgs1p participates in telomere maintenance in cells lacking telomerase. *EMBO J*, 20(4), 905-913.

[158] Cohen, H., & Sinclair, D. A. (2001). Recombination-mediated lengthening of terminal telomeric repeats requires the Sgs1 DNA helicase. *Proc Natl Acad Sci U S A* [6], 3174-3179.

[159] Neumann, A.A., & Reddel, R.R. (2002). Telomere maintenance and cancer-- look, no telomerase. *Nat Rev Cancer*, 2(11), 879-884.

[160] Dunham, M. A., Neumann, A. A., Fasching, C. L., & Reddel, R. R. (2000). Telomere maintenance by recombination in human cells. *Nat Genet*, 26(4), 447-450.

[161] Hickson, I. D. (2003). RecQ helicases: caretakers of the genome. *Nat Rev Cancer*, 3(3), 169-178.

[162] Bailey, S. M., Cornforth, M. N., Ullrich, R. L., & Goodwin, E. H. (2004). Dysfunctional mammalian telomeres join with DNA double-strand breaks. *DNA Repair (Amst)*, 3(4), 349-357.

[163] Bailey, S. M., Brenneman, M. A., & Goodwin, E. H. (2004). Frequent recombination in telomeric DNA may extend the proliferative life of telomerase-negative cells. *Nucleic Acids Res*, 32(12), 3743-3751.

[164] Laud, P. R., Multani, A. S., Bailey, S. M., Wu, L., Kingsley, J. C., Lebel, M., Pathak, S., De Pinho, R. A., & Chang, S. (2005). Elevated telomere-telomere recombination in WRN-deficient, telomere dysfunctional cells promotes escape from senescence and engagement of the ALT pathway. *Genes Dev*, 19(21), 2560-2570.

[165] Celli, G. B., Denchi, E. L., & de Lange, T. (2006). Ku70 stimulates fusion of dysfunctional telomeres yet protects chromosome ends from homologous recombination. *Nat Cell Biol 8* [8], 885-890.

[166] Opresko, P. L., Fan, J., Danzy, S., Wilson, D. M., & Bohr, V. A. (2005). Oxidative damage in telomeric DNA disrupts recognition by TRF1 and TRF2. *Nucleic Acids Res*, 33(4), 1230-1239.

[167] Cooper, M. P., Machwe, A., Orren, D. K., Brosh, R. M., Ramsden, D., & Bohr, V. A. (2000). Ku complex interacts with and stimulates the Werner protein. *Genes Dev 14* [8], 907-912.

[168] Petersen, S., Saretzki, G., & von Zglinicki, T. (1998). Preferential accumulation of single-stranded regions in telomeres of human fibroblasts. *Exp Cell Res*, 239(1), 152-160.

[169] von, T., Zglinicki, G., Saretzki, J., Ladhoff, F., d'Adda di, Fagagna., & Jackson, S. P. (2005). Human cell senescence as a DNA damage response. *Mech Ageing Dev*, 126(1), 111-117.

[170] Passos, J. F., Saretzki, G., Ahmed, S., Nelson, G., Richter, T., Peters, H., Wappler, I., Birket, M. J., Harold, G., Schaeuble, K., Birch-Machin, M. A., Kirkwood, T. B., & von Zglinicki, T. (2007). Mitochondrial dysfunction accounts for the stochastic heterogeneity in telomere-dependent senescence. *PLoS Biol*, 5(5), e110.

[171] Passos, J. F., Saretzki, G., & von Zglinicki, T. (2007). DNA damage in telomeres and mitochondria during cellular senescence: is there a connection? *Nucleic Acids Res*, 35(22), 7505-7513.

[172] Karlseder, J., Smogorzewska, A., & de Lange, T. (2002). Senescence induced by altered telomere state, not telomere loss. *Science*, 295(5564), 2446-2449.

[173] Harrigan, J. A., Wilson, D. M., 3rd Prasad, R., Opresko, P. L., Beck, G., May, A., Wilson, S. H., & Bohr, V. A. (2006). The Werner syndrome protein operates in base excision repair and cooperates with DNA polymerase beta. *Nucleic Acids Res*, 34(2), 745-754.

[174] Harrigan, J. A., Opresko, P. L., von, C., Kobbe, P. S., Kedar, R., Prasad, S. H., Wilson, , & Bohr, V. A. (2003). The Werner syndrome protein stimulates DNA polymerase beta strand displacement synthesis via its helicase activity. *J Biol Chem*, 278(25), 22686-22695.

[175] Pagano, G., Zatterale, A., Degan, P., d'Ischia, M., Kelly, F. J., Pallardo, F. V., & Koda-ma, S. (2005). Multiple involvement of oxidative stress in Werner syndrome pheno-type. *Biogerontology*, 6(4), 233-243.

[176] Labbe, A., Turaga, R. V., Paquet, E. R., Garand, C., & Lebel, M. (2010). Expression profiling of mouse embryonic fibroblasts with a deletion in the helicase domain of the Werner Syndrome gene homologue treated with hydrogen peroxide. *BMC Ge-nomics*, 11, 127.

[177] Ghosh, A., Rossi, M. L., Aulds, J., Croteau, D., & Bohr, V. A. (2009). Telomeric D-loops containing 8-oxo-2'-deoxyguanosine are preferred substrates for Werner and Bloom syndrome helicases and are bound by POT1. *J Biol Chem*, 284(45), 31074-31084.

[178] Muftuoglu, M., Wong, H. K., Imam, S. Z., Wilson, D. M., 3rd Bohr, V. A., & Opresko, P. L. (2006). Telomere repeat binding factor 2 interacts with base excision repair pro-teins and stimulates DNA synthesis by DNA polymerase beta. *Cancer Res*, 66(1), 113-124.

[179] Miller, A., Balakrishnan, L., Buncher, N. A., Opresko, P. L., & Bambara, R. A. (2012). Telomere proteins POT1, TRF1 and TRF2 augment long-patch base excision repair in vitro. *Cell Cycle*.

# Replication of Organellar Chromosomes

# Replicational Mutation Gradients, Dipole Moments, Nearest Neighbour Effects and DNA Polymerase Gamma Fidelity in Human Mitochondrial Genomes

Hervé Seligmann

Additional information is available at the end of the chapter

## 1. Introduction

The large amount of available human mitochondrial genome data from Ingman and Gyllenstein and from Ruiz-Pesini et al [see http://www.mtdb.igp.uu.se/, 1 and http://www.mitomap.org/MITOMAP, 2] enables to study in some detail the spectrum of mutations observed within this species' mitochondrion. DNA mutations have two main causes: spontaneous chemical alterations of nucleotides, from one nucleotide 'species' to another, such as hydrolytic deaminations from C->T and A to hypoxanthine, which pairs with C and leads to its replacement by G (in the following summarized as A->G); and inaccuracies by the enzymatic machinery that is responsible for the polymerization of new DNA strands from the template of the existing DNA during DNA replication. Here I explore the tendency for mutations from different genes and mutation types to be explained by the first (physicochemical), or the other (more enzymatic/biological) factor, also in relation to adaptive constraints (natural selection is weaker against DNA mutations that cause no or only conservative changes at the protein level). The relative importance of these various factors affecting mutation spectra is investigated for observed human mitochondrial mutations in relation to different types of substitutions and different genes. I also explore nearest neighbour effects on the different mutation types, though the relative contribution of this factor in relation to others is not evaluated here.

The main, presumably sole DNA replicating enzyme in vertebrate mitochondria is the DNA gamma polymerase, which evolved from a bacterial tRNA synthetase [3]. Twelve types of substitutions of one nucleotide by another nucleotide occur at different frequencies. The most frequent changes occur within each of the nucleotide families, purines (adenosine, A, and

**INTECH**
open science | open minds

guanine, G) and pyrimidines (cytosine, C, and thymidine, T), as these involve less changes in molecular structure. These four purine to purine or pyrimidine to pyrimidine mutations are called transitions, the eight mutations from one chemical group to another are called transversions. A simplistic model predicts the frequencies of all substitutions, based on the dipole moment of the nucleotides [4], for a DNA region supposed to have no function, a pseudogene [5], so that observed substitution frequencies are believed unaffected by natural selection. The partial dipole moment of a chemical bond is proportional to the distance of an electron's mean position, in the chemical bond between atoms, from mid-distance between these atoms. The dipole moment of a molecule is the product of all partial dipole moments. G and C have high dipole moments, A and T have low dipole moments [6]. The hypothesized model assumes that a high dipole moment indicates high chemical reactivity, and hence probable alteration by chemical processes. Indeed, observed frequencies of mutations in pseudogenes of one nucleotide into another nucleotide are proportional to dipole moment changes: nucleotides with low dipole moment substitute those with high dipole moment.

Independently of the dipole moment hypothesis, some spontaneous chemical reactions, deaminations, A->G and C->T, occur preferentially while the heavy DNA strand is in the single stranded state [7, 8]. This occurs mainly during replication and transcription (DNA and RNA polymerization). Distances from replication origins and for transcription, from promoters [9, 10], determine durations that different DNA regions remain single stranded, creating gradients in deaminations in genomes with asymmetric replication, such as mitochondrial genomes (reviewed in [11, 12]). Hence gene position affects transition frequencies. Site-specific mutation rates estimated by phylogenetic reconstruction suggest that mutation gradients might also exist for some transversions [13, 14], indicating that single-strandedness might affect also substitutions that are not A->G or C->T.

Here I analyse mutation patterns observed in the 13 human mitochondrial protein coding genes, to estimate relative contributions of different processes to observed mutation patterns: replicational gradients [13, 14, 15], dipole moments [6, 16], selection against mutations that alter coding properties at the protein level [17] and gamma polymerase misincorporation [18], and potential interactions between these processes. I also explore nearest neighbour effects on mutation rates. The present analyses are also original in the sense that they are based on comparative analyses of sequence data, all the data originating from a single species, and not, as previous ones (i.e. [13, 14]), from comparisons between different, frequently evolutionarily distant species.

## 2. Dipole moments and the accuracy of the DNA gamma polymerase

The estimation of process-specific contributions of different mechanisms to a given phenomenon requires considering dependence among processes. Dipole moment effects are independent of the deamination gradient model which predicts a gradient in C->T, consistent with the model of dipole moment decrease, but an A->G gradient is inconsistent. This means that one deamination gradient fits into the dipole moment model, and the other does not, making both approaches approximately unrelated.

The issue of the accuracy of the gamma polymerase is more complex. It is indeed plausible that nucleotide misinsertion results from misrecognition of nucleotides by the polymerase, the latter due to physico-chemical similarities between nucleotides. This principle is also suggested by the high average misincorporation rates resulting in transitions (447.25±375.22) as compared to misincorporation rates causing transversions (121.64±186.6), explaining 33 percent of the variation in rates between different misincorporation types [18]. If so, the absolute value of the difference between dipole moments of nucleotides (from [16]) should be inversely proportional to misincorporation rates ('kd' in [18]), high rates occurring for nucleotides with similar dipole moments. This model differs from the model of dipole moment decrease, as it deals with the absolute value of the difference between dipole moments, and not the signed difference.
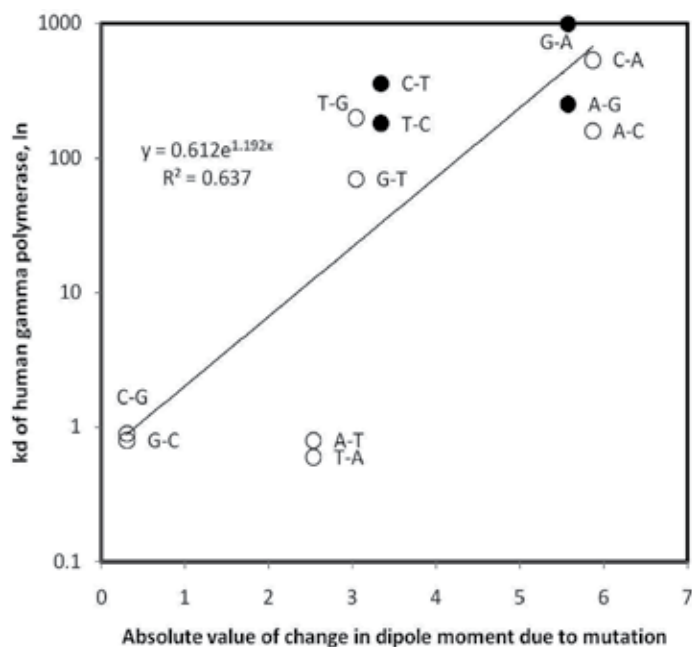


**Figure 1.** Misincorporation versus absolute difference between dipole moment of substituted and substituting nucleotide. Transitions (filled symbols) have high kds, but similar dipole moments decrease misincorporation kds.

The dipole similarity model for polymerase misincorporation rates can be dismissed at this point. Misincorporation rates increase, not decrease as expected, with absolute values of differences between dipole moments (r = 0.80, Figure 1). This unexplained association could reflect effects of other properties on misincorporation rates, properties that are inversely correlated with dipole moments. Note that after controlling for differences between transitions and transversions, the correlation shown in Figure 1 decreases (r = 0.54, not shown), yet the analysis confirms the principle that nucleotide substitutions with high kds tend to be substitutions between nucleotides with highly divergent dipole moments.

It is also possible that many nucleotide misincorporations result from the delay occurring between nucleotide recognition by the gamma polymerase and its incorporation in the elongating DNA polymer. One could suppose that some misincorporations are not due to misrecognitions, but to spontaneous mutations occurring after the nucleotide's accurate recognition by the polymerase, and before its incorporation. In that case, misincorporation rates should match the dipole moment model for decreased dipole moment: high rates are observed when substitutions decrease the dipole moment. This hypothesis cannot be ruled out, as misincorporation rates increase with the signed difference between nucleotide dipole moments (r = 0.50, not shown). Controlling for differences in kd between transitions and transversions, this positive association increases (r = 0.60, Figure 2).
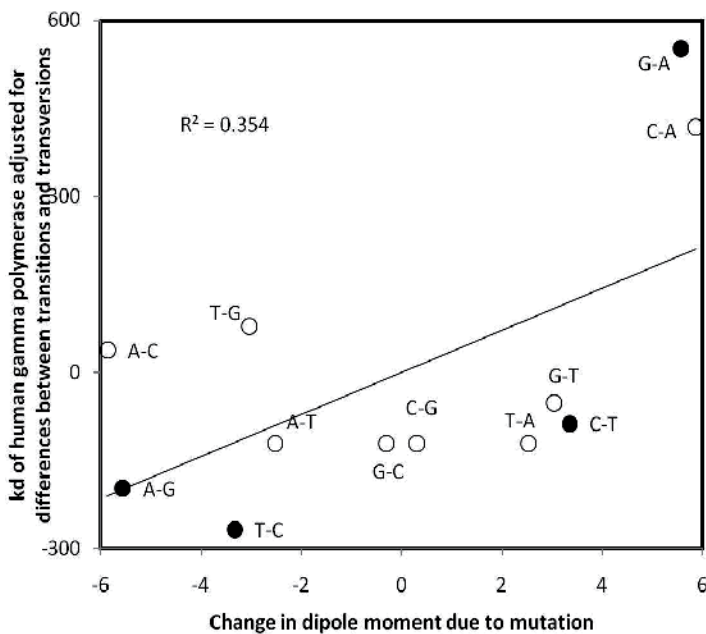


**Figure 2.** Adjusted misincorporation kd as a function of difference between dipole moments of substituted and substituting nucleotide. High kds imply dipole moment decrease.

Note that if causal interpretations of the associations in Figures 1 and 2 are relevant, it would be the dipole moment that affects kds. An alternative explanation to the trend in Figure 2 is that the gamma polymerase binds more readily nucleotides with low than high dipole moment, hence resulting in this biased misincorporation trend. Such a pattern could easily be caused by an overall relatively hydrophobic nature of the residues that constitute the polymerase's binding site (low dipole moment implying relative hydrophobicity). Even a very small bias for hydrophobic interactions would cause strong biases in analyses focusing on misincorporation rates. However, this hydrophobicity hypothesis does not seem to fit, at least in its simplistic form, what is known about the active site of the gamma polymerase according to the crystal structure published by Lee et al [19]. The active site consists of amino acids E895, Y951, R943 and Y955, among which one residue is positively charged (E,

glutamic acid), one negatively (R, arginine) and two are hydrophilic (Y, tyrosine). Note that none is classified as a hydrophobic amino acid. Hence the positive association in Figure 2 does not seem explained by active site hydrophobicity. Speculatively, electrostatic neutrality could favour misprocessing in active sites where each positive and negative charges occur, while high dipole moments would promote efficient processing.

These preliminary analyses suggest several important points on gamma polymerase fidelity: a) the causes for effects of similarity between nucleotides on misrecognition are unknown, structural similarity having effects opposite to those of dipole moment similarities; b) nucleotide properties affecting misrecognition are unknown but correlate with dipole moments; c) separating, even only conceptually, polymerase misrecognition from misincorporation, could be useful to understand polymerase accuracy; d) many misincorporations might be due to spontaneous mutations (with rates proportional to the dipole moment model) in the nucleotide occurring after accurate recognition by the polymerase, but before incorporation, resulting in misincorporation despite accurate recognition; e) alternatively, the polymerase's binding site might have in-built bias for hydrophobic misprocessing.

## 3. Selection on the gamma polymerase's misincorporation rates

Grantham [20] developed a matrix of dissimilarities based on major physico-chemical properties of amino acids (amino acid composition, polarity and molecular volume) that correlates best with amino acid replacement frequencies. From that matrix, Gojobori et al [17] estimated an average change in amino acid physico-chemical properties due to residue replacements for nucleotide substitutions in protein coding regions (see last line of table 4 in [17]). For example, A<->G substitutions have the lowest average impact, while G<->T have the greatest impact. One expects a negative association between impact on protein structure and the frequency of a nucleotide substitution. For pseudogenes, which do not code for proteins, the correlation between this average impact and the frequency of corresponding mutations (data from [17]) is weak (r = -0.33, one tailed P = 0.15), and even weaker after differences between transition and transversions have been accounted for (r = -0.18, one tailed P = 0.29). However, for mutation frequencies in coding sequences, natural selection against dysfunctional proteins has specifically decreased frequencies of non-conservative substitutions, and a strong negative correlation exists between impacts on protein structure and the frequency of a nucleotide substitution (r = -0.828, one tailed P = 0.00044). Accounting for differences between transitions and transversions does not alter qualitatively this result (r = -0.749, one tailed P = 0.0025).

Hence different misincorporations by the gamma polymerase [18] affect differently the coding properties of genes. The polymerase probably mainly adapted to avoid high impact nucleotide misincorporations. This can be tested by examining the correlation between misincorporation kds and the amino acid impact distances presented in [17], which will indicate to what extent these misincorporation rates resemble what is expected for pseudogenes (suggesting no selection occurs), or coding genes (suggesting the gamma polymerase is selected to minimize sub-

stitution impact on proteins). This correlation is negative, stronger than for pseudogenes, but weaker than for functional genes after selection (r = -0.434, one tailed P = 0.079). Controlling for differences between transitions and transversions does not alter much this result (r = -0.323, P = 0.15). The same holds after accounting for effects of dipole moments (Figure 2) on misincorporation rates: kds decrease with distances between replaced and replacing residues, but results are intermediate between mutation patterns observed for pseudogenes and genes that actually code for proteins (r = -0.44, one tailed P = 0.076).

This indicates that misincorporation rates include an adaptive component that minimizes the potential impact of nucleotide misincorporations on proteins. It is probable that a balance exists between minimizing different misincorporation rates, because the same active site in the polymerase is responsible for them. Hence the misincorporation pattern cannot be adapted to minimize all misincorporation rates, only to optimize misincorporation effects at protein levels. For frequencies of mutations observed in genes, selection affects each site (more or less) independently, hence impacts are minimized, resulting in much stronger correlations between mutation frequencies and impact at the protein level than observed for misincorporation rates, because the same active site produces the various types of misincorporations. The results indicate that this balancing effect due to interactions between different misincorporation types by the same active sites must be relatively strong in the gamma polymerase, otherwise the correlation with amino acid dissimilarities would resemble much more that found for coding genes. The matter of adaptively-tuned misincorporation rates by polymerases is nevertheless an interesting line of research that would gain from being developed further, including along the methods used here.

## 4. Gene-specific substitution matrices for human mitochondrial protein-coding genes

Misincorporation by gamma polymerases during replication is a major factor causing mutations. This factor is itself influenced by dipole moments of nucleotides, similarities between them, and greater selection pressures against specific misincorporation rates than on other rates (see previous sections). Here I examine observed mutation patterns in human mitochondrial genes.

Numbers of nucleotide substitutions for each of the 12 possible substitutions were counted from tabulations at http://www.mtdb.igp.uu.se/ [1] and http://www.mitomap.org/MITOMAP [2], separately for each gene (Table 1). Values are percentages of sites where a given mutation was observed among all sites where the substituted nucleotide mutated in that gene. The variation in that percentage within a given gene is mainly due to differences between transitions and transversions, the former dominating. Hence for further analyses, for each gene, mean percentages for transitions and transversions were calculated separately and subtracted from the observed percentages for transitions and transversions, respectively. This adjustment excludes effects due to differences between transitions and transversions in mutation percentages observed for each given gene. The two last columns in Table 1 are Pearson correlation

coefficients of percentages adjusted for differences between transitions and transversions and adjusted (along the same criterion) kd's of nucleotide misincorporations by the gamma polymerase (s), and after adjusting also for Grantham physico-chemical distances (s'). Correlation coefficients s are positive in 12 among 13 genes, a significant majority of cases according to a one tailed sign test ($P = 0.000854$). The correlation is significant ($P < 0.05$) at the level of a single gene for three genes, ND1, CO1 and AT8 (marked by asterisks in Table 1).

Results are only slightly altered after accounting for differences between transitions and transversions. Further analyses (s' in Table 1) using the residual misincorporation rates and the residual mutation percentages, calculated from their regressions with Grantham's amino acid dissimilarities do not change results much. These results show that variation in percentages of mutations of different types is to some extent due to misincorporation by the gamma polymerase, but a large part of the variation between substitution percentages remains unaccounted for. It is probable that natural selection against various mutations occurs, so that percentages in Table 1 are composites of misincorporation rates and other factors, such as selection against specific mutations. However, taking selection into account by using residuals from the regression of mutation frequencies with amino acid dissimilarities does not change patterns much. Hence further major factors affect observed mutation patterns, besides misincorporation rates and selection on coding impacts of mutations (and misincorporations).

## 5. Effects of deaminations and selection on mutation matrices

If one assumes that large parts of the variation that is not explained by the gamma polymerase's misincorporation rates in the previous analyses is due to selection, one can estimate which types of mutations are more or less prone to selection by analysing the residuals of the adjusted percentages (for each gene) from the regression with misincorporation. The line 'Res' in Table 1 indicates the number of genes for which this residual was positive, meaning that the percentage of that mutation was greater than expected from the regression with misincorporation. For two types of mutations, C->A and T->C, there were 10 such genes, which according to two tailed sign tests yields a significant tendency for observing percentages greater than expected by misincorporation ($P = 0.046$) as indicated by P in Table 1. Hence C->A and T->C are more frequent than expected by misincorporation. At least for T->C, there are two plausible explanations. T->C is a transition, and transitions cause relatively little functional effects at the level of coding properties of codons, suggesting low counterselection, hence relative over-representation (positive residuals). This explanation does not seem adequate, because the effect is not strong for other transitions (A->G, G->A and even opposite for C->T, where residuals were positive for only 2 genes ($P = 0.0095$, two tailed sign test). The latter effect on C->T is however also compatible with the second explanation for T->C. Deamination, promoted by single strandedness during replication, contributes to A->G mutations on the mitochondrial heavy strand DNA, which corresponds to T->C in Table 1 which uses the complementary light strand DNA annotation. Hence the systematic excess in T->C and systematic lack of C->T would be due to a factor that does not relate to misincorporation by the gamma polymerase, nor to selection, but presumably to the replicational

264 The Mechanisms of DNA Replication

mutation gradient of A->G. Residual analysis also indicates systematic underrepresentation of a further mutation type, G->T (P = 0.0095, two tailed sign test), a transversion that might be particularly counterselected [17]. Indeed, numbers of positive residuals tend to decrease with mean physico-chemical distances between replaced and replacing amino acids associated with these nucleotide mutations (r = -0.38, not statistically significant).

| Gene | A | | C | | G | | T | | A-C | A-G | A-T | C-A | C-G | C-T | G-A | G-C | G-T | T-A | T-C | T-G | s | s' |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ND1 | 272 | 116 | 344 | 124 | 112 | 45 | 228 | 48 | 2.8 | 88.9 | 8.3 | 13.3 | 4.7 | 82.0 | 87.8 | 6.1 | 6.1 | 7.7 | 78.2 | 14.1 | 55* | 53* |
| ND2 | 326 | 114 | 349 | 109 | 99 | 33 | 268 | 77 | 3.8 | 91.5 | 4.7 | 15.3 | 5.9 | 78.8 | 80.5 | 9.8 | 9.8 | 3.3 | 89.1 | 7.6 | -5 | -11 |
| CO1 | 419 | 121 | 462 | 121 | 250 | 59 | 410 | 97 | 5.7 | 89.5 | 4.8 | 12.7 | 4.2 | 83.1 | 91.8 | 3.3 | 4.9 | 6.7 | 90.4 | 2.9 | 55* | 47 |
| CO2 | 196 | 65 | 214 | 59 | 102 | 39 | 172 | 55 | 7.7 | 83.1 | 9.2 | 4.3 | 10.0 | 84.7 | 90.2 | 7.3 | 2.4 | 9.4 | 85.9 | 4.7 | 13 | 6 |
| AT8 | 80 | 42 | 69 | 31 | 13 | 9 | 45 | 26 | 0.0 | 95.2 | 4.8 | 11.1 | 0.0 | 88.9 | 100 | 0 | 0 | 0 | 92.9 | 7.1 | 74* | 71* |
| AT6 | 206 | 115 | 230 | 81 | 71 | 47 | 174 | 95 | 4.2 | 90.8 | 5.0 | 8.1 | 5.8 | 86.2 | 90.2 | 7.8 | 2.0 | 3.0 | 87.0 | 10.0 | 40 | 43 |
| CO3 | 210 | 87 | 249 | 70 | 116 | 44 | 209 | 69 | 11.6 | 84.9 | 3.5 | 6.3 | 3.8 | 90.0 | 93.2 | 4.6 | 2.3 | 9.5 | 87.8 | 2.7 | 45 | 37 |
| ND3 | 102 | 41 | 102 | 27 | 37 | 13 | 105 | 29 | 18.4 | 73.7 | 7.9 | 13.8 | 6.9 | 79.3 | 84.6 | 15.4 | 0 | 5.1 | 92.3 | 2.6 | 17 | 3 |
| ND4l | 84 | 27 | 92 | 19 | 36 | 12 | 85 | 23 | 7.7 | 84.6 | 7.7 | 7.7 | 0 | 92.3 | 84.6 | 0 | 15.4 | 8.7 | 78.3 | 13.0 | 20 | 47* |
| ND4 | 416 | 144 | 473 | 133 | 137 | 32 | 352 | 92 | 10.8 | 84.2 | 5.0 | 10.2 | 5.7 | 84.1 | 87.9 | 12.1 | 0 | 5.6 | 87.9 | 6.5 | 31 | 15 |
| ND5 | 518 | 207 | 580 | 183 | 190 | 49 | 416 | 117 | 7.8 | 85.5 | 5.7 | 18.7 | 18.7 | 62.6 | 85.6 | 8.1 | 6.3 | 4.9 | 90.9 | 4.2 | 19 | 3 |
| ND6 | 198 | 53 | 187 | 72 | 37 | 21 | 103 | 32 | 4.8 | 82.5 | 12.7 | 13.4 | 10.5 | 76.1 | 90.9 | 9.1 | 0 | 3.6 | 92.7 | 3.6 | 27 | -25 |
| Cytb | 326 | 142 | 391 | 141 | 137 | 67 | 287 | 95 | 4.3 | 88.7 | 7.0 | 8.6 | 6.2 | 85.2 | 89.7 | 10.3 | 0 | 1.9 | 92.5 | 5.7 | 12 | -1 |
| Res | | | | | | | | | 4 | 7 | 7 | 10 | 5 | 2 | 6 | 7 | 2 | 5 | 10 | 5 | | |
| Dssh | | | | | | | | | -29 | 16 | 3 | 38 | 23 | -39 | -13 | 30 | -17 | -63* | 35 | -9 | | |
| Pos 1 | | | | | | | | | -12 | 17 | -11 | 15 | 4 | -20 | -6 | 12 | -7 | 23 | -18 | 2 | | |
| Pos 2 | | | | | | | | | -3 | -1 | 11 | 10 | -15 | 4 | -1 | 1 | -1 | -27 | 13 | 4 | | |
| Pos 3 | | | | | | | | | -11 | -23 | 53* | 29 | 33 | -35 | -56* | 52* | 20 | -39 | 28 | -1 | | |
| Dloop | | | | | | | | | 22 | -39 | 27 | -9 | 42 | -26 | 20 | 30 | -41 | -31 | 53* | -42 | | |
| Pos 1 | | | | | | | | | -13 | 19 | -11 | 5 | 2 | -7 | 49* | -18 | -54 | 30 | -3 | -19 | | |
| Pos 2 | | | | | | | | | 9 | -20 | 36 | 41 | -37 | -4 | -19 | 13 | 13 | -21 | 33 | -34 | | |
| Pos 3 | | | | | | | | | 39 | -50 | 38 | -9 | 43 | -20 | -40 | 60* | -26 | -46 | 53* | -26 | | |

**Table 1.** Percentage of mutations observed in each human mitochondrial protein coding gene. A, C, G, T indicate the number of that nucleotide in that gene, followed by the number of sites with that nucleotide that are polymorphic. 's' is the Pearson correlation coefficient of percentages adjusted for differences between transitions and transversions and adjusted nucleotide misincorporations by the gamma polymerase (* indicates P < 0.05). The last lines (from Res on) and s' are explained in the text.

## 6. Mutation gradients across mitochondrial genomes

The previous section indicates that some mutations might be systematically more frequent than expected by misincorporations by the gamma polymerase, and suggests that mutations due to replicational deamination gradients could cause this effect. The study of mutational gradients has used different methods to compare mutation rates at different locations in the genome. Some studies infer mutation rates from phylogenetic comparisons among species of nucleotide contents at given sites (i.e. [13, 14]). Phylogeny-inferred kinetics for A->G and C->T gradients match the properties of the underlying chemical processes: the chemically faster C->T deamination saturates faster in computational analyses with duration spent single stranded than the slower A->G reaction [13, 14, 21]. Other studies infer mutation rates from gene nucleotide contents: for the C->T deamination, one expects relatively high C and low T contents in regions close to replication origin(s), and the opposite for genes with high durations spent single stranded [11, 12, 15, 21, 22, 23].

The method used here is closer to direct observation of mutations, because it compares only between genomes from the same species (*Homo sapiens* in this case). This means that one is closer to an 'instantaneous' observation of mutations. This procedure decreases numbers of undetected multiple changes. I did not use a full phylogenetic model of all human mitochondria to infer mutation rates. Data in Table 1 are for a simplified procedure that counts numbers of sites within a gene where a given type of mutation was observed and calculates the percentage of sites with that nucleotide where that mutation occurred, assuming that the most common nucleotide at any given site is the ancestral nucleotide.

Durations spent single stranded are calculated as previously [11, 12, 21, 22, 24]. I explored for replicational and transcriptional gradients (Dssh and Dloop in Table 1) for each of the 12 mutations, not only for A->G and C->T. This is because time spent single stranded might also affect other mutations, notably transversions [13]. Correlational analyses for gradients (analysis across rows, one per column in Table 1) used the residuals of mutation rates from their regression with misincorporation rates (residual analysis is across columns, one regression calculated per gene/row in Table 1), in order to exclude effects of polymerase inaccuracy on mutational gradients. However note that using the raw mutation percentage data as in Table 1, gradient analyses do not change much.

Two potential gradients in duration of singlestrandedness are considered, singlestrandedness during replication and during transcription (indicated in Table 1 by Dssh and Dloop, respectively). The last rows in Table 1 show Pearson correlation coefficients between residual mutation percentages and times spent single stranded during replication (Dssh) and transcription (Dloop). The hypothesis of singlestrandedness expects positive correlations, but this was observed only for half the cases, for each replication and transcription. There was a significant drop in T->A mutations along the replicational gradient, and a significant increase in T->C mutations along the transcription gradient (Figure 3). The latter effect is predicted by deamination gradients. Deamination gradients are also expected for G->A, but were not observed. Data in Table 1 only support the hypothesis of a deamination gradient for T->C. They cannot differentiate between replicational and transcriptional gradients. It is

notable that in this case the predicted G->A gradient is not stronger than gradients observed for other mutations. Apparently, another mutation, T->A, reacts to single strandedness, but in the direction opposite to that expected (singlestrandedness is predicted to increase mutations, not decrease them). Other, less direct methods based on phylogenetic reconstructions, perhaps fail to detect this gradient because selection, at larger evolutionary scale, might have weeded out many mutations such as the transversion T->A (this type of mutation implies non-conservative changes at the amino acid coding level), leaving mainly neutral and close to neutral ones. Indeed, transitions affect less coding properties than transversions (transitions cause on average more conservative amino acid changes than transversions). This would explain why phylogenetic comparisons detected weaker signals for transversion than transition gradients, while analyses in Table 1 for almost instantaneous mutations are apparently less affected by natural selection occurring after a mutation happened and do not show differences in gradients between transitions and transversions. These comparative data restricted to *Homo sapiens* confirm only the (heavy strand) deamination of A->G (corresponding to T->C in the annotation used here) at the level of a transcriptional gradient.
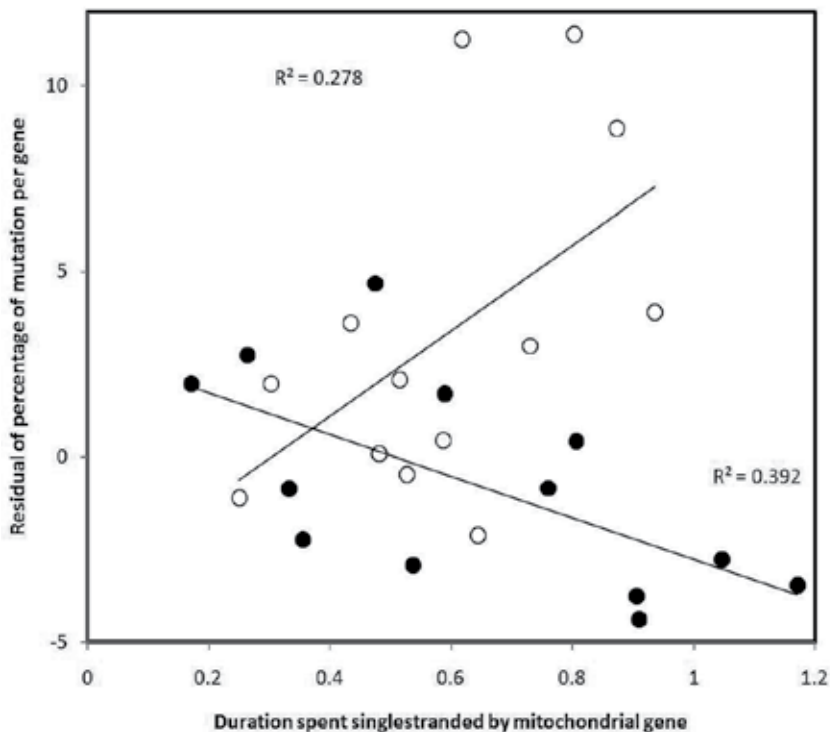


**Figure 3.** Mutations versus singlestranded during replication (T->A, filled symbols) and transcription (T->C, circles). Mutation percentages are residuals from regressions with misincorporation by the gamma polymerase, calculated based on data from Table 1.

The results suggest that mutation rates estimated from sequence comparisons within a single species reflect misincorporation rates, but barely confirm well established observations

of deamination gradients, which were based on comparisons between evolutionary more distant sequences, and on nucleotide contents of single sequences. Apparently, instantaneous mutation rates reflect misincorporation by gamma polymerase, while the effects of deamination gradients, which result from a biased cumulation of mutations, might result from long term processes and are therefore more detectable at a wider evolutionary scale.

## 7. Mutation gradients and selection at different codon positions

The issue of effects of selection on mutational gradients can also be investigated by analysing separately codon positions, as indicated for replicational and transcriptional gradients in Table 1. In terms of replicational gradients, there were no gradients detectable for any mutation at first and second codon positions, but there were three gradients, one negative (G->A) and two positive (A->T and G->C) at third codon positions. Hence these analyses confirm that replication gradients are more detectable where the mutation is synonymous or has little impact because causing a conservative amino acid change, as occurs at third codon positions, but not or much less at first and second codon positions. However, the specifically predicted deamination gradients are not detected. The opposite is observed for G->A (corresponding to C->T mutations on the heavy DNA strand), this mutation unexpectedly decreases along the singlestrandedness gradient, while an increase was expected.

Assuming a transcriptional gradient in singlestrandedness, the expected positive G->A gradient is detected for first codon positions. This is the only statistically significant gradient detected that is not at third codon position. The transcriptional gradient analyses at third codon position confirm the gradient observed for pooled codon positions for T->C, which fits the deamination gradient, and detects a gradient for G->C mutations.

Comparing the absolute values of the correlation coefficients in Table 1 for replicational and transcriptional gradients, correlations are stronger with transcriptional singlestrandedness, however this analysis does not account for the expected positive direction of the correlations of mutations with singlestrandedness. If one assumes that correlations should be positive (singlestrandedness should increase mutations), one does not detect any systematic difference between replication and transcription. The human mutation data might be better explained by transcriptional singlestrandedness, but the matter remains unclear. Deamination gradients are more detectable assuming transcriptional than replicational singlestrandedness, suggesting that deaminations observed in human sequences occurred mainly during transcription. The fact that more gradients are detected at third codon positions than at other positions indicates that selection against mutations affecting protein structure occurs and prevents detecting mutational gradients due to singlestrandedness.

## 8. Mutation gradients and misincorporations

Analyses in the previous section suggest that mutational gradients exist in mitochondria, but are less detectable at the evolutionary scale reflected by sequence variation within *Homo*

*sapiens* populations than when comparing between evolutionary more divergent sequences belonging to different species. Nevertheless, additional analyses show that replicational gradients confound effects of misincorporation by the gamma polymerase. Indeed, the column 's' in Table 1 shows that while mutation patterns in most genes overall fit the pattern predicted by misincorporation, this extent varies widely among genes (from -5 for ND2 to 73 for AT8). My first guess was that gene size (from 69 to over 600 codons, for AT8 and ND5, respectively) differences cause this. My assumption was that estimations of mutation patterns are less accurate in short genes, causing low correlations (low s) between observed mutation patterns and misincorppration rates. However, if this was true, one would expect a better match with misincorporation patterns in long genes, but surprisingly, patterns fit best in AT8: sampling inaccuracy does not explain variation in 's'.
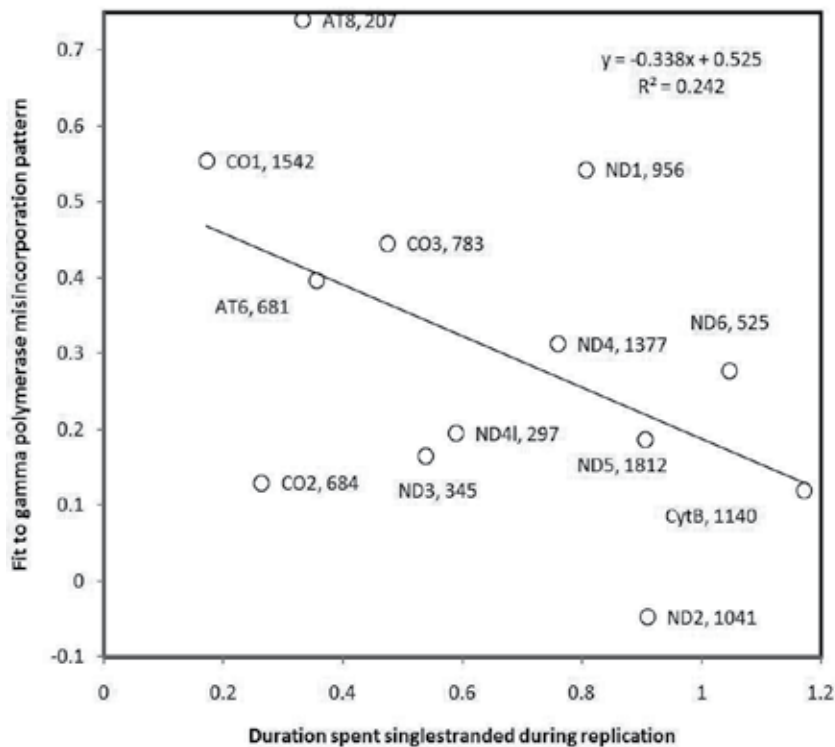


**Figure 4.** s from Table 1 as a function of singlestrandedness during replication. Mutation patterns resemble those predicted by misincorporation by the gamma polymerase in genes that remain singlestranded for a short time during replication. Values indicate gene lengths.

Replicational mutation gradients might explain variation in s between genes: mutation patterns in genes that endure short periods of singlestrandedness during replication should be least affected by replication gradients, and fit best the pattern predicted by gamma polymerase misincorporation, and vice versa (Figure 4). Indeed, s decreases with singlestrandedness during replication ($r = -0.49$, $P = 0.045$, one sided test; but there was no correlation of s with singlestrandedness during transcription, $r = -0.27$, $P < 0.10$).

Inaccurate 's' estimation due to short genes affects results in Figure 4. Short genes fit less the trend in Figure 4 than large genes (gene size is indicated in Figure 4): absolute values of residuals calculated from the regression in Figure 4 decrease with gene size (r = -0.45). Hence 21 percent of variation in s unexplained by singlestrandedness is from sampling effects. Accounting for them, the correlation in Figure 4 is r = -0.63. This means that sampling effects affect less 's' (and estimates of observed mutation rates) from Table 1 than singlestrandedness. This stresses the importance of mutational gradients despite weak results in Table 1.

Singlestrandedness during replication is an even better predictor of the fit between observed mutation patterns and gamma polymerase misincorporation when residual analyses account for each Grantham distances between replaced and replacing amino acids (s' in Table 1). This s' decreases more than s with replicational singlestrandedness (r = -0.6277, one tailed P = 0.011). Interestingly, using transcriptional singlestrandedness yields r = -0.468 (one tailed P = 0.053). Accounting for total singlestrandedness during both replication and transcription by summing both up and analysing the correlation of s' with this sum of replicational and transcriptional singlestrandedness yields r = -0.649 (one tailed P = 0.0083). In each of these analyses using s', gene size had a significant impact on residuals. Accounting for that effect systematically increased correlations between s' and replicational, transcriptional, and the combination of both singlestrandedness (r = -0.811, r = -0.68 and r = -0.89).

## 9. Mutation patterns: effects of dipole moments or gamma polymerase misincorporations?

Figure 2 shows that even after accounting for differences between transitions and transversions on misincorporation rates, differences between dipole moments of the substituted and the substituting nucleotides explain part of the variation in misincorporation rates. Hence both factors (dipole moment or misincorporation by the gamma polymerase) are confounded, and one cannot be sure which affects mutation patterns, or whether they affect each independently observed mutation patterns. For that reason I calculated residuals of adjusted kds from the regression with signed differences in dipole moments (data from Figure 2) and calculated correlations between these residuals and the adjusted mutation percentages (calculated from Table 1) for each gene. This version of s is adjusted for effects of dipole moments on misincorporation rates, and is positive for all genes. This adjusted s increased as compared to s from Table 1 in 8 (and decreased in 5) genes. Hence adjusting misincorporation for effects of dipole moments only slightly increases its fit with observed mutation patterns.

However, when examining the increase in s after adjusting for dipole moment effects in relation to replicational singlestrandedness, this increase is proportional to singlestrandedness during replication (not shown). This suggests that effects of the component of misincorporation that is independent of dipole moments increase with singlestrandedness. Hence singlestrandedness interacts with gamma polymerase fidelity. In these analyses, this fidelity is separated into a component associated with dipole moments, and a different component.

According to the analyses, it is the latter, unknown factor that increases its effects on observed mutation patterns with singlestrandedness during replication.

Similar residual analyses for dipole moments show that observed mutation patterns do not fit well with differences in dipole moments after calculating residuals from their regression with misincorporation rates (these analyses inverse between dependent and independent in Figure 2). These correlations were negative in 11 among 13 genes, suggesting a weak effect that is opposite to that expected by the hypothesis that mutations decrease dipole moments [4].

The latter analyses indicate that dipole moments affect mutation rates through their effects on misincorporation by polymerases, but not directly on spontaneous alterations of single stranded DNA. Misincorporation by gamma polymerase has at least two components, one related to dipole moments, and another one, unrelated to dipole moments. Effects of the latter on mutation patterns increase with singlestrandedness. Analyses in a previous section suggested that distinguishing between misincorporation due to nucleotide misrecognition versus misincorporation due to nucleotide alteration after accurate recognition could prove valuable. It is not clear whether the effect independent of dipole moments that increases with singlestrandedness relates to misrecognition, alteration after recognition, or a subcomponent of any of these. Hydrophobic bias (for low dipole moment) in relation to misincorporations by the gamma polymerase binding site for nucleotides is not explained by a simplistic analysis of the residues composing the active site of gamma polymerase. Nevertheless, these results indicate that the 'age' of the replication fork has some effect on its fidelity.

A similar comparison can be done between s and s' in Table 1. Here one sees that s', as compared to s, is lower than s in 11 among 13 genes. Hence gene-specific mutation patterns match misincorporation rates after accounting for differences between transitions and transversions better than after accounting, in addition, for selection against non-conservative amino acid replacements resulting from nucleotide substitutions: s' as compared to s decreases with singlestrandedness. Hence in this case, accounting for selection against non-conservative amino acid replacements improves slightly the match of observed mutations with misincorporations for genes with short singlestranded exposure, but mainly decreases that match for those with long singlestrandedness. This effect is opposite to the one reported in a previous paragraph for accounting for dipole moment effects. Accounting for the latter improves the match between mutation and misincorporation patterns with singlestrandedness, while accounting for selection decreases that match.

## 10. More evidence for complex indirect effects of mutation gradients

Sampling inaccuracy might affect estimates of s and s' in Table 1. A further potential indirect factor with opposite effect might exist. Duration spent single stranded used is for a gene's midpoint, a good approximation for short genes, but increasingly inaccurate the longer the gene. In order to evaluate this, absolute residuals of mutation percentages (Table 1) from their regressions with misincorporation rates (both adjusted for differences between transitions and transversions) are plotted versus numbers of potential sites that could mutate for that mutation type

in that gene (Figure 5). Residuals tend to decrease with sample sizes up to approximately 250 nucleotides, which corresponds to an average of a sequence of 1000 base pairs (for each muta‐ tion type, there is only one substituted nucleotide, so on average, these mutations occurred over a total sequence that is about four times longer). The absolute value of residual mutation percentages increases with sample size up from about 250 nucleotides.
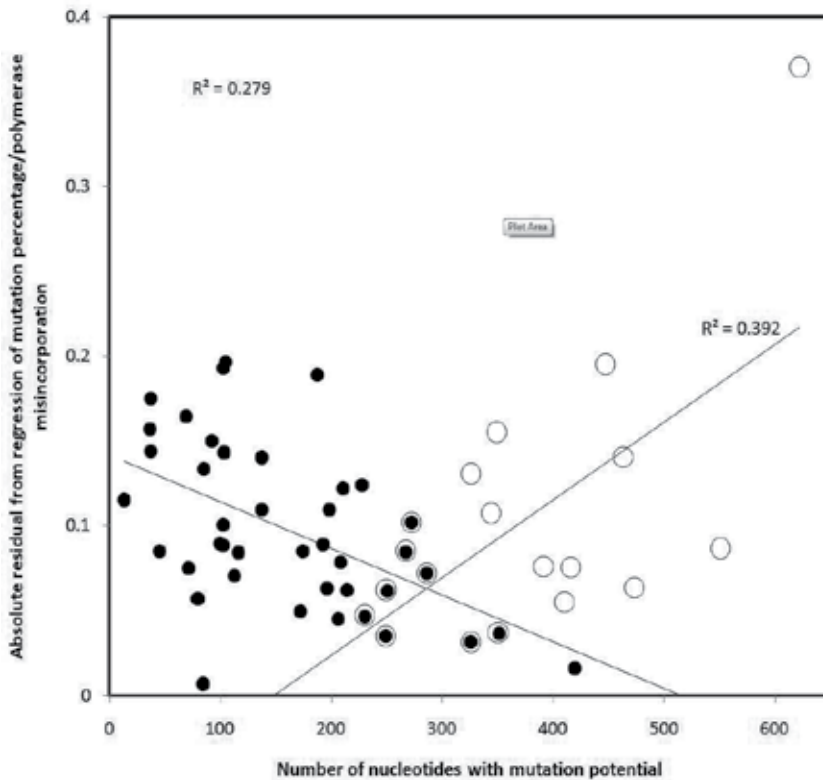


**Figure 5.** Residual mutation percentage (absolute value) from regression with gamma polymerase kds for each nu‐ cleotide in each gene, versus numbers of potentially mutating nucleotides. The decrease indicates a sampling effect: samples up to 200-300 nucleotides fit better misincorporation patterns because of sampling effects. Beyond 250, in‐ accuracy increases, perhaps because different gene regions have different mutation regimes.

The decreasing pattern is what one expects from sampling effects: up to about 1000 base pairs, longer genes enable to estimate better mutation patterns (the absolute residual is small). But for genes longer than that threshold, absolute residuals increase, hence mutation patterns tend to fit less well misincorporation as predicted by the gamma polymerase. This could be due to the mixing of regions with different singlestrandedness, which perhaps al‐ ters non-linearly mutation patterns. A similar effect where estimation inaccuracy of muta‐ tion rates decreases, then increases with sequence length exists for the correlation between rates of morphological and molecular evolution [25]. The threshold was for sequence lengths around 1200 base pairs, indicating that estimates of mutation rates (mainly from ver‐ tebrate mitochondrial protein coding sequences, as those analysed for *Homo sapiens* here) de‐

creased beyond that sequence length. It was suggested, as for Figure 5 here, that mutation patterns change with the relative position of a gene, and that for long regions, more than one mutation regime might be mixed, decreasing the accuracy of analyses. Figure 5 follows that principle, and indicates a similar threshold.

## 11. Mutational gradients after accounting for amino acid replacement impacts on proteins

Previous sections show that indirect effects of gradients in singlestrandedness on mutation patterns exist (i.e. Figures 4 and 5). Yet analyses of mutation percentages, or mutation percentages after accounting for differences between transitions and transversions, and after accounting for effects of misincorporation by gamma polymerase, do only marginally enable to detect mutation gradients with singlestrandedness, and this for any codon position. The analyses of gradients that separate codon positions indicate that natural selection might affect mutation patterns (Table 1), and could mask mutational gradients according to singlestrandedness. Selection against non-conservative amino acid replacements also affects mutation percentages. Analyses for singlestrandedness gradients did not yet account for that latter factor, in addition to misincorporation by the gamma polymerase and differences between transitions and transversions.

I calculated residuals of mutation percentages (adjusted for differences between transitions and transversions) from their regression with mean physico-chemical (Grantham's) distances between replaced and replacing amino acids resulting from that nucleotide substitution in coding sequences (for mutation percentages across all codon positions), separately for each of the protein coding genes. Hence this analysis is across columns, for each row in Table 1. Then, for each substitution type, I calculated correlations with singlestrandedness during replication, transcription, and their sum (these analyses are across rows, for each column, on residuals produced by the latter 'row' analysis across columns). This yields correlations between residual mutation rates and singlestrandedness for each mutation type. The majority of these are positive correlations (Table 2): mutation percentages (after accounting by residual analyses for differences between transitions and trasversions, misincorporation rates and Grantham distances (assumed to reflect selection against dysfunctional proteins)) increase with singlestrandedness during replication (11 among 12 cases, exception A->C mutations), transcription (11 among 12 cases, exception A->G mutations) and their sum (all cases). Hence overall, singlestrandedness promotes all types of nucleotide substitutions, not only deaminations A->G and C->T, for both replicational and transcriptional singlestrandedness. Their sum improves correlations in half the cases. Correlations were statistically significant (one tailed $P < 0.05$) for one correlation with replicational singlestrandedness (T->C), two with transcriptional singlestrandedness (C->G and T->C) and three with the sum of both (A->T, T->C and T->G).

Correlations were stronger with transcriptional singlestrandedness than replicational singlestrandedness in 7 among 12 cases, which does not indicate which among the two is the most

important factor. Possibly, singlestrandedness during replication and during transcription affect differently different substitution types, or differences are random. These analyses clearly show that after accounting for mean effects of substitutions on proteins, percentages of all types of nucleotide substitutions increase with singlestrandedness during each replication and transcription. These clear patterns were not detectable without accounting for mean nucleotide substitution impact on physico-chemical properties of coded amino acids. It seems these effects prevented detecting mutation gradients for substitutions that were not deaminations. Singlestrandedness increases at least slightly probabilities of all types of substitutions.

| Substitution | Rep | Trans | Both |
|---|---|---|---|
| A->C | -0.159 | 0.401 | 0.066 |
| A->G | 0.428 | -0.242 | 0.241 |
| A->T | 0.400 | 0.447 | 0.481* |
| C->A | 0.378 | 0.063 | 0.291 |
| C->G | 0.319 | 0.468* | 0.433 |
| C->T | 0.243 | 0.315 | 0.312 |
| G->A | 0.091 | 0.339 | 0.216 |
| G->C | 0.395 | 0.322 | 0.421 |
| G->T | 0.425 | 0.340 | 0.448 |
| T->A | 0.301 | 0.381 | 0.382 |
| T->C | 0.478* | 0.531* | 0.573* |
| T->G | 0.452 | 0.350 | 0.469* |

**Table 2.** Pearson correlation coefficients of time spent singlestranded during replication, transcription, and their sum versus substitution percentages in the 13 human mitochondrial protein coding genes adjusted for differences between transitions and transversions, misincorporation rates and for mean effect of the substitution on Grantham's physico-chemical distances between replaced and replacing amino acids.

Causes for differences in gradient strengths for different substitution types are not known. Gradients are strongest for substitutions involving a small absolute change in nucleotide dipole moment, and weakest for those where the absolute change in dipole moment is large. Speculatively, large dipole differences may affect even when singlestrandedness is short, so that no strong gradient is detectable, because the main effect is the dipole moment, independently of singlestrandedness. For small dipole moment differences, the dipole moment effect woult hence be enhanced by singlestrandedness, resulting in a gradient.

## 12. Nearest neighbour effects on mutation rates

Previous analyses of mutation patterns in human mitochondrial protein coding genes fit expectations according to several factors: misincorporation by the gamma polymerase, selec-

tion against mutations that alter amino acid properties, and dipole moments of nucleotides. A hierarchy between these factors exists. In addition, they interact: misincorporation rates are also affected by selection against non-conservative mutations; and gradients in single-strandedness affect extents by which the various factors affect mutation patterns. Only after adequate accounting for misincorporation and selection (and differences between transitions and transversions), mutation gradients along durations of singlestrandedness are cleary observed for all types of nulceotide substitutions.

| Flank | | 5' | | | | | | 3' | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | | Tot | Mut | A | C | G | T | Tot | Mut | A | C | G | T |
| | A-> | 968 | 319 | | 15 | 290 | 14 | 928 | 238 | | 19 | 193 | 26 |
| | C-> | 1037 | 289 | 22 | | 10 | 257 | 1069 | 363 | 21 | | 12 | 330 |
| | G-> | 461 | 166 | 150 | 11 | | 5 | 371 | 89 | 79 | 6 | | 4 |
| | T-> | 897 | 278 | 8 | 244 | 26 | | 1275 | 470 | 18 | 395 | 57 | |
| C | | | | | | | | | | | | | |
| | A-> | 1097 | 326 | | 36 | 260 | 30 | 1063 | 447 | | 22 | 401 | 24 |
| | C-> | 1285 | 327 | 41 | | | 270 | 1293 | 390 | 35 | | 34 | 321 |
| | G-> | 311 | 114 | 100 | 8 | | 6 | 505 | 203 | 175 | 20 | | 8 |
| | T-> | 1110 | 245 | 23 | 203 | 19 | | 981 | 286 | 17 | 240 | 29 | |
| G | | | | | | | | | | | | | |
| | A-> | 378 | 158 | | 6 | 145 | 7 | 447 | 229 | | 12 | 204 | 13 |
| | C-> | 503 | 156 | 15 | | 28 | 113 | 322 | 154 | 21 | | 8 | 125 |
| | G-> | 254 | 72 | 60 | 8 | | 4 | 256 | 72 | 68 | 2 | | 2 |
| | T-> | 227 | 83 | 6 | 67 | 10 | | 323 | 115 | 6 | 102 | 7 | |
| T | | | | | | | | | | | | | |
| | A-> | 890 | 425 | | 18 | 373 | 34 | 888 | 303 | | 20 | 271 | 12 |
| | C-> | 823 | 252 | 19 | | 21 | 212 | 1102 | 252 | 57 | | 14 | 181 |
| | G-> | 328 | 137 | 123 | 8 | | 3 | 222 | 124 | 111 | 10 | | 3 |
| | T-> | 676 | 295 | 13 | 264 | 18 | | 668 | 198 | 13 | 171 | 14 | |
| | A | | | | 40 | -54 | 1 | | | | 72* | 40 | 48* |
| | C | | | | | -37 | 36 | | | | | 43 | 60* |
| | G | | | | | 51* | | | | | | | 91* |

**Table 3.** Dinucleotide sites and mutating sites in human mitochondrial protein coding sequences, separating 5' and 3' nucleotide identity. Last 3 lines are correlations, see text.

Despite the relative complexity of factors described and affecting mutation patterns, this is not an exhaustive list of effects on mutation rates. Notably, nearest neighbour effects exist [26], where identities of nucleotide(s) flanking the mutating site affect mutation rates, as in-

dicated by the editor of this volume after reviewing a former version of this chapter. G and C, the nucleotides with the highest dipole moments, seem to increase mutation rates in various organisms along similar patterns [26-30]. This suggests a physico-chemical basis for nearest neighbour effects, possibly along the lines of dipole moment effects and the stability of DNA duplexes surrounding the mutating nucleotide [26]. These biases are strong enough to justify the need of incoporating at least the strongest nearest neighbour effect in models designed to detect natural selection on mutations [31], which is not surprising as CpG dinucleotides are disproportionately represented among sites with pathogenic polymorphisms [32,33]. Moreover, nearest neighbour effects interact with gene location and the frequency of transcription, suggesting interactions with singlestrandedness [34, 35]. Nearest neighbour analysis of mutation patterns requires large sample sizes, and therefore is unfortunately incompatible with a gene by gene analysis as a function of singlestrandedness in the context of this mitochondrial dataset.

However, even after pooling mutation data from all genes, one would ideally examine the twelve substitutions in relation to each of the 16 combinations of nucleotides at the 5' and 3' positions. Such detailed analyses are also not possible with this dataset. Nevertheless, as known to this author, nearest neighbor effects have not yet been examined in the context of mitochondrial genomes, hence even simplified analyses pooling mutations from all genes and codon positions together may still be valuable. In addition, most nearest neighbour analyses examined do not analyse substitutions in relation to their direction (they pool X->Y with X<-Y), but this can be done on this dataset. Mutation data from all genes and codon positions were pooled, and analysed each time separately in relation to the identity of their 5', and their 3' flanking nucleotide. This yields reasonable samples, and the mutation patterns can be compared according to the different flanking nucleotide identities (Table 3).

The data in Table 3 enable a number of different analyses, only one is presented here, though many others are of interest. For example, biases exist in terms of dinucleotide frequencies, between 5' or 3' flanking by the same nucleotide. I focus here on the analysis of mutation patterns. Numbers in each row in Table 3 were divided by the number of mutating sites among all possible dinucleotide sites for that category (Mut). The column Tot in Table 3, which indicates the total number of dinucleotide sites found independently of the occurrence of a mutation at that site, is indicated but not used in further analyses. The substitution matrices that result are very similar, comparing 5' and 3', and different nucleotide contexts. This is because the overwhelming majority of the variation in mutation rates is due to the difference between transitions and transversions. For that reason, effects of transitions versus transversions were accounted for by subtracting observed mutations rates from the average for transitions and transversions, respectvely, as done in previous analyses. Then these data adjusted for differences between transitions and transversions are compared between different substitution matrices, so that effects of the difference between transitions and transversions is accounted for before comparing the matrices with different neighbours.

The three last lines in Table 3 show Pearson correlation coefficients (x100) between these mutation patterns (adjusted for differences between transitions and transversions). Even after accounting for differences between transversions and transitions, substitution patterns

across different 3′ neighbouring nucleotides resemble each other: all six correlations are positive, 4 among these are statistically significant (P < 0.05, one tailed tests because positive associations are expected, see asterisks in Table 3). 3′ G and T had most similar patterns. Hence grouping of mutation patterns according to 3′ nucleotides does not follow purine/ pyrimidine nor dipole moment differences. 3′ G seems to affect most mutation patterns, hence results are probably not random also for 3′ nearest neighbour effects.
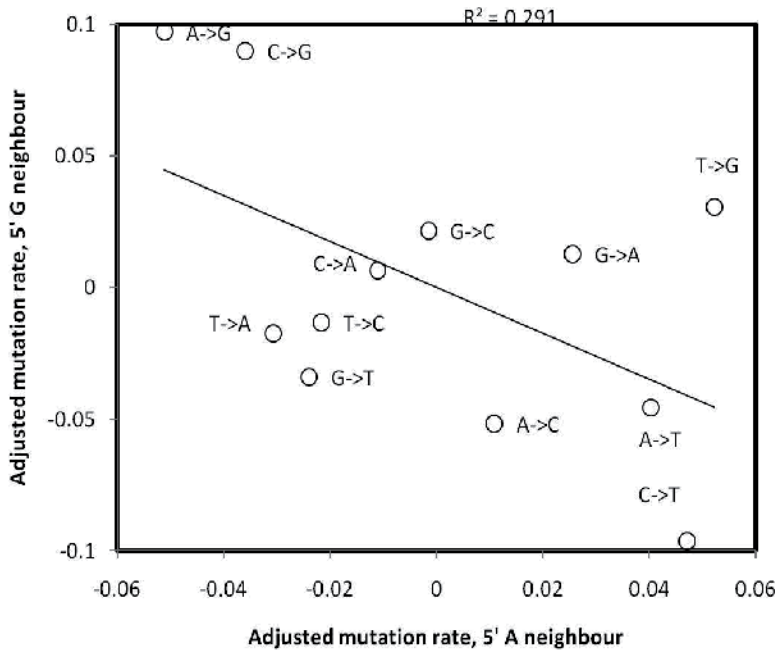


**Figure 6.** Mutation rates adjusted for differences between transition and transversions for 5′ A and G neighbours in human mitochondrial protein coding genes.

The same analysis for 5′ flanking nucleotides reveals a similar, more enhanced situation. Here, the only statistically significant association is between mutation patterns with 5′ G and T as nearest 5′ neighbour. The weak positive correlations in the 3′ context are negative in the 5′ context, one being close to statistically significant (the comparison between 5′ A and G, see Figure 6): 5′ G affects mutation rates in a way that tends to be systematically opposite to what is observed in other contexts, so that relatively high mutation rates become relative low, and vice versa. Effects of 5′ G on mutation rates are expected, considering previous reports. However, these have mainly shown effects on C->T mutations. The results here show that 5′ G has a systematic effect on all mutation types, some increasing, as expected, but others decreasing in the 5′ G context.

It is notable that the correlation matrices for 5′ and 3′ contexts (in the 3 last lines of Table 3) are very similar, if not in their values, but in their pattern: the ranks, from least to most positive correlation coefficients, are identical (Figure 7). This means that the same effects are at

work for 5′ and 3′ flanks, but that effects are stronger for 5′ flanking nucleotides. In this context, it is important to remember that the annotation used here is that of the light strand DNA in the mitochondrion, which bears the coding sequence of most genes. In the elongating light DNA strand, the 3′ nucleotide is already present before the mutating nucleotide is added, while the 5′ nucleotide is not yet there, and could not possibly have any effect. This is not compatible with a 5′ effect during replication, unless one considers that the effect is from the neighbouring nucleotide on the template heavy strand DNA. In that case, the inverse complement would have the major flanking effects, with the strongest effect by the nucleotide that is not yet complemented by the nascent strand (the 5′ of the light strand becomes the 3′ in the heavy strand), and a weaker but similar effect by the neighbouring nucleotide that is already complemented by the replication process. Along that scenario, neighbouring nucleotides would affect misincorporation rates. This scenario would be very compatible with electrostatic effects, due to dipole moments.
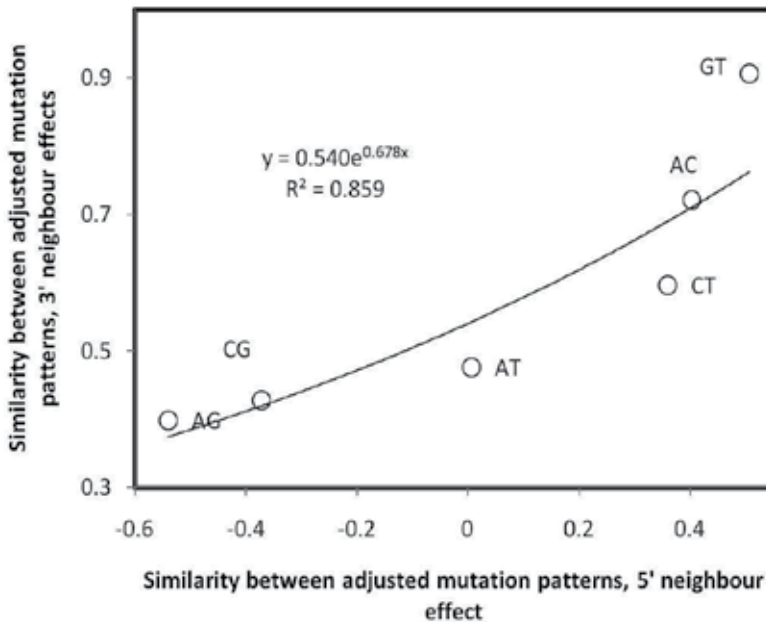


**Figure 7.** Similarities (Pearson correlation coefficients in the three last lines of Table 3) between transition versus transversion adjusted mutation patterns for 3′ neighbouring nucleotides as a function of similarities for mutation patterns found for 5′ neighbouring nucleotides (see Table 3). Letters near datapoints indicate the neighbouring nucleotides whose mutation patterns are compared.

It is notable that the 5′ G mutation pattern is very similar to the 3′ C mutation pattern as these are observed for the light strand (r = 0.87). These are the most similar mutation patterns found when comparing 5′ and 3′ mutation patterns. Because 3′ C on the light strand is 5′ G on the heavy strand, this similarity indicates that the factor at work involves both strands, always involving the 5′ G nucleotide.

Alternative explanations not involving effects on misincorporation rates, such as dipole moments and 'spontaneous' (non-enzymatic) mutations are also very plausible. The latter are more compatible with the similarities in patterns between 5′ and 3′ effects and effects on both strands, but less with the strong directional effect detected (less similar mutation patterns between 5′ than 3′ substitution patterns).

Hence strong neighbouring effects are detected on mutation patterns observed in human mitochondrial genomes, yet their cause remain unknown, and might, as for other effects on mutation patterns, have different physico-chemical causes, combined with some biological factors.

## 13. Dipole moments and retrotranscription rates by the gamma polymerase

The various analyses described show complex effects, most of them are confirmative of phenomena that have already been described. Indeed, it is quite trivial that misincorporation rates affect mutation frequencies, and that these frequencies are decreased by selection against dysfunctional proteins. Gradients in singlestrandedness as affecting mutation rates are also known, though the fact that they affect all or most mutation types is relatively original to the analyses presented here. A similar rationale relates to the original component of the results from the nearest neighbour analyses. However, the fact that so many different factors are jointly considered in the analysis of a single dateset of mutations is not the sole major originality in terms of potential mechanisms explored in this chapter.

The hypothesis that dipole moments affect misincorporation, mutation rates, and mutation gradients, is a major potential novelty. Unfortunately, when its effects are detected, these are not well understood: the main effect on misincorporation is that of absolute dipole moment change, and the bias favouring low dipole moments remains unexplained.

The suggestion by David Stuart, the editor of this volume, to examine associations between dipole moments and elongation rates in relation to the inserted nucleotide [36] yields interesting results in this respect, confirming that dipole moments affect the incorporation rate of nucleotides into nascent DNA. Results below indicate complex mechanisms, and should be considered as preliminary and with extreme caution. First, it seems that kms of incorporations of nucleotides increase with dipole moments, as one would expect if high dipole moments enable quick processing by the (charged and hydrophilic) active site of the gamma polymerase, though this effect is not statistically significant at $P < 0.05$.

However, electron singlets and or triplets of molecules can be in an 'excited' state, which modifies the dipole moment of the molecule, as calculated by Bergmann and Weiler-Feilchenfeld [6] (therein table VII) for nucleotides. Dipole moments for the excited triplet state correlate positively with nucleotide insertion rates ($r = 0.9865$, $P = 0.007$, one tailed test). Considering that more than one correlation test was done (for the regular and the two excited dipole moments), this result is not statistically very strong (especially that only 4 datapoints are involved in the analysis).

I assume that implicitly, the hypothesis developed by the editor, following my initial inter‐
est in effects of dipole moments on polymerase activity, is that if dipole moments affect nu‐
cleotide incorporation rates, discrimination against incorporation of the much more
common ribonucleotides should associate negatively with (deoxyribo)nucleotide dipole mo‐
ment. Indeed, the activity of the gamma polymerase as measured by Kasiviswanathan and
Copeland [36] includes also the (mis?) incorporation rates of ribonucleotides on the template
of DNA, and these associate negatively with nucleotide triplet excited dipole moment (r = -
0.963, one tailed P = 0.019). In addition, the rate of reverse transcription by the gamma poly‐
merase, where deoxyribonucleotides are inserted on the template of (mis)inserted
ribonucleotides correlates positively with the mean of the singlet and triplet excited dipole
moments (r = 0.99984, one tailed P = 0.00008, see Figure 8).These analyses yield notable re‐
sults, though they are not necessarily as statistically robust as they seem, due to the low
number of degrees of freedom (only four datapoints). In addition, correlations between each
of the kms and each the dipole moment, the singlet and excited, and their average was cal‐
culated, in total 12 correlations. In these cases, according to a strict Bonferroni criterion to
correct for multiple testing, to keep P <0.05 while testing 12 times a hypothesis, one should
use the threshold of P = 0.05/12= 0.0042. According to that often overconservative criterion,
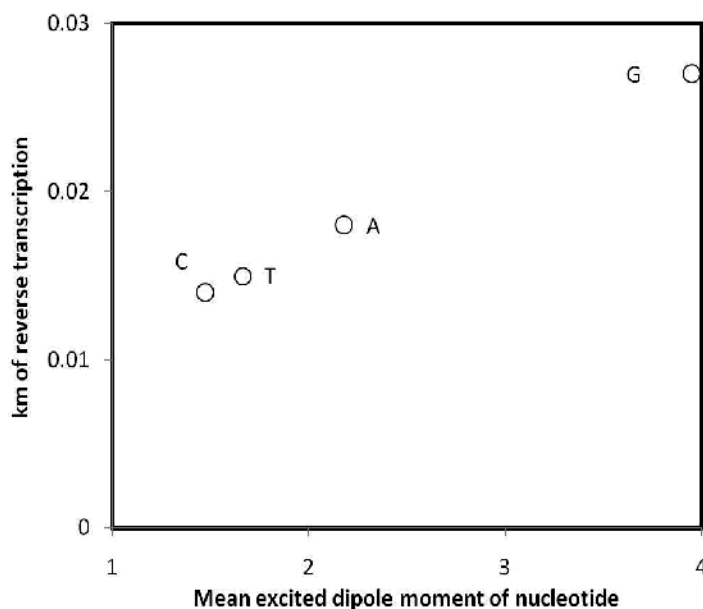only the result in Figure 8 remains statistically significant.



**Figure 8.** Rate of deoxyribonucleotide 'reverse' incorporation as a function of its mean excited dipole moment on the template of ribonucleotide.

The data nevertheless confirm the hypothesis that nucleotides are processed on the basis of
their dipole moment, where nucleotides with high dipole moments are more rapidly cor‐
rectly processed. This result might actually explain also the results obtained in earlier sec‐

tions on associations between dipole moment changes and nucleotide misrecognitions. The rate of a process, and its accuracy are frequently negatively associated. Hence if correct incorporation is proportional to dipole moments, misincorporation might be (as observed) inversely proportional, explaining that patterns in Figures 1 and 2 are opposite to predictions: the hydrophilic active site will handle correctly more rapidly a nucleotide with high dipole moment, and more probably mishandle a nucleotide with low dipole moment.

## 14. General discussion

The analyses presented above show that mutation patterns estimated from the simple comparison between sequences from a species confirm the patterns expected from experimentally determined misincorporation rates for the gamma polymerase. This is an important confirmation that comparative analyses yield trustable estimates of mutation patterns and rates. Analyses support, to lesser extents, that mutation patterns across genes are determined by durations spent singlestranded, and suggest that in order to detect such effects, comparisons involving longer evolutionary time spans than those implied by separations between different individuals from a single species are required to detect the cumulation of mutations due to singlestrandedness.

Grantham's physico-chemical distances between replaced and replacing amino acids affect misincorporation rates by the gamme polymerase, and percentages of mutations observed in protein coding genes. This suggests that natural selection to conserve protein function affects each of these two different patterns. Gradients of mutations with singlestrandedness are barely detectable without controlling for effects of Grantham distances on mutation percentages, several indirect effects are observable that indicate interactions between singlestrandedness and misincorporation patterns by the gamma polymerase. After the effects of Grantham distances on mutation percentages are accounted for by residual analyses, the expected increase in mutation percentages with singlestrandeness becomes detectable in all types of substitutions. This is a notable result, because singlestrandedness was believed until now to affect only or mainly substitutions due to deaminations (A->G and C->T). Analyses do not succeed to indicate which of replicational and transcriptional singlestrandedness is most relevant to predict mutations. Further chemical processes accounting for effects of singlestrandedness on mutation types besides deaminations have to be investigated and suggested.

It seems that gamma polymerase misincorporation patterns change with single strandedness, which may reflect the duration of activity by the replication fork's molecular 'machinery'. Only molecular experiments much more developed than those used until now could yield such results. This shows that combined analyses of bioinformatic and experimental data enable to suggest the existence of previously unknown biochemical phenomena.

Further points raised are the involvement of nucleotide dipole moments in the interactions between the nucleotide and the gamma polymerase. Analyses at this point do not yield much information beyond the fact that such effects occur. More functional hypotheses could help in this respect.

The data on mutation patterns from *Homo sapiens* do not enable to establish whether mutations cumulate during singlestrandedness due to transcription or replication. It is probable that the ratio between these two types of events that open double stranded DNA, changes with the longevity of an individual/species, where greater lifespan increases the transcription component [11]. It is probable that analyses similar to those done here based on ample mitochondrial sequence data available for other mammal species with shorter lifespans could help in this respect. Comparing results from different species would probably be fruitful. In addition, these analyses could preliminarily reveal whether misincorporation patterns by the gamma polymerases of the different species differ. This could be an exciting line of research, that could potentially link differences in mutation patterns with differences in the gamma polymerases from these species. Such analyses could yield a workable model for the efficiency and fidelity of gamma polymerase in relation to its detailed structure. It is notable that much information necessary for such analyses is already available online and only awaits the interest of enthousiastic students of molecular biology.

Nearest neighbour effects as detected for mitochondrial mutation patterns confirm what is known from previous studies on nuclear chromosomes. They also show that the 5′ G effect on mutation rates is more complex, as it affects differently different types of mutations. Unfortunately, nearest neighbour analyses require samples that are not compatible with the data at hand, so that its analysis in combination with other factors could not be done. The fact that nearest neighbour effects tend to increase (though marginally so), with the thermodynamic stability of the DNA duplex where these neighbouring effects occur, is in itself compatible with dipole moment effects as the causes for the nearest neighbour effects on mutations because nucleotide dipole moments predict duplex thermodynamic stabilities [37]. It is possible that the direct cause for this is thermodynamic stability, through the fact that regions forming stable duplexes are more able to tolerate a misinserted nucleotide. But the association between nearest neighbour effects and stability is weak, indicating that another, associated factor is at work. Possibly, it is the electrostatic effect of nucleotide dipole moments of neighbouring nucleotides on the fidelity of the gamma polymerase that causes these effects. Such effects are particularly probable, considering that the gamma polymerase active site includes two charged residues, and that nucleotide processing seems to depend to some extent on the nucleotide's dipole moments. Hence nearest neighbour effects could be due to interferences between the electrostatic fields of the active site, the incorporated nucleotide, and the nearest neighbours, especially when these nearest neighbours have high dipole moment.

Beyond effects of nucleotide dipole moments on incorporation rates, results suggest natural selection decreasing nucleotide misincorporations with high impact on protein structure. These are encouraging results that could yield further insights if similar analyses are applied to different types of polymerases.

Variation in mutation patterns for genes with different locations along singlestrandedness gradients might have explanations that differ from the ones suggested. The genome structure might be designed so that genes that cannot afford, from a functional point of view, large mutation rates, are located so as to endure little singlestrandedness. It is important to remember that this factor might interact with the results presented. It is also important to

remember that natural selection probably affects observed mutation frequencies in ways not accounted for by presented analyses. This effect might be weaker in genes located far from replication origins, and hence probably more able to tolerate mutations. Hence observed mutations would in these cases much more reflect the original processes, not confounded by effects of natural selection due to gene function.

A further point relates to the patterns observed in Figure 5, where gene length seems to affect the mutation pattern. A possible factor here is the capacity of longer sequences to form more secondary structures by self-hybridization. Considering that secondary structure protects against mutations due to singlestrandedness, this factor could hence indirectly affect mutation patterns, especially in longer genes, assuming that in some ways, genes are replicated as functional units, a possibility that cannot be ruled out *a priori*, especially if secondary structure formation is designed to involve a gene as a unit, for example in the mRNA [38]. It is also possible that secondary structures affect the function of the gamma polymerase, causing differences in misincorporation patterns between regions forming more or less secondary structures, as previous analyses possibly indicated [13, 14].

An important point to stress here is that the data that are available at this point do not limit our capacities to analyses, along multiple dimensions, the various factors that cause mutation patterns, and understand their details in relation to these factors. The computational power and statistical tools are also not limiting and close to adequate. The limiting factor is the time invested by the adequately skilled manpower, or more correctly, the financial investment to support such activity based mainly on analysing valuable molecular data of different types.

## 15. Conclusions

Combined analyses of comparative sequence data and experimentally determined gamma polymerase misincorporation data, together with models for substitutions based on nucleotide dipole moments and models for substitution impacts on protein structure reveal that observed human mitochondrial protein coding gene mutation patterns are affected in decreasing order of importance by gamma polymerase misincorporation rates, selection against non-conservative amino acid replacements, and gradients in singlestrandedness during replication and transcription. Gamma polymerase misincorporation rates are selected to optimize effects of substitutions on non-conservative amino acid replacements, and favour nucleotides with low dipole moments, suggesting hydrophobic bias in nucleotide misbinding. Further analyses confirm this: the hydrophilic active site of the gamma polymerase handles faster nucleotides with high dipole moment and mishandles more often those with low dipole moment, suggesting that process accuracy limits its rate. The wealth of results confirms known and expected patterns, and expands beyond them, revealing selection on polymerase fidelity, and spontaneous tendencies during single stranded DNA states for all substitutions, not only those previously known to react to singlestrandedness.

## Author details

Hervé Seligmann[*]

Address all correspondence to: podarcissicula@gmail.com

The National Collections of Natural History at the Hebrew University of Jerusalem, Israel

## References

[1] Ingman, M., & Gyllenstein, U. (2006). mtDB: Human mitochondrial genome data-base, a resource for population genetics and medical sciences. *Nuc Acids Res*, 34, D749-D751.

[2] Ruiz-Pesini, E., Lott, M. T., Procaccio, V., Poole, J., Brandon, M. C., Mishmar, D., Yi, C., Kreuziger, J., Baldi, P., & Wallace, D. C. (2007). An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nuc Acids Res*, 35(Database issue: D823-D828).

[3] Kaguni, L. S. (2004). DNA polymerase gamma, the mitochondrial replicase. *Ann Rev Biochem*, 73, 293-320.

[4] Seligmann, H. (2006). Error propagation across levels of organization: From chemical stability of ribosomal RNA to developmental stability. *J Theor Biol*, 242, 69-80.

[5] Li, W. H., Wu, C. I., & Luo, C. C. (1984). Nonrandomness of point mutation as reflect-ed in nucleotide substitutions in pseudogenes and its evolutionary implications. *J Mol Evol*, 21, 58-71.

[6] Bergmann, E. D., & Weiler-Feilchenfeld, H. (1973). The dipole moments of purines and pyrimidines. Chapter 1. *In: Duchesne J (ed) Physico-chemical properties of nucleic acids*, I., Academic Press, London.

[7] Frederico, L. A., Kunkel, T. A., & Shaw, B. R. (1990). A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochem*, 29, 2532-2537.

[8] Frederico, L. A., Kunkel, T. A., & Shaw, B. R. (1993). Cytosine deamination in mis-smatched base-pairs. *Biochem*, 32, 6523-6530.

[9] Francino, M. P., Chao, L., Riley, M. A., & Ochman, H. (1996). Asymmetries generated by transcription-coupled repair in enterobacterial genes. *Science*, 272, 107-109.

[10] Francino, M. P., & Ochman, H. (1993). Deamination as the basis of strandasymmetric evolution in transcribed Escherichia coli sequences. *Mol Biol Evol*, 18, 1147-1150.

[11] Seligmann, H. (2011). Mutation patterns due to converging mitochondrial replication and transcription increase lifespan, and cause growth rate-longevity tradeoffs. *In: Seligmann H. (ed.) DNA Replication-Current Advances*, Rijeka, InTech, Chapter 6, 151-180.

[12] Seligmann, H. (2012). Coding constraints modulate chemically spontaneous mutational replication gradients in mitochondrial genomes. *Curr Genomic*, 13, 37-54.

[13] Krishnan, N. M., Seligmann, H., Raina, S. Z., & Pollock, D. D. (2004). Detecting gradients of asymmetry in site-specific substitutions in mitochondrial genomes. *DNA & Cell Biol*, 23, 707-714.

[14] Krishnan, N. M., Seligmann, H., Raina, S. Z., & Pollock, D. D. Phylogenetic analysis of site-specific perturbations in asymmetric mutation gradients. *In: A. Gramada, and P.E. Bourne (eds.) Currents in Computational Molecular Biology*, ACM Press, San Diego, CA, 266-267.

[15] Seligmann, H., Krishnan, N. M., & Rao, B. J. (2006). Possible multiple origins of replication in primate mitochondria: alternative role of tRNA sequences. *J Theor Biol*, 241, 321-332.

[16] Madariaga, S. T., Contreras, J. G., & Seguel, C. G. (2005). Interaction energies in non Watson-Crick pairs: an ab initio study of G U and U U pairs. *J Chil Chem Soc*, 50, 435-438.

[17] Gojobori, T., Li, W. H., & Graur, D. (1982). Patterns of nucleotide substitution in pseudogenes and functional genes. *J Mol Evol*, 18, 360-369.

[18] Lee, H. R., & Johnson, K. A. (2006). Fidelity of the human mitochondrial DNA polymerase. *J Biol Chem*, 281, 36236-36240.

[19] Lee, Y. S., Kennedy, W. D., & Yin, Y. W. (2009). Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations. *Cell*, 139, 312-324.

[20] Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. *Science*, 185, 862-864.

[21] Seligmann, H. (2008). Hybridization between mitochondrial heavy strand tDNA and expressed light strand tRNA modulates the function of heavy strand tDNA as light strand replication origin. *J Mol Biol*, 379, 188-199.

[22] Seligmann, H., Krishnan, N. M., & Rao, B. J. (2006). Mitochondrial tRNA sequences as unusual replication origins: pathogenic implications for Homo sapiens. *J Theor Biol*, 243, 375-385.

[23] Seligmann, H. (2010). Mitochondrial tRNAs as light strand replication origins: similarity between anticodon loops and the loop of the light strand replication origin predicts initiation of DNA replication. *Biosystems*, 99, 85-93.

[24] Tanaka, M., & Ozawa, T. (1994). Strand asymmetry in human mitochondrial mutations. *Genomics*, 22, 327-335.

[25] Seligmann, H. (2010). Positive correlations between molecular and morphological rates of evolution. *J Theor Biol*, 264, 799-807.

[26] Krawczak, M., Ball, E. V., & Cooper, D. N. (1998). Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. *Am. J. Hum. Genet.*, 63, 474-488.

[27] Zhao, Z., & Boerwinkle, E. (2002). Neighboring-nucleotide effects on single nucleotide polymorphisms: a study of 2.6 million polymorphisms across the human genome. *Genome Res*, 12, 1679-1686.

[28] Zhanga, F., & Zhao, Z. (2004). The influence of neighboring-nucleotide composition on single nucleotide polymorphisms (SNPs) in the mouse genome and its comparison with human SNPs. *Genomics*, 84, 786-796.

[29] Zhao, H., Li, Q.-Z., Zeng, C.-Q., Yang, H.-M., & Yu, J. (2005). Neighboring-nucleotide effects on the mutation patterns of the rice genome. *Geno. Prot. Bioinfo.*, 3(3).

[30] Zhang, W., Bouffard, G. G., Wallace, S. S., & Bond, J. P. (2007). NISC Comparative Sequencing Program. Estimation of DNA sequence context-dependent mutation rates using primate genomic gequences. *J Mol Evol*, 65, 207-214.

[31] Suzuki, Y., Gojobori, T., & Kumar, S. (2009). Methods for incorporating the hypermutability of CpG dinucleotides in detecting natural selection operating at the amino acid sequence level. *Mol Biol Evol*, 26, 2275-2284.

[32] Cheunga, L. W. T., Leeb, Y. F., Ngb, T. W., Chingb, W. K., Khooc, U. S., Ngd, M. K. P., & Wong, A. S. T. (2007). CpG/CpNpG motifs in the coding region are preferred sites for mutagenesis in the breast cancer susceptibility genes. *FEBS Letters*, 681, 4668-4674.

[33] Antonarakis, S. E. (2006). *CpG Dinucleotides and Human Disorders. eLS*.

[34] Misawa, K. (2011). A codon substitution model that incorporates the effect of the GC contents, the gene density and the density of CpG islands of human chromosomes. *BMC Genomics*, 12, 397.

[35] Chen-L, C., Rappailles, A., Duquenne, L., Huvet, M., Guilbaud, G., Farinelli, L., Audit, B., d'Aubenton-Carafa, Y., Arneodo, A., Hyrien, O., & Thermes, C. (2010). Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res*, 20, 447-457.

[36] Kasiviswanathan, R., & Copeland, W. C. (2011). Ribonucleotid discrimination and reverse transcription by the human mitochondrial DNA polymerase. *J Biol Chem*, in press.

[37]  Seligmann, H., & Amzallag, G. N. (2002). Chemical interactions between amino acid and RNA: multiplicity of the levels of specificity explains origin of the genetic code. *Naturwissenschaften*, 89, 542-551.

[38]  Krishnan, N. M., Seligmann, H., & Rao, B. J. (2008). Relationship between mRNA secondary structure and sequence variability in chloroplast genes: possible life history implications. *BMC Genomics*, 9, 48.

Chapter 11

# The Plant and Protist Organellar DNA Replication Enzyme POP Showing Up in Place of DNA Polymerase Gamma May Be a Suitable Antiprotozoal Drug Target

Takashi Moriyama and Naoki Sato

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/51975

## 1. Introduction

Mitochondria and plastids are eukaryotic organelles that possess their own genomes. The existence of organellar genomes is explained by the endosymbiotic theory [1], which holds that mitochondria and plastids originated from $\alpha$-proteobacteria-like and cyanobacteria-like organisms, respectively [2,3]. Organellar genomes are duplicated by the replication machinery, including DNA polymerase, of the each organelle. The enzymes involved in the replication of organellar genomes are thought to be encoded by the nuclear genome and transported to the organelles after synthesis [4].

DNA polymerase $\gamma$ (Pol$\gamma$) is the enzyme responsible for replicating the mitochondrial genome in fungi and animals [5,6]. Pol$\gamma$ belongs to family A DNA polymerases, which share sequence similarity to DNA polymerase I (PolI) of *Escherichia coli*. Animal Pol$\gamma$ consists of two subunits: a large subunit with DNA polymerase and 3'-5' exonuclease activities, and a small subunit that enhances processivity and primer recognition. The activity of Pol$\gamma$ is inhibited by *N*-ethylmaleimide (NEM) and dideoxy nucleotide triphosphate (ddNTP).

In the late half of the 1960s, the presence of organellar DNA polymerase was confirmed by the measurement of DNA synthesis activity in isolated plant chloroplasts [7,8] and mitochondria of yeast and animals [9,10]. Since the 1970s, DNA polymerases have been purified from the chloroplasts and mitochondria of various photosynthetic organisms (Table 1), with biochemical data suggesting that plant organellar DNA polymerases and $\gamma$-type DNA polymerases share similarities with respect to optimal enzymatic conditions, resistance to aphidicolin (an inhibitor of DNA polymerase $\alpha$, $\delta$, and $\varepsilon$), sensitivity to NEM, molecular size, and template preference. Despite such observation, no gene encoding a homolog of Pol$\gamma$ has

been found in the sequenced genomes of bikonts, including plants and protists. Therefore, the DNA polymerase of both mitochondria and plastids in photosynthetic organisms had remained unidentified. Sakai and colleagues [11-13] isolated nucleoid-enriched fractions from chloroplasts and mitochondria of tobacco leaves. They detected DNA synthetic activity in the nucleoid fraction and showed that the apparent molecular mass of the polypeptide exhibiting the activity was similar to Klenow fragment of DNA polymerase I (PolI) in *E. coli*. After their suggestion, it was found that the genomes of bikonts, consisting of plants and protists, encode one or two copies of genes encoding a DNA polymerase having distant homology to *E. coli* PolI. Homologs of this polymerase have been isolated in several plants, algae, and ciliates. Because genes encoding this type of enzyme are present in both photosynthetic eukaryotes and protists, we proposed to call this type of DNA polymerase POP (plant and protist organellar DNA polymerase).

| Year | Organism (organelle) | Mr (kDa) | Optimal condition for enzymatic activity | | | Inhibition by NEM (mM) | 3'-5' Exo-nuclease activity |
|---|---|---|---|---|---|---|---|
| | | | pH | $MgCl_2$ (mM) | NaCl or KCl (mM) | | |
| 1973 | *Euglena gracilis* (cp)[a] | | 7.2 | 6 | 10-15 | | |
| 1979 | Wheat (mt)[b] | 110[m] | 7 | 5 | 150 | 5 | yes |
| 1980 | Cauliflower (mt)[c] | | | | 150 | 1 | |
| 1980 | Spinach (cp)[d] | 105[n] | 8-9 | 0.1-1 | 100 | 2 | |
| 1981 | Wheat (mt)[e] | 180[m] | 8 | | | no | |
| 1984 | Pea (cp)[f] | 87[m] | | 12 | 120 | 1 | no |
| 1990 | Soybean (cp & mt)[g] | 85-90[n] | 8 | | 125 | strongly | |
| 1991 | Spinach (cp)[h] | 105[n] | | | | 1 | yes |
| 1991 | *Chlamydomonas* (cp)[i] | 110[n] | | | 100 | 2 | no |
| 1993 | Chenopodium (mt)[j] | 80-90[n] | | 10 | 125 | 1 | yes |
| 1995 | Soybean (cp)[k] | | | | | | yes |
| 2002 | Pea (cp)[l] | 70[n] | 7.5 | 8 | 125 | partially | yes |

**Table 1.** Previous studies on organellar DNA polymerases with no gene identification in plants and algae. cp, chloroplast; mt, mitochondrion; NEM, *N*-ethylmaleimide. a-l: references [14-25]. [m]determined by gel filtration; [n]determined by glycerol density gradient. Reproduced from [26].

## 2. Enzymatic characteristics of POPs

The isolation of POP was first reported in rice (*Oryza sativa*) [27,28] and later in several higher plants and algae, including thale cress (*Arabidopsis thaliana*) [29,30], tobacco (*Nicotiana tabacum*) [31], red alga (*Cyanidioschyzon merolae*) [32], and a ciliate (*Tetrahymena thermophila*)

[33]. POPs typically consist of 900-1050 amino acid residues and contain 3'-5' exonuclease and DNA polymerase domains (Figure 1). In addition, POPs have an organellar targeting peptide at the *N*-terminus.
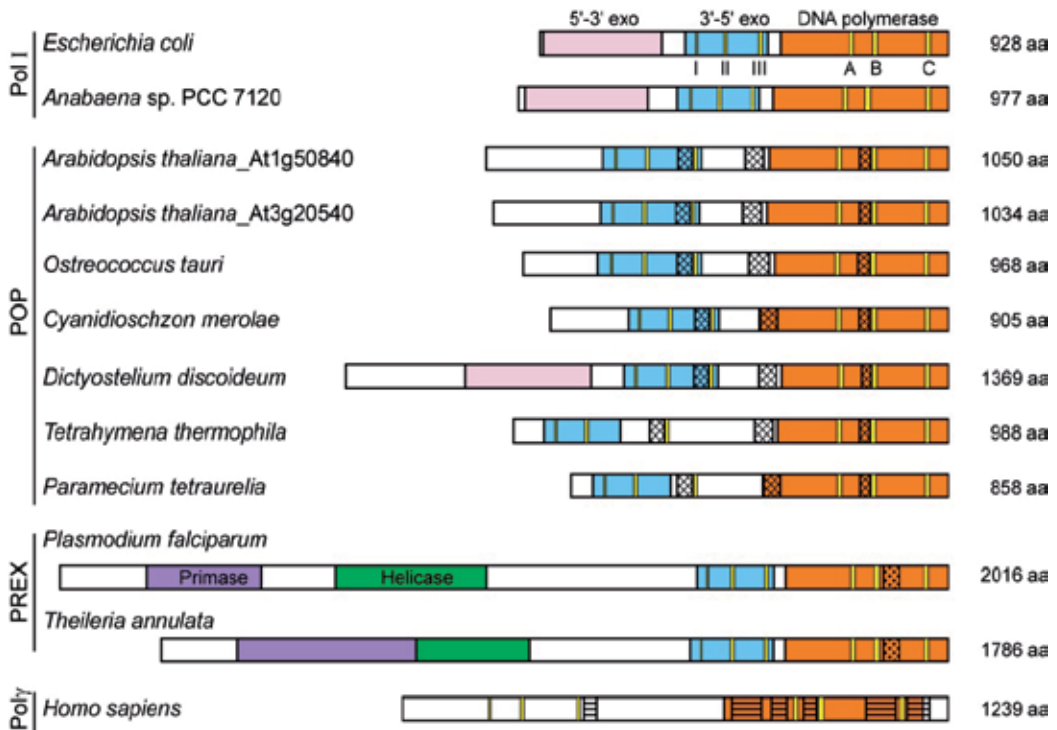


**Figure 1.** Schematic comparison of the structure of family A DNA polymerases. The colored boxes indicate domains estimated from the Pfam database: pink, 5'-3' exonuclease domain; blue, 3'-5' exonuclease domain; orange, DNA polymerase domain; purple, primase domain; green, helicase domain. Yellow boxes indicate characteristic motifs in the 3'-5' exonuclease and DNA polymerase domains. Thatched boxes represent conserved sequences in POPs. Dotted and striped boxes indicate conserved sequences in PREX and Polγ, respectively. In Polγ of *Homo sapiens*, a 3'-5' exonuclease domain was not found by Pfam, although 3'-5' exonuclease activity was reported for Polγ [6]. This figure was modified from [32] with permission of the publisher.

### 2.1. Properties of DNA polymerase activity

The properties of DNA polymerase activity of POPs have been examined using recombinant [27,28,31] or native proteins purified from *Cyanidioschyzon* and *Tetrahymena* cells [32,33]. The optimal concentrations of KCl and MgCl$_2$ for DNA polymerase activity are 50-150 and 2.5-5 mM, respectively, which roughly coincide with the values reported in previous studies for organellar DNA polymerases in plants (Table 1). POPs display the highest activity with Poly(dA)/oligo(dT) as a template. Poly(rA)/oligo(dT) could also serve as a template, indicating that POPs have reverse transcriptase activity. Polγ also exhibits reverse transcriptase activity, although the physiological importance of this activity remains to be elucidated.

## 2.2. Processivity

Processivity is defined as the number of nucleotides added by a DNA polymerase per one binding with the template DNA. POPs, in general, have high processivity values; for example, the processivity of rice recombinant GST-POP and *Cyanidioschyzon* POP is 600-900 nt and 1,300 nt, respectively [28,32]. In comparison, the Klenow fragment of *E. coli* PolI has mid-range processivity of <15 nt [28]. POPs contain three additional internal sequences relative to other family A DNA polymerases (Figure 8). The role of the two extra sequences, amino acid residues 635-674 (Insert I) and 827-852 (Insert II) positioned before motif A (Figure 8-1) and between motif A and motif B (Figure 8-2), were examined in rice POP [28]. DNA binding was decreased in Insert I and II deletion-mutant proteins, while DNA synthesis activity and processivity were decreased only in the POP protein lacking Insert I. These findings suggest that the high processivity of POPs may be due to the existence of the inserted sequences. In animals, Polγ consists of two subunits, a large subunit (PolγA) having DNA polymerase and 3'-5' exonuclease activities and a small subunit (PolγB) that enhances processivity and primer recognition [34]. Processivity of the *Drosophila* PolγA subunit is <40 nt, whereas that of Polγ holoenzyme (PolγA and PolγB) is >1,000 nt [35]. In contrast to animal Polγ, POPs display high processivity as a single subunit, and no accessory subunits of POP have been identified to date [28,32].

## 2.3. Sensitivity to inhibitors

The effects of inhibitors, such as aphidicolin, NEM, dideoxyTTP (ddTTP), and phosphonoacetate (PAA), on the DNA synthesis activity of POPs were evaluated [27,31-33]. Aphidicolin is a specific inhibitor of DNA polymerases α, δ, and ε and acts through competition with dCTP or dTTP [36,37]. The sulfhydryl reagent NEM inhibits DNA polymerases α, γ, δ, and ε [38], and has a half maximal inhibitory concentration ($IC_{50}$) of <0.1 mM for Polγ. PAA is an analog of pyrophosphate and interacts with viral DNA polymerases and reverse transcriptases at pyrophosphate binding sites to create an alternative reaction pathway [39,40]. ddTTP severely inhibits DNA polymerases β and γ, but only weakly impairs the activities of DNA polymerases δ and ε [41]. POPs are not inhibited by aphidicolin or NEM. The inhibitory effect of ddTTP differs depending on the organism, with the $IC_{50}$ ranging from 4-615 μM for POPs (Figure 2A). The activity of POPs is severely inhibited by PAA, as demonstrated by $IC_{50}$ values of 1-25 μM for several POPs (Figure 2B, C). In contrast, other family A DNA polymerases, including PolI and Polγ, are not markedly inhibited by PAA, suggesting that PAA is a useful marker for the classification of organellar DNA polymerases in unsequenced eukaryotes. T4 DNA polymerase and DNA polymerase δ of *Saccharomyces cerevisiae*, which are both family B DNA polymerases, are also not sensitive to PAA, but the respective Motif A mutants of each protein, L412M (T4 DNA polymerase) and L612M (DNA polymerase δ of *S. cerevisiae*), are inhibited by PAA [42,43]. The mechanism of inhibition by PAA has not been studied in detail for family A DNA polymerases, and the critical amino acid residues involved in sensitivity to PAA in POPs are unknown due to the limited similarity of family A and B DNA polymerases in the Motif A region.
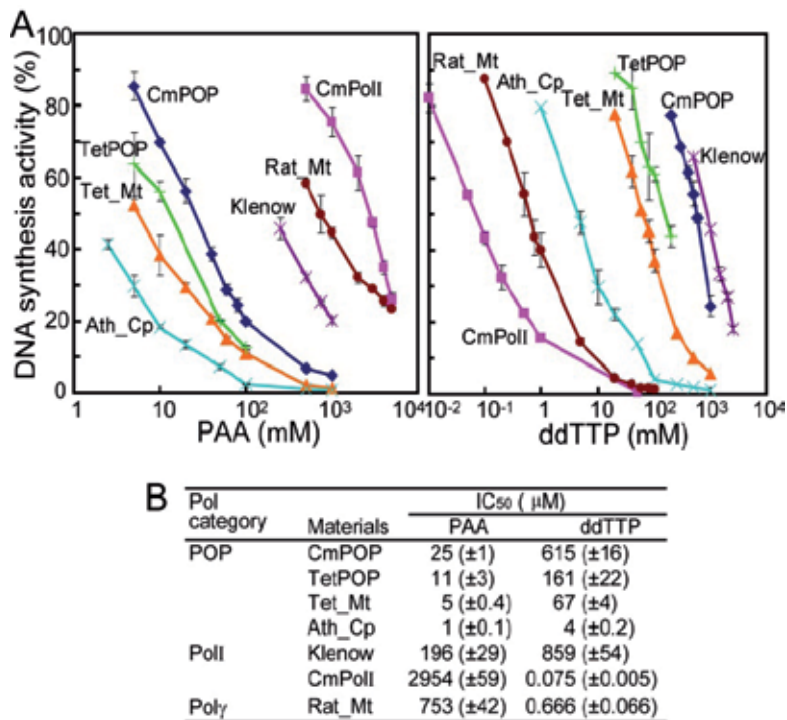
**Figure 2.** Effect of inhibitors, phosphonoacetic acid (PAA), and dideoxy TTP (ddTTP) on DNA synthesis activity (A). Half maximal inhibitory concentration (IC$_{50}$) for PAA or ddTTP (B). Tet_Mt, *Tetrahymena thermophila* mitochondria; Ath_Cp, *Arabidopsis thaliana* chloroplasts, Klenow, *Escherichia coli* PolI Klenow fragment; CmPolI, *Cyanidioschyzon merolae* PolI; Rat_Mt, rat liver mitochondria. Reproduced from [33] with permission.

## 2.4. 3'-5' Exonuclease activity

POPs have a 3'-5' exonuclease domain containing three conserved regions, Exo I, Exo II, and Exo III (Figure 1), and this exonuclease activity has been demonstrated in rice [28] and *Cyanidioschyzon* [32]. In rice POP, replacement of Asp365 with Ala in the Exo II domain abolishes nuclease activity, but has no effect on DNA polymerase activity. With regard to 3'-5' exonuclease proofreading activity, POP shows relatively high fidelity for base substitutions (10$^{-4}$ to 10$^{-5}$; [28]). The primary structure of Polγ appears to lack a 3'-5' exonuclease domain, as indicated by the low E-value of 0.17 for this domain in human Polγ determined using the motif search software Pfam (http://pfam.sanger.ac.uk/). However, Polγ possesses Exo I, Exo II, and Exo III motifs in the N-terminus (Figure 1), and exhibits 3'-5' exonuclease activity and high replication fidelity [6].

## 2.5. Subcellular localization

POP was first isolated as a plastidial DNA polymerase in rice, and its localization was confirmed by immunoblot analysis using isolated plastids [27]. Subsequent studies using GFP-fusion proteins and/or immunoblotting with isolated plastids and mitochondria

demonstrated that POPs are localized to both plastids and mitochondria in *Arabidopsis* and tobacco [31,44], and in the alga *Cyanidioschyzon* [32]. The mitochondrial localization of POP in the ciliate *Tetrahymena* was also determined by immunoblotting [33]. Figure 3 shows all of the known DNA polymerases found in the model plant *A. thaliana* and in humans. The nuclear-localized DNA polymerases involved in genome replication, DNA polymerase α, δ, and ε, are conserved in bikonts and opisthokonts, whereas the nuclear polymerases related to DNA repair differ between organisms. POP and Polγ are the sole replicational DNA polymerases in bikont or opisthokont organelles, where they also act as DNA repair enzymes.
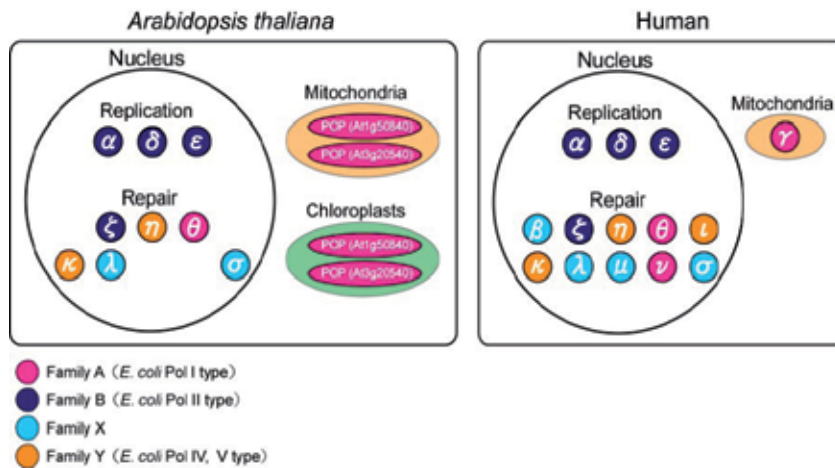


**Figure 3.** DNA polymerases of a model plant and human. Greek letters in colored circles corresponding to families indicate eukaryotic DNA polymerases alpha to sigma.

### 2.6. The role of POP *in vivo*

POPs exhibit high processivity and 3'-5' exonuclease activity, and were originally thought to function as organellar DNA replicases. This speculation was verified by analyzing *POP* mutant of *Arabidopsis* [30], whose genome encodes two *POP* genes, *At1g50840* and *At3g20540*, whose protein products are each localized to both plastids and mitochondria (Figure 3). The *At1g50840-At3g20540* double mutant was lethal, while each single mutant had a phenotype characterized as reduced DNA levels in plastids and mitochondria. In addition, only the *At3g20540* mutant displayed elevated sensitivity to ciprofloxacin, which is an inducer of DNA double-strand breaks (DSB). Together, these results show that two distinct POPs are involved in genome replication for plastids and mitochondria, and that the product of *At3g20540* also functions as a DNA repair enzyme in both organelles. In rice, the repair activity of POP was examined by a base excision repair (BER) assay using a recombinant protein, revealing that POP has 5'-deoxyribose phosphate (dRP) lyase activity [28]. Polγ also displays this repair activity [45].

# 3. Role of POPs in cell-cycle regulation

## 3.1. Organellar genome replication in plant tissues

Nuclear genomes are replicated during the DNA synthesis phase (S phase), with the daughter genomes being distributed at the mitotic phase (M phase) to maintain ploidy levels. Observations of mitochondrial DNA stained with 4',6-diamidino-2-phenylindole (DAPI) and microautoradiography using [3H]thymidine have demonstrated that the DNA content and synthesis activity in mitochondria change dramatically during cell proliferation. In the root apical meristem of geranium (*Pelargonium zonale*), mitochondrial DNA in the promeristem, which is located just above the quiescent center, maintain high levels of DNA. However, in the upper root region, located immediately below the elongation zone, mitochondria contain small amounts of DNA [46]. Similar results were reported for the root apical meristem of *Arabidopsis* [47], tobacco [48], and rice [49], shoot apical meristem of *Arabidopsis* [50], and cultured tobacco cells [48,51], in addition to plastids. In *Avena sativa*, plastid DNA is extensively replicated in small cells of shoot apical meristem. Subsequently, as the cells increase in size, plastid numbers increase, while the DNA levels within plastids decrease [52-54]. These results suggest that organellar DNA is predominantly replicated in the meristem, and that the subsequent partition of organellar DNA to daughter cells does not coincide with the synthesis of organellar DNA in cells outside of the meristem center. In multicellular plants, therefore, the replication of organelle genome is not synchronized with the cell cycle or even organellar division.

## 3.2. Expression of POP in plants

The spatial expression patterns of POPs were analyzed in *Arabidopsis* and rice by *in situ* hybridization, which revealed that *POP* genes are strongly expressed in the apical meristem of roots and shoots, leading to high POP protein levels in these tissues [27,29]. In cultivated tobacco BY-2 cells, the amount of POP transcripts and proteins increases at the initiation of plastidial and mitochondrial DNA replication [31]. These results indicate that POPs function as the organellar genome replicase.

## 3.3. Red algal cell cycle

The unicellular red alga *C. merolae* contains a single plastid and mitochondrion [55], which both have division cycles that are synchronous with the cell cycle. Synchronous cultures of *Cyanidioschyzon* have been obtained by light-dark cycles [56]. Our group has also performed synchronous culture of *C. merolae* [32,57] using an initial long dark period (30 h) to force the cells into the G1 phase (Figure 4), followed by a 6-h light/18-h dark regime with bubbling with ordinary air. However, due to the low nutrient levels, the conditions were not sufficient to drive the cell cycle (Figure 4A). Two subsequent cycles of 6-h light/18-h dark with a supply of 1% $CO_2$ enabled the cells to accumulate enough photosynthetic products to allow progression of the cell division cycle, resulting in the synchronous division of cells 4-5 h after the start of the dark period (Figure 4B, C). Therefore, this culture method can discrimi-

nate the effects of light from those of the cell cycle in photosynthetic eukaryotes, and contains the cycle in which cellular nutrient level transitions from low to high (Figure 4B).
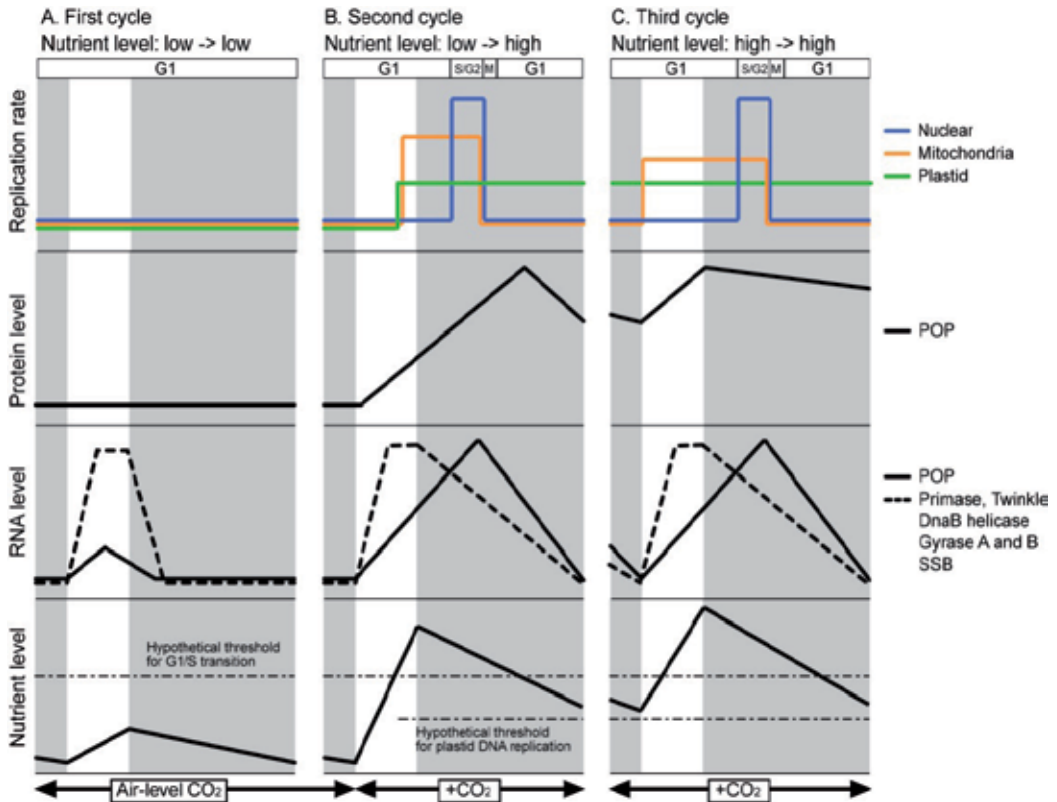


**Figure 4.** Cell cycle of *Cyanidioschyzon* and the expression of protein or mRNA related to organellar DNA replication. Three cell cycle patterns (A, B, and C) with respect to the nutrient level are shown. The nutrient level was controlled by aeration with or without the addition of $CO_2$. These drawings are based on the data taken from [57]. The shaded and white areas indicate dark and light cycles, respectively.

We have also determined the replication phases of nuclear, plastid, and mitochondrial DNA by quantitative PCR using cyanobacterial DNA as an internal standard to estimate the absolute amount of DNA (Figure 4, [57]). In the first cell cycle pattern, the level of nuclear and organellar DNA was unaltered (Figure 4A). Nuclear DNA replicated at or near the M-phase in the second and third cycles (Figure 4B, C). The replication of the mitochondrial genome was synchronized with the cell cycle to some extent, with mitochondrial DNA beginning to increase from the middle (second cycle) or beginning (third cycle) of the light phase, and doubling at or near the M-phase, as was observed for nuclear DNA (Figure 4B, C). In contrast, plastid DNA replication continued throughout the entire cell cycle, even after cell division was complete (Figure 4B, C). These results suggest that the replication of nuclear and organellar DNA is initiated after the accumulation of sufficient nutrients by photosynthesis, and that light alone does not serve as a replication signal for nuclear or organellar genomes.

Therefore, *C. merolae* cells may have two checkpoints (or thresholds) based on their nutritional state. The first checkpoint occurs during the G1/S-phase transition. Once cells overcome this point, the nuclear and organellar genomes are targeted for replication. The second threshold is specific for plastid DNA replication. After passage of the first checkpoint for G1/S transition, plastid DNA replication proceeds if the cellular nutrient level exceeds the nutritional threshold required for the replication process.

### 3.4. Expression of POP in the red algal cell cycle

We determined the expression of POP in synchronous culture of *C. merolae*. The protein level of POP was very low in the first non-dividing cycle (Figure 4A), but continued to increase from the second light period, and subsequently decreased during the dark period (Figure 4B). In the third cycle, the protein level of POP appeared constitutive during the cell cycle, although slight increases in the light phase and decreases in the dark phase were observed (Figure 4C). A small peak in the *POP* mRNA level was detected during the first light period (Figure 4A), with larger peaks appearing soon after entering the dark cycle (Figure 4B, C). The large peaks of *POP* mRNA levels correlated with the rise in mitotic indices.

The transcript level of other possible genes related to organellar DNA replication in *C. merolae* was also examined (Figure 4). Gyrase A and B, which are types of bacterial topoisomerase II, are related to both plastid and mitochondria genome replication in *C. merolae* [58] and *A. thaliana* [59]. SSB is a bacterial single-stranded DNA binding protein that is localized to mitochondria in *A. thaliana* [60]. In plants, DNA primases have not yet been isolated, although primase activity was detected in the chloroplasts of pea and the green alga *Chlamydomonas reinhardtii* [61,62]. DnaB is a bacterial replicational helicase that is encoded in the plastid genome of *C. merolae*. Twinkle is a replicational helicase and is localized to mitochondria in animals. Animal twinkle has only helicase activity; however, it is predicted that twinkle in plants and protists might have both helicase and primase activities [63]. Changes in the expression of these genes were qualitatively similar with each other, and were mainly stimulated by light. The expression pattern of these genes was also similar to that of genes related to photosynthesis, respiration, nuclear DNA repair, and ubiquitin in *C. merolae* [57]. In contrast, the expression pattern of POP transcripts was similar to that of cell cycle regulatory genes, including nuclear replicational DNA polymerase, mitotic cyclin, and mitotic cyclin-dependent kinase (CDK). Based on these findings, it appears that the replication of organellar genomes might be controlled by the expression of POP rather than that of other proteins related to organellar genome replication. Notably, the kinetics of replication differed for plastid and mitochondrial genomes; however, the regulatory mechanisms controlling the replication of the two organelles remain to be elucidated.

## 4. Possible evolutionary history of organellar DNA polymerases in eukaryotes

POP belongs to family A DNA polymerases, consisting of polymerases harboring sequence similarity to bacterial PolI, such as Polγ, DNA polymerase θ (Polθ), DNA polymerase ν (Polν), and PREX (plastid replication and repair enzyme complex, [64]). Polθ and Polν are DNA repair enzymes and are localized to the nucleus [65,66]. PREX is an apicoplast (plastid like organelle)-localized DNA polymerase in the malaria parasite *Plasmodium falciparum* and contains a DNA polymerase domain, as well as helicase and primase domains (Figure 1 and [67]). Figure 8 shows the alignment of the DNA polymerase domain of several family A DNA polymerases. Although bacterial PolI, POP, and PREX share some homology, POP and PREX contain specific sequences, and the domain structure is clearly different in each polymerase (Figure 1 and Figure 8). Polγ shows low similarity to other family A DNA polymerases, and has many Polγ specific sequences.

Figure 5 shows a phylogenetic tree of family A DNA polymerases. From the tree, it is clear that POPs belong to a well-defined clade that is evolutionarily separated from bacterial PolI. Therefore, it can be concluded that POPs did not originate from PolI of cyanobacteria nor α-proteobacteria. Although PREX may have originated from a red algal secondary endosymbiont, their origin remains unclear, because PREX do not contain POP-specific sequences (Figure 8). POPs are widely conserved in eukaryotes, including amoebozoa, that have a close relationship with opisthokonts in phylogenetic analyses, but POPs have not been detected in opisthokonts, including animals and fungi (Figure 6). This suggests that POP might have originated before the diversification of photosynthetic eukaryotes. Pathogenic protists of animals, including *Blastocystis hominis* and *Perkinsus marinus*, possess POP, while genome-unsequenced pathogens, such as the green alga *Prototheca*, are likely to have POP. Therefore, POP is expected to be a suitable target for killing these pathogens.

From the phylogenetic tree, we proposed an evolutionary model of organellar DNA polymerases (Figure 7). Initially, when the ancestor of eukaryotes acquired mitochondria, the elementary mitochondrial replicase was likely bacterial DNA polymerase III (PolIII) (1 in Figure 7A). PolIII was then replaced by a POP, and the host cell then used POP for the replication of organellar genomes (2 in Figure 7A). We presume that PolIII must have been introduced upon the endosymbiosis event, but another possibility is that an endosymbiont or a host cell had already possessed POP before endosymbiosis. But this idea is considered unlikely because no bacteria having POP have been found so far. In this respect, it is of interest to note that, based on phylogenetic analysis in family A DNA polymerases, it has been postulated that Polγ is of phage origin [69]. POP could also have been acquired from a virus. In effect, the ultimate origin of the ancestral POP is still unknown. The phylogenetic tree (Fig. 5) suggests that the closest relative of POP is Polν or Polθ, which are present in various eukaryotes. It is not impossible then that an ancestral polymerase in eukaryotic host diverged into POP, Polν and Polθ.
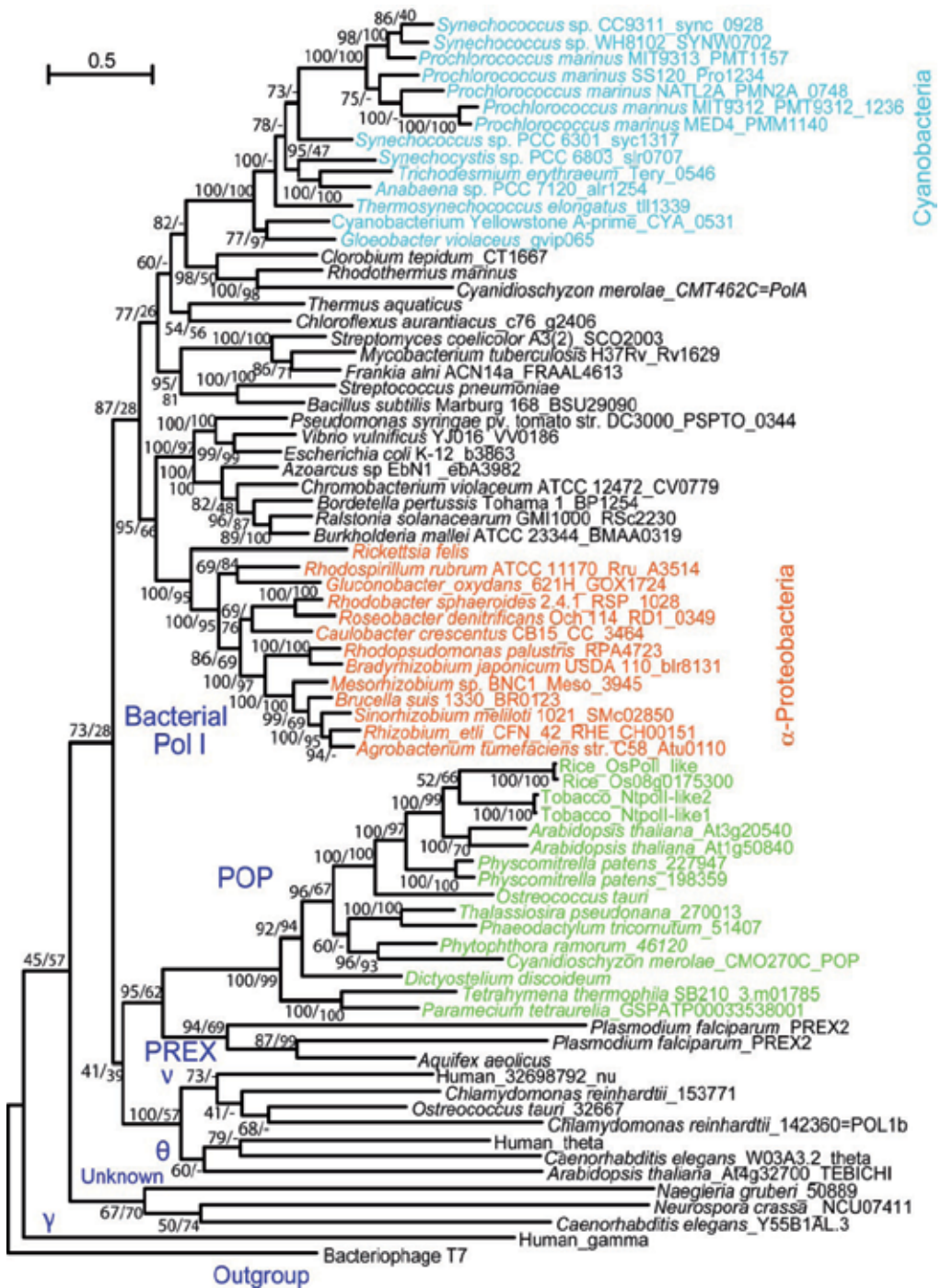
**Figure 5.** Phylogenetic tree of POPs and other family A DNA polymerases. Reproduced from [32] with permission.
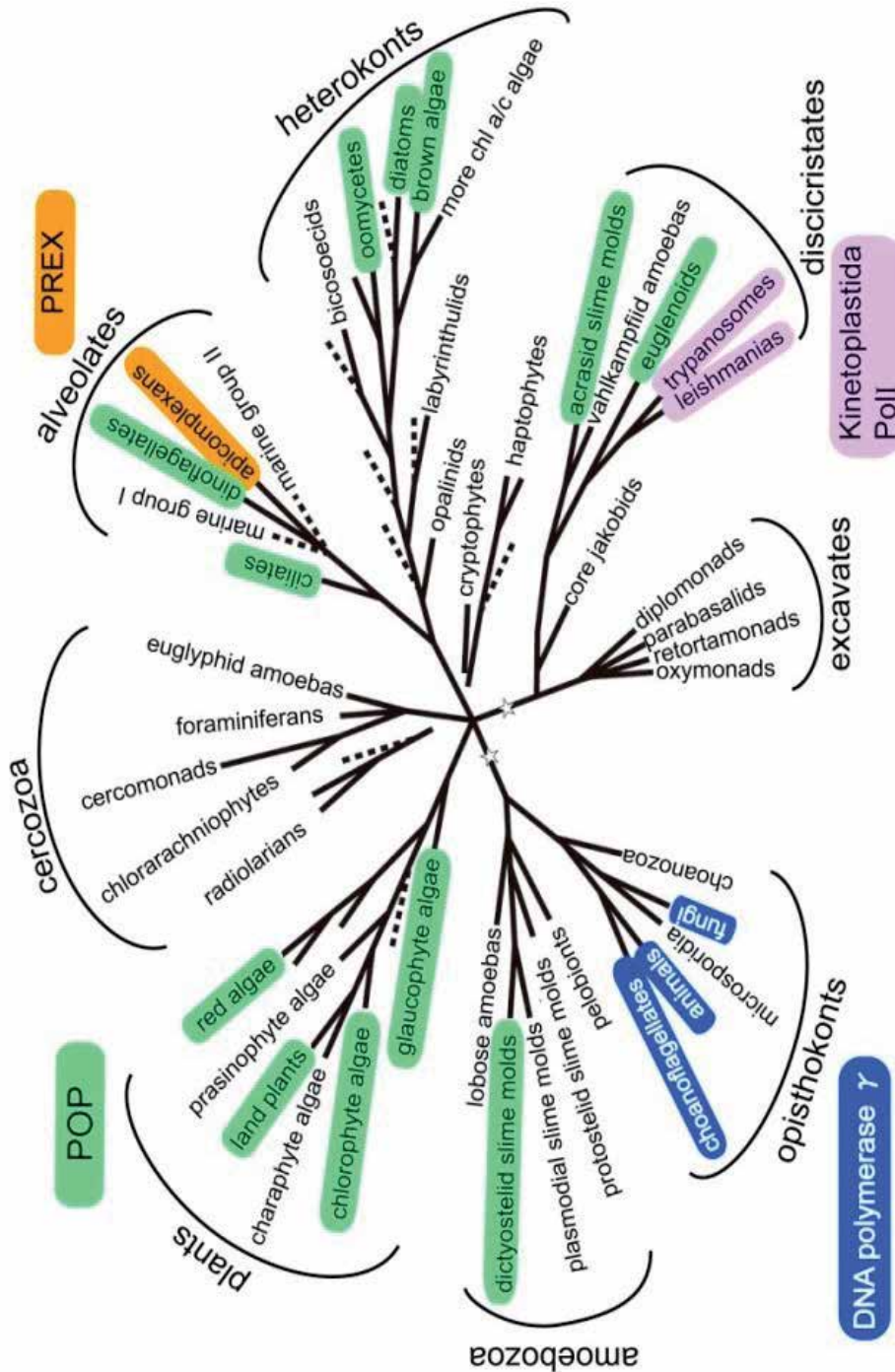
**Figure 6.** Distribution of organellar DNA polymerases in eukaryotes. Taxons containing POP, Polγ, PREX, and kineto-plastida PolI are enclosed in light green, blue, orange, and purple boxes, respectively. The tree topology in this figure was adapted from [68], and the figure was modified from [33] with permission.
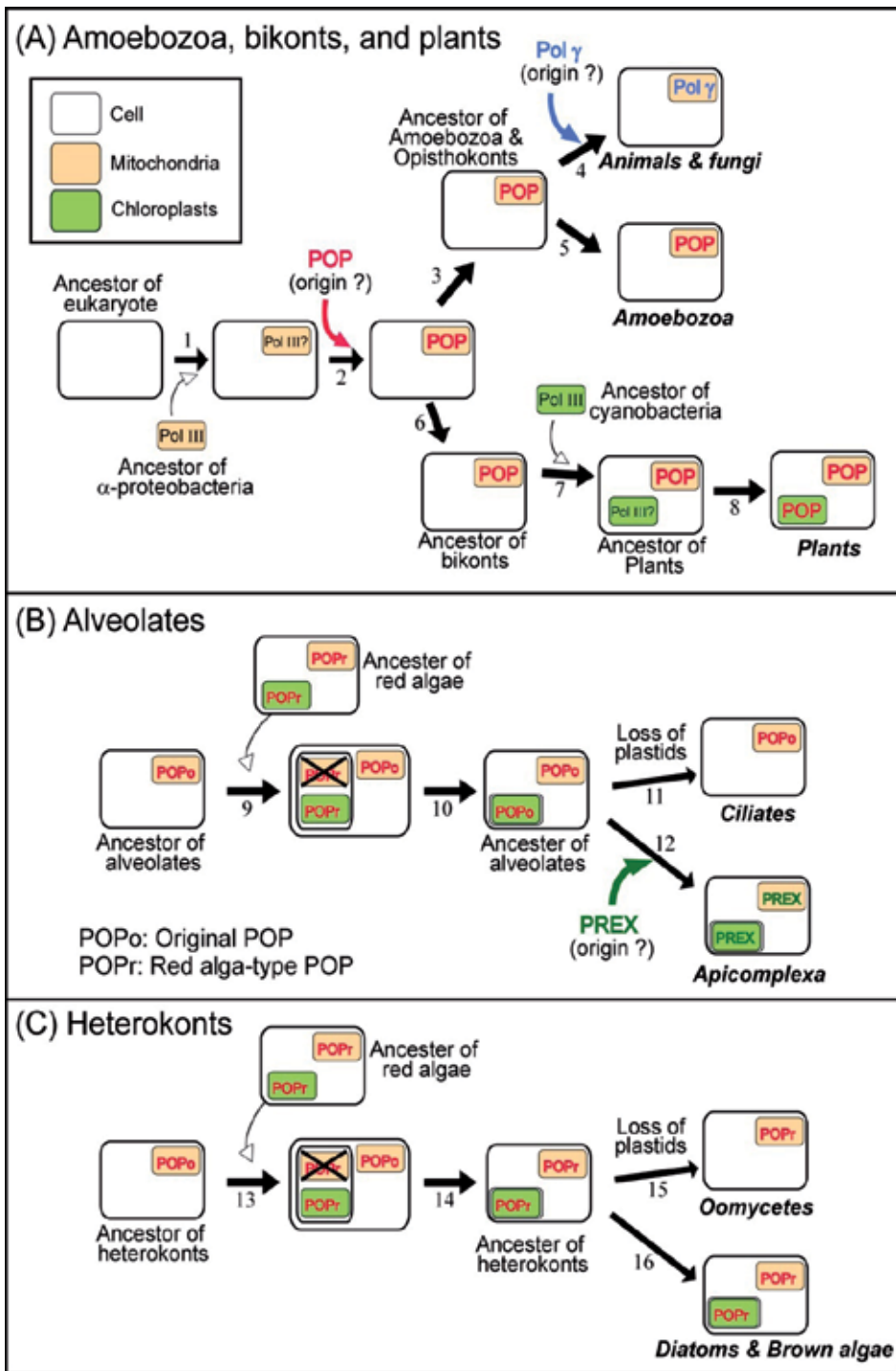
**Figure 7.** Schematic illustration of the evolution of organellar replication enzymes. The figure was modified from [33] with permission.

In the plastids of plants and algae, POP also replaced PolIII, and thus POPs are presently found in most eukaryotes (3-5 and 6-8 in Figure 7A). In opisthokonts, however, POP was replaced by Polγ, whose origin is also unknown (4 in Figure 7A). Chromalveolates, consisting of alveolates and heterokonts such as diatoms, must have had a POP for mitochondrial replication before the occurrence of secondary endosymbiosis. Phylogenetic analysis suggests that the POPs of diatoms are more closely related to red algal POP than the POPs of ciliate *Tetrahymena* (Figure 5). The original POP might have been replaced by the POP of a red algal endosymbiont in diatoms (13-16 in Figure 7C), whereas in ciliates, the original POP has been retained (9-11 in Figure 7B).

Based on the genomic data obtained to date, Polγ is found only in opisthokonts, indicating that two different polymerases cannot co-exist, at least over a long evolutionary span. The catalytic subunit of bacterial PolIII is also not encoded by eukaryotic genomes, although the PolIII gamma subunit, which functions as a clamp loader in bacteria, is conserved in land plants, such as *A. thaliana*, which has three gamma subunits, At1g14460, At2g02480, and At4g24790 [13]. One of the possible reasons why PolIII was replaced by POP may be the fact that POP is a single polypeptide enzyme, whereas PolIII consists of ten subunits. Therefore, the nuclear control of organellar DNA replication would be easier with nuclear-encoded POP. This also raises the question: why was POP replaced with Polγ? Unfortunately, although we do not have a clear answer for this question, the replacement event might be related to the mechanism of organellar genome replication. In animals, three replication modes have been proposed: the classical strand-displacement replication mode, a strand-coupled mode, and a RITOLS (ribonucleotide incorporation throughout the lagging strand) mode [70]. This contrasts with plant plastids, for which at least two modes of replication have been proposed, namely rolling circle replication via a D-loop and recombination-dependent replication [71]. Although the proposed replication modes in animals and plants remain to be confirmed, it is likely that the type of replication mode is different in the organelles of animals (opisthokonts) and plants (bikonts). In opisthokonts, the replication mode of organellar genomes of animals may have arisen before the replacement of POP with Polγ, with Polγ being a suitable enzyme for the replication process of animals. Secondary or tertiary endosymbionts do not exist among opisthokonts, a fact that may be due to differences in the organellar genome replication mode or organellar DNA polymerase type.

## 5. Conclusion and prospects

POPs have been isolated as organellar-specific DNA polymerases in a number of photosynthetic eukaryotes and ciliates. As the majority of biologists still believe that all mitochondrial replication enzymes are Polγ, the primary objective of this review was to introduce POP to the wider research community. Although both POP and Polγ are family A DNA polymerases, their primary structures are quite different from one another. However, POP and Polγ display similar DNA polymerase activities that are characteristics of replicases, including high processivity, 3′-5′ exonuclease activity, and reverse transcriptase activity. Eukaryotes containing a POP gene do not have a gene for Polγ, and *vice versa*. In our hypothesis con-

cerning the transitional evolution of organellar DNA polymerase in eukaryotes, POP was proposed to be the primary organellar replicase and was then replaced by Polγ in opistho-konts. POP might have been replaced by PREX and kinetoplastida PolI in apicomplexa and trypanosomes, respectively. Phylogenetic evidence suggests that organellar DNA polymer-ases are easily replaced, unlike nuclear replicational DNA polymerases, which are con-served in all eukaryotes.

The sensitivity of POP to DNA polymerase inhibitors clearly differs from that of Polγ. To date, POPs have been shown to be commonly inhibited by phosphonoacetate. The inhibition mechanisms remained unclear for family A DNA polymerases, including POP, although it was reported that motif A in the polymerase domain of family B DNA polymerases is in-volved in the sensitivity to phosphonoacetate [42,43]. The detailed study of the inhibitory mechanisms and structural analysis of POP are needed, although POP is likely to be con-served in pathogenic bikonts, such as the green alga *Prototheca* and chromalveolata *Blastocys-tis*. Determining the structural differences in essential enzymes between a pathogen and host, and identifying pathogen-specific enzymes with no homologues in a host may identify suitable targets for chemotherapy. Such an approach is needed for targeting the malaria par-asite. Chloroquine, mefloquine, and quinine have been used as antimalarial drugs. These re-agents inhibit the production of the malarial pigment hemozoin. In addition, dihydrofolate reductase (DHFR) of malaria parasite is inhibited by proguanil and pyrimethamine. Howev-er, drug-resistant mutants of the parasite have emerged, and a new drug and enzyme target are therefore needed [72]. An apicoplast is non-photosynthetic plastid-like organelle that contains 27-35 kb of DNA in apicomplexa, and DNA replication within apicoplasts may be a good drug target, because apicoplasts harbors various essential metabolic pathways, such as those involving fatty acids, isoprenoid, and heme [73]. In plants and protists, our knowledge of the supporting players of organellar DNA replication, such as primase, helicase, topoiso-merase, and single-stranded DNA binding protein (SSB), are limited. To understand the mechanism and regulation of replication in plastids and mitochondria, it is necessary that the composition of these enzymes in each organelle be determined. In addition, reconstitu-tion of the replicational machinery of each organellar genome should be attempted. In hu-mans, successful *in vitro* reconstitution of the mitochondrial DNA replisome, including Polγ, twinkle helicase, and SSB, was demonstrated [74]. The further development of organellar re-plisome models in plants and protists may pave the way for greater understanding of the replication mode and discovery of new antiprotozoan reagents.

In multicellular plants, genomes of organelles are replicated in meristematic tissues, but the process is not synchronous with the cell cycle or even with organellar division. In the unicel-lular red alga *C. merolae*, which contains a single plastid and mitochondrion, the expression of POP appears constitutive during the cell cycle. POP is localized in both organelles, but the kinetics of replication differs for plastid and mitochondrial genomes. Replication of the mi-tochondrial genome is synchronous with the cell cycle to a certain extent, whereas replica-tion of the plastid genome continues throughout the entire cell cycle. The organellar replication is regulated by cellular nutrient levels, and POP protein levels are closely corre-lated with nutrient levels. The mechanisms regulating the replication of plastids and mito-chondria represent a new and exciting area of research in cell biology.
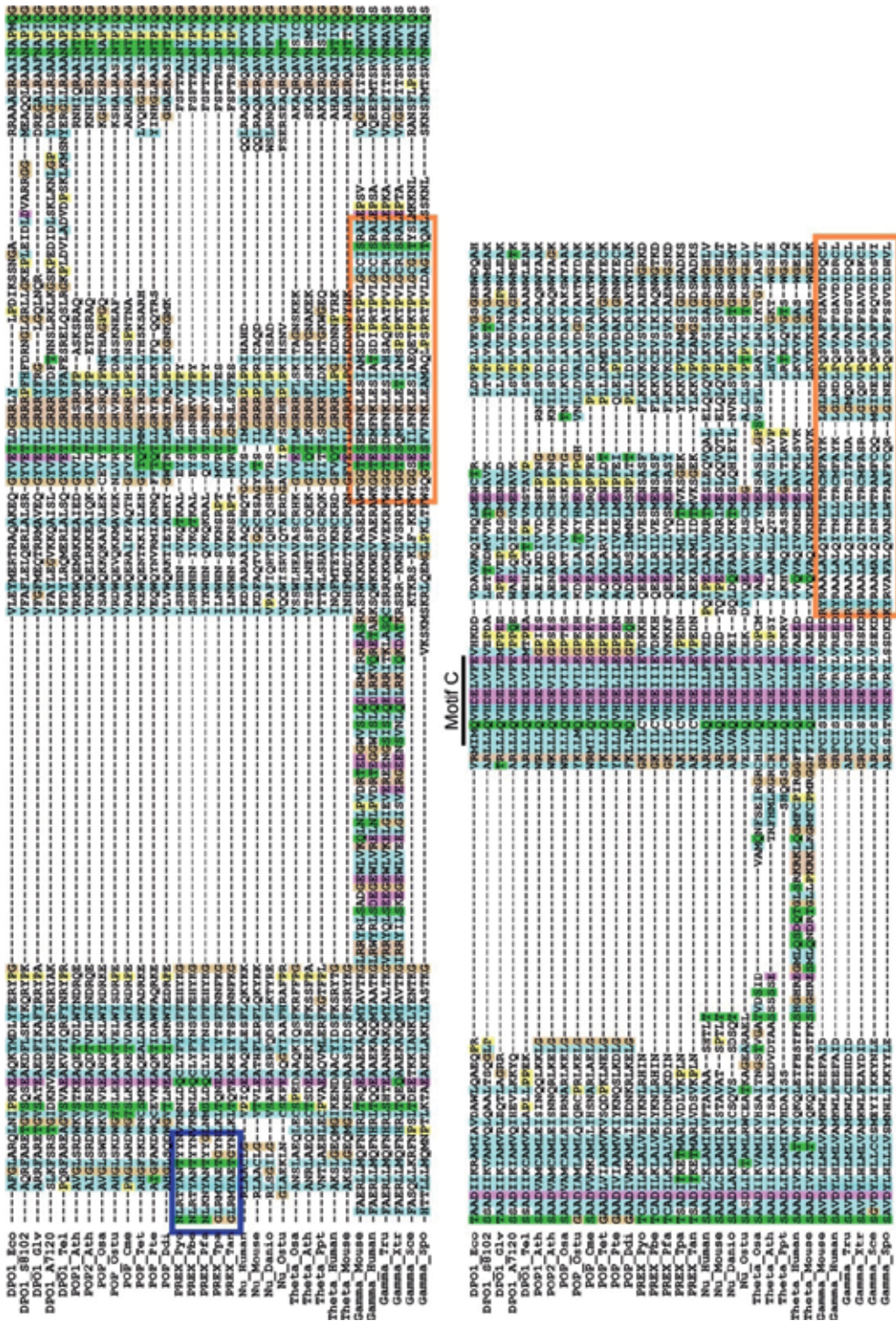
# Appendix

**Figure 8.** Alignment of the DNA polymerase domain of family A DNA polymerases. Green, blue, and orange boxes show specific sequences of POPs, DNA polymerase nu, and DNA polymerase gamma, respectively. Eco, *Escherichia coli*;

S8102, *Synechococcus* sp. WH8102; Glv, *Gloeobacter violaceus*; A7120, *Anabaena* sp. PCC 7120; Tel, *Thermosynecho-coccus elongates*; Ath, *Arabidopsis thaliana*; Osa, *Oryza sativa*; Ostu, *Ostreococcus tauri*, Cme, *Cyanidioschyzon merolae*; Tet, *Tetrahymena thermophila* SB210; Pte, *Paramecium tetraurelia*; Ddi, *Dictyostelium discoideum*; Pyo, *Plasmodium yoelii*; Pbe, *Plasmodium berghei*; Pfa, *Plasmodium falciparum*; Tpa, *Theileria parva*; Tan, *Theileria annulata*; Danio, *Danio rerio*; Ppt, *Physcomitrella patens*; Tru, *Takifugu rubripes*; Xla, *Xenopus tropicalis*; Sce, *Saccharomyces cerevisiae*; Spo, *Schyzosaccharomyces pombe*. Reproduced with permission [33].

# Acknowledgements

# Author details

Takashi Moriyama[1,2*] and Naoki Sato[1,2*]

*Address all correspondence to:

*Address all correspondence to: naokisat@bio.c.u-tokyo.ac.jp

1 Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan

2 JST, CREST, Gobancho, Chiyoda-ku, Tokyo, Japan

# References

[1] Margulis, L., & Bermudes, D. (1985). Symbiosis as a mechanism of evolution: status of cell symbiosis theory. *Symbiosis*, 1, 101-124.

[2] Douglas, S. E. (1998). Plastid evolution: Origins, diversity, trends. *Curr Opin Genet Dev*, 8(6), 655-661.

[3] Gray, M. W. (1999). Evolution of organellar genomes. *Curr Opin Genet Dev*, 9(6), 678-687.

[4] Wang, Y., Farr, C. L., & Kagni, L. S. (1997). Accessory subunit of mitochondrial DNA polymerase from Drosophila embryos. Cloning, molecular analysis, and association in the native enzyme. *J Biol Chem*, 272(21), 13640-13646.

[5] Lecrenier, N., & Foury, F. (2000). New features of mitochondrial DNA replication system in yeast and man. *Gene*, 246(1-2), 37-48.

[6]  Kaguni, L. S. (2004). DNA polymerase gamma, the mitochondrial replicase. *Annu Rev Biochem*, 73, 293-320.

[7]  Tewari, K. K., & Wildman, S. G. (1967). DNA polymerase in isolated tobacco chloroplasts and nature of the polymerized product. *Proc Natl Acad Sci U S A*, 58(2), 689-696.

[8]  Spencer, D., & Whitfeld, P. R. (1967). DNA synthesis in isolated chloroplasts. *Biochem Biophys Res Commun*, 28(4), 538-542.

[9]  Wintersberger, E. (1966). Occurrence of a DNA-polymerase in isolated yeast mitochondria. *Biochem Biophys Res Commun*, 25(1), 1-7.

[10]  Parsons, P., & Simpson, M. V. (1967). Biosynthesis of DNA by isolated mitochondria: incorporation of thymidine triphosphate-2-C-14. *Science*, 155(3758), 91-93.

[11]  Sakai, A., Suzuki, T., Nagata, N., Sasaki, N., Miyazawa, Y., Saito, C., Inada, N., Nishimura, Y., & Kuroiwa, T. (1999). Comparative analysis of DNA synthesis activity in plastid-nuclei and mitochondrial-nuclei simultaneously isolated from cultured tobacco cells. *Plant Sci*, 140(1), 9-19.

[12]  Sakai, A. (2001). In vitro transcription/DNA synthesis using isolated organelle-nuclei: Application to the analysis of the mechanisms that regulate organelle genome function. *J Plant Res*, 114(2), 199-211.

[13]  Sakai, A., Takano, H., & Kuroiwa, T. (2004). Organelle nuclei in higher plants: structure, composition, function, and evolution. *Int Rev Cytol*, 238, 59-118.

[14]  Keller, S. J., Biedenbach, S. A., & Meyer, R. R. (1973). Partial purification of a chloroplast DNA polymerase from Euglena gracilis. *Biochem Biophys Res Commun*, 50(3), 620-628.

[15]  Castroviejo, M., Tharaud, D., Tarrago-Litrak, L., & Litvak, S. (1979). Multiple deoxyribonucleic acid polymerases from quiescent wheat embryos. Purification and characterization of three enzymes from the soluble cytoplasm and one from purified mitochondria. *Biochem J*, 181(1), 183-191.

[16]  Fukasawa, H., & Chou, M. Y. (1980). Mitochondrial DNA polymerase from Cauliflower inflorescence. *Jpn J Genet*, 55(6), 441-445.

[17]  Sala, F., Amileni, A. R., Parisi, B., & Spadari, S. (1980). A gamma-like DNA polymerase in spinach chloroplasts. *Eur J Biochem*, 112(2), 211-217.

[18]  Christophe, L., Tarrago-Litvak, L., Castroviejo, M., & Litvak, S. (1981). Mitochondrial DNA polymerase from wheat embryos. *Plant Sci Lett*, 21(2), 181-192.

[19]  McKown, R. L., & Tewari, K. K. (1984). Purification and properties of a pea chloroplast DNA polymerase. *Proc Natl Acad Sci U S A*, 81(8), 2354-2358.

[20]  Heinhorst, S., Cannon, G. C., & Weissbach, A. (1990). Chloroplast and mitochondrial DNA polymerases from cultured soybean cells. *Plant physiol*, 92(4), 939-945.

[21] Keim, C. A., & Mosbaugh, D. W. (1991). Identification and characterization of a 3′ to 5′ exonuclease associated with spinach chloroplast DNA polymerase. *Biochemistry*, 30(46), 11109-11118.

[22] Wang, Z. F., Yang, J., Nie, Z. Q., & Wu, M. (1991). Purification and characterization of a gamma-like DNA polymerase from Chlamydomonas reinhardtii. *Biochemistry*, 30(4), 1127-1131.

[23] Meißner, K., Heinhorst, S., Cannon, G. C., & Börner, T. (1993). Purification and characterization of a gamma-like DNA polymerase from Chenopodium album L. *Nucleic Acids Res.*, 21(21), 4893-4899.

[24] Bailey, J. C., Heinhorst, S., & Cannon, G. C. (1995). Accuracy of deoxynucleotide incorporation by soybean chloroplast DNA polymerases is independent of the presence of a 3′ to 5′ exonuclease. *Plant Physiol*, 107(4), 1277-1284.

[25] Gaikwad, A., Hop, D. V., & Mukherjee, S. K. (2002). A 70-kDa chloroplast DNA polymerase from pea (Pisum sativum) that shows high processivity and displays moderate fidelity. *Mol Genet Genomics*, 267(1), 45-56.

[26] Moriyama, T. (2008). Studies on the DNA polymerases localized to plastids and mitochondria in plants and algae. Thesis. The University of Tokyo;.

[27] Kimura, S., Uchiyama, Y., Kasai, N., Namekawa, S., Saotome, A., Ueda, T., Ando, T., Ishibashi, T., Oshige, M., Furukawa, T., Yamamoto, T., Hashimoto, J., & Sakaguchi, K. (2002). A novel DNA polymerase homologous to Escherichia coli DNA polymerase I from a higher plant, rice (Oryza sativa L.). *Nucleic Acids Res*, 30(7), 1585-1592.

[28] Takeuchi, R., Kimura, S., Saotome, A., & Sakaguchi, K. (2007). Biochemical properties of a plastidial DNA polymerase of rice. *Plant Mol Biol*, 64(5), 601-611.

[29] Mori, Y., Kimura, S., Saotome, A., Kasai, N., Sakaguchi, N., Uchiyama, Y., Ishibashi, T., Yamamoto, T., Chiku, H., & Sakaguchi, K. (2005). Plastid DNA polymerases from higher plants, Arabidopsis thaliana. *Biochem. Biophys. Res Commun*, 334(1), 43-50.

[30] Parent, J. S., Lepage, E., & Brisson, N. (2011). Divergent roles for the two PolI-like organelle DNA polymerases of Arabidopsis. *Plant Physiol*, 156(1), 254-262.

[31] Ono, Y., Sakai, A., Takechi, K., Takio, S., Takusagawa, M., & Takano, H. (2007). NtPolI-like1 and NtPolI-like2, bacterial DNA polymerase I homologues isolated from BY-2 cultured tobacco cells, encode DNA polymerases engaged in DNA replication in both plastids and mitochondria. *Plant Cell Physiol*, 48(12), 1679-1692.

[32] Moriyama, T., Terasawa, K., Fujiwara, M., & Sato, N. (2008). Purification and characterization of organellar DNA polymerases in the red alga Cyanidioschyzon merolae. *FEBS J*, 275(11), 2899-2918.

[33] Moriyama, T., Terasawa, K., & Sato, N. (2011). Conservation of POPs, the plant organellar DNA polymerases, in eukaryotes. *Protist*, 162(1), 177-187.

[34] Wanrooij, S., & Falkenberg, M. (2010). The human mitochondrial replication fork in health and disease. *Biochim Biophys Acta*, 1797(8), 1378-1388.

[35] Williams, A. J., Wernette, C. M., & Kaguni, L. S. (1993). Processivity of mitochondrial DNA polymerase from *Drosophila* embryos. Effects of reaction conditions and enzyme purity. *J Biol Chem*, 268(33), 24855-24862.

[36] Holmes, A. M. (1981). Studies on the inhibition of highly purified calf thymus 8S and 7.3S DNA polymerase alpha by aphidicolin. *Nucleic Acids Res*, 9(1), 161-168.

[37] Weiser, T., Gassmann, M., Thömmes, P., Ferrari, E., Hafkemeyer, P., & Hübscher, U. (1991). Biochemical and functional comparison of DNA polymerases alpha, delta, and epsilon from calf thymus. *J Biol Chem*, 266(16), 10420-10428.

[38] Chavalitshewinkoon-Petmitr, P., Chawprom, S., Naesens, L., Balzarini, J., & Wilairat, P. (2000). Partial purification and characterization of mitochondrial DNA polymerase from Plasmodium falciparum. *Parasitol Int*, 49(4), 279-288.

[39] Shiraki, K., Okuno, T., Yamanishi, K., & Takahashi, M. (1989). Phosphonoacetic acid inhibits replication of human herpesvirus-6. *Antiviral Res*, 12(5-6), 311-318.

[40] Leinbach, S. S., Reno, J. M., Lee, L. F., Isbell, A. F., & Boezi, J. A. (1976). Mechanism of phosphonoacetate inhibition of herpesvirus-induced DNA polymerase. *Biochemistry*, 15(2), 426-430.

[41] Kornberg, A., & Baker, T. A. (1992). DNA Replication. 2nd ed. New York: W.H. Freeman and Company, 199.

[42] Reha-Krantz, L. J., & Nonay, R. L. (1994). Motif A of bacteriophage T4 DNA polymerase: role in primer extension and DNA replication fidelity. Isolation of new antimutator and mutator DNA polymerases. *J Biol Chem*, 269(8), 5635-5643.

[43] Li, L., Murphy, K. M., Kanevets, U., & Reha-Krantz, L. J. (2005). Sensitivity to phosphonoacetic acid: a new phenotype to probe DNA polymerase δ in Saccharomyces cerevisiae. *Genetics*, 170(2), 569-580.

[44] Christensen, A. C., Lyznik, A., Mohammed, S., Elowsky, C. G., Elo, A., Yule, R., & Mackenzie, S. A. (2005). Dual-domain, dual-targeting organellar protein presequences in Arabidopsis can use non-AUG start codons. *Plant Cell*, 17(10), 2805-2816.

[45] Pinz, K. G., & Bogenhagen, D. F. (2000). Characterization of a catalytically slow AP lyase activity in DNA polymerase γ and other family A DNA polymerases. *J Biol Chem*, 275(17), 12509-12514.

[46] Kuroiwa, T., Fujie, M., & Kuroiwa, H. (1992). Studies on the behavior of mitochondrial DNA: Synthesis of mitochondrial DNA occurs actively in a specific region just above the quiescent center in the root apical meristem of Pelargonium zonale. *J Cell Sci*, 101(3), 483-493.

[47] Fujie, M., Kuroiwa, H., Kawano, S., & Kuroiwa, T. (1993). Studies on the behavior of organelles and their nucleoids in the root apical meristem of Arabidopsis thaliana (L.) Col. *Planta*, 189(3), 443-452.

[48] Suzuki, T., Kawano, S., Sakai, A., Fujie, M., Kuroiwa, H., Nakamura, H., & Kuroiwa, T. (1992). Preferential mitochondrial and plastid DNA synthesis before multiple cell divisions in Nicotiana tabacum. *J Cell Sci*, 103(3), 831-837.

[49] Suzuki, T., Sasaki, N., Sakai, A., Kawano, S., & Kuroiwa, T. (1995). Localization of organelle DNA synthesis within root apical meristem of rice. *J Exp Bot*, 46(1), 19-25.

[50] Fujie, M., Kuroiwa, H., Kawano, S., Mutoh, S., & Kuroiwa, T. (1994). Behavior of organelles and their nucleoids in the shoot apical meristem during leaf developmet in Arabidopsis thaliana L. *Planta*, 194(3), 395-405.

[51] Okamura, S., Suzuki, T., Miyazawa, Y., Kuroiwa, T., & Sakai, A. (2002). Activation of organelle DNA synthesis during the initial phase of proliferation of BY-2 cultured tobacco cells after medium renewal. *Plant Morphol*, 14(1), 16-28.

[52] Hashimoto, H. (1986). Double ring structure around the constricting neck of dividing plastids of Avena sativa. *Protoplasma*, 135(2-3), 166-172.

[53] Possingham, J. V., Hashimoto, H., & Oross, J. (1988). Factors that influence plastid division in higher plants. In Boffey SA, Lloyd D. (ed.), *The division and segregation of organelles.*, New York: W.H. Cambridge University Press.

[54] Hashimoto, H., & Possingham, J. V. (1989). DNA Levels in Dividing and Developing Plastids in Expanding Primary Leaves of Avena sativa. *J Exp Bot*, 40(2), 257-262.

[55] Matsuzaki, M., Misumi, O., Shin-i, T., Maruyama, S., Takahara, M., Miyagishima, S., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., Yoshida, Y., Nishimura, Y., Nakao, S., Kobayashi, T., Momoyama, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, H., Oishi, K., Hayashi, H., Ohta, F., Nishizaka, S., Haga, S., Miura, S., Morishita, T., Kabeya, Y., Terasawa, K., Suzuki, Y., Ishii, Y., Asakawa, S., Takano, H., Ohta, N., Kuroiwa, H., Tanaka, K., Shimizu, N., Sugano, S., Sato, N., Nozaki, H., Ogasawara, N., Kohara, Y., & Kuroiwa, T. (2004). Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. *Nature*, 428(6983), 653-657.

[56] Suzuki, K., Ehara, T., Osafune, T., Kuroiwa, H., Kawano, S., & Kuroiwa, T. (1994). Behavior of mitochondria, chloroplasts and their nuclei during the mitotic cycle in the ultramicroalga Cyanidioschyzon merolae. *Eur J Cell Biol*, 63(2), 280-288.

[57] Moriyama, T., Terasawa, K., Sekine, K., Toyoshima, M., Koike, M., Fujiwara, M., & Sato, N. (2010). Characterization of cell-cycle-driven and light-driven gene expression in a synchronous culture system in the unicellular rhodophyte Cyanidioschyzon merolae. *Microbiology*, 156(6), 1730-1737.

[58] Itoh, R., Takahashi, H., Toda, K., Kuroiwa, H., & Kuroiwa, T. (1997). DNA gyrase involvement in chloroplast-nucleoid division in Cyanidioschyzon merolae. *Eur J Cell Biol*, 73(3), 252-258.

[59] Wall, M. K., Mitchenall, L. A., & Maxwell, A. (2004). Arabidopsis thaliana DNA gy-rase is targeted to chloroplasts and mitochondria. *Proc Natl Acad Sci U S A*, 101(20), 7821-7826.

[60] Edmondson, A. C., Song, D., Alvarez, L. A., Wall, M. K., Almond, D., Mc Clellan, D. A., Maxwell, A., & Nielsen, B. L. (2005). Characterization of a mitochondrially target-ed single-stranded DNA-binding protein in Arabidopsis thaliana. *Mol Genet Genom-ics.*, 273(2), 115-122.

[61] Nielsen, B. L., Rajasekhar, V. K., & Tewari, K. K. (1991). Pea chloroplast DNA pri-mase: characterization and role in initiation of replication. *Plant Mol Biol*, 16(6), 1019-1034.

[62] Nie, Z., & Wu, M. (1999). The functional role of a DNA primase in chloroplast DNA replication in Chlamydomonas reinhardtii. *Arch Biochem Biophys*, 369(1), 174-180.

[63] Shutt, T. E., & Gray, M. W. (2006). Twinkle, the mitochondrial replicative DNA heli-case, is widespread in the eukaryotic radiation and may also be the mitochondrial DNA primase in most eukaryotes. *J Mol Evol*, 62(5), 588-599.

[64] Mukhopadhyay, A., Chen, C. Y., Doerig, C., Henriquez, F. L., Roberts, C. W., & Bar-rett, M. P. (2009). The Toxoplasma gondii plastid replication and repair enzyme com-plex, PREX. *Parasitology*, 136(7), 747-755.

[65] Seki, M., & Wood, R. D. (2008). DNA polymerase θ (POLQ) can extend from mis-matches and from bases opposite a (6-4) photoproduct. *DNA Repair (Amst)*, 7(1), 119-127.

[66] Marini, F., Kim, N., Schuffert, A., & Wood, R. D. (2003). POLN, a nuclear PolA family DNA polymerase homologous to the DNA cross-link sensitivity protein Mus308. *J Biol Chem*, 278(34), 32014-32019.

[67] Seow, F., Sato, S., Janssen, C. S., Riehle, M. O., Mukhopadhyay, A., Phillips, R. S., Wilson, R. J., & Barrett, M. P. (2005). The plastidic DNA replication enzyme complex of Plasmodium falciparum. *Mol Biochem Parasitol*, 141(2), 145-153.

[68] Baldauf, S. L. (2003). The deep roots of eukaryotes. Science , 300(5626), 1703-1706.

[69] Filée, J., Forterre, P., Sen-Lin, T., & Laurent, J. (2002). Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. *J Mol Evol*, 54(6), 763-773.

[70] Wanrooij, S., & Falkenberg, M. (2010). The human mitochondrial replication fork in health and disease. *Biochim Biophys Acta*, 1797(8), 1378-1388.

[71] Nielsen, B. L., & Cupp, Brammer. J. (2010). Mechanisms for maintenance, replication, and repair of the chloroplast genome in plants. *J Exp Bot*, 61(10), 2535-2537.

[72] Guiguemde, W. A., Shelat-Bustos, Garcia., Diagana, J. F., Gamo, T. T., Guy, F. J., & , R. K. (2012). Global phenotypic screening for antimalarials. *Chem Biol*, 19(1), 116-129.

[73] García-Estrada, C., Prada, C. F., Fernández-Rubio, C., Rojo-Vázquez, F., & Balaña-Fouce, R. (2010). DNA topoisomerases in apicomplexan parasites: promising targets for drug discovery. *Proc Biol Sci*, 277(1689), 1777-1787.

[74] Korhonen, J. A., Pham, X. H., Pellegrini, M., & Falkenberg, M. (2004). Reconstitution of a minimal mtDNA replisome in vitro. *EMBO J*, 23(12), 2423-2429.

# Chromatin and Epigenetic Influences on DNA Replication

# Roles of Methylation and Sequestration in the Mechanisms of DNA Replication in some Members of the Enterobacteriaceae Family

Amine Aloui, Alya El May,
Saloua Kouass Sahbani and Ahmed  Landoulsi

Additional information is available at the end of the chapter

## 1. Introduction

When growing cells divide, they need to copy their genetic material and distribute it to en‐sure that each daughter cell receives one copy. This is a challenging task especially when the enormous length of the DNA compared to the cell size is considered. During DNA replica‐tion, organization of the chromosomes is even more demanding, since replication forks con‐tinuously produce new DNA. This DNA contains all the information required to build the cells and tissues of a prokaryotic or an eukaryotic organism. The exact replication of this in‐formation in any species assures its genetic continuity from generation to generation and is critical to the normal development of an individual. The information stored in DNA is ar‐ranged in hereditary units known as genes that control the identifiable traits of an organism. Discovery of the structure of DNA and subsequent elucidation of how DNA directs synthe‐sis of RNA, which then directs assembly of proteins -the so-called central dogma - were monumental achievements that marked the early days of molecular biology. However, the simplified representation of the central dogma as DNA → RNA → protein does not reflect the role of proteins in the synthesis of nucleic acids. Moreover, proteins are largely responsi‐ble for regulating DNA replication and gene expression, the entire process whereby the in‐formation encoded in DNA is decoded into the proteins that characterize various cell types. Two of these proteins are the DNA adenine methyltransferase (Dam) and the DNA-Binding Protein (SeqA).

## 1.1. The Dam methyltransferase

Methylation of DNA by the Dam methyltransferase provides an epigenetic signal that influences and regulates numerous physiological processes in the bacterial cell, including chromosome replication, mismatch repair, transposition, and transcription. A growing number of reports ascribed a role to DNA adenine methylation in regulating the mechanisms of DNA replication in diverse pathogens like in *Salmonella* Typhimurium and *Escherichia coli*, …), suggesting that DNA methylation may be a widespread and versatile regulator of this process. The Dam enzyme catalyzes the transfer of a methyl group from S -adenosyl-L-methionine (SAM) to the $N^6$ position of the adenine residue in GATC sequences, using base flipping to position the base in the enzyme's catalytic site (Figure 1).
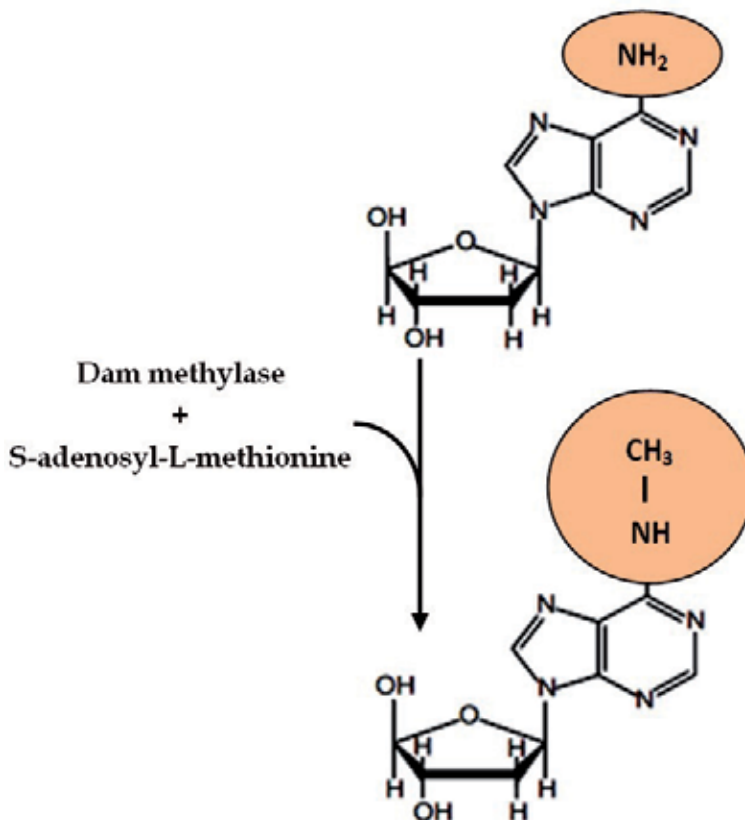
**Figure 1.** Dam catalyzes the transfer of a methyl group from S-adenosyl-L-methionine to the $N^6$ position of adenine.

The natural substrate for the enzyme is hemimethylated DNA, where one strand is methylated and the other is not. This is the configuration of DNA immediately behind the replication fork. Double stranded DNA is a better methyl receiver than denatured DNA, and there is little difference in the rate of methylation between unmethylated and hemimethylated DNA [28]. The Dam enzyme appears to have two SAM binding sites; one is the catalytic site

and the other increases specific binding to DNA, probably through an allosteric transformation [13]. Dam is thought to bind the template and to slide processing along the DNA, methylating about 55 GATC sites per binding event [60]. There are about 130 molecules of Dam per *Escherichia coli* cell, and this is considered optimal because it allows a period of time between the synthesis of the extending nucleotide chains and the methylation of the GATC sequences within them [15]. The cellular level of Dam is regulated mainly by transcription; any increases or decreases in the number of Dam molecules can profoundly alter the physiological properties of the cell.

Dam competes with two other proteins, MutH and SeqA, for hemimethylated GATC substrate sites. These two proteins act before Dam to participate in the removal of replication errors (MutH) and to form the compact and properly super coiled chromosome structure for the nucleoid (SeqA). Increasing the cellular level of Dam causes a decrease in the amount of hemimethylated DNA, and prevents these two proteins from carrying out their functions, leading to an increased mutation rate and a change in the super coiling of the chromosome, respectively [27; 42; 40]. Although Dam methylase is a highly processing enzyme, it may become less processing at GATC sites, flanked by specific DNA sequences [50]. Reduced rate of processing may allow for a competition between Dam and specific DNA- binding proteins, thus permitting the formation of non-methylated GATCs which depended on the growth phase and the growth rate, suggesting that the proteins that bind to them could be involved in gene expression or in the maintenance of chromosome structure. The unmethylated dam sites appear to be mostly [53] or completely [49] modified in strains overproducing Dam, suggesting that the enzyme competes with other DNA - binding proteins at these specific sites. In addition to the unmethylated GATC sites discussed above, persistent hemimethylated sequences have been detected in the chromosome [48; 18]. These are distinct from the transiently hemimethylated GATC sites that occur immediately behind the replication fork due to the time lag between DNA replication and Dam methylation.

## 1.2. The SeqA protein

SeqA protein was discovered in some prokaryotes as a protein involved in the methylation / hemimethylation cycle of DNA replication [41]. This protein regulates the activation of the chromosome replication origin [41]. Experiments have shown that SeqA has a high affinity to hemimethylated as compared to fully methylated DNA. It binds specifically to GATC sequences which are methylated on the Adenine of the old strand but not on the new strand. Such hemimethylated DNA is produced by progression of the replication forks and lasts until Dam methyltransferase methylates the new strand (Figure 2). It is therefore believed that a region of hemimethylated DNA covered by SeqA follows the replication fork.
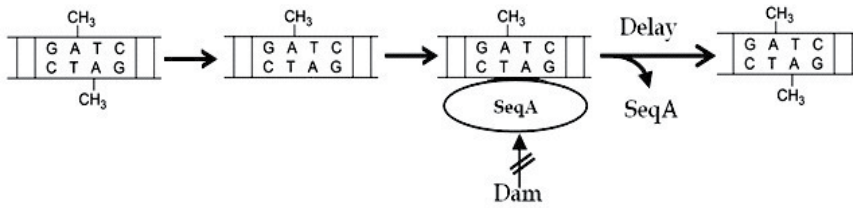
**Figure 2.** Helically phased GATC sites can be bound by SeqA when they are in the hemimethylated state. Binding of SeqA inhibits Dam methylation, maintaining the hemimethylated state for a portion of the cell cycle. Dissociation of SeqA allows Dam to methylate the hemimethylated DNA, thus generating fully methylated DNA.

Proper chromosome segregation also requires SeqA [5]. Furthermore, SeqA trails the DNA replication fork and may contribute to nucleic organization in newly replicated DNA [17; 37; 40; 67]. Aside from its roles in chromosome replication and nucleic segregation, SeqA is known to regulate the transcription of certain genes. In bacteriophage lambda, SeqA activates the $p_R$ promoter in a GATC methylation-dependent fashion. SeqA also acts as a transcriptional co-activator by facilitating binding of the *c*II transcription factor to the lambda $p_I$ and $p_{aQ}$ promoters. Competition between SeqA and the OxyR repressor for hemimethylated GATC sites has been shown to regulate phase variation in the *Escherichia coli agn43* gene [51]. These examples raised the possibility that SeqA binding to critical GATC sites may likewise regulate the expression of prokaryotic genes like in *Escherichia coli* and *Salmonella*, which are members of the *Enterobacteriaceae* family.

### 1.3. The competition of Dam methylase and SeqA for GATC sites

Newly synthesized hemimethylated DNA is also a target for Dam methylase. In fact, Dam and SeqA have been suggested to be in competition for hemimethylated DNA [34]. Experiments with unsynchronized cells indicate that the sequestration period becomes shorter upon Dam over-expression [15; 57]. SeqA binding is largely limited to hemimethylated DNA, and the action of Dam will, therefore, transform DNA into a non-target for SeqA. This protein was able to bind DNA despite Dam overproduction. Recent findings showed that SeqA bound to DNA was not actively dissociated by Dam methylase [34]. The same study showed that the SeqA protein spontaneously dissociated from bound DNA after some minutes *in vitro* and that re-binding to the same site was inhibited by methylation [34]. We reasoned that *in vivo* system, the effect of Dam overproduction on SeqA re-binding should increase with increasing distance from replication forks towards the origin of chromosomal replication. This is because the longer a SeqA molecule was bound to the DNA the more likely is its dissociation. Such an effect might be too small to be observed by visual comparison of SeqA binding patterns. This shows that Dam and SeqA are in continuous competition for GATC sites *in vivo* with SeqA being the considerably stronger competitor. Since SeqA has been shown to bind better to DNA regions with more densely packed GATC sites, we speculated that such regions would allow SeqA to be better in competing against Dam than do regions with fewer GATC sites. Recent data indicate that differences in GATC density have only minor impact on the competition of SeqA against Dam methyltransferase [32].

## 2. DNA replication

The initiation of chromosomal replication occurs only once during the cell cycle in both pro-karyotes and eukaryotes. This initiation is the first and tightly controlled step of a DNA syn-thesis. Because much of what is known about the regulation of the initiation of bacterial chromosomal replication comes from studies of *Escherichia coli* and *Salmonella enterica* sero-var Typhimurium, this review focuses mainly on regulatory mechanisms in these species.

In prokaryotic cells, DNA replication and segregation are not temporally separated process-es. Some evidence suggests that newly synthesized DNA is continuously segregated to op-posite cellular positions [45; 52]. Other work indicates that some parts of segregation may be more abrupt and domain specific [7; 21]. Coordination of DNA replication and chromosome segregation is complicated by the ability of growing with overlapping replication cycles [19; 55]. Whereas during slow growth, chromosomes are replicated in a simple pattern with only one pair of forks; replication during fast growth occurs with one pair of old and two pairs of new forks on one chromosome. (Forks are considered to be 'old forks' as soon as new forks appear at initiation.) Depending on the exact conditions, a cell can have four copies of the multi-fork chromosome and a total of 24 replication forks per cell [22; 44]. How the cell meets the obvious need for efficient organization during such extensive replication is largely unknown. However, the SeqA protein is one of the strongest candidates to contribute [41; 61]. Loss of SeqA leads to severe growth impairment during the rapid but not slow growth [16]. Biochemical studies established that SeqA binding is specific to the sequence GATC with high preference for hemimethylated over fully methylated DNA [10; 58; 11; 34]. Hemi-methylation occurs at newly replicated GATC sites which have not yet been re-methylated by the Dam methylase. A transient hemimethylation after the passage of the replication fork was found in an analysis of 10 individual GATC sites [18]. Similarly, transient binding of SeqA was detected at seven genomic sites with multiple GATC sequences [67]. Multiple DNA- bound SeqA dimers can oligomerize to form a higher order structure [25; 47]. The above findings suggested that a SeqA complex follows the replication forks, potentially in a tread milling fashion, growing at the leading end and diminishing at the tailing end. The re-duction of the SeqA bound region at the most replisome-distant GATCs would come about through the activity of Dam which turns these sites into non-targets for SeqA by its methyla-tion activity. The process described above is called DNA sequestration.

## 3. DNA methylation by Dam protein

As mentioned above, the Dam methyltransferase of *Enterobacteriaceae* methylates adenine at the $N^6$ position in GATC sequences. Methylation of DNA has multiple consequences con-cerning bacterial physiology including the regulation of chromosome replication, DNA seg-regation, mismatch repair, transposition, and transcriptional regulation. The molecular basis for the pleiotropic phenotypes associated with Dam is the differential methylation of DNA resulting in an altered affinity of regulatory DNA-binding proteins. Regulatory proteins

might preferentially bind to non -methylated DNA, thereby blocking methylation by Dam, while other proteins bind with high affinity only to hemimethylated or fully methylated DNA [65]. Therefore, it is not surprising that Dam has an impact on pathogenesis, virulence gene expression, influences DNA replication [2] and many other processes [4] in *Salmonella* microorganisms. Dam-overproducing (Dam[OP]) as well as *dam* mutant strains have been used to assess the role of DNA methylation in DNA replication. By using these strains it is possible to alter the methylation pattern in regulatory regions of genes, thereby changing the binding affinity for regulatory proteins. Although Dam[OP] does not reflect a physiologically relevant condition, as the Dam levels have been found to remain basically constant in the cell, it is a functional tool to analyze the effects of changes in DNA methylation patterns on gene expression. Strategies of this kind have been successfully used for decades in both eukaryotes and prokaryotes to perturb gene regulation for experimental purposes.

## 4. DNA sequestration by SeqA protein

Replication of the bacterial chromosomal DNA initiates only once, at a specific region known as the origin of chromosomal replication ( *oriC*), by the initiator protein DnaA. This protein interacts specifically with 9-bp non-palindromic sequences (DnaA boxes) that exists at *oriC*. To ensure that initiation at an origin occurs only once per cell cycle, specific mechanisms exist to control chromosomal replication. In one mechanism, the SeqA protein that is tightly bound to hemimethylated DNA by a mechanism known as sequestration and which recognizes GATC sequences overrepresented within *oriC* and prefers binding to hemimethylated over fully or unmethylated *oriC*.

The chromosomal DNA is methylated at adenine residues in GATC sequences by Dam methylase. Following passage of the DNA replication fork, GATC sites methylated on the top and bottom strands in a mother cell (denoted as fully methylated) are converted into two hemimethylated DNA duplexes: one methylated on the top strand and non-methylated on the bottom strand and one methylated on the bottom strand and non-methylated on the top strand due to semi-conservative replication. Most GATC sites are rapidly re-methylated by the enzyme Dam methylase and exist in the hemimethylated state for only a fraction of the cell cycle (Figure 3).

Exceptions to this are the DNA replication origin of *Escherichia coli* and *Salmonella typhimurium*, the dnaA promoter, and possibly additional GATC sites in the chromosome which bind SeqA. SeqA preferentially binds to clusters of two or more hemimethylated GATC sites spaced one to two helical turns apart (Figure 4).

In the case of *oriC*, sequestration delays re-methylation and prevents binding of the DnaA protein, which controls the initiation of DNA replication. At other sites, binding of SeqA tetramers to hemimethylated GATC sites may organize nucleic domains. Notably, the transcription profile of a SeqA mutant was found to be similar to that of a Dam overproducer strain. Based on this observation, a model was developed in which Dam and SeqA compete

for binding to hemimethylated for binding to hemimethylated DNA generated at the replication fork.
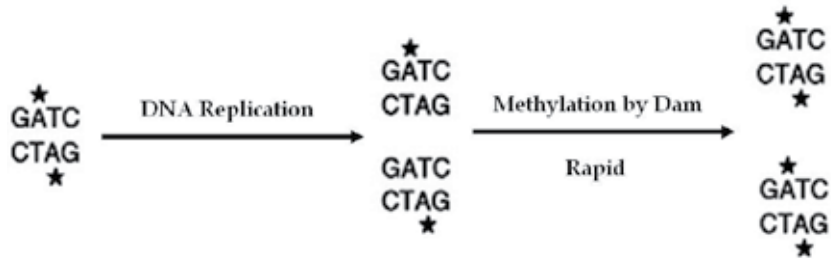


**Figure 3.** The vast majority of chromosomal GATC sites are fully methylated until DNA replication generates two hemimethylated species, one methylated on the top strand and one methylated on the bottom strand. Within a short time after replication (less than 5 minutes), Dam methylates the nonmethylated GATC site, regenerating a fully methylated GATC site.
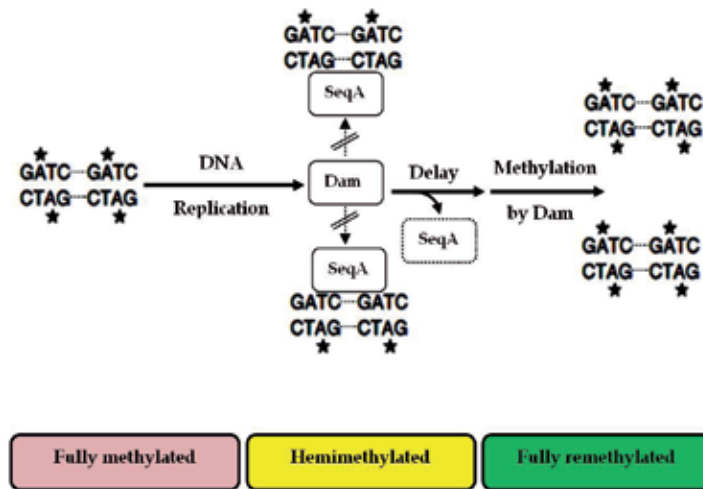


**Figure 4.** Two or more helically phased GATC sites can be bound by SeqA when they are in the hemimethylated state. Binding of SeqA inhibits Dam methylation, maintaining the hemimethylated state for a portion of the cell cycle. Dissociation of SeqA allows Dam to methylate the hemimethylated DNAs, generating fully methylated DNA.

## 4.1. Binding of SeqA to Hemimethylated GATC Sequences

As mentioned above, the adenine residues of GATC sequences are methylated on their 6 amino group by Dam methyltransferase [23; 18]. Upon replication, the GATC sequences on the newly replicated strand remain transiently unmethylated, leading to a hemimethylated state of the DNA duplex. The new strand is subsequently methylated by Dam, and the du-

plex becomes methylated on both strands. The initiation of chromosome replication at the origin of chromosomal replication *oriC*, which contains repeated GATC sequences, is tightly controlled [43; 38]. Once initiation is fired, reinitiation from the newly formed *oriC* is prevented by an *oriC* sequestration process affected by the binding of SeqA protein to the newly replicated, hemi-methylated origin [41; 34]. The hemimethylated state of the replicated *oriC* is maintained for about one-third of the cell cycle, whereas it persists in other chromosomal regions for at most 2 minutes. Further, the asynchronous and overinitiation of chromosomal replication characteristic of *seqA* mutants indicates that SeqA is a negative modulator of chromosomal initiation at *oriC* [41, 16].

Topoisomerase IV is essential for the de-catenation and segregation of replicated chromosomes at cell division [35; 36; 1; 68; 29]. Together with DNA gyrase, it also removes the positive super coils that accumulate in front of replication forks and growing mRNA transcripts. SeqA has been shown to promote the relaxation and the de-catenation activity of topoisomerase IV [33]. This appears to result from a specific interaction between topoisomerase IV and SeqA. Besides the asynchrony and overinitiation of chromosomal replication, *seqA* mutants have an aberrant nucleic structure, an increased frequency of abnormal segregation, and increased negative superhelicity of chromosomal and plasmid DNA [41; 30; 63; 64]. These findings suggest that interaction with SeqA is required for proper functioning of topoisomerase IV *in vivo*. In addition, SeqA functions as a transcriptional regulator of the bacteriophage *pR* promoter [59].

The C-terminal region of SeqA interacts via hydrogen bonds and van der Waals contacts with the major groove of DNA, with the hemimethylated A-T base pair and also with the surrounding bases and DNA backbone. The Nuclear Magnetic Resonance (NMR) structure of hemimethylated GATC revealed that it has an unusual backbone structure and a remarkably narrow major groove and suggested that this peculiar structural feature might contribute to recognition of hemimethylated GATC sites by SeqA protein [6]. To form a stable SeqA-DNA complex in the presence of competitor DNA, one SeqA tetramer binds to each of two hemimethylated GATC sequences [26] that are up to 31 bases apart on the DNA [17]. The sequential binding of SeqA tetramers to hemimethylated sites leads to the formation of higher order complexes [26]. Further, the binding of SeqA proteins to at least six adjacent hemimethylated sites induces the aggregation of free proteins onto the bound proteins, thus implying cooperative interaction between the SeqA proteins.

### 4.2. Effects of *seqA* disruption on DNA replication

As we said before, following the replication fork progression and the nascent strand synthesis, the daughter DNA becomes hemimethylated. SeqA protein binds to the hemimethylated GATC sequences (hemi-sites) and performs various roles to control the cell cycle progression. Immediately after the initiation of replication SeqA binds to the replicated *oriC* and sequesters it from remethylation and re-initiation of replication at the replicated *oriC*. SeqA tracks replication forks as a multiprotein complex and contributes to the maintenance of su-

perhelicity and de-catenation of daughter chromosomes through the stimulation of topoisomerase IV and results in a synchronous replication.

When rounds of replication are allowed to run to completion, the number of chromosomes per cell is 2n (n = 0, 1, 2, 3, etc). When initiations are asynchronous, as in *dnaA* (Ts) initiation mutants at the permissive temperature and in the *Escherichia coli dam* mutant [14; 56], the presence of a different number of chromosome equivalents (three, five, six, etc.) was detected by flow cytometry. The presence of cells containing a number of chromosomes different from 2n suggests that the *seqA* mutant has a defect in the synchrony of replication initiation. Wild type and *seqA* mutant of *Salmonella enterica* serovar Typhimurium growing exponentially in glucose–casamino acid medium were treated with rifampicin and cephalexin, which block initiation of replication and cell division respectively. Wild-type cells initiated replication synchronously (number of chromosomes per cell is 2n). The appearance of cells with chromosome numbers other than 2n indicates a moderate asynchrony of initiation.

So, the flow cytometer analysis of a *seqA* mutants has shown that replication initiation is asynchronous and can occur throughout the cell cycle, not only at the normal cell age for initiation. The most likely reason for this asynchrony phenotype is that secondary initiations occurred at newly replicated origins in *seqA* mutants, due to lack of sequestration and inadequate methylation. We showed that the initiation synchrony was dependent on intact GATC methylation sites.

This loss of synchrony affected culture growth rates and cell size distributions only slightly and suggest that *seqA* mutants have a slight defect in synchronizing replication initiation. All these results suggest that DNA sequestration plays a role in preventing the occurrence of multiple initiations at a single origin in the same replication cycle. However, using flow cytometry, we found that the asynchrony of initiation, which is one of the phenotypes of the *seqA* mutation, was returned to almost normal in a *seqA* null mutant harbouring the wild-type *seqA* gene under the control of a *tac-* promoter.

The OFF to ON phase rate was reduced in a *seqA* mutant, but much of this effect could be accounted for by a reduction in the Dam / DNA ratio caused by increased asynchronous initiation of DNA replication that occurs in the absence of SeqA, which normally sequesters *oriC* and plays a critical role in timing of DNA replication [8].

## 5. Membrane sequestration hemimethylated of *oriC*

The coordination of the synchronization of the replication initiation, the activation of the DnaA protein at *oriC*, and the cellular cycle suggested the existence of a very narrow interaction between the bacterial membrane and the SeqA protein [39]. Early studies demonstrated that membranes are capable of binding to hemimethylated *oriC in vitro* and *in vivo*, but not to fully methylated or unmethylated *oriC* [48]. While they are sequestered at the membrane, the recently replicated origins are unavailable for re-initiation and are protected from meth-

ylation by Dam methylase for an extended period. The origins remain sequestered until con-
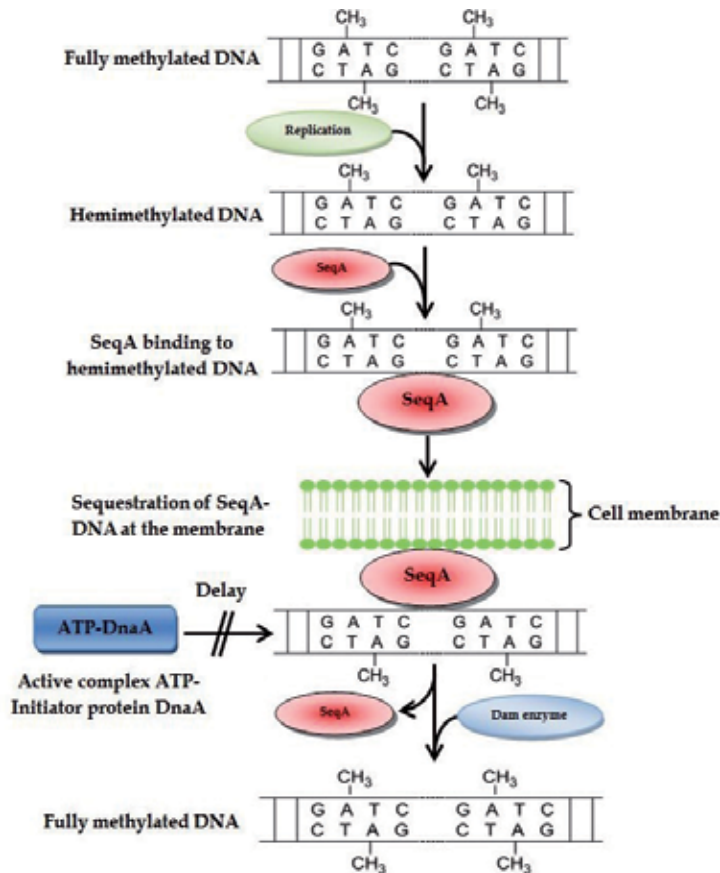ditions in the cell are no longer in a state supportive for initiation (Figure 5).



**Figure 5.** Membrane sequestration of recently replicated origins [4].

Prior to initiation of DNA replication, Dam methylase sites are fully methylated. Immediate-
ly following replication, the newly synthesized strand is unmethylated, and the resulting
hemimethylated origin is sequestered at the lipid bilayer of membrane by SeqA. This is not
accessible to replicatively active ATP–DnaA. After approximately one-third of the cell cycle,
the sequestered origin is released and methylated by Dam methylase. At this point in the
cell cycle, the levels of ATP–DnaA are not sufficient to catalyze a new round of replication.
As such, sequestration serves as a mechanism to prevent secondary initiations. Subsequent
work identified SeqA protein to be an essential factor in the *oriC* sequestration. Even though
the first steps of SeqA purificiation involve liberating SeqA from the membrane fraction of
cell lysates by treatment with high concentrations of salt and sonication, the primary se-
quence for SeqA protein does not suggest any obvious membrane-associating domains. This
is supported by the crystal structure of the C-terminal DNA-binding domain, and by bio-
chemical studies that show that the N-terminal domain serves in the aggregation of SeqA

protein into functional homotetramers [24]. Yet, there is some evidence that SeqA has an association with membranes [20; 62]. The original data that newly replicated, hemimethylated origins are sequestered at the membrane hold true. Whether the membrane sequestration of *oriC* occurs directly through the SeqA protein or through an as yet unidentified factor remains unclear.

## 6. Cooperation between "Dam" and "SeqA" in DNA Replication

In *Escherichia coli*, persistent hemimethylated sites have been detected at the origin of chromosome replication, *oriC*, and the region surrounding it [18]. This region includes the dnaA gene, which is located 43 kb from *oriC*. DnaA initiates chromosome replication by binding to *oriC* and facilitating duplex opening to load DnaB helicase and DNA polymerase III holoenzyme. The persistence of the hemimethylated state is due to the high density of GATC sequences in *oriC* (11 in 245 bp) and in the promoter region of *dnaA* (8 in 219 bp), providing multiple binding sites for the SeqA protein. The SeqA induced hemimethylated state in this region of the chromosome lasts for about one-third of the cell cycle (sequestration), but the mechanism by which it is relieved is not known. The purpose of sequestration is to prevent re-initiation from *oriC* from occurring more than once per cell cycle. For initiation to occur most efficiently, *oriC* and the *dnaA* promoter region must be fully methylated. This also contributes to ensuring that initiation occurs only once per cell cycle [9; 66]. In *Salmonella* Typhimurium, SeqA may play replication-related roles similar to those described in *Escherichia coli* [51]. In *Vibrio cholerae*, both Dam methylation and SeqA are essential [31; 54], and SeqA overproduction causes DNA replication arrest [54].

In fast-growing *Escherichia coli* or *Salmonella* Typhimurium cells, the time required for chromosome replication exceeds the doubling time. Under such conditions, *Escherichia coli* and *Salmonella* Typhimurium cells contain multiple copies of *oriC* due to initiations that occurred two or three generations ago. These origins fire simultaneously during the cell cycle, leading to synchronous initiation, which is thought to be due to the immediate release of DnaA from an origin after initiation [46]. This release will temporarily increase the DnaA / *oriC* ratio in wild-type cells for the remaining fully methylated origins. After initiation, other mechanisms ensure that DnaA is not in the proper conformation for initiation. Among these mechanisms is a reduction in the transcription of the *dnaA* gene. Sequestration by SeqA after initiation keeps the *dnaA* promoter region in a hemimethylated state, which reduces transcription initiation because the *dnaA* promoter GATC sequences need to be fully methylated for maximal expression [9].

In *dam* mutant cells, there is no sequestration by SeqA; consequently, DnaA can immediately rebind origins after the first initiation event, and initiate a second time when the concentration of the active form of DnaA is high enough. Transcription from the *dnaA* gene continues throughout the cell cycle although at a reduced level. Dam methylation, therefore, is not essential for replication initiation; rather, the cell uses methylation to discriminate between old and new origins.

# 7. Conclusion

In the *Enterobacteriaceae* family (*Salmonella* Typhimurium, *Escherichia coli*,…), DNA methylation and sequestration modulate a variety of processes such as DNA replication and transcription of certain genes. Deletion of the *dam* and / or *seqA* genes produces a variety of phenotypes ranging from replication asynchrony to virulence attenuation, indicating multiple functions for Dam and SeqA proteins in modulating gene expression, proper chromosome segregation, initiation of chromosome replication, and nucleic stabilization. Given these multiple roles, it is not surprising that these mutations are highly pleiotropic. However, the lack of Dam and / or SeqA proteins does not impair viability. Bacterial mutant strains are more sensitive to these mutations than the wild type which shows the inverse. In addition, no great difference between the mutants of *Salmonella* Typhimurium and those of some *Enterobacterial* species such as *Escherichia coli* was observed with replication asynchrony. In conclusion, the role of Dam and SeqA in the prokaryotic cellular processes such as the DNA replication is clear. So it may rely on their capacity as a global regulator of the gene expression during bacterial life, *in vitro*, in a similar manner as it does *in vivo*.

# 8. Future research

Our knowledge on the effects of Dam and SeqA proteins in *Enterobacteriaceae* family (*Salmonella* Typhimurium, *Escherichia coli*,…) has considerably improved in the last decade. This fundamental research has several implications that will prove to be useful for the development of novel therapeutic approaches. But, to date, therapeutic applications are still in their early experimental phases, but several recent studies provide promising results for future clinical developments. Over the last few years, many studies have demonstrated that *Escherichia coli* and *Salmonella typhimurium seqA* and / or *dam* mutants exhibit asynchronous DNA replication and are highly attenuated for virulence in mice and have been proposed as live vaccines. These results prove that these proteins might have a role in regulating virulence. In addition, future research must focus on the study of the decreasing virulence and the proteomic and enzymatic activities of a *seqA* and / or *dam* mutant strains. So these perspectives can be useful to more fully understand the significance of the results obtained above. Of special interests are: firstly, the growing list of genes governed by DNA methylation and sequestration in bacterial pathogens ; secondly, the finding of novel genes regulated by Dam and SeqA proteins using high throughput analysis, and, thirdly, the evidence that these proteins may regulate the expression of many unidentified genes involved in the DNA replication.

Finally, the way in which Dam and SeqA participate clearly in the DNA replication is a critical question that deserves further investigation in the near future, and may be research studies will have to identify explanations.

## Acknowledgements

## Author details

Amine Aloui[1*], Alya El May[1], Saloua Kouass Sahbani[1,2] and Ahmed  Landoulsi[1]

*Address all correspondence to: aminealoui@yahoo.fr

1 Carthage University, Faculty of Sciences of Bizerta Zarzouna, Tunisia

2 Sherbrooke University, Faculty of Medecine Québec, Canada

## References

[1] Adams, D. E., Shekhtman, E. M., Zechiedrich, E. L., Schmid, M. B., & Cozzarelli, N. R. (1992). The role of topoisomerase IV in partitioning bacterial replicons and the structure of catenated intermediates in DNA replication. *Cell*, 71, 277-88.

[2] Aloui, A., Chatti, A., May, E. L. A., & Landoulsi, A. (2007). Effect of methylation on DNA replication in Salmonella enterica serovar Typhimurium. *C R Biol*, 330, 576-580.

[3] Aloui, A., Kouass, S. S., Mihoub, M., El May, A., & Landoulsi, A. (2011). The Absence of the "GATC- Binding Protein SeqA" Affects DNA replication in Salmonella enterica serovar Typhimurium. *DNA Replication and Related Cellular Processes*, 978-9-53307-212-7.

[4] Aloui, A., Tagourti, J., May, E. L. A., Joseleau, P. D., & Landoulsi, A. (2011). The effect of methylation on some biological parameters in Salmonella enterica serovar Typhimurium. Path biol, , 59, 192-198.

[5] Bach, T., Krekling, M. A., & Skarstad, K. (2003). Excess SeqA prolongs sequestration of oriC and delays nucleoid segregation and cell division. EMBO J. , 22, 315-323.

[6]   Bae, S. H., Cheong, H. K., Cheong, C., Kang, S., Hwang, D. S., & Choi, B. S. (2003). Structure and dynamics of hemimethylated GATC sites: implications for DNA-SeqA recognition. *J. Biol. Chem.*, 278, 45987 -93 .

[7]   Bates, D., & Kleckner, N. (2005). Chromosome and replisome dynamics in E. coli: loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell*, 121, 899 -911 .

[8]   Bogan, J. A., & Helmstetter, C. E. (1997). DNA sequestration and transcription in the oriC region of Escherichia coli. *Mol. Microbiol.*, 26, 889-896.

[9]   Braun, R. E., O'Day, K., & Wright, A. (1985). Autoregulation of the DNA replication gene dnaA in E. coli K-12. *Cell*, 40, 159-169.

[10]  Brendler, T., Abeles, A., & Austin, S. (1995). A protein that binds to the 1 origin core and the oriC 13mer region in a methylation-specific fashion is the product of the host seqA gene. EMBO J. , 14, 4083-4089.

[11]  Brendler, T, & Austin, S. (1999). Binding of SeqA protein to DNA requires interaction between two or more complexes bound to separate hemimethylated GATC sequences. *EMBO J.*, 18, 2304-2310.

[12]  Brendler, T., Sawitzke, J., Sergueev, K., & Austin, S. (2000). A case for sliding SeqA tracts at anchored replication forks during Escherichia coli chromosome replication and segregation. *EMBO J.*, 19, 6249-6258.

[13]  Bergerat, A., Guschlbauer, W., & Fazakerley, G. V. (1991). Allosteric and catalytic binding of S-adenosylmethionine to Escherichia coli DNA adenine methyltransferase monitored by 3H NMR. *P Natl Acad Sci USA*, 88, 6394-7.

[14]  Boye, E., & Lobner-Olesen, A. (1990). The role of dam methyltransferase in the control of DNA replication in E. coli. *Cell*, 62, 981-9.

[15]  Boye, E., Marinus, M. G., & Løbner-Olesen, A. (1992). Quantitation of Dam methyltransferase in Escherichia coli. *J Bacteriol*, 174, 1682-1685.

[16]  Boye, E., Stokke, T., Kleckner, N., & Skarstad, K. (1996). Coordinating DNA replication initiation with cell growth: differential roles for DnaA and SeqA proteins. *Proc. Natl Acad. Sci. USA*, 93, 12206-12211.

[17]  Brendler, T., Sawitzke, J., Seerguev, K., & Austin, S. (2000). A case for sliding SeqA tracts at anchored replication forks during Escherichia coli chromosome replication and segregation. *EMBO J.*, 19, 6249-6258.

[18]  Campbell, J. L., & Kleckner, N. (1990). E. coli oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell*, 62, 967-979.

[19]  Cooper, S., & Helmstetter, C. E. (1968). Chromosome replication and the division cycle of Escherichia coli B/r. J. Mol. Biol. , 31, 519-40.

[20] d'Alencon, E., Taghbalout, A., Kern, R., & Kohiyama, M. (1999). Replication cycle dependent association of SeqA to the outer membrane fraction of E. coli. *Biochimie*, 841-846.

[21] Espeli, O., Mercier, R., & Boccard, F. (2008). DNA dynamics vary according to macrodomain topography in the E. coli chromosome. Mol. Microbiol. , 68, 1418-1427.

[22] Fossum, S., Crooke, E., & Skarstad, K. (2007). Organization of sister origins and replisomes during multifork DNA replication in Escherichia coli. EMBO J. , 26, 4514-4522.

[23] Geier, G. E., & Modrich, P. (1979). Recognition sequence of the dam methylase of Escherichia coli K12 and mode of cleavage of Dpn I endonuclease. J. Biol. Chem. , 254, 1408-13.

[24] Guarné, A., Zhao, Q., Ghirlando, R., & Yang, W. (2002). Insights into negative modulation of E. coli replication initiation from the structure of SeqA-hemimethylated DNA complex. *Nat. Struct. Biol.*, 9, 839-843.

[25] Guarné, A., Brendler, T., Zhao, Q., Ghirlando, R., Austin, S., & Yang, W. (2005). Crystal structure of a SeqA-N filament: implications for DNA replication and chromosome organization. EMBO J. , 24, 1502-1511.

[26] Han, J. S., Kang, S., Lee, H., Kim, H. K., & Hwang, D. S. (2003). Sequential binding of SeqA to paired hemi-methylated GATC sequences mediates formation of higher order complexes. *J. Biol. Chem.*, 278, 34983-9.

[27] Herman, G. E., & Modrich, P. (1981). Escherichia coli K-12 clones that overproduce dam methylase are hypermutable. *J Bacteriol*, 145, 644-6.

[28] Herman, G. E., & Modrich, P. (1982). Escherichia coli dam methylase. Physical and catalytic properties of the homogeneous enzyme. *J Biol Chem*, 257, 2605-2612.

[29] Hiasa, H., & Marians, K. J. (1996). Two distinct modes of strand unlinking during theta-type DNA replication. J. Biol. Chem. , 271, 21529-35.

[30] Hiraga, S., Ichinose, C., Niki, H., & Yamazoe, M. (1998). Cell cycle-dependent duplication and bidirectional migration of SeqA-associated DNA-protein complexes in E. coli. *Mol. Cell*, 1, 381-7.

[31] Julio, S. M., Heithoff, D. M., Provenzano, D., Klose, K. E., Sinsheimer, R. L., Low, D. A., & Mahan, M. J. (2001). DNA adenine methylase is essential for viability and plays a role in the pathogenesis of Yersinia pseudotuberculosis and Vibrio cholerae. *Infect Immun*, 69, 7610-5.

[32] Joo, S. H., Sukhyun, K., Sung, H. K., Min, J. K., & Deog, S. H. (2004). Binding of SeqA protein to hemi-methylated GATC sequences enhances their interaction and aggregation properties. J. Biol. Chem. Issue of July 16 , 279(29), 30236-30243.

[33] Kang, S., Han, J. S., Park, J. H., Skarstad, K., & Hwang, D. S. (2003). SeqA protein stimulates the relaxing and decatenating activities of topoisomerase IV. J. Biol. Chem. , 278, 48779-85.

[34] Kang, S., Lee, H., Han, J. S., & Hwang, D. S. (1999). Interaction of SeqA and Dam methylase on the hemimethylated origin of Escherichia coli chromosomal DNA replication. J. Biol. Chem. , 274, 11463-8.

[35] Kato, J., Nishimura, Y., Yamada, M., Suzuki, H., & Hirota, Y. (1988). Gene organization in the region containing a new gene involved in chromosome partition in Escherichia coli. *J. Bacteriol.*, 170, 3967-3977.

[36] Kato, J., Nishimura, Y., Imamura, R., Niki, H., Hiraga, S., & Suzuki, H. (1990). New topoisomerase essential for chromosome segregation in E. coli. *Cell*, 63, 393-404.

[37] Klungsøyr, H. K., & Skarstad, K. (2004). Positive supercoiling is generated in the presence of Escherichia coli SeqA protein. *Mol. Microbiol.*, 54, 123-131.

[38] Kornberg, A., & Baker, T. A. (1992). DNA Replication, nd Ed. 521-524, W. H. Freeman and Co., New York.

[39] Landoulsi, A., Malki, A., Kern, R., Kohiyama, M., & Hughes, P. (1990). The E. coli cell surface specifically prevents the initiation of DNA replication at oriC on hemimethylated DNA templates. *Cell*, 63, 1053-1060.

[40] Løbner-Olesen, A., Marinus, M. G., & Hanssen, F. G. (2003). Role of SeqA and Dam in Escherichia coli gene expression: a global/microarray analysis. Proc. Natl. Acad. Sci. USA. , 100, 4672-4677.

[41] Lu, M., Campbell, J. L., Boye, E., & Kleckner, N. (1994). SeqA: a negative modulator of replication initiation in E.coli. *Cell*, 413-426.

[42] Marinus, M. G., Poteete, A., & Arraj, J. A. (1984). Correlation of DNA adenine methylase activity with spontaneous mutability in Escherichia coli K-12. *Gene*, 123 EOF-5 EOF.

[43] Messer, W., & Weigel, C. (1996). Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology. Neidhardt, F. C., ed) 2nd Ed., 1580-1582, American Society for Microbiology, Washington, D. C.

[44] Morigen, , Odsbu, I., & Skarstad, K. (2009). Growth rate dependent numbers of SeqA structures organize the multiple replication forks in rapidly growing Escherichia coli. *Genes Cells*, 14, 643-57.

[45] Nielsen, H. J., Li, Y., Youngren, B., Hansen, F. G., & Austin, S. (2006). Progressive segregation of the Escherichia coli chromosome. Mol. Microbiol. , 61, 383-393.

[46] Nielsen, O., & Løbner-Olesen, A. (2008). Once in a lifetime: strategies for preventing re-replication in prokaryotic and eukaryotic cells. *EMBO Rep*, 9, 151-156.

[47] Odsbu, I., Klungsoyr, H. K., Fossum, S., & Skarstad, K. (2005). Specific N-terminal interactions of the Escherichia coli SeqA protein are required to form multimers that restrain negative supercoils and form foci. *Genes Cells*, 10, 1039-1049.

[48] Ogden, G. B., Pratt, M. J., & Schaechter, M. (1988). The replicative origin of the E. coli chromosome binds to cell membranes only when hemimethylated. *Cell*, 54, 127-35.

[49] Palmer, B. R., & Marinus, M. G. (1994). The dam and dcm strains of Escherichia coli- a review. *Gene*, 143, 1-12.

[50] Peterson, S. N., & Reich, N. O. (2006). GATC flanking sequences regulate Dam activity: evidence for how Dam specificity may influence pap expression. *J Mol Biol*, 355, 459-472.

[51] Prieto, A. I., Jakomin, M., Segura, I., Pucciarelli, M. G., Ramos-Morales, F., del Portillo, F. G., & Casadesus, J. (2007). The GATC-binding protein SeqA is required for bile resistance and virulence in Salmonella enterica Serovar Typhimurium. J. Bacteriol. , 189, 8496-8502.

[52] Reyes-Lamothe, R., Possoz, C., Danilova, O., & Sherratt, D. J. (2008). Independent positioning and action of Escherichia coli replisomes in live cells. *Cell*, 133, 90-102.

[53] Ringquist, S., & Smith, C. L. (1992). The Escherichia coli chromosome contains specific, unmethylated dam and dcm sites. P Natl Acad Sci USA, , 89, 4539-4543.

[54] Saint-Dic, D., Kehrl, J., Frushour, B., & Kahng, L. S. (2008). Excess SeqA leads to replication arrest and a cell division defect in Vibrio cholerae. *J Bacteriol*, 190, 5870-5878.

[55] Skarstad, K., Boye, E., & Steen, H. B. (1986). Timing of initiation of chromosome replication in individual Escherichia coli cells. EMBO J. , 5, 1711-1717.

[56] Skarstad, K., von, Meyenburg. K., Hansen, F. G., & Boye, E. (1988). Coordination of chromosome replication initiation in Escherichia coli: effects of different dnaA alleles. J. Bacteriol. , 170, 852-858.

[57] Skarstad, K., & Lobner-Olesen, A. (2003). Stable co-existence of separate replicons in Escherichia coli is dependent on once-per-cell-cycle initiation. EMBO J. , 22, 140-50.

[58] Slater, S., Wold, S., Lu, M., Boye, E., Skarstad, K., & Kleckner, N. (1995). E. coli SeqA protein binds oriC in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. *Cell*, 82, 927-936.

[59] Slominska, M., Wegrzyn, A., Konopa, G., Skarstad, K., & Wegrzyn, G. (2001). SeqA, the Escherichia coli origin sequestration protein, is also a specific transcription factor. Mol. Microbiol. , 40, 1371-1379.

[60] Urig, S., Gowher, H., Hermann, A., Beck, C., Fatemi, M., Humeny, A., & Jeltsch, A. (2002). The Escherichia coli dam DNA methyltransferase modifies DNA in a highly processive reaction. *J Mol Biol*, 319, 1085-1096.

[61] Waldminghaus, T., & Skarstad, K. (2009). The Escherichia coli SeqA protein. *Plasmid*, 61, 141-150.

[62] Wegrzyn, A., Wrobel, B., & Wegrzyn, G. (1999). Altered biological properties of cell membranes in Escherichia coli dnaA and seqA mutants. *Mol. Gen. Genet.*, 261, 762-769.

[63] Weitao, T., Nordstrom, K., & Dasgupta, S. (1999). Mutual suppression of mukB and seqA phenotypes might arise from their opposing influences on the Escherichia coli nucleoid structure. Mol. Microbiol. , 34, 157-168.

[64] Weitao, T., Nordstrom, K., & Dasgupta, S. (2000). Escherichia coli cell cycle control genes affect chromosome superhelicity. EMBO Rep. , 1, 494-499.

[65] Wion, D., & Casadesus, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nat. Rev. Microbiol. , 4, 183-192.

[66] Yamaki, H., Ohtsubo, E., Nagai, K., & Maeda, Y. (1988). The oriC unwinding by dam methylation in Escherichia coli. *Nucleic Acids Res*, 16, 5067-5073.

[67] Yamazoe, M. S., Adachi, S., Kanava, K. O., & Hiraga, S. (2005). Sequential binding of SeqA protein to nascent DNA segments at replication forks in synchronized cultures of E. coli. *Mol. Microbiol.*, 55, 289-298.

[68] Zechiedrich, E. L., & Cozzarelli, N. R. (1995). Roles of topoisomerase IV and DNA gyrase in DNA unlinking during replication in Escherichia coli. *Genes Dev*, 9, 2859-69.

# The Mechanisms of Epigenetic Modifications During DNA Replication

Takeo Kubota, Kunio Miyake and Takae Hirasawa

Additional information is available at the end of the chapter

## 1. Introduction

At the DNA replication step during cell division, not only fundamental information (i.e. nucleotide sequence) but also superficial information (i.e. "epigenetic" modifications) is faithfully reproduced on the newly synthesized DNA sequence. The faithful maintenance of the epigenetic pattern, which determines the gene-expression pattern of the cell, safeguards the maintenance of cell identity.

The term "epigenetics" was first used to describe "the causal interactions between genes and their products, which bring the phenotype into being" [1], and this definition initially referred to the role of the epigenetics in embryonic development, in which cells develop distinct identities despite having the same genetic information. However, today epigenetics refers to "the study of heritable changes in gene expression that occur independent of changes in the primary DNA sequence" [2]. This definition is now associated with in a wide variety of biological processes, such as genomic imprinting [3,4], inactivation of the X chromosome [5], embryogenesis [6], tissue differentiation [7], and carcinogenesis [8].

Epigenetic chemical modifications, such as DNA methylation and histone modifications, are known to be faithfully duplicated in each cell cycle and subsequently the chromatin structures are propagated through DNA replication [9]; however, little is known about how the chromatin structure is maintained during or reformed after DNA replication. Furthermore, several lines of recent evidence suggested that the superficial information on the DNA strand is more susceptible to change by environmental stress than the DNA strand itself. Therefore, for a better understanding of the DNA replication process, it is highly important and desirable, for biologists in general and molecular biology in particular, to learn about the epigenetic mechanisms.

In this chapter, we introduce the current understanding of the DNA methylation mechanism, 5-hydroxycytosine (the sixth base), histone modifications, and their significance in congenital and acquired diseases, and also discuss to which direction this field ought to proceed in the future.

## 2. DNA methylation during DNA replication

Not all genes are necessarily expressed in every cell of the organism. Most of these genes and genetic regions are programmed to remain repressed, which defines the identity of each cell. Epigenetic modifications are molecular mechanisms that can preserve the inactive state by regenerating a repressive chromatin structure on the "unnecessary genes and genomic regions" following each round of DNA replication in the cell. DNA methylation is one of the fundamental mechanisms known to be involved in this maintenance process [10].

Maintenance of such methylation pattern in DNA during replication is mediated by DNA nucleotide methyltransferase 1 (DNMT1) [11], which methylates newly synthesized CpG sequences, depending on the methylation status of the template strand (Fig. 1). A bridging protein, known as UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1), that interacts with DNMT1 and hemimethylated CpG is required to maintain the hemimethylated CpG dinucleotides pattern at the DNA fork [12,13].
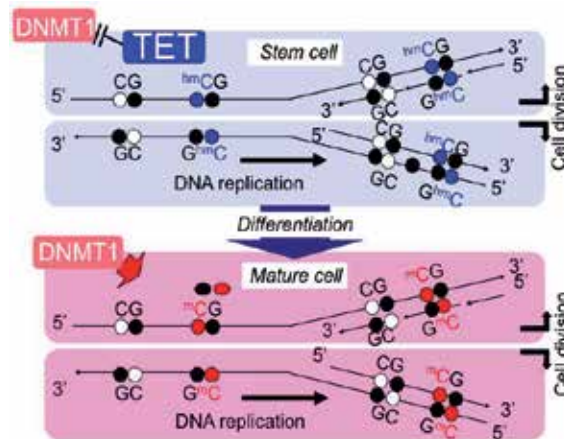


**Figure 1.** Maintenance of DNA methylation pattern with 5-hydroxymethylcytosine (upper) and 5-methylcytosine (lower) during DNA replication and cell division. 5-hydroxymethylcytosine is relatively abundantly found in embryonic stem (ES) cells and its level decreases during development due to the declining levels of *TET* expression. Cytosine, 5-hydroxymethylcytosine and 5-methylcytosine are shown in white, blue and red circles, respectively [95].

The chromatin structure, modified by DNA methylation, is not stable, but it undergoes a wave of disruption and reassembly during DNA replication. These changes in the chromatin structures influence the dynamics of DNA replication by regulating the selection of replication origin sites and their initiation timings. Interestingly, active gene promoters are often

found at these active replication origin sites. Thus, the coordination of replication and tran‐
scription is an important mechanism for the establishment and inheritance of differential
gene expression patterns during cellular differentiation [2].

DNA methylation status is also involved in determining the chromosomal replication tim‐
ing. Hypomethylation is associated with late-replication and late-replicating genomic re‐
gions are gradually demethylated with cell divisions, whereas DNA methylation of early-
replicating regions is maintained during DNA replication [14]. Moreover, DNA replication in
early S phase gets automatically repackaged with acetylated histones, whereas the regions
that replicate late in S phase assemble nucleosomes containing deacetylated histones [15].

So far several DNA nucleotide methyltransferases (DNMTs), which includes DNMT1,
DNMT2, DNMT3A, DNMT3B, DNMT3L, have been found in mammals, all of which con‐
tain a methyltransferase catalytic domain. Of these, DNMT1 is the most abundant DNMT in
differentiated cells; it has a preference for hemi-methylated DNA, and acts as a 'mainte‐
nance methylase', which allows it to efficiently methylate the hemi-methylated sites that are
generated during DNA replication. Thus, the CpG methylation pattern is maintained in the
genome after DNA replication [16]. Until recently, the biochemical and functional properties
of DNMT2 remained unknown. However, the DNMT2 is now known to act as an RNA
methylatransferase and the DNMT2-mediated methylation protects tRNAs against ribonu‐
clease cleavage in drosophila [17].

DNMT3A and DNMT3B are expressed at high levels in mouse embryonic stem (ES) cells
and at lower levels in differentiated cells. They act as '*de novo* methylases', which catalyze
the transfer of methyl groups to naked DNA, and are responsible for establishing the pat‐
tern of methylation during embryonic development [18]. Recent evidence suggests that be‐
sides playing the role as '*de novo* methylases' DNMT3A and DNMT3B may also act as
'methylation completer' and 'methylation error corrector' - by completing the methylation
process and correcting errors, respectively, left by DNMT1 - at least at highly methylated
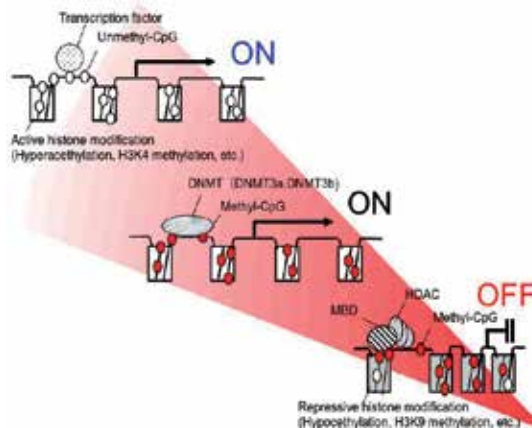DNA regions, such as imprinted regions and repetitive elements [19].



**Figure 2.** Epigenetic gene regulation based on DNA methylation, histone acetylation and histone methylation, in‐
duced by proteins including nucleotide methyltransferases (DNMTs), methyl-CpG binding domain (MBD) proteins, and
histone deacetylases (HDACs).

Once a certain site is methylated, it could then act as a candidate region where the silent chromatin is established. For this purpose, the methylated site first recruits methyl-CpG binding domain (MBD) proteins; the MBD proteins subsequently recruit histone deacetylases, histone modification proteins. In other words, MBDs, which form bridges between the methylation site and other associated proteins, are the key proteins in epigenetic regulation (Fig. 2).

So far five MBD proteins, each containing a methyl-CpG binding domain, have been reported. Among these MBD proteins, MBD1 is unique because it is capable of repressing transcription from both methylated and unmethylated promoters [20].

MBD1 associates with chromatin modifiers such as the Suv39h1-HP1 complex, and enhances DNA methylation-mediated transcriptional repression [21]. MBD1 also associates with the H3K9 methyltransferase SETDB1 [22]. During S phase, the chromatin assembly factor CAF1 recruits the MBD1-SETDB1 complex to chromatin to establish new H3K9 methylation. On the other hand, the removal of DNA methylation disrupts the formation of MBD1-SETDB1-CAF1 complex, which results in the loss of H3K9 methylation at the formerly methylated site [23].

MBD2 protein shares extensive sequence homology with MBD3. MBD2 binds to methylated CpGs and confer DNA methylation-mediated transcriptional silencing through its association with HDAC1 and HDAC2 in the NuRD chromatin remodeling complex [23]. Although *Mbd2*-null mice develops normally and remains viable and fertile [24], lack of Mbd2 affects immunological systems by inducing ectopic *IL-4* expression in undifferentiated helper T cells [25]. Lack of Mbd2 also influences X-chromosome inactivation by inducing ectopic *Xist* expression in the active X chromosome [26].

MBD3, like MBD2, is an essential subunit of the NuRD complex. It has been suggested that MBD2 and MBD3 associate with the NuRD in a mutually exclusive way, thereby forming two distinct complexes [23]. Although there is a great sequence similarity between MBD2 and MBD3, the two proteins do not perform redundant functions during early development. In contrast to *Mbd2*-mull mice which displayed a mild phenotype, *MBD3*-null embryos die on day 8.5, by failing to shut down the expression of undifferentiated cell markers such as *Oct4* and *Nanog* [27].

MBD4 is a thymine glycosylase, which acts as a DNA repair protein and targets the sites of cytosine deamination. Spontaneous hydrolytic deamination of 5mC leads to 5mCpG-TpG transitions, whereas that of non-methylated CpG leads it to UpG, and MBD4 is able to excise and repair both 'mutated' nucleotides [28]. Consistent with this observation, *Mbd4*-null mice exhibit a two to three times higher number of 5mCpG-TpG transitions, indicating that Mbd4 indeed acts to reduce the 5mCpG-TpG mutation rate [29]. More importantly, when crossed with mice carrying a germline mutation in the *Apc* (adenomatous polyposis coli) gene, *Mbd4*-null mice show accelerated tumor formation [29]. In fact, mutations in *MBD4* have been reported in various human carcinomas [30].

MeCP2 is the first MBD to be cloned [31]. As of now, MeCP2 is known to be a multifunctional nuclear protein, which is known to be involved transcriptional repression, activation of transcription, nuclear organization, and splicing [32,33]. Besides acting as a transcriptional

repressor like other MBDs, MeCP2 also acts as a splice regulator, by interacting with YB-1, a component of messenger ribonucleoprotein particles, in brain nuclear extracts [34]. Indeed, microarray splicing analysis of cerebral cortex mRNA isolated from *Mecp2*-mutant mice showed a number of aberrantly spliced genes [23]. Furthermore, MeCP2 deficiency activates L1 retrotransposition in neurons, which is possibly associated with the genomic diversity of brains [35]. Therefore, it is interesting that there exist several links between MBD-mediated repression, RNA processing and DNA-sequence diversity. It is also intriguing to find a link between epigenetic modification and its suppressive power on genetic diversity since, in addition to MeCP2, a histone modification enzyme (H3K9 methyltransferase, *ESET*) also contributes to silence retrovirus-like elements in the mammalian genome [36].

## 3. 5-hydroxymethylcytosine - the sixth base in mammalian DNA

The 5-methylcytosine (5mC) has been recognized as "the fifth base". However, early work suggested the existence of a sixth base, 5-hydroxymethylcytosine (5hmC) (Fig. 1). 5hmC was first reported in T-even bacteriophages [37] and later in mammalian cells [38]. However, the reported finding, which claimed that this modified base accounted for ~15% of total cytosines in DNA extracted from the brains of adult rats, mice and frogs, could not be reproduced [39]. The topic received only little attention for the next 30 years until 2009, when work from two research teams brought it back to life [40,41]. Actually, it was found that 5hmC accounts for 0.6%, 0.2%, 0.03% of total nucleotides in Purkinje cells, granule cells, and mouse ES cells, respectively [40,41].

The presence of 5hmC in the mammalian genome depends on pre-existing 5mC, because 5hmC is converted from 5mC with the help of TET proteins, which utilize molecular oxygen to incorporate a hydroxyl group to 5mC. TET is named after Ten-Eleven Translocation (translocation between chromosomes 10 and 11) because it is initially found as a fusion protein partner of mixed-lineage leukemia gene (*MLL*) in acute myeloid leukemia (AML) patients carrying a t(10;11)(q22;q23) translocation [42,43]. The findings that ectopic expression of TET1 in HEK293 cells lacking TET1 led to reduced levels of 5mC and increased levels of 5hmC, and that the levels of 5hmC decreased upon RNAi-mediated depletion of TET1 in ES cells indicate that TET1 is able to catalyze the conversion of 5mC and 5hmC in cultured cells [41]. Also, it has been demonstrated that TET1 is capable of acting not only on fully-methylated DNA strands but also on hemi-methylated DNA strands [41]. Furthermore, not only TET1 but also other TET proteins (TET2 and TET3) is capable of converting 5mC to 5hmC [44].

In terms of gene regulation, the significance of this 5hmC modification is similar to that of non-methylated cytosine. In other words, the 5hmC modification is associated with transcriptional activity, which is different from the 5mC modification that is associated with transcriptional repression [16]. It has recently demonstrated that TET1-binding to the promoter region (presumably 5hmC modification at this site) induces the expression of *Nanog* in ES cells and that downregulation of *Nanog* via *TET1* knockdown induces DNA methylation in the promoter region [44]. These findings indicate that the TET1 driven 5hmC modifi-

cation contribute to maintenance of the nature of un-differentiation and pluripotency of ES cells, and support a working model by which TET1 and DNMTs coordinately regulate Nanog expression.

In ES cells, high levels of TET1 block the access of DNMTs for maintained Nanog expression. On the other hand, when TET1 is downregulated in ES cells by *in vitro* differentiation, DNMTs methylate the *Nanog* promoter, which leads to the downregulation of Nanog expression and loss of ES cell identity (Fig. 1) [44]. This hypothesis is supported by a recent finding in which the chromosomes containing 5hmC are gradually reduced during the development of preimplantation embryos [45]. However, another study showed that the 5hmC level in the mouse cerebellum during development increases from 0.1% of total nucleotides at postnatal day 7 to 0.4% of total nucleotides in the adult mouse [46].

As described above, TET1 was initially identified through a rare translocation case with leukemia [42,43]. Later studies have demonstrated that deletion and mutations in *TET1*, *TET2* and *TET3* are associated with myeloid malignancies [47]. In fact, mutations found in *TET2* in myeloid cancers have been shown to impair hydroxylation of 5mC [48].

While our knowledge about 5hmC is rapidly growing, currently there is no reliable methodology available that would provide information on 5hmC at single-base-pair resolution. Although a 5hmC antibody is available for chromatin Immunoprecipitation, this method only provides some coarse information (i.e. detects presence of 5hmC but not that of 5mC in chromatin). A more sensitive method has been developed for 5hmC by capillary eletrophoresis, but this is not the one at sinlge-base-pair resolution [49]. Another method (namely, bisulfite sequencing) has proven to be a powerful tool for providing information on the methylation status at single-base-pair resolution. However, it too fails to discriminate between 5mC and 5hmC. Thus, if the bisulfite-treated DNA is used as a template for PCR analysis, cytosine will be read as thymine, whereas both 5mC and 5hmC will be read as cytosine [16]. Therefore, it is important to develop a methodology that can distinguish between 5mC and 5hmC at single-base-pair resolution in order to achieve complete understanding of the active demethylation mechanism, because TET protein-mediated 5mC oxidation may contribute to dynamic changes in global or locus-specific 5mC levels by promoting active DNA demethylation [50].

## 4. Histone modifications and DNA methylation during replication

DNA methylation and histone modifications not only occur separately, but they also work hand-in-hand at multiple levels to determine expression status, chromatin organization and cellular identity, and they are co-ordinately maintained through mitotic cell division, allowing for the transmission of parental DNA and for the histone modifications to be copied to newly replicated chromatin [51,52].

Lande-Diner et al. recently developed a *DNMT1-* knockout cell line and demonstrated that an unmethylated state, caused by the lack of DNMT1, induced deacetylation of histones H3

and H4, resulting in transcriptional activation in many genes [10]. This observation clearly indicates that DNA methylation is associated with histone deacetylation. However, this group also demonstrated that in several other genes the unmethylated state, caused by lack of DNMT, did not induce histone H3 and H4 deacetylation, resulting in transcriptional repression. In addition, late replication in S phase was observed at these loci, suggesting that the replication timing may be independent of DNA methylation [10]. Rather, histone acetylation is associated in controlling the replication timing [53].

DNA methylation is not only correlated to histone 'acetylation', but also associated with histone 'lysine methylation'. Genome-wide DNA methylation profiles suggest that DNA methylation is associated with the absence H3K4 methylation and the presence of H3K9 and H3K27 methylation [54].

In fact, DNA methylation induces histone H3K9 methylation through an MBD, thereby establishing a repressive chromatin state [55]. SETDB1, a H3K9 trimethylation (H3K9me3) methyltransferase, contains a putative MBD domain with two conserved DNA-interacting arginine residues, which are also present in the MBD domains of MBD1 and MeCP2 and are known to make direct contact with the DNA in the structures of MBD1-DNA and MeCP2-DNA complexes [56,57]. This result suggests that SETDB1 acts as an H3K9me3 'writer' in corporation with DNA methylation 'reader'. Likewise, SUV39H1/2, another H3K9me3 'writer', interacts with HP1, the H3K9me3 'reader' to create a repressed status in their recruited genomic region [58]. These are the mechanisms for propagating and maintaining repressive chromatin marks on both DNA and histones during DNA replication.

A histone methyltransferase, in turn, can direct DNA methylation to specific genomic targets by recruiting DNMTs to stably silence genes [59]; accordingly, disruption of the histone lysine methyltransferase gene with specificity for H3K4 (*MLL*) in mice not only induces the loss of H3K4 methylation but also induces de novo DNA methylation at several gene promoters [60,61]. In another study, it was shown that the lack of histone H3K9 methyltransferase induced demethylation at the imprinting center in *SNPRN* locus on the maternal chromosome, whereas the lack of DNMT1 failed to induce demethylation of histone H3K9, indicating that the modification order at this locus is histone modification followed by DNA methylation [62]. Taken together, histone methylation marks play important roles in predicting the methylation status of the genome [63].

Whereas DNMT1 is stabilized by a histone demethylase (HDM) to maintain DNA methylation [64], DNMTs can direct the local status of histone methylation patterns, recruiting MBDs and HDACs to achieve gene silencing and chromatin condensation [65,66]. Recently, DNMT3L has been shown to act as a sensor for H3K4 methylation. Thus, when methylation is absent, DNMT3L induces *de novo* DNA methylation by docking DNMT3A to the nucleosome, which is one of mechanisms by which methylated regions are newly created during the replication step [67].

The interplay of these modifications creates an epigenetic landscape that regulates the way the mammalian genome expresses itself in different cell types, developmental stages and disease states. The distinct patterns of these epigenetic modifications present in different cel-

lular states serve as a guardian of cellular identity [2]. Whereas it is well accepted that DNA methylation patterns are replicated in a semi-conservative fashion during cell division via the mechanisms discussed earlier, how histone modification patterns are similarly replicated remains to be elucidated.

## 5. Abnormalities in epigenetic mechanism and their possible inheritance

Thanks to identification of molecules that contribute to epigenetic gene regulation, we now know how that abnormalities in these molecules cause a number of congenital diseases.

The first group of diseases with abnormal epigenetic mechanism is genomic imprinting diseases [3]. Genomic imprinting is a mechanism in which only one of the two parental alleles is expressed in a gene. For example, in the case of *SNRPN* gene, the paternal allele of the *SNRPN* gene is expressed, whereas the maternal allele is suppressed by DNA methylation in normal individuals, and abnormal suppression of the normally expressing paternal allele causes a congenital obesity disease, known as Prader-Willi syndrome [4]. In the case of *UBE3A* gene, which locates adjacent to the *SNRPN* gene, the maternal allele is expressed, whereas the paternal allele is suppressed in neurons [68]; abnormal suppression of the expressing maternal allele causes a congenital epileptic disease, known as Angelman syndrome [69].

X-chromosome inactivation is another epigenetic mechanism in which only one of the two X chromosomes is activated and the other X chromosome is inactivated in females [5]. Females with aberrant X-inactivation (i.e. both two X chromosomes are activated) are thought to be embryonic lethal, since somatic clones with aberrant X-inactivation are aborted [70].

Abnormal functioning of the proteins related to epigenetic regulation also causes diseases. For example, mutations in the *DNMT3B* gene, which lead to hypomethylation at the paracentromeric chromosomal regions, cause the immunodeficiency- centromeric instability- facial anomalies (ICF) syndrome, which is characterized by immunodeficiency, centromere instability, facial abnormalities, and mild mental retardation (Fig. 3A) [71-73]. On the other hand, over-expression of *DNMTs* is associated with hypermethylation found in colorectal, breast, and hepatocellular carcinomas (Fig. 3C) [74-76]. Another example is Rett syndrome caused by *MECP2* mutations, which is characterized by seizures, ataxic gait, language dysfunction and autistic behavior [77,78]. In this disease, *MECP2* mutations induce abnormal regulation of a subset of neuronal genes [79,80] (Fig. 3B).

Besides these "DNA methylation diseases" caused by mutations in DNA methylation-related enzymes and proteins, "histone modification diseases" caused by mutations in histone modification-related enzymes have recently been reported. For example, Say-Barber-Bieseker-Young-Simpson syndrome is caused by mutations in the histone acetyltransferase gene, *KAT6B*, which is a multiple anomaly syndrome characterized by, an immobile mask-like face, abnormal narrowing of palpebral fissures (short eyelid), anomalies of the spine, ribs and pelvis, renal cysts, hydronephrosis, agenesis of the corpus callosum, and severe intellec-

tual disability [81]. Another example is Kleefstra syndrome caused by deletion or mutation in the histone H3K9 methyltransferase gene, *EHMT1*, which is characterized by childhood hypotonia, distinctive facial features, and intellectual disability with severe expressive speech delay [82].

Recently, it has been shown that short-term environmental stress could also cause aberrant epigenetic status associated with various diseases. Thus, aberrant epigenetic mechanism can not only cause congenital diseases, but can also cause acquired diseases. For example, short-term mental stress after birth, in which the mother is separated from the offspring, causes DNA hypermethylation in the promoter of the glucocorticoid receptor (*GR*) gene in the rat brain, resulting in persistent abnormal behavior [83]. Malnutrition in the fatal period is also known to induce DNA hypomethylation in the promoter of the peroxisome proliferator-activated receptor alpha (*PPARα*) gene, a so-called "thrifty gene", in the liver, which may be associated with the developmental basis of adult diseases (i.e. obesity and diabetes mellitus) [84,85] (Fig. 3D). This hypomethylation event has later been confirmed in human individuals who suffered prenatal malnutrition during the period of famine [86,87].
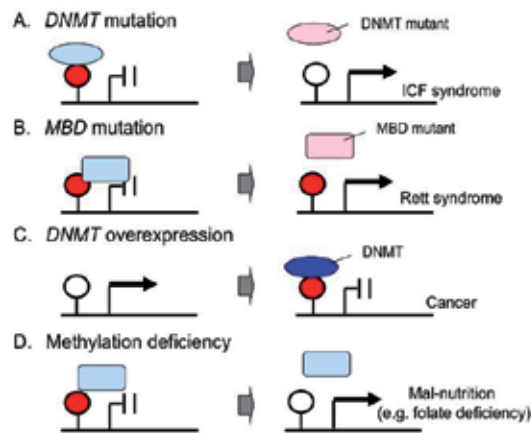
A. *DNMT* mutation — DNMT mutant — ICF syndrome

B. *MBD* mutation — MBD mutant — Rett syndrome

C. *DNMT* overexpression — DNMT — Cancer

D. Methylation deficiency — Mal-nutrition (e.g. folate deficiency)

**Figure 3.** Abnormal epigenetic patterns found in human congenital and acquired diseases.

Several lines of evidence suggested that acquired DNA methylation changes described above are transmitted to the next generation. Epigenetic marks allow the transmission of gene activity states from one cell to its daughter cells. Initially, it was assumed that epigenetic marks were completely erased and re-established in each generation. However, recent studies using several model organisms indicate that the erasing process is incomplete at some loci and so the epigenetic changes acquired in one generation are inherited by the next generation.

For example, it has been shown in mice that the mental stress caused due to maternal separation in offspring not only changes the DNA methylation status in the first generation but also in the next generation through changes in the sperms of the first generation in mice [88]. Moreover, the observed changes in the DNA methylation status altered the expression level

of corticotrophin releasing actor receptor 2 (*Crfr2*) in the brains of next generation mice, which could be associated with their abnormal behavior [88].

This phenomenon is termed "transgenerational epigenetic inheritance", which is expected to provide a biological proof for the apparent heritability of acquired characteristics [89-91].

## 6. Concluding remarks

One of the major differences between DNA sequence and epigenetic modifications is tissue specificity. Epigenetic modifications vary according to the tissue type, which consequently allows generating tissue-specific expression patterns. However, how determines the epigenetic modification (epigenomic) pattern in each tissue type is not fully understood.

Thus, it is essential to categorize epigenomic patterns in each human tissue at the nucleotide resolution [92,93]. In fact, the NIH Roadmap Epigenomics Program under the US National Center for Biotechnology Information (NCBI) and the International Human Epigenome Consortium (IHEC) have initiated the large-scale epigenomic mapping studies in order to generate epigenome maps for each human cell type for this purpose [94].

Understanding the human epigenome will be fundamental to the study of congenital and acquired diseases, and will also be invaluable for analyzing the linkage between birth defects and environmental factors. However, biological studies to understand the epigenome are in their initial phase. Further studies are necessary to elucidate the molecular mechanism by which the epigenome pattern in each cell type differs, epigenomic patterns are altered by environmental factors, and process of inheriting the epigenomic pattern from the previous generation could be avoided. The authors expect that these molecular mechanisms would hopefully be discovered by the "next generation" of researchers.

## Acknowledgements

## Author details

Takeo Kubota[*], Kunio Miyake and Takae Hirasawa

*Address all correspondence to: takeot@yamanashi.ac.jp

Department of Epigenetic Medicine, Faculty of Medicine, University of Yamanashi, Japan

# References

[1] Waddington, C. H. (1942). Epigenotype. *Endeavour*, 1, 18-20.

[2] Sharma, S., Kelly, T. K., & Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis*, 31(1), 27-36.

[3] Kubota, T., Saitoh, S., Matsumoto, T., Narahara, K., Fukushima, Y., Jinno, Y., & Niikawa, N. (1994). Excess functional copy of allele at chromosomal region 11-15 may cause Wiedemann-Beckwith (EMG) syndrome. *American Journal of Medical Genetics*, 49(4), 378-83.

[4] Kubota, T., Das, S., Christian, S. L., Baylin, S. B., Herman, J. G., & Ledbetter, D. H. (1997). Methylation-specific PCR simplifies imprinting analysis. *Nature Genetics*, 16(1), 16-7.

[5] Kubota, T., Nonoyama, S., Tonoki, H., Masuno, M., Imaizumi, K., Kojima, M., Wakui, K., Shimadzu, M., & Fukushima, Y. (1999). A new assay for the analysis of X-chromosome inactivation based on methylation-specific PCR. *Human Genetics*, 104(1), 49-55.

[6] Sasaki, H., & Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nature Reviews Genetics*, 9(2), 129-40.

[7] Sakashita, K., Koike, K., Kinoshita, T., Shiohara, M., Kamijo, T., Taniguchi, S., & Kubota, T. (2001). Dynamic DNA methylation change in the CpG island region of p15 during human myeloid development. *J Clin Invest. Journal of Clinical Investigation*, 108(8), 1195-204.

[8] Ushijima, T. (2005). Detection and interpretation of altered methylation patterns in cancer cells. *Nature Reviews Cancer*, 5(3), 223-31.

[9] Liu, Q., & Gong, Z. (2011). The coupling of epigenome replication with DNA replication. *Current Opinion in Plant Biology*, 14(2), 187-94.

[10] Lande-Diner, L., Zhang, J., Ben-Porath, I., Amariglio, N., Keshet, I., Hecht, M., Azuara, V., Fisher, A. G., Rechavi, G., & Cedar, H. (2007). Role of DNA methylation in stable gene repression. *Journal of Biological Chemistry*, 282(16), 12194-200.

[11] Cheng, X., & Blumenthal, R. M. (2008). Mammalian DNA methyltransferases: a structural perspective. *Structure*, 16(3), 341-50.

[12] Bostick, M., Kim, J. K., Estève, P. O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science*, 317(5845), 1760-4.

[13] Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M., & Koseki, H. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, 450(7171), 908-12.

[14]  Goren, A., & Cedar, H. (2003). Replicating by the clock. Nat Rev Mol Cell Biol. *Nature Reviews Molecular Cell Biology*, 4(1), 25-32.

[15]  Zhang, J., Xu, F., Hashimshony, T., Keshet, I., & Cedar, H. (2002). Establishment of transcriptional competence in early and late S phase. *Nature*, 420(6912), 198-202.

[16]  Dahl, C., Grønbæk, K., & Guldberg, P. (2011). Advances in DNA methylation: 5-hydroxymethylcytosine revisited. *Clinica Chimica Acta*, 412(11-12), 831-6.

[17]  Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., & Lyko, F. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. Genes Dev. *Genes and Development*, 24(15), 1590-5.

[18]  Bestor, TH. (2000). The DNA methyltransferases of mammals. *Human Molecular Genetics*, 9(16), 2395-402.

[19]  Jones, P. A., & Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. *Nature Reviews Genetics*, 10(11), 805-11.

[20]  Fujita, N., Shimotake, N., Ohki, I., Chiba, T., Saya, H., Shirakawa, M., & Nakao, M. (2000). Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1. *Molecular and Cellular Biology*, 20(14), 5107-18.

[21]  Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T., & Nakao, M. (2003). Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. *Journal of Biological Chemistry*, 278(26), 24132-8.

[22]  Sarraf, S. A., & Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Molecular Cell*, 15(4), 595-605.

[23]  Bogdanović, O., & Veenstra, G. J. (2009). DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma Oct*, 118(5), 549-65, Epub Jun 9.

[24]  Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A., & Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes and Development*, 15(6), 710-23.

[25]  Hutchins, A. S., Mullen, A. C., Lee, H. W., Sykes, K. J., High, F. A., Hendrich, B. D., Bird, A. P., & Reiner, S. L. (2002). Gene silencing quantitatively controls the function of a developmental trans-activator. *Molecular Cell*, 10(1), 81-91.

[26]  Barr, H., Hermann, A., Berger, J., Tsai, H. H., Adie, K., Prokhortchouk, A., Hendrich, B., & Bird, A. (2007). Mbd2 contributes to DNA methylation-directed repression of the Xist gene. *Molecular and Cellular Biology*, 27(10), 3750-7.

[27]  Kaji, K., Nichols, J., & Hendrich, B. (2007). Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. *Development*, 134(6), 1123-32.

[28] Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J., & Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature*, 401(6750), 301-4.

[29] Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P. D., Bishop, S. M., Clarke, A. R., & Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science*, 297(5580), 403-5.

[30] Riccio, A., Aaltonen, L. A., Godwin, A. K., Loukola, A., Percesepe, A., Salovaara, R., Masciullo, V., Genuardi, M., Paravatou-Petsotas, M., Bassi, D. E., Ruggeri- , B. A., Klein-Szanto, A. J., Testa, J. R., Neri, G., & Bellacosa, A. (1999). The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. *Nature Genetics*, 23(3), 266-8.

[31] Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., & Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*, 69(6), 905-14.

[32] Adams, V. H., McBryant, S. J., Wade, P. A., Woodcock, C. L., & Hansen, J. C. (2007). Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *Journal of Biological Chemistry*, 282(20), 15057-64.

[33] Chahrour, M., Jung, S. Y., Shaw, C., Zhou, X., Wong, S. T., Qin, J., & Zoghbi, H. Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science*, 320(5880), 1224-9.

[34] Young, J. I., Hong, E. P., Castle, J. C., Crespo-Barreto, J., Bowman, A. B., Rose, M. F., Kang, D., Richman, R., Johnson, J. M., Berget, S., & Zoghbi, H. Y. (2005). Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proceedings of National Academy of Science USA*, 102(49), 17551-8.

[35] Muotri, A. R., Marchetto, M. C., Coufal, N. G., Oefner, R., Yeo, G., Nakashima, K., & Gage, F. H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature*, 468(7322), 443-6.

[36] Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M. C., & Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*, 464(7290), 927-31.

[37] Wyatt, G. R., & Cohen, S. S. (1952). A new pyrimidine base from bacteriophage nucleic acids. *Nature*, 170(4338), 1072-3.

[38] Penn, N. W., Suwalski, R., O'Riley, C., Bojanowski, K., & Yura, R. (1972). The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochemical Journal*, 126(4), 781-90.

[39] Kothari, R. M., & Shankar, V. (1976). Methylcytosine content in the vertebrate deoxyribonucleic acids: species specificity. *Journal of Molecular Evolution*, 7(4), 325-9.

[40]  Kriaucionis, S., & Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, 324(5929), 929-30.

[41]  Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324(5929), 930-5.

[42]  Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., & Hayashi, Y. (2002). LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Research*, 62(14), 4075-80.

[43]  Lorsbach, R. B., Moore, J., Mathew, S., Raimondi, S. C., Mukatira, S. T., & Downing, J. R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia*, 17(3), 637-41.

[44]  Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 466(7310), 1129-33.

[45]  Inoue, A., & Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science*, 334(6053), 194.

[46]  Song, C. X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C. H., Zhang, W., Jian, X., Wang, J., Zhang, L., Looney, T. J., Zhang, B., Godley, L. A., Hicks, L. M., Lahn, B. T., Jin, P., & He, C. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnology*, 29(1), 68-72.

[47]  Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J., Kilpivaara, O., Bhat, R., Huberman, K., Thomas, S., Dolgalev, I., Heguy, A., Paietta, E., Le Beau, M. M., Beran, M., Tallman, M. S., Ebert, B. L., Kantarjian, H. M., Stone, R. M., Gilliland, D. G., Crispino, J. D., & Levine, R. L. (2009). Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood*, 114(1), 144-7.

[48]  Ko, M., Huang, Y., Jankowska, A. M., Pape, U. J., Tahiliani, M., Bandukwala, H. S., An, J., Lamperti, E. D., Koh, K. P., Ganetzky, R., Liu, X. S., Aravind, L., Agarwal, S., Maciejewski, J. P., & Rao, A. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, 468(7325), 839-43.

[49]  Krais, A. M., Park, Y. J., Plass, C., & Schmeiser, H. H. (2011). Determination of genomic 5-hydroxymethyl-2'-deoxycytidine in human DNA by capillary electrophoresis with laser induced fluorescence. *Epigenetics*, 6(5), 560-5.

[50]  Wu, H., & Zhang, Y. (2011). Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes and Development*, 25(23), 2436-52.

[51]  Lande-Diner, L., Zhang, J., & Cedar, H. (2009). Shifts in replication timing actively affect histone acetylation during nucleosome reassembly. *Molecular Cell*, 34(6), 767-74.

[52] Hashimoto, H., Vertino, P. M., & Cheng, X. (2010). Molecular coupling of DNA methylation and histone methylation. *Epigenomics*, 2(5), 657-69.

[53] Goren, A., Tabib, A., Hecht, M., & Cedar, H. (2008). DNA replication timing of the human beta-globin domain is controlled by histone modification at the origin. *Genes and Development*, 22(10), 1319-24.

[54] Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., Low, H. M., Kin, Sung. K. W., Rigoutsos, I., Loring, J., & Wei, C. L. (2010). Dynamic changes in the human methylome during differentiation. Genome Res. *Genome Research*, 20(3), 320-31.

[55] Fuks, F., Hurd, P. J., Deplus, R., & Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Research*, 31(9), 2305-12.

[56] Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., & Shirakawa, M. (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell*, 105(4), 487-97.

[57] Ho, K. L., McNae, I. W., Schmiedeberg, L., Klose, R. J., Bird, A. P., & Walkinshaw, M. D. (2008). MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Molecular Cell*, 29(4), 525-31.

[58] Grewal, S. I., & Jia, S. (2007). Heterochromatin revisited. *Nature Reviews Genetics*, 8(1), 35-46.

[59] Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H., & Shinkai, Y. (2008). G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. *EMBO Journal*, 27(20), 2681-90.

[60] Milne, T. A., Briggs, S. D., Brock, H. W., Martin, M. E., Gibbs, D., Allis, C. D., & Hess, J. L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. *Molecular Cell*, 10(5), 1107-17.

[61] Birke, M., Schreiner, S., García-Cuéllar, M. P., Mahr, K., Titgemeyer, F., & Slany, R. K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Research*, 30(4), 958-65.

[62] Xin, Z., Tachibana, M., Guggiari, M., Heard, E., Shinkai, Y., & Wagstaff, J. (2003). Role of histone methyltransferase G9a in CpG methylation of the Prader-Willi syndrome imprinting center. *Journal of Biological Chemistry*, 278(17), 14996-5000.

[63] Fan, S., Zhang, M. Q., & Zhang, X. (2008). Histone methylation marks play important roles in predicting the methylation status of CpG islands. *Biochemistry Biophysical Research Communications*, 374(3), 559-64.

[64] Wang, J., Hevi, S., Kurash, J. K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G., Gaudet, F., Li, E., & Chen, T. (2009). The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat Genet. Nature Genetics*, 41(1), 125-9.

[65] Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., & Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet. Nature Genetics*, 19(2), 187-91.

[66] Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., & Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393(6683), 386-9.

[67] Ooi, S. K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S. P., Allis, C. D., Cheng, X., & Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, 448(7154), 714-7.

[68] Albrecht, U., Sutcliffe, J. S., Cattanach, B. M., Beechey, C. V., Armstrong, D., Eichele, G., & Beaudet, A. L. (1997). Imprinted expression of the murine Angelman syndrome gene, Ube3a, in hippocampal and Purkinje neurons. *Nature Genetics*, 17(1), 75-8.

[69] Kishino, T., Lalande, M., & Wagstaff, J. (1997). UBE3A/E6-AP mutations cause Angelman syndrome. *Nature Genetics*, 15(1), 70-3.

[70] Xue, F., Tian, X. C., Du, F., Kubota, C., Taneja, M., Dinnyes, A., Dai, Y., Levine, H., Pereira, L. V., & Yang, X. (2002). Aberrant patterns of X chromosome inactivation in bovine clones. *Nature Genetics*, 31(2), 216-20.

[71] Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3), 247-57.

[72] Shirohzu, H., Kubota, T., Kumazawa, A., Sado, T., Chijiwa, T., Inagaki, K., Suetake, I., Tajima, S., Wakui, K., Miki, Y., Hayashi, M., Fukushima, Y., & Sasaki, H. (2002). Three novel DNMT3B mutations in Japanese patients with ICF syndrome. *American Journal of Medical Genetics*, 112(1), 31-7.

[73] Kubota, T., Furuumi, H., Kamoda, T., Iwasaki, N., Tobita, N., Fujiwara, N., Goto, Y., Matsui, A., Sasaki, H., & Kajii, T. (2004). ICF syndrome in a girl with DNA hypomethylation but without detectable DNMT3B mutation. *American Journal of Medical Genetics*, 129A(3), 290-3.

[74] De Marzo, A. M., Marchi, V. L., Yang, E. S., Veeraswamy, R., Lin, X., & Nelson, W. G. (1999). Abnormal regulation of DNA methyltransferase expression during colorectal carcinogenesis. *Cancer Research*, 59(16), 3855-60.

[75] Roll, J. D., Rivenbark, A. G., Jones, W. D., & Coleman, W. B. (2008). DNMT3b overexpression contributes to a hypermethylator phenotype in human breast cancer cell lines. *Molecular Cancer*, 7, 15.

[76] Saito, Y., Kanai, Y., Nakagawa, T., Sakamoto, M., Saito, H., Ishii, H., & Hirohashi, S. (2003). Increased protein expression of DNA methyltransferase (DNMT) 1 is significantly correlated with the malignant potential and poor prognosis of human hepatocellular carcinomas. *International Journal of Cancer*, 105(4), 527-32.

[77] Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, 23(2), 185-8.

[78] Chunshu, Y., Endoh, K., Soutome, M., Kawamura, R., & Kubota, T. (2006). A patient with classic Rett syndrome with a novel mutation in MECP2 exon 1. *Clinical Genetics*, 70(6), 530-1.

[79] Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., & Sun, Y. E. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science*, 302(5646), 890-3.

[80] Miyake, K., Hirasawa, T., Soutome, M., Itoh, M., Goto, Y., Endoh, K., Takahashi, K., Kudo, S., Nakagawa, T., Yokoi, S., Taira, T., Inazawa, J., & Kubota, T. (2011). The protocadherins, PCDHB1 and PCDH7, are regulated by MeCP2 in neuronal cells and brain tissues: implication for pathogenesis of Rett syndrome. *BMC Neuroscience*, 12, 81.

[81] Clayton-Smith, J., O'Sullivan, J., Daly, S., Bhaskar, S., Day, R., Anderson, B., Voss, A. K., Thomas, T., Biesecker, L. G., Smith, P., Fryer, A., Chandler, K. E., Kerr, B., Tassabehji, M., Lynch, S. A., Krajewska-Walasek, M., Mc Kee, S., Smith, J., Sweeney, E., Mansour, S., Mohammed, S., Donnai, D., & Black, G. (2011). Whole-exome-sequencing identifies mutations in histone acetyltransferase gene KAT6B in individuals with the Say-Barber-Biesecker variant of Ohdo syndrome. *American Journal of Human Genetics*, 89(5), 675-81.

[82] Kleefstra, T., Brunner, H. G., Amiel, J., Oudakker, A. R., Nillesen, W. M., Magee, A., Geneviève, D., Cormier-Daire, V., van Esch, H., Fryns, J. P., Hamel, B. C., Sistermans, E. A., de Vries, B. B., & van Bokhoven, H. (2006). Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. *American Journal of Human Genetics*, 79(2), 370-7.

[83] Weaver, I. C., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., Dymov, S., Szyf, M., & Meaney, MJ. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience*, 7(8), 847-54.

[84] Lillycrop, K. A., Phillips, E. S., Jackson, A. A., Hanson, M. A., & Burdge, G. C. (2005). Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. *Journal of Nutrition*, 135(6), 1382-6.

[85] Lillycrop, K. A., Phillips, E. S., Torrens, C., Hanson, M. A., Jackson, A. A., & Burdge, G. C. (2008). Feeding pregnant rats a protein-restricted diet persistently alters the methylation of specific cytosines in the hepatic PPAR alpha promoter of the offspring. *British Journal of Nutrition*, 100(2), 278-82.

[86] Heijmans, B. T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., Slagboom, P. E., & Lumey, L. H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proceedings of National Academy of Science USA*, 105(44), 17046-9.

[87] Tobi, E. W., Lumey, L. H., Talens, R. P., Kremer, D., Putter, H., Stein, A. D., Slagboom, P. E., & Heijmans, B. T. (2009). DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Human Molecular Genetics*, 18(21), 4046-53.

[88] Franklin, T. B., Russig, H., Weiss, I. C., Gräff, J., Linder, N., Michalon, A., Vizi, S., & Mansuy, I. M. (2010). Epigenetic transmission of the impact of early stress across generations. *Biological Psychiatry*, 68(5), 408-15.

[89] Horsthemke, B. (2007). Heritable germline epimutations in humans. *Nature Genetics*, 39(5), 573-4.

[90] Daxinger, L., & Whitelaw, E. (2010). Transgenerational epigenetic inheritance: more questions than answers. *Genome Research*, 20(12), 1623-8.

[91] Seong, K. H., Li, D., Shimizu, H., Nakamura, R., & Ishii, S. (2011). Inheritance of stress-induced, ATF-2-dependent epigenetic change. *Cell*, 145(7), 1049-61.

[92] De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J. A., Sharpe, J. A., Koch, C. M., Dunham, I., Gibbons, R. J., Wood, W. G., & Higgs, D. R. (2007). Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. *Blood*, 110(13), 4503-10.

[93] Yagi, S., Hirabayashi, K., Sato, S., Li, W., Takahashi, Y., Hirakawa, T., Wu, G., Hattori, N., Hattori, N., Ohgane, J., Tanaka, S., Liu, X. S., & Shiota, K. (2008). DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. *Genome Research Dec*, 18(12), 1969-78.

[94] Beck, S. (2010). Taking the measure of the methylome. *Nature Biotechnology*, 28(10), 1026-8.

[95] Kubota, T., Miyake, K., & Hirasawa, T. (2011). Epigenetic modifications: genetic basis of environmental stress response. *In DNA Replication / Book 1. Rijeka.*, Intech, http://www.intechopen.com/books/fundamental-aspects-of-dna-replication/epigenetic-modifications-genetic-basis-of-environmental-stress-response, accessed 15 July 2012.

# Chromatin Damage Patterns Shift According to Eu/ Heterochromatin Replication

María Vittoria Di Tomaso, Pablo Liddle,
Laura Lafon-Hughes, Ana Laura Reyes-Ábalos and
Gustavo Folle

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/51847

## 1. Introduction

In order to maintain genetic stability, strictly controlled mechanisms are essential to assure the accuracy of genetic functions. Precise genome replication and correct control of gene expression mostly *via* epigenetic mechanisms are critical in maintaining the stability of genomes. Moreover, the characteristic chromatin compartmentalization of mammalian genomes contributes to regulate the housekeeping or tissue-specific genetic activities [1, 2].

Table 1 summarizes the distinct chromatin compartments and their foremost properties. Euchromatin (*eu*: true) and heterochromatin (*hetero*: different) are two major compartments or chromatin states of the DNA originally distinguished by their isopycnotic or heteropycnotic interphase staining properties, respectively [3]. The heterochromatin compartment differentiates in both constitutive (permanent) and facultative (developmentally reorganized) states [4]. Facultative heterochromatin represents chromatin regions being facultatively inactivated (heterochromatinized) because of gene dosage compensation (i.e.: mammalian female inactive X chromosome) randomly silenced at an early stage of embryogenesis or tissue-specific gene expression. Constitutive heterochromatin consists in regions of α- and β-heterochromatin [5, 6].

Distinct features characterize the different chromatin states (Table 1). Interphase open chromatin conformation and transcriptional activity in all cell types distinguish euchromatin. Higher order chromatin compaction characterizes constitutive α- and β-heterochromatic regions while gene silencing differentiates constitutive α-heterochromatin. Tissue-specific transcriptional activity and low or high chromatin condensation, depending on gene expres-

sion, correspond to features of facultative heterochromatin [7, 6]. The mammalian genome compartmentalization can be visualized in both banded metaphase chromosomes and stained interphase nuclei.

| Compartments | Euchromatin | Facultative heterochromatin | | Constitutive heterochromatin | |
|---|---|---|---|---|---|
| Chromatin types | Euchromatin | Tissue-specific | Dosage compensation | α- heterochromatin | β-heterochromatin |
| Location in metaphase chromosomes | Light G-bands | Dark G-bands | Inactive X chromosome (Xi) | C-bands | C-bands |
| Location in interphase nuclei | Inner compartment | Peripheral compartment and chromocenters | Peripheral compartment and chromocenters | Peripheral compartment and chromocenters | Peripheral compartment and chromocenters |
| Interphase chromatin compaction | Open conformation | Low or high order compaction | High order compaction | High order compaction | High order compaction |
| Presence of genes | Housekeeping genes | Inactivated tissue-specific genes | Dosage inactivated genes | No genes | Transposable elements and heterochromatic genes |
| Gene expression and relation to chromatin state | Gene activity in euchromatic state in all cells | Tissue-specific gene activity in euchromatic state | Gene activity in euchromatic state until silencing | No gene activity | Low gene activity in heterochromatic state |
| GC or AT DNA sequences richness | GC-rich | AT-rich | GC- and AT-rich | AT-rich | AT-rich |
| Repeated DNA sequences | SINEs | LINEs | SINEs and LINEs | Tandem highly repeated DNA sequences | Tandem highly repeated DNA sequences |
| CpG island methylation | Unmethylated | Unmethylated or methylated | Methylated | Methylated | Methylated |
| Core histone tail acetylation | Hyperacetylated | Hyperacetilated or hypoacetylated | Hypoacetylated | Hypoacetylated | Hypoacetylated |
| Replication timing | Early | Early or late | Late | Late | Late or early |

**Table 1.** Distinguishing properties of chromatin compartments.

The C-banding procedure [8] produces a selective staining of specific chromosome regions, mapping at or adjacent to centromeres, telomeres or interstitial arm sites, depending on the species. Occasionally, a chromosome arm is entirely heterochromatic, such as the long arm of the Chinese hamster X chromosome (Figure 1, left). In humans, C-bands are located at centromeres and pericentric regions of all chromosomes, being conspicuous at the pericentric regions of chromosomes 1, 9 and 16 and the distal long arm of the Y chromosome (Yq) (Figure 1, right).
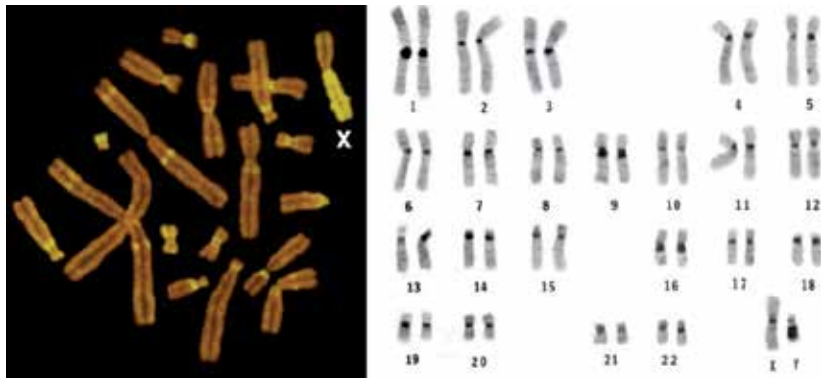
**Figure 1.** C-banding in CHO and human chromosomes. Left: C-banded metaphase of CHO9 cell line. The CHO cell line was established from a Chinese hamster ovary fibroblast culture [9] and presents a modal number of 21 chromosomes. This cell line contains eight normal and twelve rearranged autosomes with only one X chromosome. Giemsa-stained C-band regions are visualized in yellow (reflected light microscopy). The CHO X chromosome (X) shows an almost entirely heterochromatic long arm. Right: C-banded caryotype of a human peripheral lymphocyte metaphase showing centromeric, pericentric (chromosomes 1, 9 and 16) and distal Yq heterochromatic blocks.

By digestion with the proteolytic enzyme trypsin followed by Giemsa staining (G-banding procedure) [10], a pattern of alternate light and dark regions along the length of all chromosomes is obtained (light G-bands and dark G-bands, respectively). The G-band pattern is characteristic for each chromosome pair allowing their precise identification and caryotyping. Figure 2 shows the CHO9 and human G-band chromosome patterns.
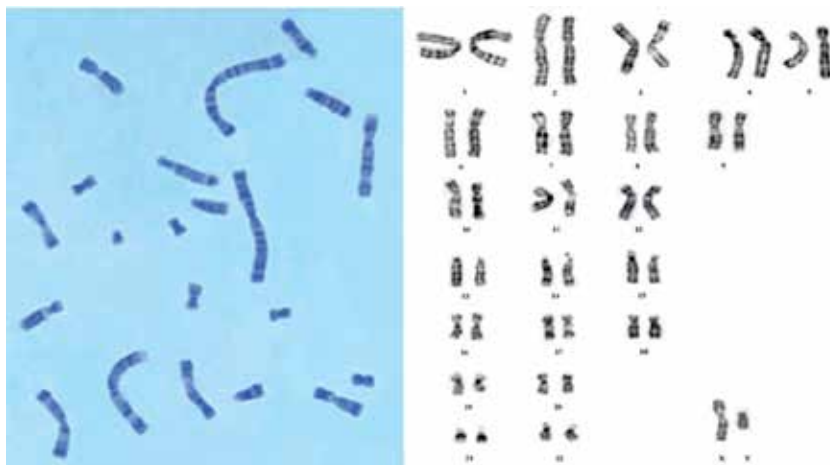


**Figure 2.** G-banded CHO9 metaphase (left) and a male human peripheral lymphocyte caryotype exhibiting G-bands (right).

C- and G-band patterns reveal the heterogeneous organization of chromatin along condensed chromosomes. C-bands enclose constitutive α- and β-heterochromatin. Regions with

ubiquitously expressed housekeeping genes (euchromatin) reside in light G-bands, while tissue-specific genes (facultative heterochromatin) dwell in dark G-bands [5, 6, 11 ].

Light and dark G-bands may reflect a differential array of SAR (**S**caffold-**A**ssociated **R**egions), composed by highly AT-rich DNA stretches binding to the chromosome scaffold. Regions of dark G-bands exhibit a tighter chromatin fiber coiling than light G-bands domains [12]. Constitutive heterochromatin has an even more dense conformation.

Moreover, euchromatic light G-bands are GC-rich and gene-dense regions, containing unmethylated CpG islands and moderately repeated **S**hort **In**terspersed **E**lements (SINE), mainly represented by Alu family sequences. Conversely, facultative heterochromatic dark G-bands are AT-rich, gene-poor and harbor hypermethylated CpG and moderately repeated family of Long Interspersed Elements (LINE) sequences. Constitutive $\alpha$-heterochromatic C-bands are the major locations of tandem non-coding highly repeated satellite DNA sequences, devoid of genes [11, 13]. However, constitutive $\beta$-heterochromatin presents inserted middle-repetitive transposable elements between the tandem repeats, some of them transcriptionally active [6]. Moreover, genes residing within regions of pericentric constitutive $\beta$-heterochromatin termed "heterochromatic genes" have been reported in *Drosophila,* mammals and plants [14, 15].

In spite of variations according to cell type or function of mammalian interphase nuclei, the corresponding chromatin of light and dark G-bands as well as C-bands is non-randomly distributed in different nuclear compartments, displaying specific chromatin conformation, molecular composition and gene expression patterns.

In most interphase cells, euchromatin (light G-bands) dwells in the inner compartment of nuclei, whereas heterochromatin (dark G-bands and C-bands) resides in the peripheral compartment, chromocenters and around nucleoli [6, 16]. Figure 3 illustrates a HeLa nucleus where the different interphase chromatin compartments can be recognized.
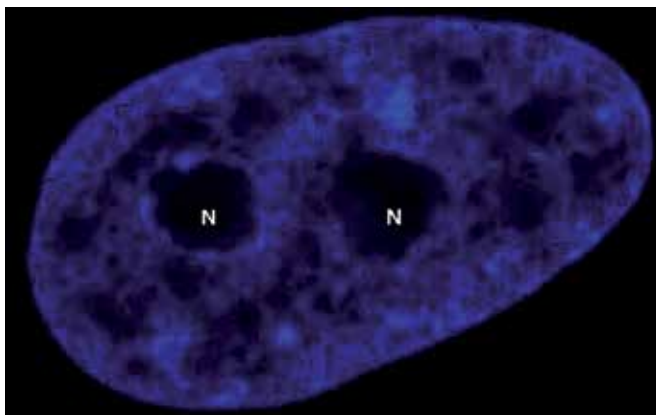


**Figure 3.** Distinct eu/heterochromatin compartments in DAPI-stained HeLa interphase nucleus. DAPI-bright regions correspond to heterochromatin and dim areas to euchromatin. N: nucleolus.

Constitutive and tissue-specific genes are only expressed in the euchromatic state. There-fore, facultative heterochromatin behaves as euchromatin in cells where its tissue-specific genes are transcribed, but holds a packed (heterochromatic) conformation when genes re-main silent.

However, some transposons and heterochromatic genes of β-heterochromatin are transcrip-tionally active in heterochromatic state suggesting that distinct epigenetic mechanisms of gene regulation and preservation of eu/heterochromatic states may exist in these regions [6, 14, 15].

Once acquired, the chromatin states are somatically maintained as stable heritable epigenet-ic states. Euchromatin remodels during mitosis and restores the original organization in ear-ly $G_1$ phase of each cell cycle. In addition, during DNA synthesis (S-phase) both euchromatin and heterocromatin transiently lose their typical condensation status recover-ing the previous folding level after replication. Establishment and maintenance of chromatin states involve post-translational modification enzymes that act coordinately to methylate CpG islands and to either acetylate, methylate, phosphorylate, ubiquitinate, poly-ADP ribo-sylate or SUMOylate the core histone tails of nucleosomes. These epigenetic changes, togeth-er with the recruitment of methyl-CpG binding proteins, ATP-dependent chromatin remodeling complexes and the association of specific non-histone proteins, such as HP1 (**H**eterochromatin **P**rotein 1) or RNAi (non-coding interference RNA), also mediate the regu-lation of DNA replication, transcription and repair [17, 18].

The N- and C-termini of H3 and H4 core histones are particularly involved in epigenetic regulation. Acetyl groups covalently added to lysines, serines or arginines of the N-terminal histone tails reduce the affinity to DNA, promoting the accessibility of chromatin remodel-ing and activating transcription factors. Therefore, histone hyperacetylation usually charac-terizes active chromatin regions. Conversely, transcriptionally silenced chromatin regions generally contain hypoacetylated histones (Table 1). For instance, H3 acetylated (ac) in ly-sine 9 (H3K9ac) is enriched at the promoter region of active genes although, it was reported that the histone H3 acetylated at lysine 4 (H3K4ac) resides in pericentric heterochromatin of *Schizosaccharomyces pombe,* playing a role in the assembly of repressive heterochromatin [19]. On the other hand, histone methylation (me) can be associated with transcriptional activa-tion or repression. For example, methylation of H3 on lysines 4, 36 or 79 (H3K4me, H3K36me, H3K79me) is associated with transcriptional activation whereas methylation of H3 on lysines 9 or 27 (H3K9me, H3K27me) and of H4 on lysine 20 (H4K20me) is involved in transcriptional repression [18]. The concerted action of acetylated and methylated histone core residues is central in creating a "histone code" which delineates distinct genomic loci that recruit factors needed for DNA remodeling, transcription, replication and repair [5, 17].

In general, methylation of CpG islands within 5'regions of genes is associated with hypoace-thylated histones, characterizing the heterochromatic state (Table 1). However, DNA meth-ylation is not exclusively related to gene silencing. It was reported that methylation of some imprinting centers can displace trans-acting repressor factors, allowing the expression of the linked imprinted genes [20].

The epigenetic mechanisms involved in the maintenance of eu/heterochromatic compartments and gene expression are connected to DNA replication. There are specific interactions between components of the replication machinery and chromatin related factors, timing the eu- or heterochromatin replication.

## 2. Replication of eu/heterochromatin compartments

Compartmentalization of vertebrate genomes cooperates in achieving the high fidelity DNA replication necessary for the accurate preservation of the genetic information throughout cell generations. DNA replication is a temporarily and spatially highly ordered and strictly regulated process, occurring during S-phase of the cell cycle, with distinct genome compartments replicating at different times. The replication timing of the genome compartments are highly conserved within consecutive cell cycles and regulated by specific epigenetic chromatin conformation domains, DNA features and transcriptional activity [21, 22, 23].

Mammalian chromosome duplication involves clusters or domains of neighboring replicons named **R**eplication **T**iming **D**omains (RTD) which synchronously start and end replication, according to a deterministic replication timing program [21, 22, 24]. When one domain completes replication, an adjacent domain successively initiates DNA synthesis [25]. Remarkably, mouse and human asynchronous replication timing may function randomly between individual replicons within a RTD and non-randomly between RTD [25]. The random firing of replication origins within a RTD generates a different replication pattern during each S-phase, but it has been reported that some origins fire preferentially and more frequently than others [26]. The RTD are stable structures of mammalian interphase nuclei, replicating and transcribing in temporal and spatial coordination [26].

Pulse labeled interphase nuclei of human, mouse and hamster cells with the base analogues 5-bromo-2'-deoxyuridine (BrdU) or 5-ethynyl-2′-deoxiuridine (EdU) demonstrated the asynchrony and specific spatial distribution of DNA replication. The early replication pattern of S-phase (ES-phase) is characterized by replication foci dispersed throughout the inner environment of the nuclei with scarce or absence of foci at the periphery or adjacent to the nucleoli. The replication pattern changes throughout the progression of S-phase. In mid S-phase (MS-phase) most foci map adjacent to the internal nuclear membrane and around nucleoli, with few foci centrally located. Lastly, late S-phase replication maps next to the nuclear envelope as well as in chromocenters and around nucleoli [16, 27]. Early S-phase and late S-phase replication patterns of CHO9 cells are illustrated in Figure 4.

In general, chromatin with transcriptional activity (euchromatin) replicates early in S-phase whereas constitutive $\alpha$-heterochromatin duplicates late. Besides, facultative heterochromatin replicates earlier if its tissue-specific genes are being expressed and later if not [6, 28] (Table 1). It has been reported that genes of mouse embryonic stem cells residing within GC-rich and LINE-poor DNA (euchromatin) do not modify their replication timing after differentiation to neural precursors, whereas genes residing in AT-rich and LINE-rich DNA revealed changes in replication timing accompanied by changes in gene expression and

chromatin folding [29]. A change of replication timing from early S to late S-phase is particularly evident in the female mammalian Xi [30]
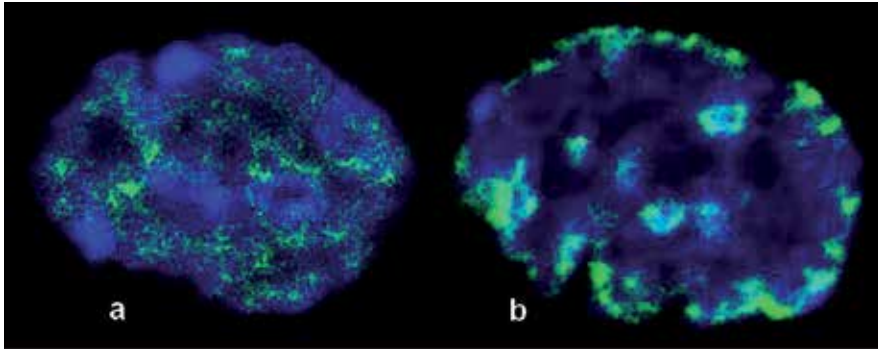


**Figure 4.** Early (ES-phase) or late (LS-phase) replication patterns of CHO9 nuclei revealed by incorporation of EdU and subsequent detection with an Alexa Fluor 488 (green) conjugated azide (Click-iT EdU imaging kit, Invitrogen). (a) ES-phase nucleus with inner compartment replication. (b) LS-phase nucleus showing replication in the peripheral compartment, chromocenters and around nucleoli.

Early replication seems to be important but not essential for gene transcription. Moreover, late replication is not an obligatory feature of heterochromatin. For example, transcriptionally active transposons of β-heterochromatin replicate late while the heterochromatic centromeres and the silent mating-type cassettes of *Schizosaccharomyces pombe* replicate in early S phase [14, 15, 31]. There are additional cases reported of early heterochromatin replication such as human telomeres [32]. and mouse pericentric heterochromatin and centromeres [33].

The early replicon clusters of higher eukaryotes alternate their replication and transcription activity. However, correlation between replication and transcription does not exist in S*accharomyces cerevisiae* [34]. Employing distinct colored fluorescent labels to recognize early replication foci and transcription foci (factories), it was shown that both labels do not colocalize. In ES-phase, actively replicating foci are transcriptionally inactive and only restart transcription after finishing replication. The replication timing is indirectly related to transcription through the assembly of a higher-order chromatin state [2]. For example, silencing of the mammalian Xi is initially reversible and only stabilizes when an identifiable higher-order chromatin configuration (Barr body) appears and replication is delayed [35].

The chromatin replication timing is reestablished early in $G_1$ phase of each cell cycle, coincident with the anchorage and positioning of chromosomal segments at specific locations within the nucleus named TDP (**T**iming **D**ecision **P**oint) [36]. Both anchorage and positioning of chromosomes are central in the organization of nuclear eu/heterochromatic compartments and the establishment of replication timing and transcriptional activity [23, 36]. Modifications in subnuclear chromatin organization are associated with changes in replication timing during development [37]. For example, the position of the immunoglobulin heavy chain locus (IgH) in B cells shows that its localization in the interphase nuclei depends on replication timing and gene activity. During early stages of B cell differentiation,

both transcribed alleles of the IgH locus are centrally located in the nucleus and replicate early. Conversely, in advanced differentiation stages the IgH locus is repositioned to the nuclear periphery, repressed and late-replicated [38].

Nonetheless, chromatin positioning at the nuclear periphery is indicative but not mandatory for gene silencing and late replication. In fact, the nuclear periphery is heterogeneous with respect to transcription. For instance, in budding yeast, nuclear pores, which mediate the transport between the nucleus and cytoplasm, enhance the transcriptional activity of genes positioned in their proximity [39]. The dosage compensation complex of the hyperacetylated *Drosophila* male X chromosome interacts with nuclear pore proteins determining its transcription up-regulation and early DNA duplication [40].

Replication clusters correspond to bands of metaphase chromosomes. Tightly coiled C-band (constitutive heterochromatin) replicates in late S-phase. Facultative heterochromatin of the dark G-bands duplicates either early or late depending on its tissue-specific expression. Early replication pattern characterizes the loosely coiled euchromatin of light G-bands. Ubiquitously expressed housekeeping genes (light G-bands) are therefore early replicating [41, 42]. Duplication timing analysis by quantitative PCR of the boundary region between G-light 13q14.3 and G-dark 13q21.1 bands showed that the G-light side of the frontier replicates early whilst the G-dark interface replicates late. However, analysis using PCR primers spaced at approximately 150 Kb intervals showed that the switch in G-light/G-dark band replication timing takes place gradually from early-mid to late S-phase over a 1-2 Mb region [43]. The DNA segments corresponding to large regions between early and late-S phase replication timing domains are termed TTR (Timing Transition Regions) [44].

A correlation between replication timing and epigenetic modification of chromatin has also been shown. Early replication domains are related to specific combination of changes in histone lysine residues (H3K9Ac, H3K27Ac, H3K4me, H3K36me and H3K79me) associated with transcriptional activity. On the other hand, the repressive epigenetic modifications (H3K9me, H3K27me and H4K20me) are linked to late replication [18].

Chromatin epigenetic changes occurring throughout DNA replication may provide a replication timing mechanism (firing early or late replication origins) in the direction of maintaining specific chromatin expression patterns [45]. It was reported that histone hypoacetylation is needed to preserve normal heterochromatin replication dynamics [46] and that histone hyperacetylation may increase the efficiency of replication origins, advancing the replication timing of distinct genomic regions [47]. For instance, removal of acetyl groups by HDAC (**H**istone **DeAC**etylase) contributes to mantain late replication at imprinted loci [48] and the generation of neocentromeres [49].

Several proteins, including CpG island-methylating DNMT (**DN**A **M**ethyl **T**ransferase), core histone tail-methylating HMT (**H**istone **M**ethyl **T**ransferase) and HP1 (**H**eterochromatin-associated **P**rotein), colocalize with late replicating DNA regions [45]. HP1 binds to heterochromatin, facilitating the extension of the repressive H3K9me modification [50] and hence delaying replication timing by supporting heterochromatin conformation. HP1 could facilitate the late firing of replication origins within heterochromatin [51]. Furthermore, muta-

tions of DNMT result in earlier replication of normally late replicating DNA. For instance, patients with mutations in the Dnmt3b gene (coding protein DNMT3b) have hypomethylated CpG islands in the Xi chromosome, which replicates at an earlier S-phase stage despite the presence of XIST (**X**-**I**nactive **S**pecific **T**ranscript) RNA [52]. Accordingly, changes in either DNA or histone methylation status in concert with histone acetylation patterns may promote open or tight chromatin conformations and thus modifications in the firing of replication origins and/or replication rates [47].

In mammals, several distinct discrete or diffuse genomic sequence motifs can potentially act as **O**rigin **R**eplication **I**dentification (ORI), where a large number of proteins bind to load replication complexes. A protein complex, named the pre-**R**eplication **C**omplex (pre-RC) associates with potential replication origins in $G_1$ phase. This complex includes the **O**rigin **R**ecognition **C**omplex (ORC), which recognizes the replication origins, the helicase MCM2-7 (**M**ini **C**hromosome **M**aintenance 2-7), and other essential factors. Early firing ORI demonstrated to be rich in MCM proteins. Besides, MCM could be more efficient in early firing than in late firing ORI suggesting that heterochromatin could repress MCM activities [53, 54].

Accessibility of replication initiation factors to redundant or discrete replication origins may be regulated by its nuclear localization in relation to chromatin states. For example, the early replicating α-globin locus is located within a light G-band. Deletions that juxtapose the α-globin locus next to a region of late replicating telomeric condensed heterochromatin (repositioning this locus to the nuclear periphery), delay the initiation of α-globin replication by restricting the access of initiation factors to the ORI [55].

There is a complex cell cycle intra-S checkpoint involving the ATR/CHK1-related network in metazoas and ATR/Rad53 in *Saccharomyces cerevisiae* that controls replication asynchrony. The transition from early to late S-phase replication (mid-S replication pause) is coupled with the activation of the intra S-phase checkpoint at mid S-phase which inhibits the initiation of late replicons. It has been reported that inhibition of CHK1 generates earlier firing of a late-firing subset of ORI [56, 57]. Accordingly, the checkpoint function may play a role in regulating replication asynchrony and S-phase progression [25, 58].

Both DNA and histone methylation can affect replication timing *via* the ATR/CHK1 control pathway. There is a complex and so far not completely understood relationship between checkpoint function and epigenetic modifications (DNA methylation, histone methylation and histone acetylation) in the regulation of replication origins firing during S-phase [47, 59].

Following pre-RC loading to ORI, a protein pre-**I**nitiation **C**omplex (pre-IC) assembles upon MCM proteins together with factors required for loading replicative polymerase. The chromatin association of pre-RC and pre-IC is asynchronous, allowing pre-RC inhibition and pre-IC activation (from S-phase initiation toward the end of mitosis) by the cell cycle CDK proteins (**C**yclin-**D**ependent **K**inases). This regulation licenses replication to occur at a specific time, only once per cell cycle, and ensures that cell cycle cannot progress until checkpoints are satisfied. In *Xenopus laevis* and mammalian cells there is an additional system to

control licensing by means of the geminin protein, which also inhibits pre-RC. Degradation of geminin at the end of mitosis is essential for a new license of replication [56, 60].

Completion of replication is necessary for entire chromosome condensation. *Drosophila* ORC mutants unable to complete S-phase have defects not only in DNA replication (with some euchromatic regions replicating even later than heterochromatin) but also in cell cycle progression and chromatin condensation [61]. Although some levels of chromosome condensation occur in the absence of a complete replication cycle, mitotic chromosomes are shorter and thicker than in wild type *Drosophila*. Even though ORC is principally involved in the initiation of DNA replication, additional roles in mitotic chromosome condensation, centromere function as well as the establishment and maintenance of gene silencing and heterochromatin have been suggested [61, 62, 63].

## 3. Eu/heterochromatin replication and distribution of genetic damage

The S-phase of the cell cycle has proved to be very sensitive to genetic damage. S-phase has been considered as one of the sources of genomic instability. There are several lines of evidence that correlate genomic instability with chromosomal aberrations (CA), birth defects and infertility [64]. Besides, oncogene activation or tumor suppressor gene repression can arise as a consequence of primary DNA damage or CA [65]. Several authors have reported the colocalization of induced CA breakpoints (BP) (sites of chromosomal breaks in a CA) with regions harboring fragile sites, oncogenes or cancer-associated CA [66-72].

The human genome holds long stretches of AT-rich sequences as well as inverted, mirror or direct tandem repeats, prone to be arranged in unusual DNA secondary structures that may inhibit replication. The presence of secondary structures, unstable single-stranded or non-replicated regions could lead to chromosome fragility expressed as gaps or breaks in metaphase chromosomes [73, 74].

DNA replication in mammals slows down significantly when the 1-2 Mb regions of TTR are replicated [57]. It was reported that after replication of euchromatic light G-bands, the replication fork stalls at TTR of the interband regions, restarting DNA synthesis at the adjacent dark G-band after a mid S-phase pause [6]. This interband region devoid of replication origins is often replicated by means of a single replication fork [75]. Such genomic segments could generate damage-prone regions that frequently overlap with DNA fragile sites [43, 76]. For example, the common fragile site FRA3B is devoid of replication origins and thus completes replication very late in S-phase [77]. In addition, it was observed that mutation rates increase with the distance from replication origins [78, 79].

Furthermore, it was reported an increase in mutation rate as S-phase advances. Early replicating housekeeping genes are more conserved than later replicating tissue-specific genes [57, 80]. Genes corresponding to mutational hot spots involved in speciation and adaptive radiation response are late replicating [57]. CpG methylation status of late replicating regions may contribute to the rise in mutation rate mostly due to $5^{me}$CpG substitutions [81, 82].

### 3.1. Eu/heterochromatin replication and induced-damage distribution in a mitotic chromosome model

DNA lesions trigger a **D**NA **D**amage **R**esponse (DDR) characterized by activation of cell cycle checkpoints, damage sensor proteins, DNA repair mechanisms and apoptotic pathways [83, 84]. The DNA **D**ouble-**S**trand **B**reak (DSB) is the critical DNA lesion involved in CA production [85]. DSB can be generated by DNA-damaging agents or spontaneously through the endogenous production of reactive oxygen species (ROS) or cellular processes such as DNA replication, repair, transposition or mitotic recombination. Agents inducing DSB and CA are named clastogens. The S-phase independent clastogens, like ionizing radiation and the radiomimetic agent bleomycin, directly induce DSB. Conversely, S-phase dependent clastogens such as UV-C and alkylating compounds need the intervention of DNA repair and replication in order to generate DSB, which could ultimately lead to CA. Hence, DNA replication constitutes a relevant step in the transformation of DNA lesions into CA. Besides, some clastogenic agents such as the anti-topoisomerase II cleavable complex trappers behave as S-phase independent clastogens. Eukaryotic topoisomerases II alleviate tensional DNA stress by the generation of a DNA topoisomerase II complex (cleavable complex) within which the topoisomerase II component introduces transient breaks in both DNA strands (DSB) allowing the DNA to pass through the breaks [86]. Drugs that act by trapping cleavable complexes hamper the resealing of DSB produced by topoisomerase II and, as a consequence, DNA DSB persist [87, 88].

As shown in Figure 1, the CHO9 X-chromosome exhibits an almost entire constitutive heterochromatic long arm (Xq) with the exception of a medial secondary constriction. Besides, Xq replicates in late S-phase whereas the euchromatin of the short arm (Xp) and the Xq secondary constriction duplicates during early S-phase (Figure 5) [89, 90]. Differential replication timing of Xp and Xq of CHO cells provided a valuable experimental model to analyze the relationship between eu/heterochromatin DNA replication and CA induced by different types of clastogens: UV-C light, the methylating agent methylmethane sulphonate (MMS) and the anti-topoisomerase II inhibitor etoposide (a cleavable complex trapper) in BrdU pulse-labeled CHO9 chromosomes [91, 92].

CHO9 cells were treated with MMS (20 mM) or etoposide (20 µM) and simultaneously exposed to 30 mM BrdU (40 min) or otherwise exposed to UV-C (30 J/m2; 0.1 J/m2/s) and immediately labeled with BrdU (40 min). Incorporation of BrdU in Xp or Xq was disclosed by immunolabeling either treated or control CHO9 metaphases with anti-BrdU antibodies coupled to FITC. The relationship between replication timing, chromatin conformation and genetic damage was investigated by mapping induced BP in Xp and Xq in cells treated both in early and late S-phase [91, 92].

Examples of CA induced by MMS, etoposide and UV-C in replicating CHO9 Xp or Xq are shown in Figure 5. Figure 6 illustrates Xp/Xq distribution of etoposide, UV-C and MMS-induced BP in relation to replication.

The application of $\chi^2$ test to analyze the association between Xp/Xq replication pattern and Xp/Xq BP localization showed that when Xp replicates, BP produced by either MMS, UV-C

or etoposide clustered in Xp. On the other hand, during Xq replication, BP induced by the clastogens concentrated in Xq [91, 92] (Figure 6).
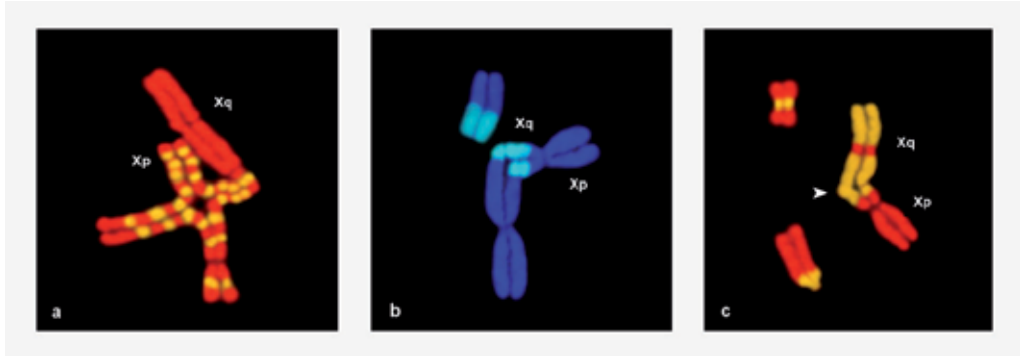


**Figure 5.** Illustrations of CA involving CHO9 Xp or Xq induced by (a) MMS, (b) etoposide, or (c) UV-C in (a) early (Xp replication) or (b and c) late (Xq replication) S-phase. Different types of CA are shown: (a) symmetric quadrirradial affecting Xp; (b) asymmetric quadrirradial with acentric fragment involving Xq; (c) duplication-deletion in Xq (arrow). Chromosomes exhibit BrdU immunolabeling (yellow) and either PI (red) or DAPI (blue) counterstaining.
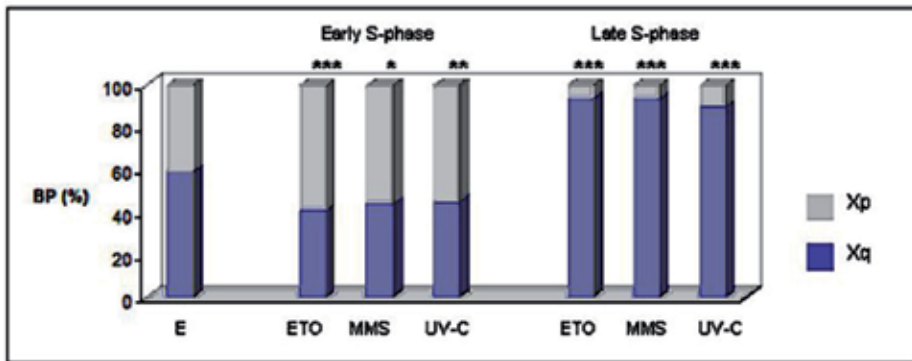


**Figure 6.** Bar diagram illustrating CHO9 X chromosome BP distribution induced by etoposide (ETO, 20 μM), methylmethane sulphonate (MMS, 20 mM) and UV-C (30 J/m²; 0.1 J/m²/s) in Xp (grey) and Xq (blue) during early (ES-phase) and late (LS-phase) cell cycle phases. The bar on the left side (E) indicates the expected frequencies of induced BP according to Xp and Xq relative length.

Since UV-C and MMS are S-phase dependent clastogens, the observed predominance of BP produced in Xp or Xq according to replication timing could be explained based on their requirement of DNA synthesis to produce CA. DNA base damage induced by MMS as well as cyclobutane pyrimidine dimers (CPD) and 6-4 photoproducts (6-4 PP) produced by UV-C are preferentially repaired through **B**ase **E**xcision **R**epair (BER) and **N**ucleotide **E**xcision **R**epair (NER) mechanisms, respectively. Both repair systems create an excision repair **S**ingle-**S**trand **B**reak (SSB) intermediate at the site of DNA lesion which is then filled by DNA

repair synthesis [93]. If DNA replication initiates with an excision repair SSB intermediate, another SSB can be generated in the complementary DNA strand, thus forming a DSB [94, 95]. Additionally, CPD, 6-4 PP or base damage in a single strand (unrepaired before DNA replication) may stall the replication fork and as a result, may produce a SSB in the opposite DNA strand [96, 97]. Furthermore, two nearby SSB in each DNA strand may behave as a DSB [98]. The DSB generated could be ultimately processed and transformed in CA [91, 92].

Nonetheless, the preferential location of CA in replicating Xp or Xq during etoposide treatment (independently of its eu/heterochromatic states) may occur due to the inhibition of topoisomerase II activity during DNA synthesis [87, 88]. The local unraveling and subsequent rewinding of eu or heterochromatin regions undergoing replication require topoisomerase II activities to alleviate DNA torsional stress [86]. Etoposide stabilizes DNA-topoisomerase II cleavable complex and hinders the resealing of DSB introduced by the enzyme generating the accumulation of DSB unable to reach resolution. In addition, chromatin unwinding during replication may turn DNA more accessible to S-independent and S-dependent chemical agents including etoposide and MMS, respectively [91, 92].

### 3.2. Eu/heterochromatin replication and primary induced-damage distribution in interphase nuclei

Few minutes after exposure of mammalian cells to DSB-inducing agents, the nucleosomal histone variant H2AX is phosphorylated at serine 139 (humans) or 129 (mouse) of C-terminal tails reaching a peak of phosphorylation 30 min later. H2AX phosphorylation (named γH2AX) initiates around the induced DSB and spreads through a large chromatin region (~2000 H2AX molecules) flanking the lesion, which can be visualized as discrete γH2AX foci in interphase nuclei and mitotic chromosomes by means of specific fluorochrome-conjugated antibodies [99].

γH2AX is involved in the DDR by coordination with other damage response proteins to recruit signaling, remodeling, checkpoint and repair proteins. At sites of DSB, the DNA-PK (**DNA D**ependent **P**rotein **K**inase) binds to activate the **N**on **H**omologous **E**nd **J**oining (NHEJ) DSB repair pathway. If DSB are produced after replication, RAD51 and BRCA2 are recruited to DSB sites initiating the **H**omologous **R**ecombination repair pathway (HR). Simultaneously, the sensing complex MRN (MRE11, RAD50, NBS1) associates to DSB, facilitating the recruitment and activation (auto-phosphorylation) of ATM (**A**taxia **T**elangiectasia **M**utated), MDC1, BRCA1 and 53BP1 [100].

ATM, ATR (**AT**M- and **R**ad3-related) and DNA-PK are members of the phosphatidylinositol 3-kinase-like family of serine/threonine protein kinases that phosphorylate H2AX. Unlike ATM, which appears to be mainly activated by DSB, ATR seems to be activated by induced SSB and the excision repair SSB intermediates generated during DNA repair. Since ATR activation was observed in replicating cells, it was suggested that the blockage of replication forks by SSB is required to initiate ATR-mediated phosphorylation of H2AX. Besides, it was reported that stalled replication forks may also trigger H2AX phosphorylation when bulky lesions (i.e.: CPD and 6-4 PP) collide with replication forks [101, 102].

NBS1, MDC1, 53BP1, and BRCA1 may all function as mediators and amplifiers of the DDR, recruiting diverse repair and checkpoint proteins (including ATM and ATR) and generating an amplification loop that also extends H2AX phosphorylation [99]. 53BP1 can bind directly to H3K79me and H4K20me accumulated at sites of DSB collaborating with a global chromatin unwinding following the formation of DSB in concert with other proteins like TIP60 (member of an histone acetyltransferases family) and KAP1 [103, 104, 105].

Several immunofluorescence studies have demonstrated that induced-γH2AX foci are located preferentially within euchromatic regions of the genome, suggesting that heterochromatin could be refractory to γH2AX foci formation. Employing immuno-FISH to analyze radiation induced-DSB (γH2AX foci) in chromatin regions with known chromatin compaction (human chromosome 18 versus chromosome 19; RIDGE versus anti-RIDGE region of human chromosome 11), it has been observed that condensed regions of gene-poor chromatin are less susceptible to DSB induction compared with decondensed, gene-rich chromatin [106-109].

Different hypothesis have been raised to explain the non-homogeneous distribution of γH2AX foci in nuclei. The highly condensed state or abundance of binding proteins may reduce the accesibility of chemical DNA damaging agents to heterochromatin. Besides, since condensed chromatin is less hydrated than euchromatin, a lower amount of free radicals could be induced by radiation [110]. Furthermore, compact heterochromatin could contain a lower proportion of H2AX isoform or be less accessible to kinases due to compaction or protein coating [106]. Additionally, a wave of chromatin unwinding starting at DSB sites and spreading throughout the entire chromatin was described (as a result of KAP1 phosphorylation by ATM kinase) implying that the preferential location of γH2AX foci in decondensed chromatin perhaps reflects chromatin reorganization [105, 111-113].

Finally, a short-range migration of DSB from packed chromatin toward specific decondensed DSB repair domains could also take place [106, 110]. Using carbon ion microirradiation to induce DSB combined to a modified TUNEL assay to directly visualize these lesions and γH2AX immunodetection, a bending of the linear ion-induced γH2AX track around heterochromatic regions was observed [114]. The γH2AX foci migration from the interior to the periphery of heterochromatin appears to initiate within 20 min post-irradiation and be almost complete 1 h after damage induction. The decondensation of heterochromatin at sites of ion hits possibly promotes the movement of DSB to peripheral regions of lower chromatin density where repair may potentially proceed [114].

To assess the influence of replication in the distribution of chromatin damage, we analyzed the localization of bleomycin-induced γH2AX foci in relation to replication of eu- or heterochromatin interphase compartments in 5-ethynyl-2'deoxiuridine (EdU) pulsed-labeled CHO9 nuclei. Bleomycin (BLM) is a radiomimetic S-independent clastogen that induces oxidative damage, SSB and mainly DSB as well as a rapid phosphorylation of H2AX [115].

Asynchronously growing CHO9 cultures were pulse-exposed (30 min) to EdU (controls) or simultaneously (30 min) treated with BLM (40 μg/ml). Early and late replication regions and γH2AX foci were detected with an azide conjugated to Alexa Fluor 488 (Click-iT EdU, Invi-

trogen) and mouse anti-γH2AX (Abcam) followed by Cy3-conjugated antimouse antibodies, respectively. Single-cell z-stacks from control (n=25) and treated (n=63) nuclei were obtained by confocal microscopy and processed with Image J software. Using binary masks for each channel, the relation (ratio) between the percentage of damaged (γH2AX) area in replicating chromatin (EdU) area and the percentage of damaged area in the whole nuclear area (DAPI) was calculated for each nucleus. Finally, the arithmetic mean of the ratios corresponding to early S (n=30) and late S (n=33) nuclei was calculated.
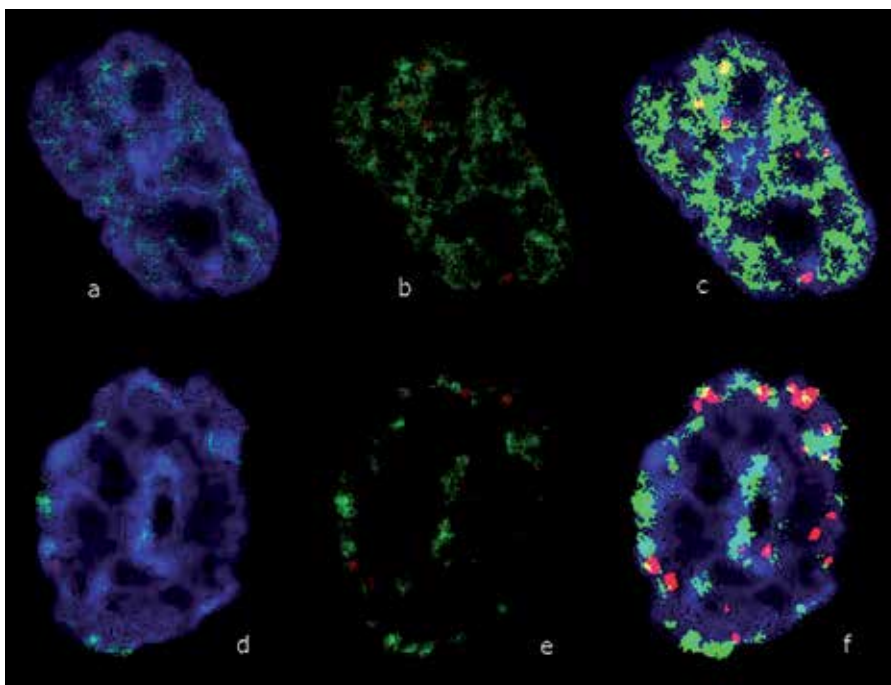


**Figure 7.** Distribution of BLM induced-γH2AX foci revealed by immunolabelling (Cy3; red) in early (top) or late (bottom) S-phase CHO9 nuclei. Replicating patterns were obtained by EdU incorporation and chemical detection (azide-Alexa Fluor 488; green). Nuclei were counterstained with DAPI (blue). Early S (a-c) and late S replicating nuclei are shown. Panels (a, d) and (d, e) contain DAPI/γH2AX/EdU and γH2AX/EdU merged images, respectively. Panels (c) and (f) illustrate binary masks of red (γH2AX) and green (EdU) channels overlaying the respective DAPI images.

Preliminary results (arithmetic mean of the ratios: 1.57 in early S- and 1.45 in late S-nuclei) suggest a bias in damage distribution towards replicating areas (~50 % higher than expected) probably due to local unwinding of chromatin down to naked DNA in both eu- and heterochromatin during DNA synthesis. Chromatin decondensation may increase the susceptibility to DNA damage as well as the accessibility of kinases that phosphorylate H2AX. Noteworthy, detailed visual analysis of fluorescent images or the corresponding binary masks in both early and late S-phase revealed that these results were not due to a large amount of γH2AX foci dwelling within replicating area and few of them outside. Instead, γH2AX foci recurrently mapped to the interfaces between replicating and non-replicating regions (Figure 7; Liddle P, unpublished observations).

The fact that in late-replicating cells γH2AX foci tend to map to the boundaries of replicating compartments (Figure 7, panels d-f) may be due to repositioning of damaged sites to less condensed peripheral heterochromatin regions, as it has been suggested in other models [112, 113]. However, this peculiar distribution of γH2AX foci in replicating/ non-replicating interfaces was also observed in early S-phase when the less compact euchromatin replicates (Figure 7, panels a-c). In this respect, BLM-induced DNA lesions could preferentially map at the damage-prone TTR located at the boundaries of early and late replicating compartments.

## 4. Conclusions

We assayed the influence of eu/heterochromatin replication timing in the distribution of chromatin induced damage using two different approaches: (1) the analysis of UV-C, MMS and etoposide-induced BP in Xp or Xq replicating CHO9 X mitotic chromosome and; (2) the analysis of primary BLM-induced damage (γH2AX foci) in CHO9 early and late replicating interphase nuclei. Our findings support the assumption that induced damage patterns shift according to eu- or heterochromatin replication. The asynchronic replication of eu- or heterochromatin compartments could influence the distribution of primary DNA lesions and CA, prevailing in replicating chromatin regions, irrespective of its eu- or heterochromatic state. Thus, eu/heterochromatin replication timing seems to play an overriding role in the production and localization of chromosome damage in S-phase cells.

## Acknowledgments

## Author details

María Vittoria Di Tomaso, Pablo Liddle, Laura Lafon-Hughes, Ana Laura Reyes-Ábalos and Gustavo Folle

*Address all correspondence to: marvi@iibce.edu.uy

Department of Genetics, Instituto de Investigaciones Biológicas Clemente Estable, Montevideo, Uruguay

# References

[1] Gilbert DM. Replication timing and transcriptional control: beyond cause and effect. Current Opinion in Cell Biology 2002;14 377-383.

[2] Gilbert DM, Gasser SM. Nuclear structure and DNA replication. In: DePamphilis ML. (ed.) DNA replication and human disease. New York: Cold Spring Harbor Laboratory Press; 2006. p175-196.

[3] Heitz E. Das Heterochromatin der Moose. Jahrbuch der Wissenschafltichen Botanik 1928;69 762-818

[4] Brown SW. Heterochromatin. Science 1966;151 417-425.

[5] Craig JM. Heterochromatin-many flavours, common themes. BioEssays 2004;27 17-28.

[6] Holmquist GP, Ashley T. Chromosome organization and chromatin modification: influence on genome function and evolution (Review). Cytogenetic and Genome Research 2006;114(2) 96-125.

[7] Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, Bickmore WA. Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. Cell 2004;118 555-566.

[8] Arrighi FE, Hsu TC. Localization of heterochromatin in human chromosomes. Cytogenetics and Cell Genetics 1971;10 81-86.

[9] Puck TT, Cieciura SJ, Robinson A. Genetics of somatic mammalian cells. III. Long term cultivation of euploid cells from human and animal subjects. The Journal of Experimental Medicine 1958;108 954-956.

[10] Drets ME, Shaw MW. Specific banding patterns of human chromosomes. Proceedings of the National Academy of Sciences of the United States of America 1971;68 2073-2077.

[11] Holmquist GP. DNA sequences in G-bands and R-bands. In: Adolph K W (ed.) Chromosomes and Chromatin. Boca Raton: CRC Press; 1988; p76-121.

[12] Saitoh Y, Laemmli UK. Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. Cell 1994;76 609-622.

[13] Korenberg JR, Rykowski MC. Human genome organization: Alu, Lines, and the molecular structure of metaphase chromosome bands. Cell 1988;53 391-400.

[14] Yasuhara JC, Wakimoto BT. Oxymoron no more: the expanding world of heterochromatic genes. Trends in Genetics 2006;22(6) 330-338.

[15] Dimitri P, Caizzi R, Giordano E, Accardo MC, Lattanzi G, Biamonti G. Constitutive heterochromatin: a surprising variety of expressed sequences. Chromosoma 2009;118 419-435.

[16] Sadoni N, Langer S, Fauth C, Bernardi G, Cremer T, Turner BM, Zink D. Nuclear organization of mammalian genomes: polar chromosome territories build up functionally distinct higher order compartments. The Journal of Cell Biology 1999;146 1211-1226.

[17] Jenuwein T, Allis CD. Translating the histone code. Science 2001;293 1074-1080.

[18] Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheime E. Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 2005;122 517-527.

[19] Kim JH, Workman JL. Histone acetylation in heterochromatin assembly. Genes and Development 2010;24 738-740.

[20] Ohlsson R, Renkawitz R, Lobanenkov V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends in Genetics 2001; 17 520-527.

[21] Woodfine K, Fiegler H, Beare DM, Collins JE, McCann OT, Young BD, Debernardi S, Mott R, Dunham I, Carter NP. Replication timing of the human genome. Human Molecular Genetics 2004;13 191-202.

[22] Gilbert DM. Evaluating genome-scale approaches to eukaryotic DNA replication. Nature Reviews Genetics 2010;11 673-684.

[23] Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, Dalton S, Gilbert D M. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Research 2010;20 761-770.

[24] Farkash-Amar S, Lipson SD, Polten A, Goren A, Helmstetter C, Yakhini Z,. Simon I. Global organization of replication time zones of the mouse genome. Genome Research 2008;18 1562-1570.

[25] Maya-Mendoza A, Petermann E, Gillespie DA, Caldecott KW, Jackson DA. Chk1 regulates the density of active replication origins during the vertebrate S phase. The EMBO Journal 2007;26 2719-2731.

[26] Labit H, Perewoska I, Germe T, Hyrien O, Marheineke K. DNA replication timing is deterministic at the level of chromosomal domains but stochastic at the level of replicons in Xenopus egg extracts. Nucleic Acids Research 2008;36 5623-5634.

[27] Ma H, Samarabandu J, Devdhar RS, Acharya R, Cheng PC, Meng C, Berezney R. Spatial and temporal dynamics of DNA replication sites in mammalian cells. The Journal of Cell Biology 1998;143 1415-1425.

[28] Jeon Y, Bekiranov S, Karnani N, Kapranov P, Ghosh S, MacAlpine D, Lee C, Hwang DS, Gingeras TR, Dutta A. Temporal profile of replication of human chromosomes. Proceedings of the National Academy of Sciences of the United States of America 2005;102 6419-6424.

[29] Hiratani I, Leskovar A, Gilbert DM. Differentiation-induced replication-timing changes are restricted to AT-rich/long interspersed nuclear element (LINE)-rich iso-

chores. Proceedings of the National Academy of Sciences of the United States of America 2004;101(48) 16861-16866.

[30] Subramanian PS, Chinault AC. Replication timing properties of the human HPRT locus on active, inactive and reactivated X chromosomes. Somatic Cell and Molecular Genetics 1997;23 97-109.

[31] Kim SM, Dubey DM, Huberman JA. Early-replicating heterochromatin. Genes and Development 2003;17 330-335.

[32] Wright WE, Tesmer VM, Liao ML, Shay JW. Normal human telomeres are not late replicating. Experimental Cell Research 1999;251 492-499.

[33] Holló G, Keresõ J, Praznovszky T, Cserpán I, Fodor K, Katona R, Csonka E, Fatyol K, Szeles A, Szalay AA, Hadlaczky G. Evidence for a megareplicon covering megabases of centromeric chromosome segments. Chromosome Research 1996;4 240-247.

[34] Schübeler D, Scalzo D, Kooperberg C, van Steensel B, Delrow J, Groudine MGenome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing. Nature Genetics 2002;32 438-442.

[35] Wutz A, Jaenisch R. A shift from reversible to irreversibleX inactivation is triggered during ES cell differentiation. Molecular Cell 2000;5 695-705.

[36] Dimitrova DS, Gilbert DM. The spatial position and replication timing of chromosomal domains are both established in early G1-phase. Molecular Cell 1999;4 983-993.

[37] Arney KL, Fisher AG. Epigenetic aspects of differentiation. Journal of Cell Science 2004;117 4355-4363.

[38] Zhou J, Ermakova OV, Riblet R, Birshtein BK, Schildkraut CL. Replication and subnuclear location dynamics of the immunoglobulin heavy-chain locus in B-lineage cells. Molecular Cell Biology 2002a;22 4876-4889.

[39] Taddei A, Van Houwe G, Hediger F, Kalck V, Cubizolles F, Schober H, Gasser SM. Nuclear pore association confers optimal expression levels for an inducible yeast gene. Nature 2006; 441 774-778.

[40] Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, Schelder M, Vermeulen M, Buscaino A, Duncan, K, Mueller J, Wilm M, Stunnenberg HG, Saumweber H, Akhtar A. Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. Molecular Cell 2006;21 811-823.

[41] Goldman MA, Holmquist GP, Gray MC, Caston LA, Nag A. Replication timing of mammalian genes and middle repetitive sequences. Science 1984;224 686-692.

[42] Craig JM, BickmoreWA. Chromosome bands-flavours to savour. BioEssays 1993;15 349-354.

[43] Takebayashi S, K. Sugimura, T. Saito, C. Sato, Y. Fukushima, H. Taguchi, and Okumura K. Regulation of replication at the R/G chromosomal band boundary and peri-

centromeric heterochromatin of mammalian cells. Experimental Cell Research 2005;304 162-174.

[44]  Méndez J. Temporal regulation of DNA replication in mammalian cells. Critical Reviews in Biochemistry and Molecular Biology 2009;44 343-351.

[45]  McNairn AJ, Gilbert DM. Epigenomic replication: linking epigenetics to DNA replication. BioEssays 2003;25 647-656.

[46]  Casas-Delucchi CS, van Bemmel JG, Haase S, Herce HD, Nowak D, Meilinger D, Stear JH, Leonhardt H, Cardoso MC. Histone hypoacetylation is required to maintain late replication timing of constitutive heterochromatin. Nucleic Acids Research 2011;10 1-11.

[47]  Unnikrishnan A,. Gafken PR, Tsukiyama T. Dynamic changes in histone acetylation regulate origins of DNA replication. Nature Structural and Molecular Biology 2010;17 430–437.

[48]  Bickmore WA, Carothers AD. Factors affecting the timing and imprinting of replication on a mammalian chromosome. Journal of Cell Science 1995;108 2801-2809.

[49]  Craig JM, Wong LH, Lo AWI, Earle E, Choo KHA. Centromeric chromatin pliability and memory at a human neocentromere. The EMBO Journal 2003;12 3109-3121.

[50]  Cowell IG, Aucott R, Mahadevaiah Sk, Burgoyne PS, Huskisson N, Bongorini S, Prantera G, Fanti L, Pimpinelli S, Wu R, Gilbert DM, Shi W, Fundele R, Morrison H, Jeppesen P, Singh PB. Heterochromatin, HP1 and methylation at lysine 9 of histone H3 in animals. Chromosoma 2002;111 22-36.

[51]  Hediger F, Gasser SM. Heterochromatin protein 1: don't judge the book by its cover! Current Opinion in Cell Biology 2006;16 143-150.

[52]  Hansen RS, Stoger R, Wijmenga C, Stanek AM, Canfield TK, et al. 2000 Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. Human Molecular Genetics 2000;9 2575-2587.

[53]  Rhind N, Yang SC, Bechhoefer J. Reconciling stochastic origin firing with defined replication timing. Chromosome Research 2010;18 35-43.

[54]  Yang, S. C., N. Rhind, and J. Bechhoefer. Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. Molecular Systems Biology 2010;6:404.

[55]  Smith ZE, Higgs DR. The pattern of replication at a human telomeric region (16p13.3): its relationship to chromosome structure and gene expression. Human Molecular Genetics 1999;8 1373-86.

[56]  Costanzo V, Shechter D, Lupardus PJ, Cimprich KA, Gottesman M, Gautier J. An ATR- and Cdc7-dependent DNA damage checkpoint that inhibits initiation of DNA replication. Molelular Cell 2003;11 203-213.

[57] Herrick J. Genetic variation and DNA replication timing, or why is there late replicat-ing DNA? Evolution 2011;65 3031-3047.

[58] Niida H, Katsuno Y, Banerjee B, Hande MP, Nakanishi M. Specific role of Chk1 phos-phorylations in cell survival and checkpoint activation. Molecular Cell Biology 2007; 27 2572-2581.

[59] Karnani N, Dutta A. The effect of the intra-S-phase checkpoint on origins of replica-tion in human cells. Genes and Development 2011;25 621-633.

[60] Machida YJ, Hamlin JL, Dutta A. Right place, right time, and only once: replication initiation in metazoans. Cell 2005;123 13-24.

[61] Loupart ML, Krause SA, Heck MMS. Aberrant replication timing induces defective chromosome condensation in Drosophila ORC2 mutants. Current Biology 2000;10 1547-1556.

[62] Pflumm MF. The role of DNA replication in chromosome condensation. BioEssays 2002;24 411-418.

[63] Sasaki T, Gilbert DM. The many faces of the origin recognition complex. Current Opinion in Cell Biology 2007;19 337-343.

[64] Gardner RJM, Sutherland GR. Chromosome Abnormalities and Genetic Counseling. New York: Oxford University Press; 1966.

[65] Mitelman F. Patterns of chromosome variation in neoplasia. In: Obe G, Natarajan AT (ed.) Chromosomal Aberrations: Basic and Applied Aspects. Berlin: Springer-Verlag; 1990; p 86-100.

[66] De Braekeleer M. Fragile sites and chromosomal structural rearrangements in human leukemia and cancer. Anticancer Research 1987;7 417-422.

[67] Yunis JJ, Soreng AL, Bowe AE. Fragile sites are target of diverse mutagens and carci-nogens. Oncogene 1987;1 59-69.

[68] Hecht F. Fragile sites, cancer chromosome breakpoints and oncogenes all cluster in light G bands. Cancer Genetics and Cytogenetics 1988; 31 17-24.

[69] Barrios L, Miró R, Caballín MR, Fuster C, Guedea F, Subias A, Egozcue J. Cytogenet-ic. effects of radiotherapy breakpoint distribution in induced chromosome aberra-tions. Cancer Genetics and Cytogenetics 1989;41 61-70.

[70] Porfirio B, Tedeschi B, Vernole P, Caporossi D, Nicoletti B The distribution of Msp I-induced breaks in human lymphocyte chromosomes and its relationship to common fragile sites. Mutation Research 1989;213 117-124.

[71] Tedeschi B, Porfirio B, Caporossi D, Vernole P, Nicoletti B. Structural chromosomal rearrangements in Hpa II-treated human lymphocytes. Mutation Research 1991;248 115-121.

[72] Folle G, Liddle P, Lafon-Hughes L, Di Tomaso M. Close encounters: RIDGEs, hyperacetylated chromatin, radiation breakpoints and genes differentially expressed in tumors cluster at specific human chromosome regions. Cytogenetics and Genome Research 2010; 128 17-27.

[73] Gollin SM. Mechanisms leading to chromosomal instability. Seminars in Cancer Biology 2005;15 33-42.

[74] Folle GA. Nuclear architecture, chromosome domains and genetic damage. Mutation Research-Reviews in Mutation Research 2008;658 172-183.

[75] Ermakova OV, Nguyen LH, Little RD, Chevillard C, Riblet R, Ashouian N, Birshtein BK, Schildkraut CL. Evidence that a single replication fork proceeds from early to late replicating domains in the IgH locus in a Non-B cell line. Molecular Cell 1999;3 321-330.

[76] Durkin S, Glover T. Chromosome fragile sites. Annual Review of Genetics 2007;41 169-192.

[77] Letessier A, Millot GA, Koundrioukoff S, Lachagès AM, Vogt N, Hansen RS, Malfoy B, Brison O, Debatisse M. Cell-types pecific replication initiation programs set fragility of the FRA3B fragile site. Nature 2011;470 120-123.

[78] Flynn KM, Vohr SH, Hatcher PJ, Cooper VS. Evolutionary rates and gene dispensability associate with replication timing in the archaeon Sulfolobus islandicus. Genome Biology and Evolution 2010;2 859-869.

[79] Mugal CF,. Wolf JB, von Grünberg HH, Ellegren H. Conservation of neutral substitution rate and substitutional asymmetries in mammalian genes. Genome Biology and Evolution 2010;2 19-28.

[80] Kaufman DG, Cohen SM, Chastain PD. Temporal and functional analysis of DNA replicated in early S phase. Advances in Enzyme Regulation 2011;51 257-271.

[81] Elango N, Kim SH, Vigoda E, Yi SV. Mutations of different molecular origins exhibit contrasting patterns of regional substitution rate variation. PLoS Computational Biology 2008;4(2) e1000015. doi:10.1371/journal.pcbi.1000015

[82] Walser JC, Furano AV. The mutational spectrum of non-CpG DNA varies with CpG content. Genome Research 2010;20 875-882

[83] Nyberg KA, Michelson RJ, Putnam CW, Weinert TA. DNA damage and replication checkpoints. Annual Review of Genetics 2002;36 617-656

[84] Di Tomaso MV, Martínez-López W, Méndez- Acuña L, Lafon-Hughes L, Folle GA. Factors leading to the induction and conversion of DNA damage into structural chromosomal aberrations. In: Miura S, Nakano S (ed.). Progress in DNA Damage Research. New York: Nova Publisher; 2008; p30-40.

[85] Obe G, Johannes C, Schulte-Frohlinde D. DNA double-strand breaks induced by sparsely ionizing radiation and endonucleases as critical lesions for cell death, chro-

mosomal aberrations, mutations and oncogenic transformation. Mutagenesis 1992;7
3-12.

[86]  Champoux JJ. DNA topoisomerases: structure, function, and mechanism. Annual Review of Biochemistry 2001;70 369-413.

[87]  Palitti F, Mosesso P, Di Chiara D, Schinoppi A, Fiore M, Bassi L. Use of antitopoisomerase drugs to study the mechanisms of induction of chromosomal damage. In: Chromosomal Alterations: Origin and Significance. Obe G, Natarajan AT (ed.) Berlin: Springer-Verlag; 1994; p103-115.

[88]  Degrassi F, Fiore M, Palitti F. Chromosomal aberrations and genomic instability induced by topoisomerase-targeted antitumor drugs. Current Medicinal Chemistry-Anti-Cancer Agents 2004;4 317-325.

[89]  Ray M, Mohandas T. Proposed banding nomenclature for the Chinese hamster chromosomes (Cricetulus griseus). In: Report of the Committee on Chromosome Markers. Hamerton JL (ed.) Cytogenetics and Cell Genetics 1976;16 83-91.

[90]  Schmid W, Leppert MF. Rates of DNA synthesis in heterochromatic and euchromatic segments of the chromosome complements of two rodents. Cytogenetics 1969;8 125-135.

[91]  Di Tomaso MV, Martínez-López W, Folle GA, Palitti F. Modulation of chromosome damage localisation by DNA replication timing. International Journal of Radiation Biology 2006;82 877-886.

[92]  Di Tomaso MV, Martínez-López W, Palitti F. Asynchronously replicating eu/heterochromatic regions shape chromosome damage Cytogenetics and Genome Research 2010;1281(1-3) 111-117.

[93]  De Boer J, Hoeijmakers JH. Nucleotide excision repair and human syndromes. Carcinogenesis 2000;21 453-460.

[94]  García CL, Carloni M, de la Pena NP, Fonti E, Palitti F. Detection of DNA primary damage by premature chromosome condensation in human peripheral blood lymphocytes treated with methylmethanesulfonate. Mutagenesis 2001;16 121-125.

[95]  Christmann M, Roos WP, Kaina B. DNA methylation damage formation, repair and biological consequences. In Obe G, Vijayalaxmi, (ed.) Chromosomal Alterations: Methods, Results and Importance in Human Health. Heidelberg: Springer Verlag; 2007; p99-121.

[96]  Bender MA, Griggs HG, Walker PL. Mechanisms of chromosomal aberration production. I. Aberration induction by ultraviolet light. Mutation Research 1973;20 387-402.

[97]  Natarajan AT, Obe G, van Zeeland AA, Palitti F, Meijers M, Vergaal-Immerzeel EAM. Molecular mechanisms involved in the production of chromosomal aberrations. II. Utilization of [96]Neurospora endonuclease for the study of aberration pro-

duction by X-rays in G1 and G2 stages of the cell cycle. Mutation Research 1980;69 293- 305.

[98]   Bradley M. Double-strand breaks in DNA caused by repair of damage due to ultra-violet light. Journal of Supramolecular Structure and Cellular Biochemistry 1981;(16) 337-343.

[99]   Rogakou EP, Boon C, Redon C, Bonner WM. Megabase chromatin domains involved in DNA double-strand breaks in vivo. The Journal of Cell Biology 1999;146(5) 905-916.

[100]   Li L, Zou L. Sensing, signaling, and responding to DNA damage: organization of the checkpoint pathways in mammalian cells. Journal of Cellular Biochemistry 2005;4(2) 298-306.

[101]   Marti TM, Hefner E, Feeney L, Natale V, Cleaver JE. H2AX phosphorylation within the G1phase after UV irradiation depends on nucleotide excision repair and not DNA double-strand breaks. Proceedings of the National Academy of Sciences of the United States of America 2006;103 9891-9896.

[102]   Hanasoge S, Ljungman M. H2AX phosphorylation after UV irradiation is triggered by DNA repair intermediates and is mediated by the ATR kinase. Carcinogenesis 2007;28 2298-2304.

[103]   Huyen Y, Zgheib O, Ditullio RA Jr, Gorgoulis VG, Zacharatos P, Petty TJ. Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature 2004;432 406-411.

[104]   Botuyan MV, Lee J, Ward IM, Kim JE, Thompson JR, Chen J, Mer G. Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell 2006;7 1361-1373.

[105]   Noon AT, Shibata A, Rief N, Löbrich M, Stewart GS, Jeggo PA, Goodarzi AA. 53BP1-dependent robust localized KAP-1 phosphorylation is essential for heterochromatic DNA double-strand break repair. Nature Cell Biology 2010;12 177-184.

[106]   Cowell IG, Sunter NJ, Singh PB, Austin CA, Durkacz BW, Tilby MJ. GammaH2AX foci form preferentially in euchromatin after ionising-radiation. PLoS One 2007;2 e1057.

[107]   Kim JA, Kruhlak M, Dotiwala F, Nussenzweig A, Haber J E. Heterochromatin is re-fractory to $\gamma$-H2AX modification in yeast and mammals. The Journal of Cell Biology 2007;178(29) 209-218.

[108]   Falk M, Lukásová E, Kozubek S. Chromatin structure influences the sensititivty of DNA to $\gamma$-radiation. Biochimica et Biophysica Acta 2008;1783 2398-2414.

[109]   Vasireddy RS, Karagiannis TC, Assam EO. -radiation-induced ⊚H2AX formation oc-curs preferentially in actively transcribing euchromatic loci. Cellular and Molecular Life Sciences 2010;67 291-294.

[110] Falk M, Lukásová E, Kozubek S. Higher-order chromatin structure in DSB induction, repair and misrepair. Mutation Research 2010;704 88-100.

[111] Ziv Y, Bielopolski D, Galanty Y, Lukas C, Taya Y, Schultz DC, Lukas J, Bekker-Jensen S, Bartek J, ShilohY. Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. Nature Cell Biology 2006;8 870-876.

[112] Kruhlak MJ, Celeste A, Dellaire G, Fernandez-Capetillo O, Muller WG, et al. (2006) Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks. The Journal of Cell Biology 2006;172 823-834.

[113] Cann K L, Dellaire G. Heterochromatin and the DNA damage response: the need to relax1. Biochemistry and Cell Biology 2011;89 45-60.

[114] Jakob B, Splinter J, Conrad S, Voss KO, Zink D, Durante M, Löbrich M, Taucher-Scholz1 G. DNA double-strand breaks in heterochromatin elicit fast repair protein recruitment, histone H2AX phosphorylation and relocation to euchromatin. Nucleic Acids Research 2011;39(15) 6489-6499.

[115] Chen J, Stubbe J Bleomycins: Toward better therapeutics. Nature Reviews Cancer 005;5 102-112.

# A Histone Cycle

Douglas Maya, Macarena Morillo-Huesca,
Lidia Delgado Ramos, Sebastián Chávez and
Mari-Cruz Muñoz-Centeno

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/53993

## 1. Introduction

Each time a cell divides it must duplicate its DNA content and segregate it equally between two daughter cells. Once a cell has decided to replicate its DNA, hundreds of different proteins must carefully interact with each other in a very orchestrated process. Defects in any of these steps can lead to cell death and genetic instability and have been shown to be present in many human diseases including cancer [1]. Each step of DNA replication must take place in a correct spatial and temporal window, so cells have evolved complex regulatory networks allowing an efficient regulation of this process. One important feature of eukaryotic cells is that DNA is strongly associated with histones, basic proteins that wrap DNA around octameric structures called nucleosomes. The association of DNA and nucleosomes is commonly known as chromatin.

Nucleosomes are, among others, one of the principal differences between eukaryotic and prokaryotic DNA. Unlike bacteria, eukaryotic cells are not able to live without DNA packed into chromatin [2]. Replication involves dramatic changes in the whole chromatin landscape, since nucleosomes must be removed transiently from the front of the replication machinery and repositioned after it. Nucleosome disassembly and assembly involves the action of chromatin remodeling factors, proteins able to destabilize interactions between histones and DNA allowing the interaction of other complexes with DNA. Restoration of chromatin after the replication bubble is a very important step because nucleosomes are not repetitive units of information and contain a specific epigenetic signature or code [3]. In order to ensure enough histones for the nascent DNA, cells must increase the pool of free histones. In human cells, each passage through S-phase requires the synthesis and assembly of almost 30 million nucleosomes that are synthesized mainly during S-phase and are rapidly packaged

to DNA. Histone production is very tightly coupled to DNA synthesis and is rapidly shut off when replication finishes or is halted by treatment with mutagenic agents. Histone regulation is very important and accumulation of free histones in the cell has been shown to be highly deleterious for the cell and lead to chromosome loss [2].

Chromatin is a structural barrier for replication but can also play an important role in the regulation of some of the steps within. In this chapter, we will focus on how the chromatin landscape influences DNA replication and show that histones and DNA must adapt to each other in order to ensure a correct genomic duplication. We will describe the influence that chromatin plays at the different stages of DNA replication and then jump to the accurate control that cells exert on histone levels during the cell cycle. We will finally show different situations that uncouple DNA replication from histone deposition and synthesis, and discuss if chromatin state can influence the decision of cells to replicate their DNA or not.

# 2. Replication from a chromatin point of view

## 2.1. Chromatin influences early steps of replication

The initiation of DNA replication in any organism requires a series of proteins able to recruit and ultimately load two hexameric DNA helicases. These proteins are able to unwind DNA, a process required to start replication. In eukaryotic cells the pre-initiation complex is formed by two MCM2-7 rings that are loaded in an inactive form next to the Origin Recognition Complex (ORC). The MCM helicase must be activated by the sequential action of Dbf4 Dependent Kinase (DDK) and Cyclin Dependent Kinase (CDK), and the addition of other accessory proteins. In mammalian cells, 30,000 to 50,000 origins are sequentially activated each time a cell divides [4].

The nature of the sequence and the structure of replication origins is still a matter of debate, and most higher eukaryotes lack a specific consensus sequence for ORC binding. Origins of DNA Replication Initiation (ORIs) are normally regions of DNA sequence rich in AT that contain a nucleosome-free region (NFR) [5] [6] and it has been suggested that the chromatin environment is important for the establishment of the ORC complex [7] [8]. In *Drosophila* follicle cells histones that localize around ORIs are hyperacetylated and changes in the acetylation levels of these histones affects ORC binding [9].

As ilustrated in figure 1, methylation of histone H4 has also been shown to be important for ORC recruitment and artificial tethering of the methyltransferase SET8 to a random locus promotes ORC1 binding [10]. Once ORC is bound to DNA, proteins CDC6 and CDT1 help to load the two MCM2-7 helicases to DNA [11]. Loading could also be influenced by the acetylation of histone H4, since CDT1 is able to recruit the histone acetyltransferase HBO1 to the ORC and enhances the recruitment of MCM2-7 to the origin [12, 13] [14]. Nevertheless, some ORC and MCM subunits are acetylated by Hbo1 in yeast and could therefore be the real targets of this enzyme [14]. Although all origins that are selected are able to load the pre-initiation complex, only one out of every ten will fire. Regulation of firing depends on

the activation of the MCM helicases by sequential phosphorylation of some of its subunits by DDK and CDK kinases, that allow the recruitment of CDC45 and the GINS complex [15]. Once these proteins are loaded, the rest of the replication machinery is recruited and replication starts.
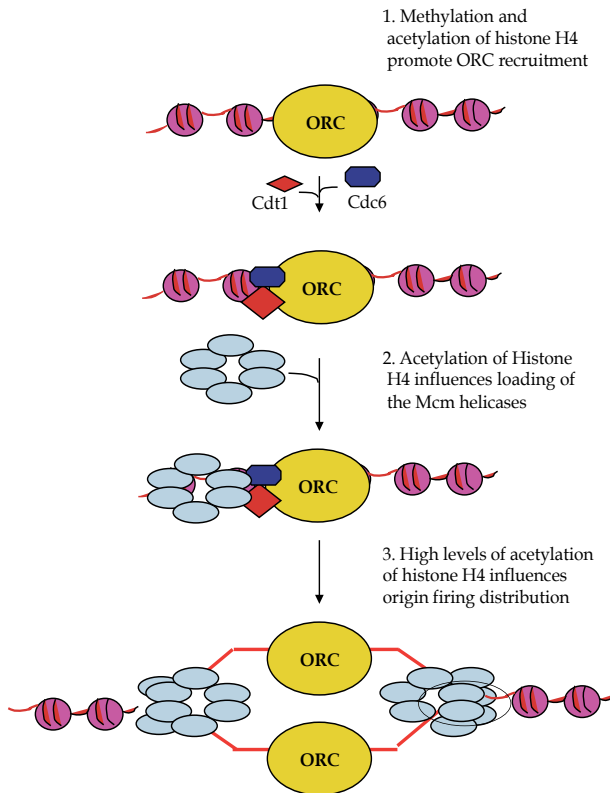


**Figure 1. Chromatin influences DNA replication fork establishment.** Schematic representation of the different steps during the assembly and activation of the replication fork machinery that are influenced by histone modifications. 1.0RC recruitment is influenced by methylation and acetylation levels of histone I-I4. 2.Acetylation of histone H4 by the Hbo1 might influence loading of the two Mcm2-7 helicases. 3. H4K 16 aoetytation is related to the timing of origin firing distribution.

Two interesting observations have lead to the hypothesis that chromosome architecture can also be important for origin usage. One of them is that origins seem to be organized into clusters of 5 to 10 origins that fire simultaneously [18]. Cohesins are enriched next to origins and depletion of the RAD21 cohesin subunit greatly reduces the number of active origins. This ring-like complex is able to wrap two chromatin fibers and creates a chromatin loop. It has been suggested that this spatial organization of chromatin could define replication domains that are activated synchronously [16, 17]. In agreement with this hypothesis, analysis of the oriGNA13 replication origin of hamster cells shows that active origins localize close to the base of chromatin loops [18]. The second feature that directly relates chromatin structure

and replication is that genome replication does not take place in a single and continuous way. Domains containing several megabases of contiguous DNA are replicated earlier than others [19] and this replication timing is somehow correlated with acetylation levels of histone H4 at the K16 residue [20].

Chromatin influences the recruitment and the activity of different elements of the replication machinery. Once this machinery has been set up and is fully active, fork progression must now cope with the fact that approximately every 192 bp there is a nucleosome that must be displaced from DNA in order to continue with replication.

## 2.2. Nucleosome reorganization around the replication fork

Replication fork progression involves many proteins that interact closely with DNA. Electron micrography of replicating SV40 mini-chromosomes has shown that 300bp ahead of the replication fork, DNA remains *naked*, or at least contains nucleosomes that are unstable when compared to a canonical nucleosome [21] [22]. MCM progression in mammalian cells suggests that chromatin is decondensed in front of the replication fork [23] and artificial tethering of Cdc45 to DNA is able to promote a partial decondensation of chromatin without DNA synthesis [24]. This initial decondensation could be related to an increase in the mobility of histone H1 due to its phosphorylation by the cyclin A-CDK2 complex. It is still unclear if nucleosome disruption in front of the replication fork is due to specific chromatin remodeling in front of the fork or to the passage of the replication machinery itself [3].

Nucleosome disassembly and reassembly are processes quite well described for transcription. The efficiency of these processes is largely dependent on chromatin-remodeling complexes, proteins able to interact with and change the stability of chromatin, allowing the transcription machinery to interact with DNA. There are many different chromatin-remodeling complexes and all of them are possible candidates for nucleosome eviction during replication. The fact that chromatin disassembly and assembly occur in such a small spatial window makes it very difficult to distinguish between the complexes required for one or the other process. There are two major complexes that could be involved in H2A/H2B displacement during replication: FACT and NAP1, and another two for H3/H4: Asf1 and CAF1.

The FACT complex is composed of two main subunits, SPT16 and SSRP1 (Pob3 in *S.cerevisiae*), and plays a key role in nucleosome reorganization during transcription elongation. FACT function has been mostly related to the displacement of an H2A/H2B dimer during transcription, but it has also been proposed that displacement could be an indirect effect of nucleosome reorganization by this complex [25, 26]. There are many different items of evidence suggesting that FACT plays a role as a histone chaperone during DNA replication. FACT is required for DNA replication in *Xenopus* extracts, is present at human replication origins [27] and has been co-purified as part of the replication fork progression complex in yeast [28]. The other H2A/H2B histone chaperone candidate is Nap1, which has been shown to disassemble nucleosomes in vitro when combined with the RSC complex [29, 30]. Once H2A/H2B dimers are displaced, the H3/H4 tetramer is now accessible for an H3/H4 chaperone.

To date, it is not known if H2A/H2B dimers removed during replication are recycled. On the contrary, it is well established that the original H3/H4 tetramer present in front of the replication fork machinery is restored after the replication fork in a random semi-conservative process [31]. The fact that the H3/H4 tetramer is recycled suggests that a histone chaperone must disassemble this tetramer in front of the replication machinery and reassemble it after. One good candidate for this process is Asf1. This protein along with Chromatin Assembly Factor 1 (CAF1) plays a key role in deposition of new H3/H4 following passage of the replication fork. Asf1 binds PCNA and replication factor C [32], and can also bind the MCM helicase complex through histones H3 and H4. Upon a replication fork progression block, Asf1 can be found associated with post-translationally modified H3/H4 histones, which most likely belong to the parental chromatin [33].
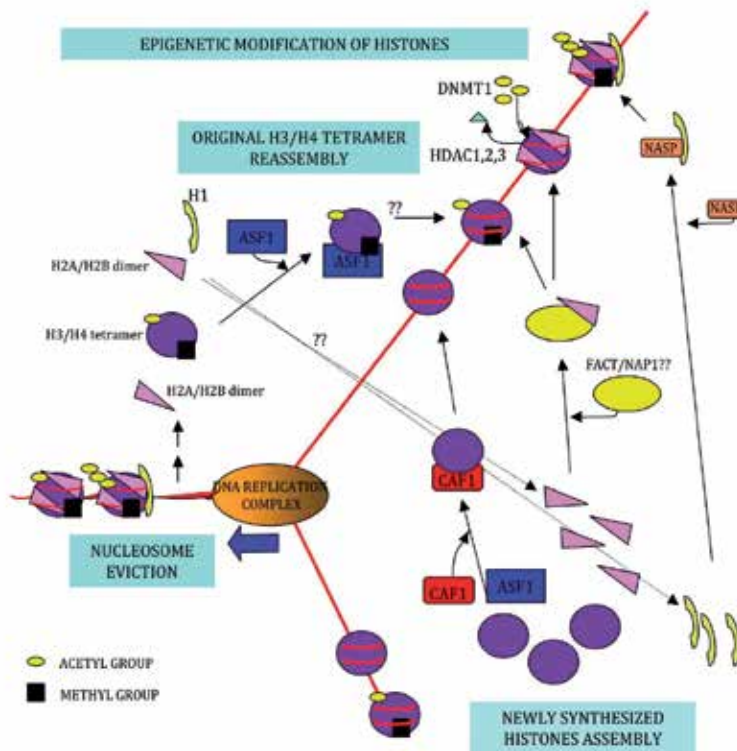


**Figure 2. Nucleosome reorganization around the replication fork**. Representation of the different nucleosome reorganization events that take place during replication fork progression. In order to simplify the figure, the DNA replication machinery and other accessory proteins that are important during fork progression are not shown. Interrogation marks are used when the protein/s involved in such process remain unknown or when the pathway has not been directly demonstrated.

Deposition of histone octamers occurs as soon as DNA is long enough to wrap around a nucleosome [21]. Since one H3/H4 tetramer is recycled after DNA replication, one new H3/H4 tetramer must be deposited on the other strand. The mechanism of reposition of the original

H3/H4 tetramer remains unclear but probably involves Asf1 (see previous paragraph). Incorporation of the new H3/H4 tetramer on the other hand is more defined and involves the action of CAF1 and Asf1. CAF1 is recruited to both leading and lagging strands by the proliferating cell nuclear antigen PCNA. Depletion of CAF1 produces a clear decrease in the assembly of new chromatin during replication [34], activates the DNA Damage Checkpoint (DDR), and stalls replication forks [35], suggesting that efficient chromatin repositioning after replication is important for replication fork progression. Asf1 plays a role in this process as a histone pool protein that delivers H3/H4 dimers to CAF1. After the H3/H4 tetramer is assembled, H2A/H2B becomes incorporated into chromatin in a process that probably involves FACT or NAP1. Finally, H1 protein is incorporated to allow further compacting of chromatin. Incorporation of H1 is probably mediated by the NASP protein and is required for efficient S-phase progression [36].

One interesting feature recently described is that the chromatin landscape influences the length of the Okazaki fragments synthesized at the lagging strand during DNA replication [37]. Due to the 5′to 3′ polarity of DNA polymerase, synthesis of DNA in the lagging strand is discontinuous and generates short fragments of DNA named Okazaki fragments. These fragments must then join to form a unique DNA strand in a process known as maturation. Okazaki fragment maturation requires the sequential action of the flap endonuclease 1 (FEN1) and DNA ligase I. This group has demonstrated that the ligation junctions of Okazaki fragments are preferentially located in the nucleosome midpoint. The length of Okazaki fragments depends on the chromatin behind the replication fork and mutations that impair chromatin repositioning increase the average size of Okazaki fragments. According to their model, Pol · runs into the nucleosome assembled into the previous Okazaki fragment and this triggers termination, flap processing and ligation.

### 2.3. Chromatin maturation and centromere formation

After nucleosome incorporation to DNA, two major processes must take place: chromatin maturation and centromere formation. Histones start to acquire certain modifications in their tails as soon as they are repositioned to DNA. Nascent chromatin is highly acetylated and must be deacetylated and methylated to reach a more compact state. Deacetylation normally takes place by the histone deacetylases HDAC1-3 and DNA methylation by the DNA methyltransferase 1 (DNMT1). In addition to chromatin compacting, there are also some specific post-transcriptional modifications that must be acquired to establish a specific epigenetic code that is transmitted to daughter cells. Maintaining this "epigenetic memory" of daughter cells is important and has implications during cell differentiation (23). Restoration of all these marks does not take place exclusively in replication and can also take place during mitosis or even in daughter cells [38, 39]. Replication of the chromatin near the centromere is also vital to ensure an efficient segregation during mitosis. This heterochromatin presents a specific variant of histone H3 named CENP-A, which is essential for the efficient binding of the kinetochore during mitosis. The kinetochore is a huge structure that attaches to centromeric DNA and mediates the interaction of chromosomes with the mitotic spindle

and their movement to the spindle poles during mitosis [40]. Accurate segregation of chromosomes relies on the correct formation of the spindle apparatus.

CENP-A (also known as CENH3) is an essential protein that replaces histone H3 at centromeric DNA. This protein is highly divergent among different species but is functionally well conserved since the homologue protein of *S. cerevisiae*, Cse4 is able to complement human cells lacking CENP-A or *vice versa* [41]. In human cells, CENP-A is not assembled on to DNA just after DNA replication and CENP-A-containing nucleosomes are interspersed with canonical nucleosomes during replication of centromeres [42, 43]. This organization promotes the folding of centromeric chromatin into a unique structure during metaphase, in which all the nucleosomes containing CENP-A are facing the external side of chromosomes. This structure allows kinetochore assembly and microtubule attachment and favours sister chromatid cohesion [44]. Once chromosomes are separated, CENH3 is fully positioned on centromeric chromatin during the period between telophase and G1 in a process that is dependent on the transient incorporation of Mis18 and KNL2 in anaphase [45]. The incorporation of CENP-A to centromeric chromatin is mediated at least in part by the HJURP protein (Scm3 in *S. cerevisiae*) and is related to low levels of acetylation of the K16 residue of histone H4. Defects in the proper incorporation of this histone variant can lead to cell death, genetic instability and chromosome loss [46-48]. There is also a subset of specific proteins important to prevent the deposition of CENH3 containing nucleosomes out of the centromeric DNA. In *S. cerevisiae*, the ubiquitin E3 ligase Psh1 prevents the spread of Cse4 containing nucleosomes out of the centromere [49] [50]. The absence of both CAF1 and HIRA also leads to the presence of this type of nucleosomes in euchromatic regions in both *S. cerevisiae* and *S. pombe* and has been shown to cause genetic instability [51]. Finally, several papers point out that a proper homeostasis between H3 and CENH3 histones is important for the distribution of this centromeric variant and for efficient chromosome segregation [52, 53].

## 3. From gene to protein, histones are highly regulated

It is clear that there is a strong interdependency between DNA replication and chromatin reorganization. Nucleosomes are more than structural bricks for DNA, and require the modification of specific residues or the substitution of certain histone variants for others to maintain a correct epigenetic state. Once cells have decided to replicate their DNA, an increase in the abundance of histone proteins is required to pack the new genome that is about to be generated. Histone genes are among the most highly cell-cycle-regulated genes [54] because cells need to ensure a high demand of histones during replication, but must make sure that these levels are quickly down-regulated when replication slows down or is blocked, to avoid the deleterious effects of free histones on cell survival.

Canonical histone proteins can be regulated at transcriptional, post-transcriptional, translational and post-translational levels. The importance of each pathway on histone metabolism largely depends on the organism. In *S. cerevisiae* for example, the transcriptional regulation has been traditionally shown to play a major role in histone regulation, while in mammalian

cells, post-transcriptional and translational mechanisms seem to be more important. Nevertheless, it is clear that all organisms try to produce histones exclusively during the replicative S-phase and more specifically only when replication is actively taking place.

### 3.1. Histones are regulated from the beginning: transcriptional regulation

Histone transcription is tightly regulated during the cell cycle. In some organisms like *S. cerevisiae*, transcription of histones can only be detected in late G1 and during DNA replication [55]. In higher eukaryotes, however, histones mRNAs can be found at all stages but increase as cells enter S-phase [56]. Expression of all canonical histones must be stoichiometric and several studies show that an imbalance between the different histone subtypes can be highly deleterious for the cell [52, 53, 57].

In metazoans, entry into S-phase increases the expression of replication-dependent histone genes three to five-fold [58]. Histone genes are clustered and each cluster normally contains at least one copy of the five canonical histones. Although transcription of all histones is carefully coordinated, no obvious common sequence element has been found at their promoters. Nevertheless, common elements can be found for some particular histone variants, like the Octamer-binding Transcription Factor (OTF1) for H2B promoters [59] or the Coding Region Activating Sequences (CRAS) in H2A, H3 and H4 genes [60]. Activation of histone gene transcription requires the Nuclear Protein Ataxia-Telangiectasia (NPAT), which is essential for S-phase progression [61]. This protein normally locates next to the *Histone locus bodies* and is phosphorylated by cyclin E-CDK2 at the beginning of S-Phase. Phosphorylation persists through S-phase and increases histone gene transcription [62, 63].

Transcriptional regulation of histones in *S. cerevisiae* is largely dependent on the integrity of the HIR complex. This complex is conserved from yeast to humans and has been shown to play a role in both of them in replication-independent chromatin assembly. In yeast, this complex is composed of the three subunits Hir1-3 and Hpc2. Deletion of any of the subunits leads to a de-repression of histones outside of S-phase [64]. Histone genes are grouped in 4 clusters, and each of them express simultaneously H2A/H2B or H3/H4 from a bidirectional promoter. These promoters contain upstream activating sequences (UAS) required for the recruitment of activators such as Spt10 and SBF [65]. Three of these four clusters also contain a negative regulatory site (NEG or CCR) able to maintain these genes in a repressed state in cell cycle phases outside of late G1 and S-phase and under replication stress conditions [64, 66, 67]. Deletion of the negative regulatory site is able to de-repress the *HTA1-HTB1* histone locus and allow expression outside of S-phase. The mechanism of repression is not completely understood but involves changes at the chromatin structure creating a repressive chromatin, which depends on the HIR complex, proteins Rtt106, Yta7 and Asf1, and the chromatin remodelling complex RSC. Two recent reports coming from the same group have shed some light on how repressive chromatin switches to active chromatin (in terms of transcription). The first one involves the degradation of Yta7 mediated by a phosphorylation event that involves Casein Kinase II (CKAII) and the cyclin-dependent kinase Cdc28. Degradation of Yta7 allows the efficient expression of histone mRNAs during S-phase

through a mechanism that could involve transcription elongation efficiency [68]. The second report is related to the cell cycle regulation of Spt21, an activator of histone gene expression [69]. Spt21 protein levels outside S-phase are regulated by proteolysis, in a mechanism that depends on a complex formed by the Anaphase Promoting Complex (APC) with Cdh1 during G1, and on APC-Cdc20 during G2 and M (Brenda Andrews, EMBO transcription meeting 2012).

It has been recently shown that the HIR complex is conserved through evolution [70, 71] [55]. In humans, this complex is formed by three proteins: HIRA, Ubinuclein1 and Cabin1. The role described for HIRA in humans has been mostly associated with chromatin assembly of the transcriptional histone variant H3.3 in cooperation with ASF1 [72]. Nevertheless, several studies suggest that this complex could also play an important role in metazoan histone regulation. Ectopic over-expression of HIRA is able to repress histone gene transcription and block S-phase progression in human cells. This protein localizes to histone gene clusters in an immunofluorescence essay [73]. Cyclin E-CDK2 and cyclin A-CDK2 can phosphorylate HIRA, and this phosphorylation is inhibited by cyclin inhibitor p21, which has been shown to be important for repression of histone synthesis upon replication stress [74] [69]. HIRA could therefore be acting as a repressor of histone gene expression outside of S-phase regulated through phosphorylation by the cyclin E-CDK2. In this model, phosphorylation by the cyclin E-CDK2 could switch histone expression by activating NPAT and inactivating HIRA.

### 3.2. Once they are transcribed: post-transcriptional and translational regulation

Mammalian histone mRNAs lack introns and do not have a poly(A) tail as do most mRNAs. Instead, they contain a special 3′UTR sequence that forms a stem-loop structure [54]. Histone clusters localize to specific Cajal Bodies that are enriched in factors required for expression (NPAT) and maturation (U7 snRNA) of histone mRNA named *histone locus bodies* [75]. Maturation requires the formation of the 3′end through an endonucleolytic cleavage that has been shown to be important for transcription termination [76, 77]. Cleavage takes place between the stem-loop and the histone downstream element (HDE). The machinery involved in this process uses some common elements from the processing machinery of polyadenylated mRNAs but has also some specific components like SLBP, the Sm-like proteins (LSM1-11), the U7 snRNA and ZFP100. Additional information on maturation of histone mRNAs can be found in a nice review published some years ago [54]. The Stem Loop Binding Protein, SLBP, is one of the most important proteins for post-transcriptional and translational regulation of histone mRNAs and accompanies histone mRNA throughout its life.

SLBP is the only known cell cycle regulated protein of all the histone processing machinery. This protein starts accumulating during late G1 and is degraded at the end of S-phase by the phosphorylation of two threonine residues that target it for degradation [78]. There are three major roles for SLBP on histone regulation: 1. Allow an efficient cleavage during mRNA maturation 2. Facilitate circularization of histone mRNAs, required for their efficient translation by polyribosomes [79, 80] and 3. Increase histone mRNA stability preventing degradation of histone mRNAs by the 3′hExo [81]. Nevertheless, histone mRNA are still degraded when

SLBP is artificially present at constitutive levels at the end of S-phase [78] or when DNA replication is inhibited [82], indicating that although this protein has a major contribution to histone mRNA stability, it is not able to prevent degradation itself.

Canonical histone mRNAs in lower eukaryotes and plants are polyadenylated. The fact that these transcripts lack any known specific structure at their 3′end and have a short half-life has lead to the conclusion that regulation in these organisms mostly takes place at a transcriptional level. Nevertheless, there is quite a lot of recent evidence that strongly suggests the importance of the post-transcriptional regulation of histone mRNAs in *S. cerevisiae*. Several reports implicate some of the components of the exosome in the specific degradation of the H2B transcripts [83]. One year ago, Herrero and Moreno revealed the importance of the SM-like protein Lsm1 in histone mRNA degradation. Mutants lacking Lsm1 are sensitive to DNA damaging drugs and histone over-expression, and show a defect in histone mRNA degradation under replication stress conditions [84]. This protein is part of the Lsm1-7-Pat1 mRNA degradation complex, which has an important role in histone mRNA degradation under replication stress conditions in human cells. Lsm1-7-Pat1 has been shown to bind preferentially mRNAs carrying U-tracts in human cells, and oligoadenylated over polyadenylated mRNAs in yeast [85, 86]. Upon DNA replication arrest, histone mRNAs in human cells suffer an oligouridylation process acquiring a terminal oligo U-tract required for an efficient degradation by this complex [87]. Uridylation of mRNA has not been detected to date in *S. cerevisiae* but there is a recent report showing that the average length of the poly(A) tail of the yeast H2B histone mRNA is quite short compared to other transcripts. The length of this poly(A) is cell cycle-dependent and seems to decrease as cells exit G1 and progress through S-phase up to G2, when some of the transcripts completely lack a poly(A) tail [88]. This difference in length opens a possible explanation as to how the Lsm1-7 yeast complex preferentially recognizes yeast histone mRNAs over other transcripts to degrade them at the end of S-phase.

### 3.3. Last frontier of histone regulation: controling protein levels

In addition to the tight regulation of histone mRNA levels, a mechanism able to control histone protein levels was described some years ago [89]. To date, this pathway has only been described in the yeast *S. cerevisiae* and involves the action of the yeast homologue of CHK2, Rad53. Rad53 plays an important role in the DNA Damage Response and has been shown to be essential upon DNA damage or replication stress [1, 90]. Histone degradation involves the direct action of Rad53 along with the E2 ubiquitin ligases (UL) Ubc4 and Ubc5 and the E3 UL Tom1 [91]. This complex is able to degrade histones in a mechanism that involves tyrosine phosphorylation and poly-ubiquitylation, before their proteolysis by the proteasome. Histone degradation is independent on the central DNA damage checkpoint signal, since it does not depend on other kinases involved in the DDR like Mec1 (ATM) or Tel1 (ATR). Further studies in higher eukaryotes need to be done to confirm if this pathway is conserved in all eukaryotes.
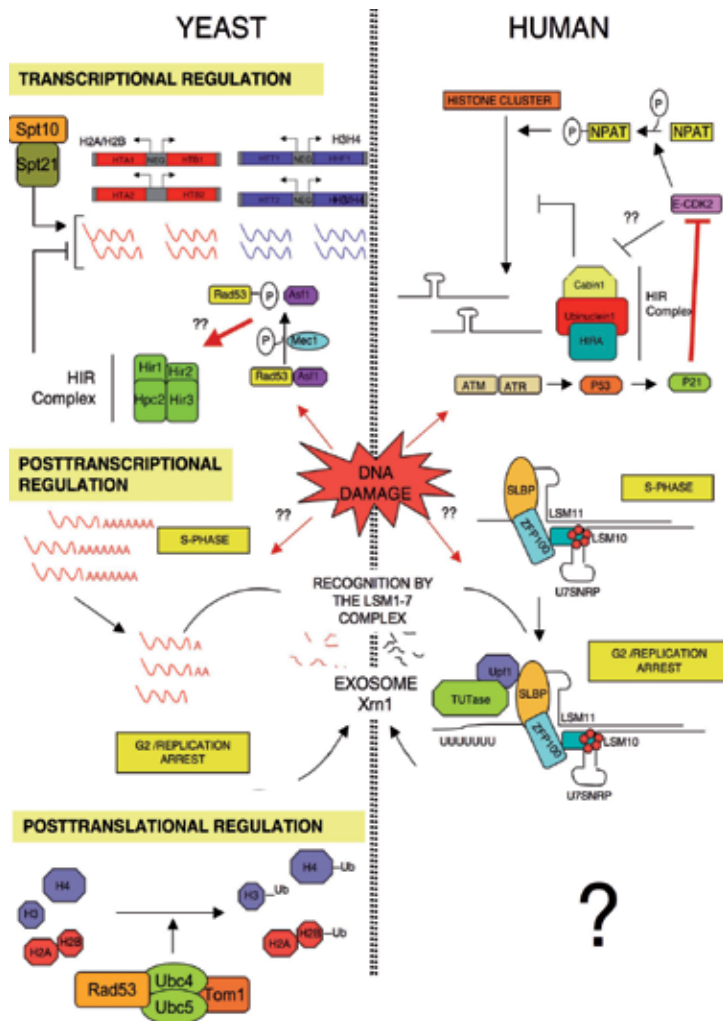
**Figure 3. Regulation of histone levels in *S.cerevisiae* and *H.sapiens*.** Different mechanisms able to control histone levels in *S.cerevisiae* and *H.sapiens*. Arrows normally indicate a positive effect on the pathway and straight lines a negative One. Interrogation marks are used when the protein/s involved in such process remain unknown or when the pathway has not been directly demonstrated. The big interrogation mark shown for post-translational regulation in the *H.sapiens* column, remarks that this pathway has not been demonstrated to date in human cells.

## 4. Histones: Enough to pack but not too much

Histone levels are regulated as soon as transcription of its mRNA starts. On top of the normal cell cycle regulation, additional mechanisms are able to block histone production when replication slows down or is completely blocked. Eukaryotic cells are unable to live without histones [92] and inhibition of histone deposition behind the replication fork blocks DNA synthesis and activates the DNA Damage Response (DDR) [35]. Eukaryotes seem to have

evolved to a situation in which histones must not be free in the cell and DNA must not be free of histones. In this last part of the chapter, we will focus on how cells cope with situations that break this bidirectional relationship.

## 4.1. Harmful effects of free histones

Histones are basic proteins that can bind-specifically to negatively charged molecules. Reconstitution experiments show that a slight excess of histones over DNA is sufficient to promote chromatin aggregation, probably due to the neutralization of negative charged DNA. In yeast, high levels of histones increase chromosome loss and enhance DNA damage sensitivity [57, 89]. Defects in histone degradation during replication stress or DNA damage severely decreases cell viability [84]. Free histones show electrostatic interactions with some cellular macromolecules carrying the opposite charge such as RNA molecules. Additionally, an excess of free histones can saturate and inhibit the activity of some histone modifying enzymes, and change the expression pattern of almost 240 genes [93]. Two different studies in the yeasts *S. cerevisiae* and *Schizosaccharomyces pombe* have demonstrated the importance of a correct balance between histone H3 and the centromeric variant Cse4 (CENPA) for efficient chromosome segregation. H3 can compete with Cse4 in the assembly of centromeric chromatin and this competition largely depends on a correct balance between levels of H3 and H4 [52, 53]. Cells must therefore not only prevent the accumulation of free histones but also ensure a correct homeostasis between canonical and other histone variants. Once cells have decided to initiate replication, any problem that unbalances replication fork progression with histone levels can potentially lead to an increase in the abundance of free histones. In order to prevent this, there is an additional pathway linked to the DDR able to block histone synthesis under DNA replication stress conditions or replication fork arrest.

## 4.2. The DNA Damage Response (DDR): Coupling DNA and histone synthesis

DDR is probably one of the most well characterized checkpoints in the cell and is normally activated whenever a cell senses DNA damage. Activation leads to the sequential action of a cascade of kinases that block or delay cell cycle progression to allow the cell to correct the damage. If damage cannot be repaired, human cells enter the apoptosis program and die [94]. Proper functioning of this pathway is essential for genome integrity and mutations in most of the branches of this path are linked to cancer and other diseases. DDR is able to block cells at G1, S and G2/M [95]. In human cells, two kinases ATM and ATR (Tel1 and Mec1 respectively in *S. cerevisiae*) play a major role in the activation of the DDR. ATM has been directly involved in the activation of a mechanism that ultimately leads to repression of histone expression.

In human cells, activation of histone gene transcription requires NPAT phosphorylation by the cyclin E-CDK2 complex at the beginning of S-Phase. Activation of NPAT is essential for S-phase progression and histone expression. Repression of histone synthesis upon DNA damage requires the activation of ATM, which leads to the sequential activation of p53 and then p21. p21 is able to block the activity of the cyclin E-CDK2 complex. Inhibition of this complex leads to a progressive dephosphorylation of NPAT, which no longer localizes to

histone clusters to activate transcription [96]. One interesting hypothesis that remains to be tested is if this cascade could also lead to histone repression by a change in the activity or location of HIRA, the human homologue of the HIR complex, at histone promoters (see previous paragraph about transcriptional regulation of histones). DNA damage also promotes post-transcriptional degradation of histone mRNAs. Treatment of cells with hydroxyurea (HU) increases oligouridylation of histone mRNAs in a process that depends on Upf1. Upf1 binds SLBP and helps to recruit a 3′ Terminal Uridylyl Transferase (TUT-ase) required for oligourydilation. These 3′ oligo(U) tails are recognized by the Lsm1–7 complex that triggers mRNA degradation through the exosome and Xrn1 [87]. How Upf1 is recruited to histone mRNAs upon DNA damage remains unknown.

Regulation of histone levels upon DNA damage in *S. cerevisiae* shows some common regulatory elements with human cells, and suggests the existence of a conserved mechanism. Post-transcriptional regulation also depends on the Lsm1-7 complex. It is not clear how this complex recognizes histone mRNAs but it could be related to the poly (A) tail-length (see post-transcriptional regulation of histones). Post-translational regulation by the Rad53 histone degradation pathway has not been directly addressed during the DDR, but taking into account the role of this protein in both pathways, it is reasonable to think that Rad53 could be important to destroy the population of translated histones when replication is halted. There are no NPAT homologues described in yeast and negative regulation during the DDR depends on the integrity of the HIR complex. The repressive structure formed to block transcription on histone promoters also requires Asf1 and Rtt106 among others. Although there is a lot of information about the formation of the repressive structure created at the promoter [97], the first steps by which DNA damage triggers histone repression remain largely unknown. There is some data nevertheless that suggest that Asf1 and Rad53 could play a role in this process.

Asf1 is able to form a very stable complex with Rad53. Upon activation of the DNA damage response, Mec1 phosphorylates Rad53 and this phosphorylation dissociates the stable Asf1-Rad53 complex. This mechanism has been linked to a possible connection between checkpoint activation and DNA repair since Asf1 plays a role in chromatin remodeling during DNA repair [98]. Rad53 can also be found in a hypophosphorylated form in normal conditions during G1, G2 and M, stages at which histone transcription is repressed. This phosphorylation seems to depend on Cdc28, the yeast functional homologue of human CDK1 and CDK2. Asf1 is able to co-immunoprecipitate with all the subunits of the HIR complex. This complex has been related to replication-independent nucleosome assembly and *in vitro* data prove that it is able to assembly nucleosomes to a DNA template [99]. Mutants lacking Asf1 have higher levels of histone mRNA and show defects in S-phase progression [100]. We have recently seen in our lab that mutants lacking the kinase activity of Rad53 also have enhanced levels of these mRNAs (unpublished results). Taking into account the close relationship that Asf1 plays with both Rad53 and the HIR complex, it is possible to think that the dissociation of Rad53 and Asf1 during DNA damage could be important for the efficient repression of histone transcription.

### 4.3. Generation of free histones in the cell

How can free histones be generated during a normal cell cycle? Taking into account the tight regulation of histone levels, such situations may seem unlikely. There are two scenarios in which it is possible to think that histone supply and DNA replication can be unbalanced during a normal cell cycle. In the first one, this situation could arise from differences between the rate of DNA synthesis and histone supply during replication. Early S-phase cells use more replication forks than late S-phase cells [101, 102] and lesions in DNA or replication stress also affect the speed of the replication fork [103-106]. The second scenario in which a cell can encounter free histones would take place during the G2 stage of the cell cycle. Given the importance of a balanced ratio between histone H3 and CENPA in chromosome segregation, once cells have finished replication, all free histones that are not positioned should be quickly degraded. It is possible to think that an imbalance between these two types of histones could sometimes take place in actively replicating cells and opens a simple explanation to why most cancer cells have a high incidence rate of chromosome loss [107].

### 4.4. Transcription as a source of free histones

Transcription of a chromatin template also requires nucleosomes to be disassembled and reassembled after the passage of RNA polymerase II (RNA Pol II). Outside of the S-phase, transcribed chromatin is probably the major potential source of free histones. These free histones could arise due to minor imbalances between histone supply and demand during chromatin reassembly. One very well described essential factor involved in RNA pol II transcription is the FACT complex [108, 109]. This complex is able to stimulate RNA Pol II-dependent transcription elongation through chromatin *in vitro* [110, 111] and also *in vivo* [112-114]. FACT is able to bind H3/H4 tetramers and H2A/H2B dimers [115, 116] and it has been shown that the integrity of at least one of its subunits, Spt16, is important for an efficient reassembly of the original H3 and H4 histones evicted during transcription [117]. Our group, in collaboration with others, demonstrated two years ago that dysfunction in chromatin reassembly during transcription due to defects in the Spt16 protein generates an accumulation of free histones in yeast. Combination of this mutant with a kinase dead version of Rad53 *(rad53K$^{227}$A)*, unable to efficiently degrade histones, increases the accumulation of free histones and greatly impairs viability of this mutant in a checkpoint-independent way [118]. Deletion of one of the two-histone clusters for H2A-H2B expression is able to partially suppress the growth defect of this mutant and increasing H2A-H2B expression has the opposite effect. There is a strong correlation between histone levels and viability defects of the chromatin reassembly mutant of Spt16. This defect is not exclusive for Spt16, since Spt6, another chromatin remodeling factor involved in H3-H4 repositioning during transcription, also has a strong negative interaction with *rad53K227A*. Chromatin reassembly defects can lead to the generation of free histones evicted from chromatin during transcription, a new source of histones potentially toxic for the cell. Rad53 negatively interacts with many different proteins involved in chromatin-related processes and could have an important function in maintaining chromatin structure in yeast [119]. Some of these interactions are with factors that have

only been exclusively involved to date in chromatin related processes during transcription [118, 119] suggesting that Rad53 could play an important role in the degradation of histones when chromatin is not correctly reassembled during transcription.

## 4.5. Can the state of chromatin influence the decision of cells to initiate DNA replication?

Histones are able to affect DNA replication right from the beginning; the state of the chromatin influences the timing and organization of origin firing. Replication fork progression also depends on the correct histone deposition behind the replication machinery, since defects in CAF1 lead to checkpoint activation and block cells in S-phase. The state of chromatin is able therefore to influence DNA replication. Work done in our lab, suggests that chromatin state might also influence the decision of cells to enter or not replication during the G1/S transition in the *S. cerevisiae*.

The commitment to a new round of cell division takes place towards the end of the G1 phase of the cell cycle in a process called START in yeast, and Restriction Point in mammals [120]. In yeast, this is the main regulatory event of the G1 phase of the cell cycle and involves an extensive transcriptional program driven by transcription factors SBF (Swi4-Swi6) and MBF (Mbp1-Swi6) [121, 122]. MBF and SBF activation depends on the cyclin/cyclin-dependent-kinase (CDK) complex Cln3-Cdc28. This complex phosphorylates Whi5, the negative regulator of START, thus promoting its release from SBF (Swi4-Swi6). Activation of MBF-dependent transcription by Cln3-Cdc28 acts through a mechanism independent of Whi5, involving the phosphorylation of Mbp1 [123]. Activation of these two complexes results in the accumulation of G1 (Cln1 and Cln2) and the early S-phase cyclins (Clb5 and Clb6), which promote in last term S-phase entry [124]. The kinase activity of Cln1,2-Cdc28 triggers the degradation of cyclin-dependent kinase inhibitor Sic1 which no longer inhibits the S phase-promoting complex Clb5,6-Cdc28 [125, 126].

FACT plays a role in maintaining the integrity of the chromatin structure during transcription [127-129] but has also been related to a G1/S cell cycle defect in yeast in a genetic screen to identify *cdc* (cell division cycle) mutants. This cell cycle defect had been linked initially to a general transcription defect of the three G1 cyclins Cln1-Cln3 [130] and later, to a possible important role of FACT in the transcription of *CLN1* and *CLN2* [131]. We recently described that this G1 defect is also due to a transcriptional downregulation of the cyclin Cln3 at the promoter level. Surprisingly, FACT seems not to be directly involved in the transcriptional regulation of this cyclin, since it is not recruited to the promoter at START when *CLN3* levels are maximal (D. Stillman unpublished results). One rather unexpected but interesting result is that this cell cycle defect shows a direct correlation with histone levels. Decreasing the H2A-H2B histone pool diminishes the cell cycle accumulation of this mutant while blocking the efficient degradation of histones has an additive effect. This defect is not exclusive for FACT mutants, since an Spt6 mutant also shows cell cycle defects at the G1/S transition [118]. Moreover, in yeast, a structural mutant of histone H4 in a region important for the interaction between the H3-H4 tetramer and the two H2A-H2B dimers completely mimics the cell cycle defects of the Spt16 mutant [132]. Defects in the chromatin structure seem to be connected somehow to the G1/S transition. Our group has speculated that cells might be

able to sense the chromatin state before entering a new round of replication. This mechanism would act at least in part through a transcriptional repression of the cyclin Cln3 mRNA. Although our first results pointed out that free histones could be the signal that triggers this G1/S transition defect, new results obtained by our lab show that this regulation could be more complex and also involve the chromatin structure itself (unpublished results).

## 5. Conclusion

In eukaryotes and also some archaebacteria, DNA forms a nucleoprotein complex called chromatin, which allows extensive compaction of genomic DNA in the limited space of the nucleus. This traditional view of chromatin as simple building-bricks has substantially changed since the nucleosome hypothesis was proposed [133, 134]. Cells have evolved a unique and complex machinery to cope with the fact that most of the processes involving DNA are going to need to interact with and probably modify chromatin first. Chromatin acts as a new step of regulation and carries an epigenetic specific code that in some cases can be as important for the cell as the one contained on DNA. In addition, cells must also carefully balance the levels of histones during chromatin formation to avoid the generation of free histones, in order to prevent their deleterious effects.

## Acknowledgments

## Author details

Douglas Maya[*], Macarena Morillo-Huesca, Lidia Delgado Ramos, Sebastián Chávez and Mari-Cruz Muñoz-Centeno

*Address all correspondence to: dmaya@us.es

Department of Genetics, Faculty of Biology, University of Seville, Spain

# References

[1] Jones RM, Petermann E. Replication fork dynamics and the DNA damage response. Biochem J. 2012 Apr 1;443(1):13-26.

[2] Gunjan A, Paik J, Verreault A. Regulation of histone synthesis and nucleosome assembly. Biochimie. 2005 Jul;87(7):625-635.

[3] Alabert C, Groth A. Chromatin replication and epigenome maintenance. Nat Rev Mol Cell Biol. 2012 Mar;13(3):153-167.

[4] Mechali M. Eukaryotic DNA replication origins: many choices for appropriate answers. Nat Rev Mol Cell Biol. 2010 Oct;11(10):728-738.

[5] Xu J, Yanagisawa Y, Tsankov AM, Hart C, Aoki K, Kommajosyula N, et al. Genome-wide identification and characterization of replication origins by deep sequencing. Genome Biol. 2012 Apr 24;13(4):R27.

[6] Gao F, Luo H, Zhang CT. DeOri: a database of eukaryotic DNA replication origins. Bioinformatics. 2012 Jun 1;28(11):1551-1552.

[7] Sclafani RA, Holzen TM. Cell cycle regulation of DNA replication. Annu Rev Genet. 2007;41:237-280.

[8] Dohrmann PR, Sclafani RA. Novel role for checkpoint Rad53 protein kinase in the initiation of chromosomal DNA replication in Saccharomyces cerevisiae. Genetics. 2006 Sep;174(1):87-99.

[9] Aggarwal BD, Calvi BR. Chromatin regulates origin activity in Drosophila follicle cells. Nature. 2004 Jul 15;430(6997):372-376.

[10] Tardat M, Brustel J, Kirsh O, Lefevbre C, Callanan M, Sardet C, et al. The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. Nat Cell Biol. 2010 Nov;12(11):1086-1093.

[11] Kouzarides T. Chromatin modifications and their function. Cell. 2007 Feb 23;128(4):693-705.

[12] Miotto B, Struhl K. HBO1 histone acetylase is a coactivator of the replication licensing factor Cdt1. Genes Dev. 2008 Oct 1;22(19):2633-2638.

[13] Miotto B, Struhl K. HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin. Mol Cell. 2010 Jan 15;37(1):57-66.

[14] Iizuka M, Matsui T, Takisawa H, Smith MM. Regulation of replication licensing by acetyltransferase Hbo1. Mol Cell Biol. 2006 Feb;26(3):1098-1108.

[15] Remus D, Diffley JF. Eukaryotic DNA replication control: lock and load, then fire. Curr Opin Cell Biol. 2009 Dec;21(6):771-777.

[16]  MacAlpine HK, Gordan R, Powell SK, Hartemink AJ, MacAlpine DM. Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Res. 2010 Feb;20(2):201-211.

[17]  Guillou E, Ibarra A, Coulon V, Casado-Vela J, Rico D, Casal I, et al. Cohesin organizes chromatin loops at DNA replication factories. Genes Dev. 2010 Dec 15;24(24): 2812-2822.

[18]  Courbet S, Gay S, Arnoult N, Wronka G, Anglana M, Brison O, et al. Replication fork movement sets chromatin loop size and origin choice in mammalian cells. Nature. 2008 Sep 25;455(7212):557-560.

[19]  Gilbert DM, Takebayashi SI, Ryba T, Lu J, Pope BD, Wilson KA, et al. Space and time in the nucleus: developmental control of replication timing and chromosome architecture. Cold Spring Harb Symp Quant Biol. 2010;75:143-153.

[20]  Schwaiger M, Stadler MB, Bell O, Kohler H, Oakeley EJ, Schubeler D. Chromatin state marks cell-type- and gender-specific replication of the Drosophila genome. Genes Dev. 2009 Mar 1;23(5):589-601.

[21]  Sogo JM, Stahl H, Koller T, Knippers R. Structure of replicating simian virus 40 minichromosomes. The replication fork, core histone segregation and terminal structures. J Mol Biol. 1986 May 5;189(1):189-204.

[22]  Gasser R, Koller T, Sogo JM. The stability of nucleosomes at the replication fork. J Mol Biol. 1996 May 3;258(2):224-239.

[23]  Kuipers MA, Stasevich TJ, Sasaki T, Wilson KA, Hazelwood KL, McNally JG, et al. Highly stable loading of Mcm proteins onto chromatin in living cells requires replication to unload. J Cell Biol. 2011 Jan 10;192(1):29-41.

[24]  Alexandrow MG, Hamlin JL. Chromatin decondensation in S-phase involves recruitment of Cdk2 by Cdc45 and histone H1 phosphorylation. J Cell Biol. 2005 Mar 14;168(6):875-886.

[25]  Belotserkovskaya R, Oh S, Bondarenko VA, Orphanides G, Studitsky VM, Reinberg D. FACT facilitates transcription-dependent nucleosome alteration. Science. 2003 Aug 22;301(5636):1090-1093.

[26]  Xin H, Takahata S, Blanksma M, McCullough L, Stillman DJ, Formosa T. yFACT induces global accessibility of nucleosomal DNA without H2A-H2B displacement. Mol Cell. 2009 Aug 14;35(3):365-376.

[27]  Hertel L, De Andrea M, Bellomo G, Santoro P, Landolfo S, Gariglio M. The HMG protein T160 colocalizes with DNA replication foci and is down-regulated during cell differentiation. Exp Cell Res. 1999 Aug 1;250(2):313-328.

[28]  Tan BC, Chien CT, Hirose S, Lee SC. Functional cooperation between FACT and MCM helicase facilitates initiation of chromatin DNA replication. EMBO J. 2006 Sep 6;25(17):3975-3985.

[29] Ito T, Bulger M, Kobayashi R, Kadonaga JT. Drosophila NAP-1 is a core histone chaperone that functions in ATP-facilitated assembly of regularly spaced nucleosomal arrays. Mol Cell Biol. 1996 Jun;16(6):3112-3124.

[30] Lorch Y, Maier-Davis B, Kornberg RD. Chromatin remodeling by nucleosome disassembly in vitro. Proc Natl Acad Sci U S A. 2006 Feb 28;103(9):3090-3093.

[31] Xu M, Long C, Chen X, Huang C, Chen S, Zhu B. Partitioning of histone H3-H4 tetramers during DNA replication-dependent chromatin assembly. Science. 2010 Apr 2;328(5974):94-98.

[32] Franco AA, Lam WM, Burgers PM, Kaufman PD. Histone deposition protein Asf1 maintains DNA replisome integrity and interacts with replication factor C. Genes Dev. 2005 Jun 1;19(11):1365-1375.

[33] Groth A, Corpet A, Cook AJ, Roche D, Bartek J, Lukas J, et al. Regulation of replication fork progression through histone supply and demand. Science. 2007 Dec 21;318(5858):1928-1931.

[34] Hoek M, Stillman B. Chromatin assembly factor 1 is essential and couples chromatin assembly to DNA replication in vivo. Proc Natl Acad Sci U S A. 2003 Oct 14;100(21): 12183-12188.

[35] Ye X, Franco AA, Santos H, Nelson DM, Kaufman PD, Adams PD. Defective S phase chromatin assembly causes DNA damage, activation of the S phase checkpoint, and S phase arrest. Mol Cell. 2003 Feb;11(2):341-351.

[36] Finn RM, Browne K, Hodgson KC, Ausio J. sNASP, a histone H1-specific eukaryotic chaperone dimer that facilitates chromatin assembly. Biophys J. 2008 Aug;95(3): 1314-1325.

[37] Smith DJ, Whitehouse I. Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature. 2012 Mar 22;483(7390):434-438.

[38] Pesavento JJ, Yang H, Kelleher NL, Mizzen CA. Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. Mol Cell Biol. 2008 Jan;28(1):468-486.

[39] Lanzuolo C, Lo Sardo F, Diamantini A, Orlando V. PcG complexes set the stage for epigenetic inheritance of gene silencing in early S phase before replication. PLoS Genet. 2011 Nov;7(11):e1002370.

[40] Westermann S, Drubin DG, Barnes G. Structures and functions of yeast kinetochore complexes. Annu Rev Biochem. 2007;76:563-591.

[41] Wieland G, Orthaus S, Ohndorf S, Diekmann S, Hemmerich P. Functional complementation of human centromere protein A (CENP-A) by Cse4p from Saccharomyces cerevisiae. Mol Cell Biol. 2004 Aug;24(15):6620-6630.

[42] Jansen LE, Black BE, Foltz DR, Cleveland DW. Propagation of centromeric chromatin requires exit from mitosis. J Cell Biol. 2007 Mar 12;176(6):795-805.

[43]  Schuh M, Lehner CF, Heidmann S. Incorporation of Drosophila CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. Curr Biol. 2007 Feb 6;17(3):237-243.

[44]  Black BE, Bassett EA. The histone variant CENP-A and centromere specification. Curr Opin Cell Biol. 2008 Feb;20(1):91-100.

[45]  Fujita Y, Hayashi T, Kiyomitsu T, Toyoda Y, Kokubu A, Obuse C, et al. Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1. Dev Cell. 2007 Jan;12(1):17-30.

[46]  Choy JS, Mishra PK, Au WC, Basrai MA. Insights into assembly and regulation of centromeric chromatin in Saccharomyces cerevisiae. Biochim Biophys Acta. 2012 Jul; 1819(7):776-783.

[47]  Stoler S, Rogers K, Weitze S, Morey L, Fitzgerald-Hayes M, Baker RE. Scm3, an essential Saccharomyces cerevisiae centromere protein required for G2/M progression and Cse4 localization. Proc Natl Acad Sci U S A. 2007 Jun 19;104(25):10571-10576.

[48]  Dunleavy EM, Roche D, Tagami H, Lacoste N, Ray-Gallet D, Nakamura Y, et al. HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell. 2009 May 1;137(3):485-497.

[49]  Hewawasam G, Shivaraju M, Mattingly M, Venkatesh S, Martin-Brown S, Florens L, et al. Psh1 is an E3 ubiquitin ligase that targets the centromeric histone variant Cse4. Mol Cell. 2010 Nov 12;40(3):444-454.

[50]  Ranjitkar P, Press MO, Yi X, Baker R, MacCoss MJ, Biggins S. An E3 ubiquitin ligase prevents ectopic localization of the centromeric histone H3 variant via the centromere targeting domain. Mol Cell. 2010 Nov 12;40(3):455-464.

[51]  Lopes da Rosa J, Holik J, Green EM, Rando OJ, Kaufman PD. Overlapping regulation of CenH3 localization and histone H3 turnover by CAF-1 and HIR proteins in Saccharomyces cerevisiae. Genetics. 2011 Jan;187(1):9-19.

[52]  Castillo AG, Mellone BG, Partridge JF, Richardson W, Hamilton GL, Allshire RC, et al. Plasticity of fission yeast CENP-A chromatin driven by relative levels of histone H3 and H4. PLoS Genet. 2007 Jul;3(7):e121.

[53]  Au WC, Crisp MJ, DeLuca SZ, Rando OJ, Basrai MA. Altered dosage and mislocalization of histone H3 and Cse4p lead to chromosome loss in Saccharomyces cerevisiae. Genetics. 2008 May;179(1):263-275.

[54]  Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet. 2008 Nov;9(11):843-854.

[55]  Amin AD, Vishnoi N, Prochasson P. A global requirement for the HIR complex in the assembly of chromatin. Biochim Biophys Acta. 2012 Mar;1819(3-4):264-276.

[56]  Jaeger S, Barends S, Giege R, Eriani G, Martin F. Expression of metazoan replication-dependent histone genes. Biochimie. 2005 Sep-Oct;87(9-10):827-834.

[57] Meeks-Wagner D, Hartwell LH. Normal stoichiometry of histone dimer sets is necessary for high fidelity of mitotic chromosome transmission. Cell. 1986 Jan 17;44(1): 43-52.

[58] Marzluff WF, Duronio RJ. Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. Curr Opin Cell Biol. 2002 Dec;14(6):692-699.

[59] Fletcher C, Heintz N, Roeder RG. Purification and characterization of OTF-1, a transcription factor regulating cell cycle expression of a human histone H2b gene. Cell. 1987 Dec 4;51(5):773-781.

[60] Kaludov NK, Pabon-Pena L, Hurt MM. Identification of a second conserved element within the coding sequence of a mouse H3 histone gene that interacts with nuclear factors and is necessary for normal expression. Nucleic Acids Res. 1996 Feb 1;24(3): 523-531.

[61] Ye X, Wei Y, Nalepa G, Harper JW. The cyclin E/Cdk2 substrate p220(NPAT) is required for S-phase entry, histone gene expression, and Cajal body maintenance in human somatic cells. Mol Cell Biol. 2003 Dec;23(23):8586-8600.

[62] Ma T, Van Tine BA, Wei Y, Garrett MD, Nelson D, Adams PD, et al. Cell cycle-regulated phosphorylation of p220(NPAT) by cyclin E/Cdk2 in Cajal bodies promotes histone gene transcription. Genes Dev. 2000 Sep 15;14(18):2298-2313.

[63] Zhao J, Kennedy BK, Lawrence BD, Barbie DA, Matera AG, Fletcher JA, et al. NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription. Genes Dev. 2000 Sep 15;14(18):2283-2297.

[64] Spector MS, Raff A, DeSilva H, Lee K, Osley MA. Hir1p and Hir2p function as transcriptional corepressors to regulate histone gene transcription in the Saccharomyces cerevisiae cell cycle. Mol Cell Biol. 1997 Feb;17(2):545-552.

[65] Eriksson PR, Ganguli D, Clark DJ. Spt10 and Swi4 control the timing of histone H2A/H2B gene activation in budding yeast. Mol Cell Biol. 2011 Feb;31(3):557-572.

[66] Osley MA, Gould J, Kim S, Kane MY, Hereford L. Identification of sequences in a yeast histone promoter involved in periodic transcription. Cell. 1986 May 23;45(4): 537-544.

[67] Freeman KB, Karns LR, Lutz KA, Smith MM. Histone H3 transcription in Saccharomyces cerevisiae is controlled by multiple cell cycle activation sites and a constitutive negative regulatory element. Mol Cell Biol. 1992 Dec;12(12):5455-5463.

[68] Kurat CF, Lambert JP, van Dyk D, Tsui K, van Bakel H, Kaluarachchi S, et al. Restriction of histone gene transcription to S phase by phosphorylation of a chromatin boundary protein. Genes Dev. 2011 Dec 1;25(23):2489-2501.

[69]  Dollard C, Ricupero-Hovasse SL, Natsoulis G, Boeke JD, Winston F. SPT10 and SPT21 are required for transcription of particular histone genes in Saccharomyces cerevisiae. Mol Cell Biol. 1994 Aug;14(8):5223-5228.

[70]  Balaji S, Iyer LM, Aravind L. HPC2 and ubinuclein define a novel family of histone chaperones conserved throughout eukaryotes. Mol Biosyst. 2009 Mar;5(3):269-275.

[71]  Banumathy G, Somaiah N, Zhang R, Tang Y, Hoffmann J, Andrake M, et al. Human UBN1 is an ortholog of yeast Hpc2p and has an essential role in the HIRA/ASF1a chromatin-remodeling pathway in senescent cells. Mol Cell Biol. 2009 Feb;29(3): 758-770.

[72]  Tagami H, Ray-Gallet D, Almouzni G, Nakatani Y. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell. 2004 Jan 9;116(1):51-61.

[73]  Nelson DM, Ye X, Hall C, Santos H, Ma T, Kao GD, et al. Coupling of DNA synthesis and histone synthesis in S phase independent of cyclin/cdk2 activity. Mol Cell Biol. 2002 Nov;22(21):7459-7472.

[74]  Hall C, Nelson DM, Ye X, Baker K, DeCaprio JA, Seeholzer S, et al. HIRA, the human homologue of yeast Hir1p and Hir2p, is a novel cyclin-cdk2 substrate whose expression blocks S-phase progression. Mol Cell Biol. 2001 Mar;21(5):1854-1865.

[75]  Nizami Z, Deryusheva S, Gall JG. The Cajal body and histone locus body. Cold Spring Harb Perspect Biol. 2010 Jul;2(7):a000653.

[76]  Chodchoy N, Pandey NB, Marzluff WF. An intact histone 3'-processing site is required for transcription termination in a mouse histone H2a gene. Mol Cell Biol. 1991 Jan;11(1):497-509.

[77]  Gu X, Marzluff WF. 3' Processing and termination of mouse histone transcripts synthesized in vitro by RNA polymerase II. Nucleic Acids Res. 1996 Oct 1;24(19): 3797-3805.

[78]  Zheng L, Dominski Z, Yang XC, Elms P, Raska CS, Borchers CH, et al. Phosphorylation of stem-loop binding protein (SLBP) on two threonines triggers degradation of SLBP, the sole cell cycle-regulated factor required for regulation of histone mRNA processing, at the end of S phase. Mol Cell Biol. 2003 Mar;23(5):1590-1601.

[79]  Sanchez R, Marzluff WF. The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro. Mol Cell Biol. 2002 Oct;22(20): 7093-7104.

[80]  Gorgoni B, Andrews S, Schaller A, Schumperli D, Gray NK, Muller B. The stem-loop binding protein stimulates histone translation at an early step in the initiation pathway. RNA. 2005 Jul;11(7):1030-1042.

[81]  Huang Y, Gattoni R, Stevenin J, Steitz JA. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. Mol Cell. 2003 Mar;11(3):837-843.

[82] Whitfield ML, Kaygun H, Erkmann JA, Townley-Tilson WH, Dominski Z, Marzluff WF. SLBP is associated with histone mRNA on polyribosomes as a component of the histone mRNP. Nucleic Acids Res. 2004;32(16):4833-4842.

[83] Reis CC, Campbell JL. Contribution of Trf4/5 and the nuclear exosome to genome stability through regulation of histone mRNA levels in Saccharomyces cerevisiae. Genetics. 2007 Mar;175(3):993-1010.

[84] Herrero AB, Moreno S. Lsm1 promotes genomic stability by controlling histone mRNA decay. EMBO J. 2011 May 18;30(10):2008-2018.

[85] Chowdhury A, Mukhopadhyay J, Tharun S. The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. RNA. 2007 Jul;13(7):998-1016.

[86] Tharun S, He W, Mayes AE, Lennertz P, Beggs JD, Parker R. Yeast Sm-like proteins function in mRNA decapping and decay. Nature. 2000 Mar 30;404(6777):515-518.

[87] Mullen TE, Marzluff WF. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. Genes Dev. 2008 Jan 1;22(1):50-65.

[88] Beggs S, James TC, Bond U. The PolyA tail length of yeast histone mRNAs varies during the cell cycle and is influenced by Sen1p and Rrp6p. Nucleic Acids Res. 2012 Mar;40(6):2700-2711.

[89] Gunjan A, Verreault A. A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in S. cerevisiae. Cell. 2003 Nov 26;115(5):537-549.

[90] Navadgi-Patil VM, Burgers PM. Cell-cycle-specific activators of the Mec1/ATR checkpoint kinase. Biochem Soc Trans. 2011 Apr;39(2):600-605.

[91] Singh RK, Kabbaj MH, Paik J, Gunjan A. Histone levels are regulated by phosphorylation and ubiquitylation-dependent proteolysis. Nat Cell Biol. 2009 Aug;11(8):925-933.

[92] Gunjan A, Paik J, Verreault A. The emergence of regulated histone proteolysis. Curr Opin Genet Dev. 2006 Apr;16(2):112-118.

[93] Singh RK, Liang D, Gajjalaiahvari UR, Kabbaj MH, Paik J, Gunjan A. Excess histone levels mediate cytotoxicity via multiple mechanisms. Cell Cycle. 2010 Oct 15;9(20):4236-4244.

[94] Hanel W, Moll UM. Links between mutant p53 and genomic instability. J Cell Biochem. 2012 Feb;113(2):433-439.

[95] Ubezio P, Lupi M, Branduardi D, Cappella P, Cavallini E, Colombo V, et al. Quantitative assessment of the complex dynamics of G1, S, and G2-M checkpoint activities. Cancer Res. 2009 Jun 15;69(12):5234-5240.

[96]  Su C, Gao G, Schneider S, Helt C, Weiss C, O'Reilly MA, et al. DNA damage induces downregulation of histone gene expression through the G1 checkpoint pathway. EMBO J. 2004 Mar 10;23(5):1133-1143.

[97]  Eriksson PR, Ganguli D, Nagarajavel V, Clark DJ. Regulation of histone gene expression in budding yeast. Genetics. 2012 May;191(1):7-20.

[98]  Emili A, Schieltz DM, Yates JR, 3rd, Hartwell LH. Dynamic interaction of DNA damage checkpoint protein Rad53 with chromatin assembly factor Asf1. Mol Cell. 2001 Jan;7(1):13-20.

[99]  Green EM, Antczak AJ, Bailey AO, Franco AA, Wu KJ, Yates JR, 3rd, et al. Replication-independent histone deposition by the HIR complex and Asf1. Curr Biol. 2005 Nov 22;15(22):2044-2049.

[100]  Sutton A, Bucaria J, Osley MA, Sternglanz R. Yeast ASF1 protein is required for cell cycle regulation of histone gene transcription. Genetics. 2001 Jun;158(2):587-596.

[101]  Berezney R, Dubey DD, Huberman JA. Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. Chromosoma. 2000 Mar;108(8):471-484.

[102]  Leonhardt H, Rahn HP, Weinzierl P, Sporbert A, Cremer T, Zink D, et al. Dynamics of DNA replication factories in living cells. J Cell Biol. 2000 Apr 17;149(2):271-280.

[103]  Paulovich AG, Hartwell LH. A checkpoint regulates the rate of progression through S phase in S. cerevisiae in response to DNA damage. Cell. 1995 Sep 8;82(5):841-847.

[104]  Santocanale C, Diffley JF. A Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication. Nature. 1998 Oct 8;395(6702):615-618.

[105]  Feijoo C, Hall-Jackson C, Wu R, Jenkins D, Leitch J, Gilbert DM, et al. Activation of mammalian Chk1 during DNA replication arrest: a role for Chk1 in the intra-S phase checkpoint monitoring replication origin firing. J Cell Biol. 2001 Sep 3;154(5):913-923.

[106]  Tercero JA, Diffley JF. Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. Nature. 2001 Aug 2;412(6846):553-557.

[107]  Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature. 1998 Dec 17;396(6712):643-649.

[108]  Reinberg D, Sims RJ, 3rd. de FACTo nucleosome dynamics. J Biol Chem. 2006 Aug 18;281(33):23297-23301.

[109]  Formosa T. FACT and the reorganized nucleosome. Mol Biosyst. 2008 Nov;4(11):1085-1093.

[110]  Orphanides G, LeRoy G, Chang CH, Luse DS, Reinberg D. FACT, a factor that facilitates transcript elongation through nucleosomes. Cell. 1998 Jan 9;92(1):105-116.

[111]  Pavri R, Zhu B, Li G, Trojer P, Mandal S, Shilatifard A, et al. Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell. 2006 May 19;125(4):703-717.

[112] Mason PB, Struhl K. The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. Mol Cell Biol. 2003 Nov;23(22):8323-8333.

[113] Saunders A, Werner J, Andrulis ED, Nakayama T, Hirose S, Reinberg D, et al. Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. Science. 2003 Aug 22;301(5636):1094-1096.

[114] Jimeno-Gonzalez S, Gomez-Herreros F, Alepuz PM, Chavez S. A gene-specific requirement for FACT during transcription is related to the chromatin organization of the transcribed region. Mol Cell Biol. 2006 Dec;26(23):8710-8721.

[115] Stuwe T, Hothorn M, Lejeune E, Rybin V, Bortfeld M, Scheffzek K, et al. The FACT Spt16 "peptidase" domain is a histone H3-H4 binding module. Proc Natl Acad Sci U S A. 2008 Jul 1;105(26):8884-8889.

[116] VanDemark AP, Xin H, McCullough L, Rawlins R, Bentley S, Heroux A, et al. Structural and functional analysis of the Spt16p N-terminal domain reveals overlapping roles of yFACT subunits. J Biol Chem. 2008 Feb 22;283(8):5058-5068.

[117] Jamai A, Puglisi A, Strubin M. Histone chaperone spt16 promotes redeposition of the original h3-h4 histones evicted by elongating RNA polymerase. Mol Cell. 2009 Aug 14;35(3):377-383.

[118] Morillo-Huesca M, Maya D, Munoz-Centeno MC, Singh RK, Oreal V, Reddy GU, et al. FACT prevents the accumulation of free histones evicted from transcribed chromatin and a subsequent cell cycle delay in G1. PLoS Genet. 2010 May;6(5):e1000964.

[119] Pan X, Ye P, Yuan DS, Wang X, Bader JS, Boeke JD. A DNA integrity network in the yeast Saccharomyces cerevisiae. Cell. 2006 Mar 10;124(5):1069-1081.

[120] Hartwell LH, Kastan MB. Cell cycle control and cancer. Science. 1994 Dec 16;266(5192):1821-1828.

[121] Costanzo M, Nishikawa JL, Tang X, Millman JS, Schub O, Breitkreuz K, et al. CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. Cell. 2004 Jun 25;117(7):899-913.

[122] de Bruin RA, McDonald WH, Kalashnikova TI, Yates J, 3rd, Wittenberg C. Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. Cell. 2004 Jun 25;117(7):887-898.

[123] Ferrezuelo F, Colomina N, Futcher B, Aldea M. The transcriptional network activated by Cln3 cyclin at the G1-to-S transition of the yeast cell cycle. Genome Biol. 2010;11(6):R67.

[124] Wittenberg C. Cell cycle: cyclin guides the way. Nature. 2005 Mar 3;434(7029):34-35.

[125] Schneider BL, Yang QH, Futcher AB. Linkage of replication to start by the Cdk inhibitor Sic1. Science. 1996 Apr 26;272(5261):560-562.

[126]  Schwob E, Bohm T, Mendenhall MD, Nasmyth K. The B-type cyclin kinase inhibitor p40SIC1 controls the G1 to S transition in S. cerevisiae. Cell. 1994 Oct 21;79(2): 233-244.

[127]  Formosa T, Ruone S, Adams MD, Olsen AE, Eriksson P, Yu Y, et al. Defects in SPT16 or POB3 (yFACT) in Saccharomyces cerevisiae cause dependence on the Hir/Hpc pathway: polymerase passage may degrade chromatin structure. Genetics. 2002 Dec; 162(4):1557-1571.

[128]  Kaplan CD, Laprade L, Winston F. Transcription elongation factors repress transcription initiation from cryptic sites. Science. 2003 Aug 22;301(5636):1096-1099.

[129]  Cheung V, Chua G, Batada NN, Landry CR, Michnick SW, Hughes TR, et al. Chromatin- and transcription-related factors repress transcription from within coding regions throughout the Saccharomyces cerevisiae genome. PLoS Biol. 2008 Nov 11;6(11):e277.

[130]  Lycan D, Mikesell G, Bunger M, Breeden L. Differential effects of Cdc68 on cell cycle-regulated promoters in Saccharomyces cerevisiae. Mol Cell Biol. 1994 Nov;14(11): 7455-7465.

[131]  Takahata S, Yu Y, Stillman DJ. The E2F functional analogue SBF recruits the Rpd3(L) HDAC, via Whi5 and Stb1, and the FACT chromatin reorganizer, to yeast G1 cyclin promoters. EMBO J. 2009 Nov 4;28(21):3378-3389.

[132]  Santisteban MS, Arents G, Moudrianakis EN, Smith MM. Histone octamer function in vivo: mutations in the dimer-tetramer interfaces disrupt both gene activation and repression. EMBO J. 1997 May 1;16(9):2493-2506.

[133]  Olins AL, Olins DE. Spheroid chromatin units (v bodies). Science. 1974 Jan 25;183(4122):330-332.

[134]  Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. Science. 1974 May 24;184(4139):868-871.

# Replicating – DNA in the Refractory Chromatin Environment

Angélique Galvani and Christophe Thiriet

Additional information is available at the end of the chapter

## 1. Introduction

The replication of DNA is a process found throughout the prokaryotic and the eukaryotic kingdoms. Although the basic aim of this process is the duplication of the genetic informa‐tion, the mechanisms leading to replication are different in prokaryotes and in eukaryotes. A major divergence between the two kingdoms corresponds to the nature of the substrate of the replication process [1]. Indeed, while the genetic information in prokaryotic cell is recov‐ered in the nucleoid, the eukaryotic genome is found in the nucleus and the genetic material is associated with proteins. The tight interaction of the DNA molecule with proteins forms the chromatin, and for replication as well as for the other cellular processes that require the access to the genetic material, the chromatin is the actual substrate [2]. This organization of the eukaryotic genome in chromatin generates additional constraints to enzymatic activities. Therefore, it is required for the replication machinery to over-rule the refractory environ‐ment of chromatin.

Although the arrangement of the genetic material with proteins is an inhibitory environ‐ment, it is also required for packaging the molecule of DNA within the confined nuclear vol‐ume and for organizing the genome. Therefore, defects in the genetic material packaging affect genome stability and cell viability. Importantly, as replication results in the doubling of DNA, it is required for the cell to synthesize DNA-associated proteins and to form chro‐matin. This process known as replication-coupled chromatin assembly implies the copy of the epigenetic information carried by the histone proteins [3].

In the present chapter, we define the general features of chromatin, primarily on the ba‐sis of the fundamental sub-unit, the nucleosome, and the constraints that this structure generates for creating a refractory environment to replication. In addition to the view of the single nucleosome, as chromatin can be viewed as a polymer of nucleosomes which

are highly ordered, the impediment of the replication machinery induced by higher chromatin order is discussed. Although replication activity should be inhibited by the chromatin, we review the mechanisms developed by eukaryotic cells to over-rule this non-permissive environment. Genetic experiments have shown that chromatin structure is essential for cell viability. We review the data providing evidence that the genome stability is, at least partly, inherent to chromatin assembly during replication, and the histone requirement in this process.

## 2. Chromatin: From the nucleosome sub-unit to the higher order structure

The basic chromatin sub-unit is the nucleosome, which is composed of the association of histone proteins with DNA [4]. The histone proteins are the most abundant nuclear proteins and are divided in four classes, H2A, H2B, H3 and H4, respectively. We distinguish in the histone protein two regions, the histone fold domain which is involved in the histone-histone and histone-DNA interactions, and the histone tail domain located at the N-terminal part of the protein, which is unstructured and extends out of the nucleosome [2, 5](Figure 1A). The association of the histones via their fold domain is highly conserved throughout the eukaryotic kingdom. Indeed, H3 is always associated with H4 and H2A with H2B forming therefore heterocomplexes H3/H4 and H2A/H2B (Figure 1B, upper panel). The histone pairing is done by three helixes of the fold domain of two histone counterparts which adopt a specific 'handshake' structure. The first high resolution crystal structure of the histone octamer in absence of DNA revealed that the histone octamer was organized in a tripartite structure wherein the H3/H4 complex formed a central tetramer which is flanked by two H2A/H2B dimers [6, 7](figure 1B, lower panel). Interestingly, while the histone fold domains were clearly resolved in the crystal, the unstructured tail domains were unseen. Although the histone octamer arrangement in presence of DNA confirmed the tripartite structure of the histone octamer, details of the edge of histone tails revealed the exit of these unstructured domains from the nucleosome [8].

It has been believed that the basic nature of the histones allowed the neutralization of the DNA phosphodiester backbone. However, the structure of the nucleosome at 1.9 Å resolution substantially improved the clarity of the electron density and revealed the presence of over 3000 water molecules and 18 ions [9]. The water molecules within the nucleosome promote the formation of hydrogen-bond bridges between the histone and the DNA molecule, like balls in a ball-bearing. Therefore, the water molecules enable the accommodation of intrinsic DNA conformational variation and promote the nucleosome mobility by limiting the rigidity of the nucleoprotein complex. The nucleosome crystal structures provided important information on the interactions between the histones and showed that the histone-DNA association is not only due to electrostatic interactions between the positively charge histones and the negatively charge DNA as it was primarily believed.
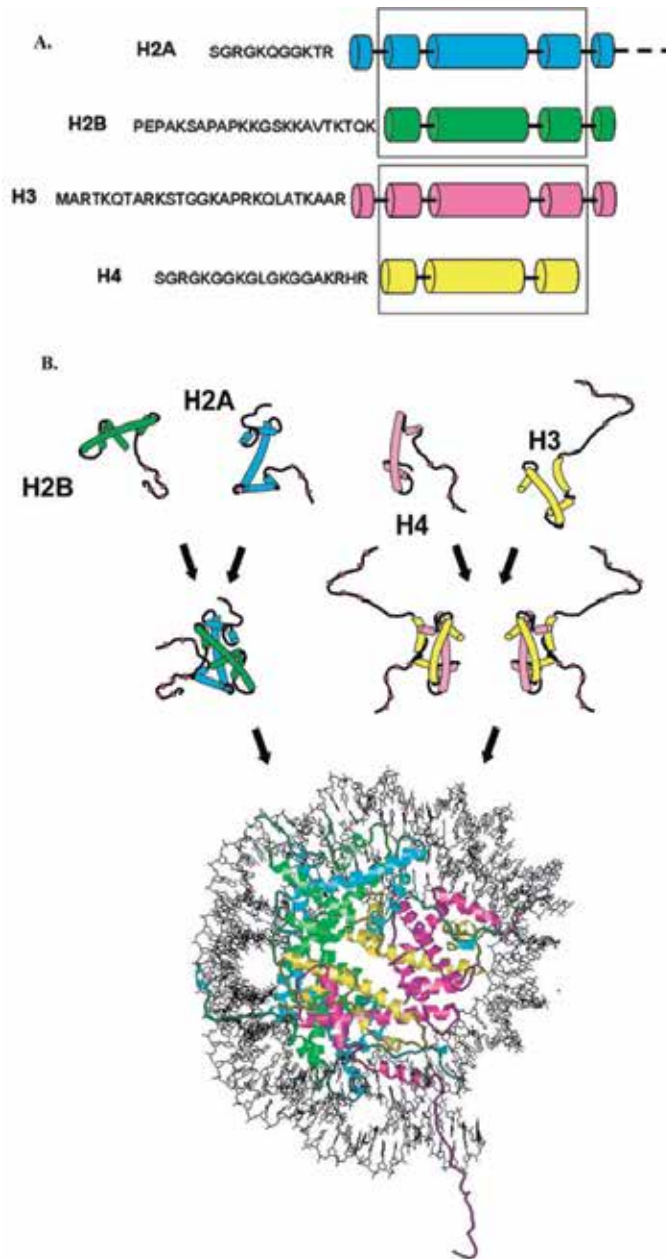
**Figure 1.** Histones and nucleosome formation: (A) schematic representation of the core histones. The boxes indicate the helixes of the histone fold domain, which is involved in the histone-histone interactions between H2A/H2B, and H3/H4. The amino-acid sequences correspond to the conserved sequence of the unstructured histone tail domain. (B) Individual core histones H2A (green), H2B (blue), H3 (yellow) and H4 (magenta) first heterodimerize to form the H2A/H2B and the H3/H4 complexes. The different complexes can either under different stringencies or with the help of histone chaperones associate together to form the nucleosome composed of a central tetramer of H3/H4 flanked by two heterodimers of H2A/H2B, and wrapped by 146 base pairs of DNA.

The demonstration of the labile interactions between the DNA molecule and the histone octamer was performed by the development of an elegant biochemical approach examining the accessibility of specific DNA sites within the nucleosomal DNA [10, 11]. In these experiments, the authors used a known nucleosome-positioning DNA sequence from the 5S gene, and by directed mutagenesis, restriction sites were generated at precise position within the DNA sequence. Nucleosome core particles were reconstituted with the different DNA sequences and purified by sucrose gradient centrifugation. The accessibility of the specific DNA sequences was examined as a function of time by adding to the nucleosome core particles the restriction enzymes. The quantitative analyses of the digested nucleosomal DNA reflect the accessibility of precise positions within the nucleosome core particles corresponding to the loss of histone-DNA contacts. Interestingly, the results revealed that DNA sequences engaged in the histone-DNA interactions are accessible to the restriction enzymes, and the accessibility gradually decreased when the restriction site is placed at proximity of the diad axis [12]. It was thus proposed that within the nucleosome core particle, dissociation of the histone-DNA contacts enables the transient exposure of DNA stretches to the solvent. Using a similar strategy, Widom and colleagues have also examined the contribution of the histone tail domains in the accessibility of nucleosomal DNA [13]. The results revealed that the removal of the histone tail domains leads to up to 14-fold increase in the site exposure within the nucleosomal DNA. Therefore, the tail domains within the nucleosome are also involved in the stabilization of DNA-histone fold domain interactions possibly by repressing the intrinsic dynamic nature of DNA.

The packaging of DNA in the nucleosome is a dynamic structure in conformational equilibrium, transiently exposing stretches of DNA off the histone surface, as demonstrated in model systems. Importantly, the binding of linker histone nearby the dyad axis to DNA restricts the flapping of the arms of DNA at the entry and at the exit of the nucleosome [14]. Although the analyses of the nucleosome behavior are very informative on the potential mobility of the nucleosome, it is obvious that the nucleosome is not recovered as a single subunit in living cell but rather found as a nucleosome polymer. Thus, the mobility of a considered monomer is possibly modulated by the surrounding nucleosomes. The analyses of a dinucleosome template generated from the 5S gene revealed a spontaneous mobility of the core histones which is restricted by the presence of the linker histone [15]. To better understand the function of the histones in the chromatin folding, it was required to examine templates that contained more than one or two nucleosomes. Using defined oligonucleosome models systems, the molecular mechanisms through which the histones modulated the chromatin folding were investigated [16]. These experiments revealed that the core histone tails play a critical function in the chromatin folding, as demonstrated by the removal of the tail domains *in vitro* [17, 18]. Interestingly, analyses of histone acetylation mimics on the chromatin fiber folding exhibited effects on the self-association properties of model nucleosome arrays, which depended upon the histone carrying the acetylation mimics and the number of mimics within the nucleosomes [19]. Such *in vitro* approaches using reconstituted nucleosomes systems are performed under particular pH and salt conditions. Additionally, acetylated histones increase chromatin solubility. Even if this can potentially biased the results, these investigations provide important features for understanding the physico-chemi-

cal parameters that facilitates or relieves the folding chromatin. But to date, the actual arrangement of the nucleosomes in the fiber is not yet well-determined. Nonetheless, experimental data have enabled to propose different models, the solenoid model and the zig-zag model, and it is possible that both models are juxtaposed in the nucleus [20, 21].

## 3. Relieving the chromatin inhibition

The ordered structure of chromatin represents the primary barrier to access the genetic information. On the basis of *in vitro* studies, the linker histones are proposed to be involved in the high-ordered chromatin structures [22]. Although the linker histone is not essential in protozoans [23, 24], the knock-out experiments in mouse revealed critical functions [25]. Indeed, in higher eukaryotes, the linker histones are composed of about eight subtypes which can compensate each other in some extend. However, upon the deletion of three subtypes, the synthesis compensation fails and embryonic lethality is observed. To attempt to gain insight in the function of the linker histone, analyses of the histone modifications have been carried out and reported a correlation between the cell cycle and the phosphorylation of the C-terminal tail domain [26, 27]. Surprisingly, while the genetic analyses revealed that preventing the phosphorylation of linker histone affects the chromatin organization leading to an increase of the nuclear volume, a raise in the linker histone phosphorylation was also detected in mitosis [28, 29]. Nonetheless, at the G1/S phase transition, linker histone is also found as substrate of cyclin-dependant kinase Cdk2, wherein the phosphorylation of the C-terminal tail leads to a relaxation of chromatin structure which might facilitate DNA replication [30, 31]. More recently, knock-down experiments of the linker histone in the slime mold *Physarumpolycephalum* showed a significantly faster rate of genome duplication, which was caused by a lost in the regulation of replication origin firing rather than the increase in the replication fork propagation [32]. Clearly, it has been evidenced that the linker histone affect the compaction of chromatin into the nucleus, and its release is required for the initial transition from non-permissive to permissive chromatin, but the actual mechanisms remain unclear.

Undoubtedly, if the primary inhibition for DNA replication is the higher levels of chromatin structure, relieving the high order of chromatin leaves the core histones associated with DNA, which is still an impediment for DNA accessibility. Thus, the next step is the release of the parental core histones to allow replication machinery to process all along the DNA molecule. To reach this goal, several concerns have to be taken into account. A bevy of studies have attempted to address the segregation of parental histones during replication, but the results are often controversial and many questions still need to be addressed. The fate of parental nucleosomes deals mainly with two overlapping key questions : do they dissociate from DNA during replication ? and, how are they transferred behind the replication fork ?

*In vitro* studies claimed chromatin replication without histone displacement. Initially showed in prokaryotic *in vitro* system [33], same conclusions were drawn from eukaryotic system studies [34]. In contrast, other studies evidenced that parental nucleosomes dissoci-

ate from DNA [35, 36]. The main argument for a non-displacement was that radioactively-labeled histone octamers are not reassembled onto a large excess of competitor DNA templates, suggesting that they do not dissociate from initial DNA matrix [34]. The idea that nucleosomes could partially relax to allow the passage of DNA processing machineries without complete dissociation is a matter of intense debate in the chromatin field, where the problematic of DNA accessibility is essential for most chromatin activities including replication, transcription and repair. Regarding replication, although no definite model can be drawn, it is commonly believed that disrupted parental nucleosomes are bound to specific protein chaperones which would transfer the core histone building blocks behind the replication fork.

The tripartite structure of the histone octamer implies that the removal of the H3/H4 from the nucleosome is associated with a displacement of the histone dimers H2A/H2B. However, two hypotheses could be postulated for lost of the nucleosomal structure, either the entire octamer is evicted or this is performed by the successive displacement of the different building blocks composing the histone octamer. Experimental approach for studying parental histone segregation implies the possibility of discriminating the old pool of histones and the new one [37]. By preventing the synthesis of new histones using translation inhibitors, like cycloheximide and puromycine, would enable the analysis of parental nucleosome transfer, though such treatments impair replication progression. Still, one can argue that as the replication process requires a tight regulation of the histone supply, impairing this regulation profoundly impact the replication leading to the replication fork blocks. Thus, most conclusions from these experiments have to be taken with caution. Original studies using this approach coupled with micrococcale digestion (enzyme allowing specific digestion of internucleosomal DNA) revealed that the size of the fragments obtained were consistent with DNA size protected by the histone octamer. So it was originally proposed that the parental nucleosomes are dissociated ahead of the replication fork and transferred behind with no detectable intermediate. Whether the experimental design led to artifacts remains likely.

Importantly, several studies using different approaches have demonstrated a distinct mobility for the H2A/H2B and the H3/H4 in living cells [36, 38]. On the basis of the different motions of the H2A/H2B and the H3/H4, one can reasonably believe that the octamer building blocks dissociate during cellular processes. Moreover, *in vitro* experiments for reconstituting or destabilizing nucleosome revealed the presence of basic heterocomplexes of H3/H4 tetramer and H2A/H2B dimer [37]. At physiological conditions, the heterotetramer H3/H4 prepared from chromatin and in absence of histone chaperones is the most stable form of the complex in solution [39]. Even if it has been claimed that a very transient dimeric state can exist, the absence of demonstration of the H3/H4 dimers led to the anchored view that parental nucleosomes split into two H2A/H2B dimers and a H3/H4 tetramer, and are then reassembled behind the fork, with the central tetramer H3/H4 deposited first [40, 41].

The simplest view regarding the dissociation of the parental core histone from DNA could be that the driving force of the replication fork progression is sufficient for overriding the histone-DNA interactions by the only action of replication specific proteins as helicases [42]. This model involves that core histones in presence of DNA spontaneously form nucleosomal

structures with a tripartite organization. Unfortunately, *in vitro* experiments demonstrated that such arrangement of the histone octamer required either high salt concentrations or chaperone proteins to assist the proper loading of the histones in a tripartite structure [43]. A more comprehensive view was provided by a study by Groth *et al* [44] showing that the major H3/H4 histone chaperone ASF1 (Anti-Silencing Factor 1) forms a complex with the putative replicative helicase MCM2-7 (Minichromosome Maintenance Complex), via a H3/H4 bridge. On the basis of the *in vitro* capability of ASF1 to assemble chromatin, it has been proposed that this chaperone might be involved in the recycling and the transfer of parental H3/H4 histones directly coordinated by the DNA replication process.

Concerning H2A/H2B, picture is even less clear. Chaperones, like NAP1 (Nck-associated protein 1) and FACT (Facilitates Chromatin Transcription) might be involved. The heterodimeric complex FACT, a chromatin-modifying factor initially described to promote nucleosome rearrangement during RNA polymerase II-driven transcription through H2A/H2B dimer destabilization [45], was shown to be involved in DNA replication. FACT interacts with DNA polymerase $\alpha$, and in human with the MCM helicase to act on DNA unwinding [46]. Recently, a conditional knock-out of one of the FACT subunit in DT40 chicken cells (Structure-Specific Recognition Protein 1, SSRP1) showed apparent impairment in replication fork progression [47]. Even if the precise mechanisms are still to be elucidated because this complex interacts with H2A/H2B and H3/H4 in multiple ways, the synergized action of histone chaperones and replication actors is actually an attractive model of coordinated nucleosome eviction/reassembly and DNA replication during S-phase.

It is known for a long time that chromatin assembly is an ATP-dependent process [48], so it is not surprising that ATP-dependent chromatin remodeling factors have been implicated in the release of the chromatin structure. Most studies focused on nucleosome movement during transcription, but strong arguments of their involvement during replication exist. The ISWI-class of ATP-dependent remodeling family interacts with several proteins in complexes, among them ACF1 (ATP-utilizing Chromatin assembly and remodeling Factor) and WSTF (Williams syndrome transcription factor). Depletion experiments demonstrated that ACF1 is critical for efficient DNA replication of highly condensed regions of mouse cells [49], and that WSTF, targeted to replication foci via its interaction with the processivity factor PCNA (Proliferating Cell Nuclear Antigen), promotes DNA replication by preventing premature maturation of chromatin [50].

## 4. Reforming chromatin behind the replication fork

Chromatin reassembly behind the replication fork is a rapid process. Electron microscopic studies and psoralen cross-linked nucleosome used, have clearly shown random distribution of the nucleosomal structures on both strand of the nascent DNA, with no apparent free-DNA [35]. By blocking protein synthesis with different inhibitors, it was demonstrated that half of the nucleosome pool came from random segregation of recycled parental ones, whereas the other half came from newly synthesized histones. In proliferating cells, the histone biosynthe-

sis is coupled with the cell cycle progression. The vast majority of histones (the canonical histones) are massively produced at the beginning of the S phase, mainly by transcriptional activation of histone genes and improvement of pre-mRNA processing and stability, that begins during G1 phase (reviewed in [51, 52]). Through a feedback regulation reducing drastically the half-life of histone mRNAs, the amount of proteins then decreased at the end of S-phase until the baseline level is reached. However, experiments using replication blocking agents showed distinct synthesis profiles between H3/H4 and H2A/H2B, illustrating that specific level of regulation may exist [53]. Some specific histones (histone variants), used for deposition and exchange of nucleosomes outside of the S-phase (replication-independent chromatin assembly), are produced throughout the cell cycle. Although this aspect presents a great interest, the present chapter focuses on the regulation of the canonical histone proteins at the onset of DNA replication (for reviews about histone variants see [54, 55]).

Once the histones are synthesized, they are rapidly delivered to the site of replication and assembled into chromatin. Because these proteins are highly basic proteins, histones tend to promptly bind non-specifically to nucleic acids with a higher affinity to RNA than DNA, and they do not spontaneously form nucleosomes. To allow correct transfer into the nucleus and efficient deposition onto DNA, histone chaperones play a dual function, they neutralize the histone charge to prevent the formation of aggregates and they address the histones at precise locations within the nucleus [56].

The supply of histone is a tightly regulated process. Any events leading to replicational stress (as DNA damage for example) disturb the fine balance between histone supply and demand and have deleterious effects on the cell. Histone chaperone have critical roles in regulating this process. Consistently, deletion of the major histone H3/H4 chaperones CAF-1 (Chromatin Assembly Factor 1) or ASF1 in various organisms impair S-phase progression [57, 58]. In human, it was shown that ASF1 exists in a highly mobile soluble pool that buffered the histone excess [59]. In the budding yeast *S. cerevisiae*, ASF1 depletion impairs cell cycle progression and generates chromosome instability [60]. In this organism, it was shown that the up-regulation of the amount of histone in the cells leads to the degradation of the excess histones by a Rad53 kinase-dependent mechanism [61].

### 4.1. Transport into the nucleus

The nuclear import of the histone complexes is among the first levels of regulation. Several groups have attempted to define the mechanisms by which the histone supply might be regulated. The role of specific domains within newly synthesized histones essential for transport (and also formation of nascent chromatin) was first addressed using powerful genetic approaches in the yeast *S. cerevisiae.* Pioneer studies performed in budding yeast emphasized the essential role of both N-terminal H2A/H2B tails for cell viability (reviewed in [62]). Fusion protein experiments using fluorescent tracers led to the assumption that nuclear localization signals (NLS) are present in the N-terminal non-structured domain of histone proteins, and their interaction with karyopherin or importins would promote nuclear import of newly synthesized histones [63, 64]. Nevertheless, incorporation experiments of exogenous histones in the slime mold *Physarum polycephalum* showed that H2A/H2B dimers

lacking both tail regions still localized to the nucleus. It was thus concluded that the tails of H2A/H2B are dispensable for nuclear import. However, the chromatin assembly analyses revealed that at least one tail is necessary for the deposition of the dimer complex into chromatin [65]. Conversely, studies using a similar strategy of incorporation of exogenous histones in *Physarum* to examine the fate of the H3/H4 complexes exhibited a function of the amino-terminal domains in nuclear import. Indeed, the histone H3/H4 dimers lacking H4 tail are inefficiently imported, while H3 tail was found dispensable in this process, but impaired nucleosome assembly coupled to replication [66].

By extending out of the nucleosomal structure, the exposed N-terminal regions of histones are subjected to active post-translational modifications. These marks, when imposed on assembled histones, have been shown to impact on the overall nature of the chromatin [67]. Newly synthesized histones are also characterized by a specific pattern of post-translational modifications, imposed in the cytoplasm shortly after synthesis. For example, newly synthesized H4 are diacetylated at lysine 5 and 12 by the holoenzyme HAT1 (Histone Acetyl Transferase 1), and these acetyl groups are rapidly removed after the assembly of histones into chromatin [68]. Despite the conservation of the H4 diacetylation throughout the evolution, the actual function in histone nuclear import and/or chromatin assembly remains undetermined. In Drosophila embryos, the RCAF complex comprises ASF1, acetylated H3K14, and diacetylated H4K5K12 [60] and in human, the CAC complex is composed of diacetylated H4K5K12 and CAF-1 [69]. This highlights an essential role of this dual signature for the formation of a complex between H3/H4 and the major chaperones associated to replication. However, as revealed by the co-crystal structure, ASF1 interacts with the C-terminal region of H3 [70], so the precise role of the post-translational modifications is not obvious. Strikingly, all described chaperones so far do not interact with the unstructured tails of histones. To conclude, even if the requirement of the amino-terminal regions of the histones has been evidenced for the assembly of chromatin and/or regulation of histones, their precise involvements in the overall process still necessitate investigations.

### 4.2. Mechanism of chromatin reassembly

Albeit the two DNA strands run in opposite directions, the progression of the replication fork is unidirectional. To reconcile that, during the replication process one daughter strand is synthesized continuously (the leading strand) whereas the other (the lagging strand) is build by short stretches of DNA named Okazaki fragments, ligated afterwards. Does this particular mode of duplication have an impact on parental nucleosomes segregation ? Even if adjacent "old" histones tend to segregate together, no clear preference for the leading or lagging strand have been demonstrated, mainly because the studies did not clearly discriminate the two strands. A recent study suggests that nucleosome positioning onto the lagging-strand could determined the length of Okazaki fragment in *S. cerevisiae*, via interaction with the enzyme polymerase pol δ, responsible for the extension of the nascent DNA chain through the 5′ end of an Okazaki fragment [71]. By purifying Okazaki fragments, and performing the alignment onto the yeast genome, they demonstrated that they strikingly mapped with nucleosome positions. Once again, these experiments nicely illustrated the coupling between the DNA replication and the chromatin assembly.

The apparent higher sensitivity to nuclease digestion of newly synthesized chromatin compared to bulk chromatin suggests that new chromatin is not completely mature. Even though it was shown that specific post-translational modifications carried by newly synthesized histones and the absence of linker H1 histone could at least partially outline a more relaxed chromatin state, the reasons for the detection of the greater DNA accessibility in replicated chromatin remain actually elusive.

Newly synthesized H3/H4 are sequestered into the cytoplasm by ASF1, probably through interaction with several other chaperones, like the histone acetyltransferase HAT1, heat-shock proteins as HSC70 (Heat Shock Cognate 70 kDa protein), HSP90 (Heat Shock Protein 90), and NASP (Nuclear Autoantigenic Sperm Protein). The recent involvement of NASP as part of a cytosolic H3/H4 histone buffering complex is surprising, as this protein was initially described as an H1 chaperone [72, 73]. Indeed, in the nucleus ASF1 synergize with CAF1 via direct interaction with the p60 subunit. CAF1 was described to promote chromatin assembly *in vitro* [74]. This evolutionary conserved trimeric protein complex is recruited to site of DNA synthesis through interaction of the p150 subunit with the replication processivity factor PCNA, linking again chromatin assembly to replication fork progression [58]. As for parental histones, pioneer experiments using pulse-labeled histones suggested a sequential deposition of newly synthesized histones, with a H3/H4 tetramer assembled first, follow by the deposition of two H2A/H2B dimers.

The deposition model of nucleosomes, based on the stable tetrameric nature of histone H3/H4, was recently revisited [75]. Tagami and colleagues purified predeposition chromatin assembly complexes from HeLa cells stably expressing epitope-tagged histone H3.1 isoform (the replicative histone). The analyses of the immunoprecipitated tagged histones from purified nucleosomes and from the predeposition complexes showed that whereas about 50% of H3 in the nucleosomal fraction contained the epitope tag, all the histone complex in the predeposition complexes were tagged. It was thus concluded that H3/H4 complex is deposited onto DNA as dimer rather than tetramer. Biochemical, crystallographic and NMR analyses of ASF1 in complex with H3 (and sometimes H4) confirmed the dimeric nature of H3/H4 bound to the chaperone [70, 76, 77]. Furthermore, the structural data pointed out that the H3/H4 heterodimer binds ASF-1 at critical residues for H3/H3 interaction in the nucleosome, thus physically blocking the formation of a H3/H4 heterotetramer. This model has been reinforced by mutations of amino acids at critical regions. The dimeric nature of H3/H4 is also supported by a paper analyzing the composition of centromeric nucleosomes in the fruit fly *Drosophila*. At this particular genomic location, the nucleosome would exist in interphase as a stable tetramer, as a complex of single copy of CenH3-H2A-H2B and H4 has been identified [78].

## 5. Concluding remarks

The semi-conservative mode of replication of DNA ensures that the genetic information is faithfully transmitted to the daughter cells after mitosis. In higher eukaryotes, as the DNA is replicated, the chromatin environment has to be removed and subsequently restored. Here,

we have reviewed an overview of the actual mechanisms that can sustain this operation. The studies described and cited in this chapter are based upon different experimental approaches, which might potentially present caveats inherent to the experimentations. Even though profound advancements have been reported during the past few years to clarify the factors involved in the transport and delivery of histones, basic concerns still have to be unraveled.

It is generally believed that the histone post-translational modifications impact chromatin structure and the chromatin activities through the recruitment of different effectors and modulators. Beside the mechanistic comprehension of the process of DNA replication in the chromatin context, underlying question addressed is how the chromatin organization and the information carried by histones are maintained or altered during replication. Indeed, the demonstration of the link between chromatin replication and cell differentiation suggests that the S-phase is a window of great opportunity for modulating the epigenetic regulations in a genetic program. However, in this context, the alterations of the chromatin structure and the histone modifications have not yet been fully elucidated.Three models can emphasize the nucleosome reorganization behind replication fork (Figure 2): (A) the entire parental octamer is transferred to form nucleosome and newly synthesized histones fill up the gaps. (B) The parental nucleosome splits into building blocks composed of a tetramer of H3/H4 and dimers of H2A/H2B, and the blocks are redistributed onto the two strands of DNA. The new histones are utilized for achieving the formation of the octamer. (C) The recently advanced dimeric nature of H3/H4 paved a new avenue for future investigations. The splitting of the tetramer could lead to mixed nucleosome, composed of parental and new histones.
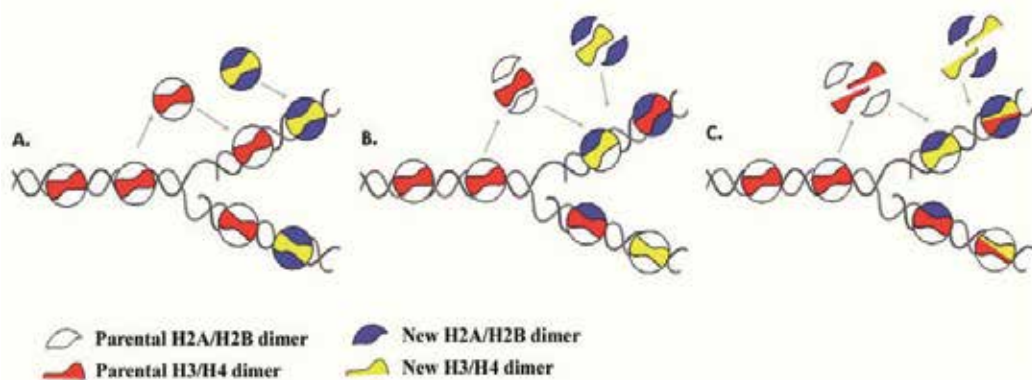


**Figure 2.** Working models of nucleosome reorganization during DNA replication. (A) Parental nucleosome is transferred as intact unit, without disruption of the octamer, leading to nucleosome fully constituted either of old or of new histones. (B) Parental nucleosome splits into H3/H4 tetramer and H2A/H2B dimers. In this model, new and old H3/H4 cannot coexist in the same nucleosome behind the replication fork. (C) Parental nucleosome splits into H3/H4 and H2A/H2B dimers, leading to mixed nucleosomes composed of old and new histones in each nucleosome building block.

In any considered model, the epigenetic information associated with the histone marks need to be copied from parental histones to newly synthesized ones. Concerning DNA methyla-

tion, the inheritance is a better-characterized process. In mammals, this modification mainly occurs on CpG dinucleotide (a cytosine followed by a guanine). The anti-parallelism of the DNA molecule, and the semi-conservative mode of DNA replication, ensure that the PCNA-interacting DNA methyltransferase DNMT1 easily copy the parental pattern onto the virgin daughter strand. To date, the mechanisms of the histone modification inheritance remains unclear. Most likely, future works in the field will attempt to address this issue.

## Acknowledgements

## Author details

Angélique Galvani[1,2,3] and Christophe Thiriet[1*]

*Address all correspondence to: Christophe.Thiriet@univ-nantes.fr

1 CNRS, UFIP (FRE 3478), Univ. of Nantes, Epigenetics: Proliferation and Differentiation, F-44322 Nantes, France

2 CNRS, UMR7216 Epigenetics and Cell Fate, F-75013 Paris, France

3 Univ Paris Diderot, Sorbonne Paris Cité, F-75013 Paris, France

## References

[1]  Mechali, M. (2001). Replicating. DNA in the Refractory Chromatin Environment. Nat Rev Genet, 2, 640-645.

[2]  van Holde KE(1989). Chromatin. Springer-Verlag, New-York.

[3]  Polo, S. E., & Almouzni, G. (2006). Replicating. DNA in the Refractory Chromatin Environment. Curr Opin Genet Dev, 16, 104-111.

[4]  Felsenfeld, G. ((1978). ) Chromatin. Nature ., 271, 115-122.

[5]  Pruss, D., Hayes, J. J., & Wolffe, A. P. (1995). Nucleosomal anatomy--where are the histones? Bioessays, 17, 161-170.

[6]  Arents, G., Burlingame, R. W., Wang, B. C., Love, W. E., & Moudrianakis, E. N. (1991). Replicating. DNA in the Refractory Chromatin Environment. Proc Natl Acad Sci U S A, 88, 10148-10152.

[7] Arents, G., & Moudrianakis, E. N. (1993). Replicating. DNA in the Refractory Chromatin Environment. Proc Natl Acad Sci U S A, 90, 10489-10493.

[8] Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Replicating. DNA in the Refractory Chromatin Environment. Nature, 389, 251-260.

[9] Davey, Sargent. D. F., Luger, K., Maeder, A. W., & Richmond, T. J. (2002). Replicating. DNA in the Refractory Chromatin Environment. J Mol Biol, 319, 1097-1113.

[10] Polach, K. J., & Widom, J. (1999). Replicating. DNA in the Refractory Chromatin Environment. Methods Enzymol, 304, 278-298.

[11] Protacio, R. U., Polach, K. J., & Widom, J. (1997). Replicating. DNA in the Refractory Chromatin Environment. J Mol Biol, 274, 708-721.

[12] Polach, K. J., & Widom, J. (1995). Replicating. DNA in the Refractory Chromatin Environment. J Mol Biol, 254, 130-149.

[13] Polach, K. J., Lowary, P. T., & Widom, J. (2000). Replicating. DNA in the Refractory Chromatin Environment. J Mol Biol, 298, 211-223.

[14] Lee KM & Hayes JJ. (1998). Replicating. DNA in the Refractory Chromatin Environment. *Biochemistry*, 37, 8622-8628.

[15] Ura, K., Hayes, J. J., & Wolffe, A. P. (1995). Replicating. DNA in the Refractory Chromatin Environment. Embo J, 14, 3752-3765.

[16] Hansen, J. C., van Holde, K. E., & Lohr, D. (1991). Replicating. DNA in the Refractory Chromatin Environment. J Biol Chem, 266, 4276-4282.

[17] Fletcher TM & Hansen JC(1995). Replicating. DNA in the Refractory Chromatin Environment. J Biol Chem, 270, 25359-25362.

[18] Tse, C., & Hansen, J. C. (1997). Replicating. DNA in the Refractory Chromatin Environment. *Biochemistry*, 36, 11381-11388.

[19] Wang, X., & Hayes, J. J. (2008). Replicating. DNA in the Refractory Chromatin Environment. Mol Cell Biol, 28, 227-236.

[20] Woodcock, C. L., & Dimitrov, S. (2001). Replicating. DNA in the Refractory Chromatin Environment. Curr Opin Genet Dev, 11, 130-135.

[21] Schlick, T., Hayes, J., & Grigoryev, S. (2012). Replicating. DNA in the Refractory Chromatin Environment. J Biol Chem, 287, 5183-5191.

[22] Carruthers, L. M., Bednar, J., Woodcock, C. L., & Hansen, J. C. (1998). Replicating. DNA in the Refractory Chromatin Environment. *Biochemistry*, 37, 14776-14787.

[23] Shen, X., Yu, L., Weir, J. W., & Gorovsky, . (1995). Replicating. DNA in the Refractory Chromatin Environment. *Cell*, 82, 47-56.

[24] Shen, X., & Gorovsky, . (1996). Replicating. DNA in the Refractory Chromatin Environment. *Cell*, 86, 475-483.

[25]  Fan, Y., Nikitina, T., Morin-Kensicki, E. M., Zhao, J., Magnuson, T. R., Woodcock, C. L., & Skoultchi, A. I. (2003). Replicating. DNA in the Refractory Chromatin Environment. Mol Cell Biol, 23, 4559-4572.

[26]  Dou, Y., Mizzen, Abrams. M., Allis, C. D., & Gorovsky, . (1999). Replicating. DNA in the Refractory Chromatin Environment. Mol Cell, 4, 641-647.

[27]  Dou, Y., & Gorovsky, . (2002). Replicating. DNA in the Refractory Chromatin Environment. Proc Natl Acad Sci U S A, 99, 6142-6146.

[28]  Bradbury, E. M., Inglis, R. J., Matthews, H. R., & Sarner, N. (1973). Replicating. DNA in the Refractory Chromatin Environment. Eur J Biochem, 33, 131-139.

[29]  Bradbury EM, Inglis RJ, Matthews HR & Langan TA. (1974). Replicating. DNA in the Refractory Chromatin Environment. *Nature*, 249, 553-556.

[30]  Alexandrow MG & Hamlin JL(2005). Replicating. DNA in the Refractory Chromatin Environment. J Cell Biol, 168, 875-886.

[31]  Contreras, A., Hale, T. K., Stenoien, D. L., Rosen, J. M., Mancini, , & Herrera, R. E. (2003). Replicating. DNA in the Refractory Chromatin Environment. Mol Cell Biol, 23, 8626-8636.

[32]  Thiriet, C., & Hayes, J. J. (2009). Replicating. DNA in the Refractory Chromatin Environment. J Biol Chem, 284, 2823-2829.

[33]  Bonne-Andrea, C., Wong, M. L., & Alberts, . (1990). Replicating. DNA in the Refractory Chromatin Environment. *Nature*, 343, 719-726.

[34]  Krude, T., & Knippers, R. (1991). Replicating. DNA in the Refractory Chromatin Environment. Mol Cell Biol, 11, 6257-6267.

[35]  Sogo, J. M., Stahl, H., Koller, T., & Knippers, R. (1986). Replicating. DNA in the Refractory Chromatin Environment. The replication fork, core histone segregation and terminal structures. J Mol Biol, 189, 189-204.

[36]  Jackson, V. (1990). Replicating. DNA in the Refractory Chromatin Environment. *Biochemistry*, 29, 719-731.

[37]  Annunziato AT(2005). Split decision: what happens to nucleosomes during DNA replication? J Biol Chem, 280, 12065-12068.

[38]  Kimura, H., & Cook, P. R. (2001). Replicating. DNA in the Refractory Chromatin Environment. J Cell Biol, 153, 1341-1353.

[39]  Baxevanis AD, Godfrey JE & Moudrianakis EN(1991). Associative behavior of the histone (H3-H4)2 tetramer: dependence on ionic environment. Biochemistry, 30, 8817-8823.

[40]  Gruss, C., & Sogo, J. M. (1992). Chromatin replication. Bioessays, 14, 1-8.

[41] Gruss, C., Wu, J., Koller, T., & Sogo, J. M. (1993). Replicating. DNA in the Refractory Chromatin Environment. Embo J, 12, 4533-4545.

[42] Ramsperger, U., & Stahl, H. (1995). Replicating. DNA in the Refractory Chromatin Environment. Embo J, 14, 3215-3225.

[43] Wilhelm, F. X., Wilhelm, M. L., Erard, M., & Duane, M. P. (1978). Replicating. DNA in the Refractory Chromatin Environment. Nucleic Acids Res, 5, 505-521.

[44] Groth, A., Corpet, A., Cook, A. J., Roche, D., Bartek, J., Lukas, J., & Almouzni, G. (2007). Regulation of replication fork progression through histone supply and demand. Science, 318, 1928-1931.

[45] Belotserkovskaya, R., Oh, S., Bondarenko, V. A., Orphanides, G., Studitsky, V. M., & Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science, 301, 1090-1093.

[46] Tan, B. C., Chien, C. T., Hirose, S., & Lee, S. C. (2006). Functional cooperation between FACT and MCM helicase facilitates initiation of chromatin DNA replication. EMBO Jsj.emboj.7601271., 25, 3975-3985.

[47] Abe, T., Sugimura, K., Hosono, Y., Takami, Y., Akita, M., Yoshimura, A., Tada, S., Nakayama, T., Murofushi, H., Okumura, K., et al. (2011). The histone chaperone facilitates chromatin transcription (FACT) protein maintains normal replication fork rates. J Biol Chem, 286, 30504-30512.

[48] Glikin, G. C., Ruberti, I., & Worcel, A. (1984). Chromatin assembly in Xenopus oocytes: in vitro studies. Cell, 37, 33-41.

[49] Collins, N., Poot, R. A., Kukimoto, I., Garcia-Jimenez, C., Dellaire, G., & Varga-Weisz, P. D. (2002). An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. Nat Genet, 32, 627-632.

[50] Poot, R. A., Bozhenok, L., van den, Berg. D. L., Hawkes, N., & Varga-Weisz, P. D. (2005). Chromatin remodeling by WSTF-ISWI at the replication site: opening a window of opportunity for epigenetic inheritance? Cell Cycle, 4, 543-546.

[51] Marzluff WF & Duronio RJ(2002). Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. Curr Opin Cell Biol, 14, 692-699.

[52] Gunjan, A., Paik, J., & Verreault, A. (2005). Regulation of histone synthesis and nucleosome assembly. Biochimie, 87, 625-635.

[53] Loidl, P., & Grobner, P. (1987). Histone synthesis during the cell cycle of Physarum polycephalum. Synthesis of different histone species is not under a common regulatory control. J Biol Chem, 262, 10195-10199.

[54] Talbert, P. B., & Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol, 11, 264-275.

[55] Hardy, S., & Robert, F. (2010). Random deposition of histone variants: A cellular mistake or a novel regulatory mechanism? Epigenetics, 5, 368-372.

[56] Hamiche, A., & Shuaib, M. (2012). Chaperoning the histone H3 family. Biochim Biophys Acta, 1819, 230-237.

[57] Mousson, F., Ochsenbein, F., & Mann, C. (2007). The histone chaperone Asf1 at the crossroads of chromatin and DNA checkpoint pathways. Chromosoma, 116, 79-93.

[58] Hoek, M., & Stillman, B. (2003). Chromatin assembly factor 1 is essential and couples chromatin assembly to DNA replication in vivo. Proc Natl Acad Sci U S A, 100, 12183-12188.

[59] Groth, A., Ray-Gallet, D., Quivy, J. P., Lukas, J., Bartek, J., & Almouzni, G. (2005). Human Asf1 regulates the flow of S phase histones during replicational stress. Mol Cell, 17, 301-311.

[60] Tyler, J. K., Adams, C. R., Chen, S. R., Kobayashi, R., Kamakaka, R. T., & Kadonaga, J. T. (1999). The RCAF complex mediates chromatin assembly during DNA replication and repair. Nature, 402, 555-560.

[61] Gunjan, A., & Verreault, A. (2003). A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in S. cerevisiae. Cell, 115, 537-549.

[62] Ejlassi-Lassallette, A., & Thiriet, C. (2012). Replication-coupled chromatin assembly of newly synthesized histones: distinct functions for the histone tail domains (1) (1) This article is part of Special Issue entitled Asilomar Chromatin and has undergone the Journal's usual peer review process. Biochem Cell Biol.

[63] Mosammaparast, N., Jackson, K. R., Guo, Y., Brame, C. J., Shabanowitz, J., Hunt, D. F., & Pemberton, L. F. (2001). Nuclear import of histone H2A and H2B is mediated by a network of karyopherins. J Cell Biol, 153, 251-262.

[64] Mosammaparast, N., Guo, Y., Shabanowitz, J., Hunt, D. F., & Pemberton, L. F. (2002). Pathways mediating the nuclear import of histones H3 and H4 in yeast. J Biol Chem, 277, 862-868.

[65] Thiriet, C., & Hayes, J. J. (2001). A novel labeling technique reveals a function for histone H2A/H2B dimer tail domains in chromatin assembly in vivo. Genes Dev, 15, 2048-2053.

[66] Ejlassi-Lassallette, A., Mocquard, E., Arnaud, M. C., & Thiriet, C. (2011). H4 replication-dependent diacetylation and Hat1 promote S-phase chromatin assembly in vivo. Mol Biol Cell, 22, 245-255.

[67] Kouzarides, T. (2007). Chromatin modifications and their function. Cell, 128, 693-705.

[68] Annunziato AT & Hansen JC(2000). Role of histone acetylation in the assembly and modulation of chromatin structures. Gene Expr, 9, 37-61.

[69] Verreault, A., Kaufman, P. D., Kobayashi, R., & Stillman, B. (1996). Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. Cell, 87, 95-104.

[70] English CM, Adkins MW, Carson JJ, Churchill ME & Tyler JK(2006). Structural basis for the histone chaperone activity of Asf1. Cell, 127, 495-508.

[71] Smith, D. J., & Whitehouse, I. (2012). Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature, 483, 434-438.

[72] Ransom, M., Dennehey, B. K., & Tyler, J. K. (2010). Chaperoning histones during DNA replication and repair. Cell, 140, 183-195.

[73] Hondele, M., & Ladurner, A. G. (2011). The chaperone-histone partnership: for the greater good of histone traffic and chromatin plasticity. Curr Opin Struct Biol, 21, 698-708.

[74] Smith, S., & Stillman, B. (1989). Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. Cell, 58, 15-25.

[75] Tagami, H., Ray-Gallet, D., Almouzni, G., & Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell, 116, 51-61.

[76] Agez, M., Chen, J., Guerois, R., van Heijenoort, C., Thuret, J. Y., Mann, C., & Ochsenbein, F. (2007). Structure of the histone chaperone ASF1 bound to the histone H3 C-terminal helix and functional insights. Structure, 15, 191-199.

[77] Natsume, R., Eitoku, M., Akai, Y., Sano, N., Horikoshi, M., & Senda, T. (2007). Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. Nature, 446, 338-341.

[78] Dalal, Y., Wang, H., Lindsay, S., & Henikoff, S. (2007). Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. PLoS Biol5, e218.

# Telomeres

# Telomeres: Their Structure and Maintenance

Radmila  Capkova Frydrychova and James M. Mason

Additional information is available at the end of the chapter

## 1. Introduction

Telomeres are essential nucleoprotein structures at the ends of eukaryotic chromosomes. They play several essential roles preserving genome stability and function, including distinguishing chromosome ends from DNA double stranded breaks (DSBs) and maintenance of chromosome length. Due to the inability of conventional DNA polymerases to replicate the very end of a chromosome, sometimes known as the end replication problem, chromosome ends shorten with every round of DNA replication. In the absence of special telomere maintenance mechanisms this telomere shortening leads to replicative senescence and apoptosis. Several telomere maintenance mechanisms have been identified; these are reflected in several known types of telomeres. In most eukaryotes telomeres comprise a tandem array of a short, 5-8 bp, well conserved repeat unit, and telomere length is maintained by telomerase, a specialized reverse transcriptase that carries its own RNA template and adds telomeric sequences onto chromosome ends [1]. Nevertheless, in some organisms the array of short telomeric sequence motifs has been replaced with less conventional sequences, such as satellite sequences or transposable elements. The telomeres of such organisms are maintained through homologous recombination or through transposition of the mobile elements [2,3]. These different telomere types present distinct difficulties for chromosome end protection. Telomeres maintained by telomerase are protected by a proteinaceous telomere cap, termed shelterin, that recognizes chromosome ends in a DNA sequence specific manner, while telomeres with long terminal repeat units are protected by a cap, termed terminin, that binds to chromosome ends independently of DNA sequence.

## 2. The structure of telomeric DNA: "usual" and "unusual" telomeres

The most common telomere structure found across the whole eukaryotic tree is a simple telomeric repeat of the form $(T_xA_yG_z)_n$ generated by telomerase. For example, the sequence in

unikonts generally, including animals, fungi and amoebozoa, is $T_2AG_3$, while in most plants and green algae it is $T_3AG_3$. Within these broad generalizations, however, there are exceptions. Some species seem to have lost the canonical telomeric motif altogether. We will mention a few examples here, then describe one of these examples in more detail.

## 2.1. Chromalveolata

The terminal sequence motif seems to be quite variable among the Chromalveolates, while still adhering to the consensus telomeric motif (Figure 1). Apicomplexa species use three different motifs [4-6], and ciliates use two [4,7]. Dinoflagellates use $T_3AG_3$ [8], similar to plants and green algae, while diatoms use $T_2AG_3$ [9], similar to unikonts. Photosynthetic species in the Chromalveolates are derived from the engulfment of a red alga. The resulting nucleomorphs retain the algal linear chromosomes and telomeres that are very different. The cryptomonad, *Guillardia theta*, for example, uses $T_3AG_3$ in its nucleus and $(AG)_7A_2G_6A$ in its nucleomorph [11,12].
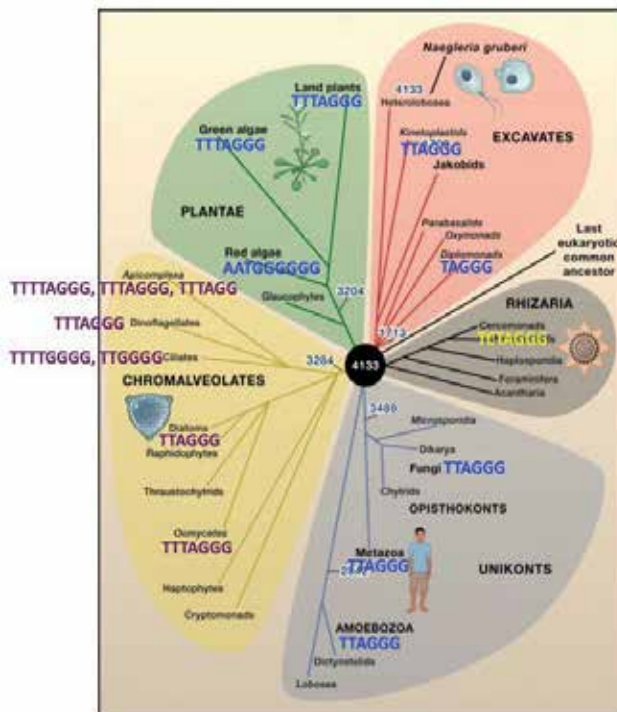


**Figure 1.** Diagram showing five major eukaryotic supergroups and representative telomeric motifs. These groups are shown to have diverged from a single latest common ancestor, because the evolutionary relationships are not known. Trees connecting the major taxa within these supergroups are shown, but the branch lengths are arbitrary. Representative telomeric motifs are shown for the major subtaxa. In some cases, two or three representative motifs are known for one of these taxa, as shown. Exceptions to these representations are discussed in the text. The figure was modified from [10] with permission.

It seems likely that the telomere binding proteins in these organisms are either different in the two intracellular bodies, or do not bind in a DNA sequence specific manner.

## 2.2. Plantae

Among the Plantae, land plants and green algae mostly use $T_3AG_3$ as a telomeric motif, while the red algae have a very different sequence at their chromosome ends. The red alga *Cyanidioschyzon merolae*, for example, uses $A_2TG_6$ [13]. While telomeres in most green algae conform to the telomeric motif of this kingdom, the order Chlamydomonadales includes species that carry the telomeric motifs $T_4AG_3$, $T_3AG_3$ and $T_2AG_3$, apparently independently of phylogeny as determined by the 18S rDNA sequence [14]. Some species of this order lack all three of these telomeric motifs and carry unknown DNA sequences at their chromosome ends. It is possible that the 18S rDNA sequence does not represent an accurate reflection of phylogeny or the telomeric motif is quite variable in this order. In either case, it seems that sequence specific binding by telomeric proteins may have eased in this order.

Similarly, while most land plants retain the canonical $T_3AG_3$ telomeric motif, telomeres in a few orders differ from this structure. Within the monocot order Asparagales some species of Alliaceae have switched to the sequence $T_2AG_3$, and others appear to have lost the canonical telomeric sequence completely. It has been proposed that the telomeres of these latter Alliaceae species are maintained through transposition of mobile elements or through homologous recombination between the satellite sequences [15,16]. In the eudicot order Solanales the canonical telomeric motif as well as telomerase are absent from several genera of the family Solanaceae [17-20]. The actual telomeric sequence and compensation mechanism in this group of plants, however, remain unknown.

## 2.3. Unikonta

The $T_2AG_3$ telomeric sequence is found widely among the unikonts (Figure 1). While this is generally true within the fungi, representatives of two classes, Schizosaccharomycetes and Saccharomycetes, use variable, degenerate telomeric sequences that may result from replication infidelity or slippage [12]. In *Saccharomyces cerevisiae*, for example, the repeat motif is $TG_{1-3}$.

Similarly, $T_2AG_3$ is found widely at chromosome ends among metazoans. The animal phylum Arthropoda, however, uses the sequence $T_2AG_2$ at telomeres, and its sister phylum Tardigrada lacks both of these telomeric motifs [21]. Insects are the largest class of arthropods, and even here individual insect taxa may have different forms of the canonical sequence or even unrelated telomeric sequences. Insects seem to have lost the canonical arthropod telomeric motif several times [22,23]. In some cases, such as the coleopteran superfamily Tenebrionoidea, the arthropod repeat has been replaced by a similar motif, in this case $TCAG_2$ [24], while in many other instances the new telomeric DNA sequence has not been identified.

Insects of the orders Diptera, Mecoptera and Siphonaptera (superorder Antliophora) do not carry a canonical telomeric DNA sequence at their chromosome ends [23,25]; nor do they have a telomerase gene [26], indicating that telomerase was lost some 260-280 Mya. Even so, Diptera is one of the most successful insect orders, with some 152,000 species [27]. This

suggests that telomerase and the canonical telomeric DNA sequences generated by telomerase, *per se*, are not critical for evolutionary survival. It is possible telomerase is expendable, as long as the telomere capping complex is compatible with whatever terminal DNA sequence is present on chromosome ends. When the primary pathway for telomere replication is defective, an alternative backup mechanism can restore telomere function. It was documented in yeast. Yeast mutants lacking telomerase showed the progressive telomere loss and, although the majority of the cells died, a minor subpopulation survived via homologous recombination [28].

Long satellite sequences have been reported in nematoceran species. Chromosome tips of several *Chironomus* species (infraorder Culicomorpha) consist of large, 50-200 kb, blocks of complex, tandemly repeated sequences that have been classified into subfamilies based on sequence similarities. Different telomeres display different sets of subfamilies, and the distribution of subfamilies differs between individuals within a species. The variation of the satellite sequences supports the proposal that telomeres in *Chironomus* are elongated by a gene conversion mechanism involving these long blocks of complex repeat units [29-32]. A similar situation has been found in *Anopheles gambiae* (infraorder Culicomorpha) using a plasmid fortuitously inserted into the complex telomeric sequences at the tip of chromosome 2L. The telomere carrying the plasmid was found to engage in frequent recombination events that resulted in extension of the terminal array [33,34]. Recently, a similar case was reported in *Rhynchosciara americana* (infraorder Bibionomorpha). Tandem arrays of short repeats, 16 and 22 bp in length, were found to extend to chromosome ends [35]. Although telomere elongation could not be assayed in this case, it seems likely that the mechanism is similar to that seen in other nematoceran species. In many respects, these complex arrays resemble subtelomeric sequences [36], suggesting a possible mechanism for telomere formation upon the loss of telomerase.

Telomere structures have only been examined in a single brachyceran genus, *Drosophila* (infraorder Muscomorpha). Telomeric DNA sequences consist of long arrays of non-long terminal repeat (LTR) retrotransposons and are thus very different from those found in Nematocera. These telomeric transposons resemble long interspersed elements (LINEs) found in mammals, but have some differences that may reflect their telomere-specific 'lifestyle.' Three families of telomeric elements have been described in *Drosophila melanogaster* (subgenus Sophophora), *HeT-A*, *TART* and *TAHRE* [2]; in all cases these elements are attached to the chromosome by their 3' oligo(A) tails. Many of the elements are truncated at the 5' end, possibly due to the end replication problem. *HeT-A* transposons are about 6 kb in length and make up about 80-90% of the elements found at chromosome ends. They are atypical LINE-like elements in three respects: the 3' untranslated region (UTR) comprises about 3 kb or half the length of the element; the transcriptional promoter is at the 3' end of the element to prevent loss when the element is present at the chromosome terminus with its 5' end exposed to incomplete DNA replication of linear DNA; and an open reading frame (ORF) coding for a reverse transcriptase is absent. *TART* elements are about 10 kb in length and make up about 10-20% of the telomeric retrotransposons. They are also unusual elements, but in some ways that differ from *HeT-A*: they also have an unusually long 3' UTR; they have a relatively strong antisense promoter of unknown function and a pair of perfect

long non-terminal repeats that may be important for replication [37,38]; they make a reverse transcriptase, but the encoded Gag-like protein is unable to target telomeres in the absence of the *HeT-A* Gag [39]. *TAHRE* elements closely resemble *HeT-A*, except they encode their own reverse transcriptase. Thus, while *TAHRE* seems to be the only one of the three elements capable of independent transposition, it is by far the least abundant, comprising only 1% of the telomeric retrotransposons.

*HeT-A* and *TART* elements have also been found in *Drosophila virilis* (subgenus Drosophila) Although there is little sequence homology across species, the two types of retrotransposons can be recognized by their telomeric locations and unusual structures, as described above [40,41]. Given the difficulty in finding homology between evolutionarily related telomeric elements within the *Drosophila* genus, finding similar elements in other brachyceran species based on homology alone is unlikely. Thus, it is not known when these targeted transpositions took over the role of telomere maintenance from homologous recombination.

Human telomeres have been shown to form a large terminal loop dependent on the presence of a 3' G strand overhang at the telomeric end. This 3' end is tucked back into the double-stranded DNA as a loop, termed a t-loop [42]. Similar t-loops may also be formed in yeast [43].

## 3. Proteins associated with telomeres

The telomere cap, a multiprotein structure at chromosome end ensuring stability and integrity of the genome, was revealed by early cytological observations of chromosomal rearrangements after exposure to ionizing radiation [44]. The telomere cap allows cells to distinguish their natural chromosome ends from DSBs, thus protecting the chromosome termini from inadvertent DNA damage response (DDR) activities. Defects in the cap, or DSBs elsewhere in the genome, lead to activation of cell cycle checkpoints followed by DDR mechanisms. A consequence of inappropriate DSB repair are end-to-end fusions of chromosomes, i.e. formation of ring chromosomes or dicentric linear chromosomes, followed by chromosome breakage, which results in genomic instability and loss of cellular viability [45,46]. Although, in this context, telomeres perform the same essential function across phyla, cap proteins of diverse organisms are less conserved that one might expect. Even within a single taxonomic class, such as mammals, telomeric proteins display less conservation than other chromosomal proteins [47]. In mammals the telomere-specific cap complex has been termed 'shelterin' (Figure 2). The six-protein complex is formed by double-stranded TTAGG repeat-binding proteins TRF1 and TRF2, which recruit TIN2 and TPP1. The latter proteins make a bridge between the TRF proteins and G-overhang-binding protein, POT1. The sixth protein is the TRF2-interacting protein RAP1 [46,48,49]. A characteristic of shelterin proteins is specific and exclusive association with telomeric DNA, where they are permanently present throughout the cell cycle and serve as platform for a transient and dynamic recruitment of a number of telomere-associated factors, referred to as non-shelterin telomeric proteins. These non-shelterin proteins are required for telomere protection and replication but also have nu-

merous nontelomeric functions. Examples include DDR proteins that are commonly involved in DSB repair through nonhomologous end joining (NHEJ) or homologous recombination (HR), such as ATM, ATR and Ku70/80, which associate with TRF1 and TRF2, and the MRN complex, composed of the MRE11, RAD50 and NBS1 (MRN) proteins, which associates with TRF2 [50-55]. Another protein associated with TRF2 is Apollo, an exonuclease important for recreating the 3' overhang [51,56]. The binding of shelterin proteins and formation of a functional cap require a terminal DNA array of specific sequence and of satisfactory length.

Analysis of deleterious events at shelterin-free telomeres revealed six pathways for end protection [57]. The primery protection by shelterin is against classical NHEJ and unwanted activation of ATM and ATR signaling. Additionally, shelterin provides a defense against alternate NHEJ, HR and 5' end resection. Another protective layer is achieved through the Ku70/80 heterodimer or 53BP1. 53BP1 minimizes resection but only at telomeres eliciting a DNA damage signal. Ku70/80 blocks alternate NHEJ and HR at telomeres independent of a DNA damage signal [57].
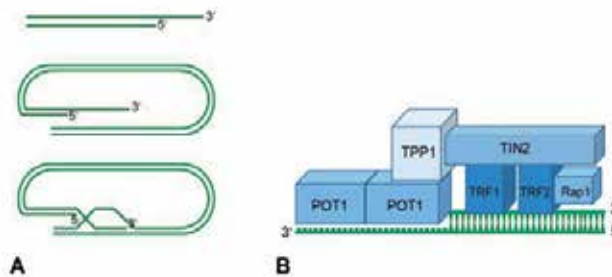


**Figure 2.** A. The telomere forms a t-loop structure characterized by invasion of the 3' overhang into a double stranded telomeric DNA. B. Six proteins, TRF1, TRF2, TPP1, POT1, TIN2, and RAP1 form a dedicated telomere-protection protein complex in humans [48,49,58].

Telomeres in *Saccharomyces cerevisiae* are protected by two separate protein complexes. One is the Rap1/Rif1/Rif2 complex, which localizes to double-stranded telomeric DNA. The other is the Cdc13/Stn1/Ten1 (CST) complex, which is targeted to the single-stranded G-overhangs through sequence-specific binding of Cdc13. Defects in the CST complex result in degradation of the C-stand and activation of DDR mechanisms [47]. As with shelterin, CST interacts with numerous proteins required for telomere function. Some evolutionary conservation in the protein composition of the cap is expected, for instance similarities to CST and shelterin are observed in telomeric proteins of numerous organisms. This is documented for mammalian CST, which, although not involved in telomere capping, facilitates telomere replication and, if impaired, leads to catastrophic telomeric defects [59]. Another example is Ver, a component of the *Drosophila* cap with weak structural similarities to Stn1 [60,61].

A multiprotein capping complex in *Drosophila*, termed 'terminin,' is an analog of mammalian shelterin [61]. One major difference between these two complexes is that terminin does

not bind to a specific telomeric DNA sequence. Rather limited information is available about the structure and function of four known terminin proteins, HOAP, Moi, Ver, and HipHop. As with shelterin, terminin proteins localize specifically to telomeres and appear to function only at telomeres. HOAP is encoded by the *cav* gene [62]; Moi is a HOAP-binding protein encoded by *moi* [63,64]; Ver is structurally homologous to STN1 and is encoded by *ver* [60]; and HipHop is a HP1-HOAP interacting protein [65]. Assembly of the terminin complex requires strict dependencies. For example, the binding of HOAP and HipHoP to telomeres is interdependent, loss of one protein reduces binding of the other [65]; HOAP is required for Ver and Moi localization [61]. The terminin complex seems to occupy a broad region covering a more than 10 kb from the chromosome termini [65]. As with shelterin proteins, defects in terminin proteins lead to frequent telomeric fusions.

As there is no specific telomeric DNA sequence in Drosophila, terminin binding to telomeric DNA is sequence-independent, which makes a substantial difference between mammalian and Drosophila telomeres. In contrast to mammals, the complete loss of a Drosophila telomere does not definitely mean inescapable damage to genome stability and cell death, because under the right circumstances the telomere cap can be formed *de novo* as on any broken chromosome end and perform there the same protective functions as the regular telomere. This demonstrates that the telomeric retrotransposons, although important for telomere elongation, are not required as an unique assembly platform for cap formation [2,66,67].

Similar to shelterin, terminin presents a docking site for binding of additional proteins, called non-terminin capping proteins. Although not exclusively located at telomeres and having some telomere-unrelated functions, these proteins are required for the capping function and, in many cases, facilitate terminin assembly. There are several known non-terminin proteins; most of them were identified because their mutants display frequent telomeric fusions [61]. The best characterized is HP1a that is encoded by *Su(var)205*. The presence of HP1 at telomeres is required for HOAP binding, which reveals the importance of HP1 for terminin assembly. As in mammals, other non-terminin proteins are DNA repair factors: the *Drosophila* homologs of the ATM kinase and proteins of the MRN complex. Defects in the MRN complex lead to reduction of HOAP and HP1 at telomeres and frequent telomeric fusions. Through its effects on the binding of HOAP and possibly other terminin components, the MRN complex seems to be essential for the terminin formation [61,68,69]. ATM prevents telomeric fusions, and defects in this protein partially affect HP1/HOAP localization [70-72]. Woc is a zinc-finger protein preventing telomeric fusions, but acting independently of HP1, HOAP, and RAD50 [73]. UbcD1 is an E2 ubiquitin conjugating enzyme. It has been suggested that UbcD1-mediated ubiquitination of telomeric proteins is an essential post-translational modification ensuring their proper function [61,74]. In contrast to non-terminin and non-shelterin proteins that are largely conserved, a comparison between shelterin and terminin reveals no obvious homology in protein composition. Loss of conservation between shelterin and terminin proteins may correspond the evolutionary stage when a Antliophoran ancestor lost telomerase-based telomere elongation and had to evolve a sequence-independent protection of chromosome ends and acquire a new elongation system.

A highly condensed chromatin structure is a common characteristic of telomeres from yeast to man. Usually telomeres are heterochromatic, and the heterochromatic properties are thought to play an important role in telomeric function [75-77]. Telomeric chromatin is the source of telomeric position effect (TPE), a silencing of transgenes inserted into telomeres or their vicinity [78]. Besides the cap region, *Drosophila* telomeres contain two distinct chromatin domains: a subtelomeric region of repetitive DNA, termed TAS (telomere associated sequence), exhibiting features that resemble heterochromatin, and a terminal array of retrotransposons with euchromatic characteristics [79]. The *Drosophila* TAS region is, in contrast to retrotransposon array, the source of TPE [79,80]. Although organized into a heterochromatic structure, the vertebrate TTAGGG sequence remains unmethylated due to the lack of a appropriate cytosine substrate. The subtelomeric region is, in contrast, heavily methylated by DNA methyltranferases DNMT1, DNMT3a and DNMT3b [81]. Both in vertebrates and *Drosophila*, telomeric and subtelomeric regions are enriched in histone H3 methylated at lysine 9 (H3K9me), mediated by a H3K9-specific histone methyltransferase and HP1.

## 4. Telomeric replication and its difficulties

Based on DNA and protein composition, telomeres are typical heterochromatin, so their replication should correspond with a common paradigm of late heterochromatin replication. Based on early microscopic studies, it is generally accepted that DNA replication at early stages of S phase is associated with expressed genes, whereas repressed tissue-specific genes or heterochromatic regions are replicated during the late stages of replication [82-84]. The late replication seems to be common, but definitely is not universal [85]. Replication of human telomeres takes place throughout S phase, and specific telomeres tend to replicate at defined stages, some replicating early and others late [86]. The pattern of replication timing seems to be conserved between homologous chromosomes and does not vary between cells of different individuals. Although no correlation was found with telomere length or telomerase activity, a strong association was observed with nuclear localization. Late-replicating telomeres show a preferential association with the nuclear periphery, while early-replicating telomeres are preferentially located near the nuclear center [86]. A different situation was found in budding yeast, *Saccharomyces cerevisiae*, where early telomere replication correlates with short telomeric length and telomerase activity [87,88]. In fission yeast, *Schizosaccharomyces pombe*, telomere replication corresponds to S/G2 phase [85,89].

Because of the repetitive nature of telomeric DNA, telomeres present a significant problem for their replication. Spontaneous replication fork regression in telomeric DNA *in vitro* was determined to be 41% higher than seen in non-repeated DNA [90]. The obstacles during replication may lead to formation of cruciform intermediates, resulting in unwanted recombination events, amplifications or deletions [90,91]. Most of the telomere is replicated by a standard replication fork, however, to achieve efficient telomere replication a number of additional steps are needed. The process requires cooperation between standard replication factors and telomeric proteins, DDR proteins and numerous additional fac-

tors [47]. Examples of additional proteins are RecQ-type helicases that are present at replication forks in addition to standard helicases and are shown to unwind structures similar to chickenfoot intermediates [90,92]. Cooperation of replication factors with shelterin proteins is also documented. TRF1 mutants showed a reduction in replication efficiency, suggesting that TRF1 promotes efficient replication of telomeric DNA by preventing fork stalling [93]. Similarly, Taz1, a TRF1 homolog in fission yeast, has been shown to prevent fork stalling [94]. Another example is mammalian CTC1; deletion of CTC1 results in increased loss of leading C-strand telomeres, dramatic telomere loss and accumulation of excessive single-stranded telomeric DNA [95].

In yeast, the replication of telomeres is initiated in subtelomeric regions, and the replication fork moves towards the chromosome termini [96]. In mammalian cells, the origin of telomeric replication and direction is ambiguous.

After the replication fork reaches the chromosome terminus, the lagging strand gains a 3' overhang due to the removal of the primer for the terminal Okazaki fragment. At the same time C-strand specific resection occurs by nucleases Exo1 and/or Dna2 to produce a G-overhang on the leading strand [89]. If active, telomerase elongates the G-overhangs by addition of new telomeric repeats. Telomerase action is followed by complementary C-strand synthesis by DNA pol $\alpha$. The process is terminated by additional processing to remove the RNA primer and to leave a 40-400 nucleotide G-overhang. The timing of the events differs between species. In human cells, telomere replication occurs at the same time as telomerasemediated extension, and fill-in synthesis of C- strand is delayed until S/G2. Budding yeast shows tight coupling between G-strand extension and C-strand synthesis [89,91].

## 5. The mechanisms of telomeric elongation and their regulation

Telomerase is a ribonucleoprotein reverse transcriptase that utilizes its protein subunit (TERT in mammals, Est2p in *S. cerevisiae*) to elongate the 3' end of telomeric DNA using an internal RNA subunit (TR) as a template [97-99]. Telomerase activity is related to cell proliferation status: it is high in actively cycling cultures and low in quiescent differentiated cells [100]. Telomerase is not detected in human mature sperm or unfertilized eggs, but after fertilization telomerase is rapidly activated. A dramatic increase is observed in blastocysts, but during later stages of gestation telomerase activity declines. In the 16-week fetus Wright [101] showed high levels of telomerase in liver and intestine; detectable activity in lung, skin, muscle, adrenal glands, and kidney; and very weak or no activity in brain, bone or placental extracts. Most somatic cells in adults show no telomerase activity, as enzyme activity is limited to specific types of proliferative cells, such as embryonic, stem and epithelial cells, the germline, or cells of the hematopoetic system [102,103]. Telomerase activity is highly regulated. Reactivation of telomerase is associated with tumor development, and conversely, insufficient telomerase activity is linked to stem cell diseases, such as dyskeratosis congenita and aplastic anemia [104-106].

Telomerase is regulated through genetic, epigenetic and environmental factors: TERT and TR transcription, posttranscriptional and posttranslational modifications of TERT, and telomerase recruitment and processivity [104]. TERT promoter activity has been studied extensively, and numerous transcription factors have been found to interact with TERT. TERT transcription is, for instance, activated by the oncogene c-Myc and suppressed by the tumor-suppressor WT1 (Wilm's tumor suppresor). Misregulation of TERT through the c-Myc or WT1 pathways is associated with telomerase reactivation in cancer cells [107,108]. Although transcription of TERT is the major determinant of telomerase activity, TERT transcript levels do not always correlate with enzyme activity. Posttranslational phosphorylation may regulate telomerase activity, as may telomerase degradation through ubiquitination, as the half-life of telomerase activity was approximately 24 hours [109]. In human cells the POT1-TPP1 complex was found to be a key regulator of telomerase processivity [110-113].

Little is known about the regulation of telomere length in *Drosophila*, where two modes of telomere elongation have been described: transposition of telomeric elements and gene conversion. The process of telomeric transposition is composed of several steps: 1. transcription of the telomeric elements, 2. export of retroelement transcripts from the nucleus to the cytoplasm, 3. translation, 4. nuclear re-import of the retroelement transcripts together the retroelement proteins, 5. recognition of chromosome ends, and 6. target-primed reverse transcription, which attaches the 3' oligo(A) tails of the elements to chromosome termini [2]. Transposition of these elements to chromosome termini does not depend on a specific DNA sequence at the target site and together with the loss of telomeric DNA results in tandem arrays of mixed complete and 5' truncated elements [2]. The regulation of telomere elongation may be on the level of retroelement transcription and/or accessibility of the chromosome ends for new retroelement attachments. A variety of proteins have been identified to play a role in *Drosophila* telomere capping, however, only a few proteins are known to function in telomere elongation. HP1 was found to have a dual role in telomere protection and telomere length control. Compared to wild-type, heterozygotous *Su(var)205* mutants displayed much longer telomeres associated with a dramatic increases in retroelement transcription and transposition [114-116]. The regulation of retroelement transcription by HP1 was observed along the terminal retrotransposon array, thus this HP1 function is not limited to the telomere cap [117]. No, or only minor, changes were observed in telomere length or retroelement transcription in mutants of genes involved in telomere capping, such as *cav*, *moi*, *ver* or *atm* [61,117], which may indicate that terminin does not control telomere length. Another gene regulating telomere length is *prod*. Although *prod* mutants showed elevated levels of *HeT-A* transcripts, no change in telomere length was observed, suggesting that elevated retroelement transcription is not sufficient for telomere length growth [118]. Similar data were observed for members of rasiRNA (repeat-associated small interfering RNA) pathway *aub* (*aubergine*) and *Spn-E*. Their mutants displayed higher *HeT-A* transcript levels [119], albeit without any significant increase in telomere length (our unpublished data). In parallel with telomerase activity, transcription of telomeric elements is observed only in proliferating cells, such as embryonic cells, cells of imaginal discs, testis and ovaries [120,121].

Telomere length is maintained through an interplay between telomere maintenance mechanisms and telomere shortening events. Based on human research it has been proposed that telomerase activity and telomere length are modulated by different endogenous and exogenous factors, such as emotional or physical stress, health conditions and aging [102]. However, the prime factor in telomeric shortening may well be oxidative stress. Due to a high content of guanines, telomeres are particularly vulnerable to oxidative damage, and the impact of oxidative stress on telomeric length has been proposed to be much larger than the end-replication problem [122]. Endogenous oxidative stress is associated with several cellular processes, such as the mitochondrial OXPHOS system and inflammation. Mitochondrial dysfunction-induced reactive oxygen species and hyperoxia *in vitro* lead to accelerated telomere shortening and reduced proliferative lifespan of cultured somatic cells [123]. Thus, short telomeric length in humans appears to be linked to the limited proliferative capacity of normal somatic cells, and it is likely that telomeric shortening is one of the key events related to cellular senescence and organismal aging. As telomeres shorten with age, telomere length is considered as a biomarker of aging and a forecaster of longevity [102].

# 6. Conclusion

The ends of all linear chromosomes face the same difficulties regardless of their structures. Chromosome ends are not replicated completely by the standard replication machinery, resulting in loss of sequence and a 3' overhang on half of the replication products. Early eukaryotes may have solved the end replication problem by co-opting a reverse transcriptase that had arisen in a retrotransposable element [124] and using it to generate arrays of a simple repeat unit. They then solved the end protection problem by engineering long 3' overhangs on all termini, which could then loop around and tuck into the double stranded telomeric region and coating the terminal arrays with proteins that recognize the product of the reverse transcriptase. This combination of telomere maintenance by telomerase and protection by CST/shelterin served eukaryotes well and has been amazingly stable for more than a billion years.

Depending on how strictly shelterin recognizes the telomeric motif, the sequence may be conserved over long expanses of time, as in unikonts. If shelterin is less strict in recognizing the telomerase-generated motif, this sequence may have more latitude to vary, as in chromalveolates. If the protective telomere cap completely loses its ability to recognize the telomeric sequence, telomerase and the canonical telomeric motif may be lost. Many eukaryotes, including yeast and humans use unequal homologous recombination/gene conversion as a backup telomere maintenance system. It appears that in some species of plants and animals telomerase has been lost, and gene conversion has taken over as the primary mechanism to maintain chromosome length, with the eventual loss of telomeric motif. Chromosome length maintenance and end protection must be maintained through all of this. The evolution of new telomere structures, therefore, requires a delicate interplay between these two functions, as well as other telomeric functions that may be less well understood, such as heterochromatin formation and meiotic chromosome pairing.

## Acknowledgements

## Author details

Radmila  Capkova Frydrychova[1*] and James M. Mason[2*]

*Address all correspondence to: radmila.frydrychova@hotmail.com

1 Institute of Entomology, Czech Republic

2 National Institute of Environmental Health Sciences, USA

## References

[1] Greider, C.W., & Blackburn, E. H. (1996). Telomeres, telomerase and cancer. *Sci Am*, 274(2), 92-97.

[2] Mason, J. M., Frydrychova, R. C., & Biessmann, H. (2008). Drosophila telomeres: an exception providing new insights. *Bioessays*, 30(1), 25-37.

[3] Biessmann, H., & Mason, J. M. (2003). Telomerase-independent mechanisms of telomere elongation. *Cell Mol Life Sci*, 60(11), 2325-2333.

[4] Zakian, V. A. (2012). Telomeres: The beginnings and ends of eukaryotic chromosomes. *Exp Cell Res*, 318(12), 1456-1460.

[5] Sohanpal, B. K., Morzaria, S. P., Gobright, E. I., & Bishop, R. P. (1995). Characterisation of the telomeres at opposite ends of a 3 Mb Theileria parva chromosome. *Nucleic Acids Res*, 23(11), 1942-1947.

[6] Liu, C., Schroeder, A. A., Kapur, V., & Abrahamsen, M. S. (1998). Telomeric sequences of Cryptosporidium parvum. *Mol Biochem Parasitol*, 94(2), 291-296.

[7] Klobutcher, L. A., Swanton, M. T., Donini, P., & Prescott, D. M. (1981). All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proc Natl Acad Sci U S A*, 78(5), 3015-3019.

[8] Alverca, E., Cuadrado, A., Jouve, N., Franca, S., & Moreno Diaz., de la Espina. (2007). Telomeric DNA localization on dinoflagellate chromosomes: structural and evolutionary implications. *Cytogenet Genome Res*, 116(3), 224-231.

[9]  Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature*, 456(7219), 239-244.

[10] Koonin, E. V. (2010). Preview .The incredible expanding ancestor of eukaryotes. *Cell*, 140(5), 606-608.

[11] Gilson, P. R., & McFadden, G. I. (2002). Jam packed genomes--a preliminary, comparative analysis of nucleomorphs. *Genetica*, 115(1), 13-28.

[12] Wellinger, R. J., & Sen, D. (1997). The DNA structures at the ends of eukaryotic chromosomes. *Eur J Cancer*, 33(5), 735-749.

[13] Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., & Maruyama, S. (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga Cyanidioschyzon merolae. *BMC Biol*, 5, 28.

[14] Fulneckova, J., Hasikova, T., Fajkus, J., Lukesova, A., Elias, M., & Sykorova, E. (2012). Dynamic evolution of telomeric sequences in the green algal order Chlamydomonadales. *Genome Biol Evol*, 4(3), 248-264.

[15] Pich, U., Fuchs, J., & Schubert, I. (1996). How do Alliaceae stabilize their chromosome ends in the absence of TTTAGGG sequences? *Chromosome Res*, 4(3), 207-213.

[16] Schubert, I., Rieger, R., Fuchs, J., & Pich, U. (1994). Sequence organization and the mechanism of interstitial deletion clustering in a plant genome (Vicia faba). *Mutat Res*, 325(1), 1-5.

[17] Fajkus, J., Kovarik, A., Kralovics, R., & Bezdek, M. (1995). Organization of telomeric and subtelomeric chromatin in the higher plant Nicotiana tabacum. *Mol Gen Genet*, 247(5), 633-638.

[18] Peska, V., Sykorova, E., & Fajkus, J. (2008). Two faces of Solanaceae telomeres: a comparison between Nicotiana and Cestrum telomeres and telomere-binding proteins. *Cytogenet Genome Res*, 122(3-4), 380-387.

[19] Sykorova, E., Lim, K. Y., Fajkus, J., & Leitch, A. R. (2003). The signature of the Cestrum genome suggests an evolutionary response to the loss of (TTTAGGG)n telomeres. *Chromosoma*, 112(4), 164-172.

[20] Watson, J. M., & Riha, K. (2010). Comparative biology of telomeres: where plants stand. *FEBS Lett*, 584(17), 3752-3759.

[21] Vitkova, M., Kral, J., Traut, W., Zrzavy, J., & Marec, F. (2005). The evolutionary origin of insect telomeric repeats, (TTAGG)n. *Chromosome Res*, 13(2), 145-156.

[22] Frydrychova, R., & Marec, F. (2002). Repeated losses of TTAGG telomere repeats in evolution of beetles (Coleoptera). *Genetica*, 115, 179-187.

[23] Frydrychova, R., Grossmann, P., Trubac, P., Vitkova, M., & Marec, F. (2004). Phylogenetic distribution of TTAGG telomeric repeats in insects. *Genome*, 47(1), 163-178.

[24]  Mravinac, B., Mestrovic, N., Cavrak, V. V., & Plohl, M. (2011). TCAGG, an alternative telomeric sequence in insects. *Chromosoma*, 120(4), 367-376.

[25]  Mason, J. M., & Biessmann, H. (1995). The unusual telomeres of Drosophila. *Trends Genet*, 11(2), 58-62.

[26]  Sasaki, T., & Fujiwara, H. (2000). Detection and distribution patterns of telomerase activity in insects. *Eur J Biochem*, 267(10), 3025-3031.

[27]  Wiegmann, B. M., Trautwein, M. D., Winkler, I. S., Barr, N. B., Kim, J. W., Lambkin, C., et al. (2011). Episodic radiations in the fly tree of life. *Proc Natl Acad Sci U S A*, 108(14), 5690-5695.

[28]  Lundblad, V., & Blackburn, E. H. (1993). An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell.*, 73(2), 347-360.

[29]  Nielsen, L., & Edstrom, J. E. (1993). Complex telomere-associated repeat units in members of the genus Chironomus evolve from sequences similar to simple telomeric repeats. *Mol Cell Biol*, 13(3), 1583-1589.

[30]  Nielsen, L., Schmidt, E. R., & Edstrom, J. E. (1990). Subrepeats result from regional DNA sequence conservation in tandem repeats in Chironomus telomeres. *J Mol Biol*, 216(3), 577-584.

[31]  Cohn, M., & Edstrom, J. E. (1992). Chromosome ends in Chironomus pallidivittatus contain different subfamilies of telomere-associated repeats. *Chromosoma*, 101(10), 634-640.

[32]  Biessmann, H., Walter, M. F., & Mason, J. M. (1997). Drosophila telomere elongation. *Ciba Found Symp*, 211, 53-67.

[33]  Biessmann, H., Kobeski, F., Walter, M. F., Kasravi, A., & Roth, C. W. (1998). DNA organization and length polymorphism at the 2L telomeric region of Anopheles gambiae. *Insect Mol Biol*, 7(1), 83-93.

[34]  Roth, C. W., Kobeski, F., Walter, M. F., & Biessmann, H. (1997). Chromosome end elongation by recombination in the mosquito Anopheles gambiae. *Mol Cell Biol*, 17(9), 5176-5183.

[35]  Madalena, C. R., Amabis, J. M., & Gorab, E. (2010). Unusually short tandem repeats appear to reach chromosome ends of Rhynchosciara americana (Diptera: Sciaridae). *Chromosoma*, 119(6), 613-623.

[36]  Pryde, F. E., Gorham, H. C., & Louis, E. J. (1997). Chromosome ends: all the same under their caps. *Curr Opin Genet Dev*, 7(6), 822-828.

[37]  George, J. A., Traverse, K. L., DeBaryshe, P. G., Kelley, K. J., & Pardue, M. L. (2010). Evolution of diverse mechanisms for protecting chromosome ends by Drosophila TART telomere retrotransposons. *Proc Natl Acad Sci U S A*, 107(49), 21052-21057.

[38] Sheen, F. M., & Levis, R. W. (1994). Transposition of the LINE-like retrotransposon TART to Drosophila chromosome termini. *Proc Natl Acad Sci U S A*, 91(26), 12510-12514.

[39] Rashkova, S., Karam, S. E., Kellum, R., & Pardue, M. L. Gag proteins of the two Drosophila telomeric retrotransposons are targeted to chromosome ends. *J Cell Biol*, 159(3), 397-402.

[40] Casacuberta, E., & Pardue, M. L. (2003). HeT-A elements in Drosophila virilis: retrotransposon telomeres are conserved across the Drosophila genus. *Proc Natl Acad Sci U S A*, 100(24), 14091-14096.

[41] Casacuberta, E., & Pardue, M. L. (2003). Transposon telomeres are widely distributed in the Drosophila genus: TART elements in the virilis group. *Proc Natl Acad Sci U S A*, 100(6), 3363-3368.

[42] Greider, C. W. (1999). Telomeres do D-loop-T-loop. *Cell*, 97(4), 419-422.

[43] Tomaska, L., Willcox, S., Slezakova, J., Nosek, J., & Griffith, J. D. (2004). Taz1 binding to a fission yeast model telomere: formation of telomeric loops and higher order structures. *J Biol Chem*, 279(49), 50764-50772.

[44] Muller, H. J., Herskowitz, I. H., Abrahamson, S., & Oster, I. I. (1954). A Nonlinear Relation between X-Ray Dose and Recovered Lethal Mutations in Drosophila. *Genetics*, 39(5), 741-749.

[45] Riha, K., Heacock, M. L., & Shippen, D. E. (2006). The role of the nonhomologous end-joining DNA double-strand break repair pathway in telomere biology. *Annu Rev Genet*, 40, 237-277.

[46] Capper, R, Britt-Compton, B, Tankimanova, M, Rowson, J, Letsolo, B, & Man, S. (2007). The nature of telomere fusion and a definition of the critical telomere length in human cells. *Genes Dev*, 21(19), 2495-2508.

[47] Linger, B. R., & Price, C. M. (2009). Conservation of telomere protein complexes: shuffling through evolution. *Crit Rev Biochem Mol Biol*, 44(6), 434-446.

[48] Palm, W., & de , L. T. (2008). How shelterin protects mammalian telomeres. *Annu Rev Genet*, 42, 301-334.

[49] Chen, L. Y., Liu, D., & Songyang, Z. (2007). Telomere maintenance through spatial control of telomeric proteins. *Mol Cell Biol*, 27(16), 5898-909.

[50] d'Adda di, F. F., Teo, S. H., & Jackson, S. P. (2004). Functional links between telomeres and proteins of the DNA-damage response. *Genes Dev*, 18(15), 1781-1799.

[51] Diotti, R., & Loayza, D. (2011). Shelterin complex and associated factors at human telomeres. *Nucleus*, 2(2), 119-135.

[52] Longhese, M. P. (2008). DNA damage response at functional and dysfunctional telomeres. *Genes Dev*, 22(2), 125-140.

[53] Celli, G. B., Denchi, E. L., & de, L. T. (2006). Ku70 stimulates fusion of dysfunctional telomeres yet protects chromosome ends from homologous recombination. *Nat Cell Biol*, 8(8), 885-890.

[54] Denchi, E. L., & de, L. T. (2007). Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature*, 448(7157), 1068-1071.

[55] Lamarche, B. J., Orazio, N. I., & Weitzman, M. D. (2010). The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett*, 584(17), 3682-3695.

[56] Wu, P., van Rooney, O. M., & de , S. L. T. (2010). Apollo contributes to G overhang maintenance and protects leading-end telomeres. *Mol Cell*, 39(4), 606-617.

[57] Sfeir, A., & de , L. T. (2012). Removal of shelterin reveals the telomere end-protection problem. *Science*, 336(6081), 593-597.

[58] Greider, C. W. (1999). Telomeres do D-loop-T-loop. *Cell*, 97(4), 419-422.

[59] Gu, P, Min, J. N., Wang, Y, Huang, C, Peng, T, & Chai, W. (2012). CTC1 deletion results in defective telomere replication, leading to catastrophic telomere loss and stem cell exhaustion. *EMBO J*, 31(10), 2309-2321.

[60] Raffa, G. D., Raimondo, D., Sorino, C., Cugusi, S., Cenci, G., Cacchione, S., et al. (2010). Verrocchio, a Drosophila OB fold-containing protein, is a component of the terminin telomere-capping complex. *Genes Dev*, 24(15), 1596-1601.

[61] Raffa, G. D., Ciapponi, L., Cenci, G., & Gatti, M. (2011). Terminin: a protein complex that mediates epigenetic maintenance of Drosophila telomeres. *Nucleus*, 2(5), 383-391.

[62] Cenci, G., Siriaco, G., Raffa, G. D., Kellum, R., & Gatti, M. (2003, Jan). The Drosophila HOAP protein is required for telomere capping. *Nat Cell Biol*, 5(1), 82-84.

[63] Raffa, G. D., Siriaco, G, Cugusi, S, Ciapponi, L, Cenci, G, & Wojcik, E. (2009). The Drosophila modigliani (moi) gene encodes a HOAP-interacting protein required for telomere protection. *Proc Natl Acad Sci U S A*, 106(7), 2271-2276.

[64] Komonyi, O., Schauer, T., Papai, G., Deak, P., & Boros, I. M. (2009). A product of the bicistronic Drosophila melanogaster gene CG31241, which also encodes a trimethyl-guanosine synthase, plays a role in telomere protection. *J Cell Sci*, Mar 15, 122(Pt 6), 769-774.

[65] Gao, G., Walser, J. C., Beaucher, M. L., Morciano, P., Wesolowska, N., Chen, J., et al. HipHop interacts with HOAP and HP1 to protect Drosophila telomeres in a sequence-independent manner. *EMBO J*, 29(4), 819-829.

[66] Biessmann, H., Kasravi, B., Bui, T., Fujiwara, G., Champion, L. E., & Mason, J. M. (1994, Apr). Comparison of two active HeT-A retroposons of Drosophila melanogaster. *Chromosoma*, 103(2), 90-98.

[67] Capkova, F. R., Biessmann, H., & Mason, J. M. (2008). Regulation of telomere length in Drosophila. *Cytogenet Genome Res*, 122(3-4), 356-364.

[68] Ciapponi, L., Cenci, G., & Gatti, M. (2006, Jul). The Drosophila Nbs protein functions in multiple pathways for the maintenance of genome stability. *Genetics*, 173(3), 1447-1454.

[69] Ciapponi, L., Cenci, G., Ducau, J., Flores, C., Johnson-Schlitz, D., & Gorski, M. M. (2004). The Drosophila Mre11/Rad50 complex is required to prevent both telomeric fusion and chromosome breakage. Curr Biol Aug 10; , 14(15), 1360-1366.

[70] Oikemus, S. R., McGinnis, N, Queiroz-Machado, J, Tukachinsky, H, Takada, S, Sunkel, C. E., et al. (2004, Aug). Drosophila atm/telomere fusion is required for telomeric localization of HP1 and telomere position effect. *Genes Dev*, 18(15), 1850-1861.

[71] Oikemus, S. R., Queiroz-Machado, J., Lai, K., Mc Ginnis, N., Sunkel, C., & Brodsky, M. H. (2006, May). Epigenetic telomere protection by Drosophila DNA damage response pathways. *PLoS Genet*, 2(5), e71.

[72] Bi, X., Wei, S. C., & Rong, Y. S. Telomere protection without a telomerase; the role of ATM and Mre11 in Drosophila telomere maintenance. *Curr Biol*, 14(15), 1348-1353.

[73] Raffa, G. D., Cenci, G., Siriaco, G., Goldberg, M. L., & Gatti, M. The putative Drosophila transcription factor woc is required to prevent telomeric fusions. *Mol Cell*, 20(6), 821-831.

[74] Cenci, G., Rawson, R. B., Belloni, G., Castrillon, D. H., Tudor, M., Petrucci, R., et al. UbcD1, a Drosophila ubiquitin-conjugating enzyme required for proper telomere behavior. *Genes Dev*, 11(7), 863-875.

[75] Canudas, S, Houghtaling, B. R., Bhanot, M, Sasa, G, Savage, S. A., Bertuch, A. A., et al. A role for heterochromatin protein 1gamma at human telomeres. *Genes Dev*, 25(17), 1807-19.

[76] Iglesias, N., Redon, S., Pfeiffer, V., Dees, M., Lingner, J., & Luke, B. (2011, Jun). Subtelomeric repetitive elements determine TERRA regulation by Rap1/Rif and Rap1/Sir complexes in yeast. *EMBO Rep*, 12(6), 587-593.

[77] Luke, B., & Lingner, J. TERRA: telomeric repeat-containing RNA. *EMBO J*, 28(17), 2503-2510.

[78] Schoeftner, S, & Blasco, M. A. A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. *EMBO J*, 28(16), 2323-2336.

[79] Biessmann, H., Prasad, S., Semeshin, V. F., Andreyeva, E. N., Nguyen, Q., Walter, M. F., et al. (2005, Dec). Two distinct domains in Drosophila melanogaster telomeres. *Genetics*, 171(4), 1767-1777.

[80] Kurenova, E., Champion, L., Biessmann, H., & Mason, J. M. (1998, Nov). Directional gene silencing induced by a complex subtelomeric satellite from Drosophila. *Chromosoma*, 107(5), 311-320.

[81] Schoeftner, S., & Blasco, M. A. (2010, Apr). Chromatin regulation and non-coding RNAs at mammalian telomeres. *Semin Cell Dev Biol*, 21(2), 186-193.

[82] Su, T. T. Heterochromatin replication: better late than ever. *Curr Biol*, 20(23), R 1018-R1020.

[83] Leach, T. J., Chotkowski, H. L., Wotring, M. G., Dilwith, R. L., & Glaser, R. L. (2000, Sep). Replication of heterochromatin and structure of polytene chromosomes. *Mol Cell Biol*, 20(17), 6308-16.

[84] Gilbert, D. M., Takebayashi, S. I., Ryba, T., Lu, J., Pope, B. D. , Wilson, K. A., et al. (2010). Space and time in the nucleus: developmental control of replication timing and chromosome architecture. *Cold Spring Harb Symp Quant Biol*, 75, 143-153.

[85] Kim, S. M., Dubey, D. D., & Huberman, J. A. Early-replicating heterochromatin. *Genes Dev*, 17(3), 330-335.

[86] Arnoult, N., Schluth-Bolard, C., Letessier, A., Drascovic, I., Bouarich-Bourimi, R., Campisi, J., et al. (2010, Apr). Replication timing of human telomeres is chromosome arm-specific, influenced by subtelomeric structures and connected to nuclear localization. *PLoS Genet*, 6(4), e1000920.

[87] Bianchi, A., & Shore, D. Molecular biology Refined view of the ends. *Science*, 320(5881), 1301-1302.

[88] Bianchi, A., & Shore, D. Early replication of short telomeres in budding yeast. *Cell*, 128(6), 1051-1062.

[89] Stewart, J. A., Chaiken, M. F., Wang, F., & Price, C. M. (2012). Maintaining the end: roles of telomere proteins in end-protection, telomere replication and length regulation. Mutat Res Feb 1 , 730(1-2), 12-19.

[90] Fouche, N., Ozgur, S., Roy, D., & Griffith, J. D. (2006). Replication fork regression in repetitive DNAs. *Nucleic Acids Res*, 34(20), 6044-6050.

[91] Sampathi, S., & Chai, W. (2011, Jan). Telomere replication: poised but puzzling. *J Cell Mol Med*, 15(1), 3-13.

[92] Yang, Q., Zhang, R., Wang, X. W., Spillare, E. A., Linke, S. P., Subramanian, D., et al. (2002). The processing of Holliday junctions by BLM and WRN helicases is regulated by. J Biol Chem Aug 30 , 277(35), 31980-31987.

[93] Sfeir, A., Kosiyatrakul, S. T., Hockemeyer, D., Mac Rae, S. L., Karlseder, J., Schildkraut, C. L., et al. (2009). Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. CellJul 10 , 138(1), 90-103.

[94] Miller, K. M., Rog, O., & Cooper, J. P. (2006). Semi-conservative DNA replication through telomeres requires Taz1. NatureApr 6 , 440(7085), 824-828.

[95] Gu, P, Min, J. N., Wang, Y, Huang, C, Peng, T, Chai, W, et al. CTC1 deletion results in defective telomere replication, leading to catastrophic telomere loss and stem cell exhaustion. *EMBO J*, 31(10), 2309-2321.

[96] Raghuraman, M. K., Winzeler, E. A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., et al. (2001). Replication dynamics of the yeast genome. Science Oct 5 , 294(5540), 115-121.

[97] Greider, C. W., & Blackburn, E. H. The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. Cell (1987). Dec 24 , 51(6), 887-98.

[98] Greider, C. W. (2006). Telomerase RNA levels limit the telomere length equilibrium. *Cold Spring Harb Symp Quant Biol*, 71, 225-229.

[99] Greider, C. W., & Blackburn, E. H. (2004). Tracking telomerase. CellJan 23 2 , 116, S 83-S 86.

[100] Greider, C. W. (1994). Mammalian telomere dynamics: healing, fragmentation shortening and stabilization. Curr Opin Genet Dev Apr; , 4(2), 203-211.

[101] Wright, W. E., Piatyszek, M. A., Rainey, W. E., Byrd, W., & Shay, J. W. (1996). Telomerase activity in human germline and embryonic tissues and cells. *Dev Genet*, 18(2), 173-179.

[102] Zhu, H., Belcher, M., & van der Harst, P. (2011, May). Healthy aging and disease: role for telomere biology? *Clin Sci (Lond)*, 120(10), 427-440.

[103] Le , S., Zhu, J. J., Anthony, D. C., Greider, C. W., & Black, P. M. (1998, May). Telomerase activity in human gliomas. *Neurosurgery*, 42(5), 1120-1124.

[104] Cifuentes-Rojas, C., & Shippen, D. E. (2012). Telomerase regulation. Mutat Res Feb 1 , 730(1-2), 20-7.

[105] Nelson, N. D., & Bertuch, A. A. (2012). Dyskeratosis congenita as a disorder of telomere maintenance. Mutat Res Feb 1 , 730(1-2), 43-51.

[106] Armanios, M. Y., Chen, J. J., Cogan, J. D., Alder, J. K., Ingersoll, R. G., Markin, C., et al. (2007). Telomerase mutations in families with idiopathic pulmonary fibrosis. N Engl J Med Mar 29 , 356(13), 1317-1326.

[107] Greenberg, R. A., O'Hagan, R. C., Deng, H., Xiao, Q., Hann, S. R., Adams, R. R., et al. (1999). Telomerase reverse transcriptase gene is a direct target of c-Myc but is not functionally equivalent in cellular transformation. Oncogene Feb 4 , 18(5), 1219-1226.

[108] Kyo, S., Takakura, M., Fujiwara, T., & Inoue, M. (2008, Aug). Understanding and exploiting hTERT promoter regulation for diagnosis and treatment of human cancers. *Cancer Sci*, 99(8), 1528-1538.

[109] Holt, S. E., Wright, W. E., & Shay, J. W. (1997, Apr). Multiple pathways for the regulation of telomerase activity. *Eur J Cancer*, 33(5), 761-766.

[110] Loayza, D., & de, LT. (2003). POT1 as a terminal transducer of TRF1 telomere length control. Nature Jun 26 , 423(6943), 1013-1018.

[111]   Wang, F., Podell, E. R., Zaug, A. J., Yang, Y., Baciu, P., Cech, T. R., et al. (2007). The POT1-TPP1 telomere complex is a telomerase processivity factor. Nature Feb 1 , 445(7127), 506-510.

[112]   Zaug, A. J., Podell, E. R., Nandakumar, J., & Cech, T. R. (2010). Functional interaction between telomere protein TPP1 and telomerase. Genes Dev Mar 15 , 24(6), 613-22.

[113]   Zaug, A. J., Podell, E. R., & Cech, T. R. (2005). Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. Proc Natl Acad Sci U S A Aug 2 , 102(31), 10864-10869.

[114]   Savitsky, M., Kravchuk, O., Melnikova, L., & Georgiev, P. (2002, May). Heterochromatin protein 1 is involved in control of telomere elongation in Drosophila melanogaster. *Mol Cell Biol*, 22(9), 3204-3218.

[115]   Perrini, B., Piacentini, L., Fanti, L., Altieri, F., Chichiarelli, S., & Berloco, M. (2004). HP1 controls telomere capping, telomere elongation, and telomere silencing by two different mechanisms in Drosophila. Mol Cell Aug 13 , 15(3), 467-476.

[116]   Frydrychova, R. C., Biessmann, H., Konev, A. Y., Golubovsky, M. D., Johnson, J., Archer, T. K., et al. (2007, Jul). Transcriptional activity of the telomeric retrotransposon HeT-A in Drosophila melanogaster is stimulated as a consequence of subterminal deficiencies at homologous and nonhomologous telomeres. *Mol Cell Biol*, 27(13), 4991-5001.

[117]   Frydrychova, R. C., Mason, J. M., & Archer, T. K. (2008, Sep). HP1 is distributed within distinct chromatin domains at Drosophila telomeres. *Genetics*, 180(1), 121-131.

[118]   Torok, T., Benitez, C., Takacs, S., & Biessmann, H. (2007, Apr). The protein encoded by the gene proliferation disrupter (prod) is associated with the telomeric retrotransposon array in Drosophila melanogaster. *Chromosoma*, 116(2), 185-195.

[119]   Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A., & Gvozdev, V. (2006). Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline. Genes Dev Feb 1 , 20(3), 345-354.

[120]   George, J. A., & Pardue, M. L. (2003, Feb). The promoter of the heterochromatic Drosophila telomeric retrotransposon, HeT-A, is active when moved into euchromatic locations. *Genetics*, 163(2), 625-635.

[121]   Walter, M. F., & Biessmann, H. (2004, May). Expression of the telomeric retrotransposon HeT-A in Drosophila melanogaster is correlated with cell proliferation. *Dev Genes Evol*, 214(5), 211-9.

[122]   Oikawa, S., Tada-Oikawa, S., & Kawanishi, S. (2001). Site-specific DNA damage at the GGG sequence by UVA involves acceleration of telomere shortening. Biochemistry Apr 17 , 40(15), 4763-8.

[123]   Barja, G. (2002). Rate of generation of oxidative stress-related damage and animal longevity. Free Radic Biol Med Nov 1 , 33(9), 1167-1172.

[124] Belfort, M., Curcio, M. J., & Lue, N. F. (2011). Telomerase and retrotransposons: reverse transcriptases that shaped genomes. Proc Natl Acad Sci U S A Dec 20 , 108(51), 20304-20310.

# Telomere Shortening Mechanisms

Andrey Grach

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/55244

## 1. Introduction

Telomeres are the terminal regions of the linear chromosomes of eukaryotes, which are composed of telomeric DNA and associated specific telomeric proteins. In most kinds of organisms, telomeric DNA is presented by a large number of repetitive, strictly defined short nucleotide sequences, such as: TTAGGG (in vertebrates), TTTAGGG (in the majority of terrestrial plants) and TTGGGG (in the ciliated infusoria Tetrahymena), etc. Although telomeric proteins differ among different groups of organisms they perform similar functions, which mainly consist of telomere length regulation and their protection against degradation (Grach, 2009). For a long time, it was considered that telomeres did not code RNA molecules and thus proteins. Subsequently, it was found that RNA is still transcribed from telomeres but that it did not encode any proteins. Further studies showed that this RNA plays an important role in telomere length regulation and chromatin reorganisation during both development and cell differentiation (Azzalin et al., 2007). In spite of the fact that telomeres do not code proteins, they also perform very important functions, the main role of which is to maintain the stability and functionality of the cellular genome. Among these are the protection of chromosomes from fusion with each other (Blackburn, 2001), participation in mitotic and meiotic chromosome segregation (Conrad et al., 1997; Dynek and Smith, 2004; Kirk et al., 1997), the stabilisation of broken chromosome ends (Pennaneach et al., 2006), their attachment to the nuclear envelope (Hediger et al., 2002; Podgornaya et al., 2000), influencing gene expression (Baur et al., 2001; Pedram et al., 2006), counting the quantity of cell divisions (Allsopp et al., 1992; Kurenova & Mason, 1997; Olovnikov, 1973), and also an original buffer function (Olovnikov, 1973). The latter consists of the protection of the mRNA coding regions of chromosomes from the end replication problem. The end replication problem consists of the impossibility of the full reproduction of the previous length of linear DNA ends on the leading telomeres during of the S-phase of the cell division cycle. It is caused by peculiarities in their structure and the functioning of the DNA replication machinery. As a result the telomeric regions of chromo-

somes in daughter cells are shortened by several tens of nucleotides at each cell division (Lingner et al., 1995). In addition to the end replication problem, the telomere repair problem can also play a role in telomere length shortening. This problem in turn can be divided into the end repair problem and the shelterin-mediated telomere repair problem. The end repair problem includes the incomplete repair of DNA ends and direct damage-mediated telomere shortening and it can occur at the extreme ends of chromosomes. The incomplete repair of DNA ends consists in the inability of a repair system to complete the repair of damage if it occurred at the extreme ends of telomeres, since repair proteins cannot work correctly on the brink of a template and, as a result, they will also be shortened. Direct damage-mediated telomere shortening is based on the fact that, in some cases, the repair of damage at the extreme ends of chromosomes cannot even begin, in contrast to the incomplete repair of DNA ends at which the repair process begins but is not fully completed. It can generally be invoked by the fact that the breaks occurring on the extreme ends of chromosomes lead to the complete separation of the terminal DNA section and as a result repair system proteins are not able to even partially restore such damage. The consequence of this - and also of the subsequent actions of nucleases, which restore previous telomere ends' configuration - is the DNA ends shortening again. The shelterin-mediated telomere repair problem consists of the inability of the proteins involved in DNA damage response to detect and repair the damage occurring at telomeres due to the fact that the telomeric proteins in combination with telomeric DNA form a special structure on the telomeres called a telomeric loop (t-loop) that directly blocks DNA damage response proteins, as well as they block various DNA repair pathways themselves that is especially actual for the uncapped telomere condition when t-loop is not yet formed. This ultimately leads to the accumulation of damage and telomere shortening, and occurs along the entire length of telomeres where there is a telosome organisation and not just at their extreme ends as is the case with the end repair problem.

Telomere shortening is closely related to the replicative potential of cells and their lifespan. Thus, in accordance with A. M. Olovnikov's telomere theory of aging, when the telomere length approaches a certain critical level the cells stop dividing and begin ageing and are exposed to apoptosis upon reaching that level (Olovnikov, 1973). This fact has been confirmed experimentally in a number of studies (Allsopp et al., 1992; Allsopp et al., 1995; Aubert & Lansdorp, 2008). Besides playing a key role in aging, telomeres are also have great significance for carcinogenesis, as some cells with shortened telomeres acquire mechanisms to bypass the aging program and gain (among other characteristics) the ability to maintain telomere length and hence to "unlimited" quantity of divisions (Desmaze et al., 2003; Londoño-Vallejo, 2008; Stewart & Weinberg, 2006). The ability to elongate telomeres in vertebrates can be realised by means of two known mechanisms. The first and the most widespread mechanism among tumours provides for the use of a special enzyme called telomerase. It is a ribonucleoprotein enzyme consisting of a catalytic subunit, a telomerase RNA molecule and several additional components. Joining in with the ends of telomeres, its catalytic subunit uses reverse transcription of RNA, which is a part of telomerase to elongate a G-rich chain of telomeric DNA, which corresponds to the 3'-end regions of the DNA. Further, a C-rich chain corresponding to 5'-end DNA is synthesised on a template of a significantly elongated G-rich chain by a regular DNA polymerase reaction. As a result, the telomere ends gain the same structure as they had prior

to the telomerase action but they become much longer in this case (Blackburn & Collins, 2011; Dong et al., 2005; Testorelli, 2003). The second mechanism, which is found in a minority of neoplasm types, is accomplished by recombination-mediated telomere replication and belongs to the alternative lengthening of telomeres mechanisms (ALT) (Grach, 2011a; Grach, 2011b; Henson et al., 2002; Muntoni & Reddel, 2005; Stewart, 2005). Besides the elongation of telomeres in tumour cells, telomerase also has a high activity in stem and germ cells, thereby providing them a high proliferative capacity (Meeker & Coffey, 1997). Meanwhile, its activity is low or absent in normal somatic cells, making their replicative capacity strictly limited (Rhyu, 1995). As for ALT, it is usually repressed in normal cells by telomeric proteins and certain other factors (Grach, 2011b).

In recent years, the study of telomeres has becomes increasingly popular among scientists who are engaged both with different branches of molecular biology as well as with the most distant problems in the whole of modern medicine. Such heightened interest in their study is first of all caused by the fact that telomeres perform very important functions in the maintenance of eukaryotic cell genome normal functionality. Besides this, telomeres define the replicative capacity of cells and play a key role in their aging and transformation processes which make these end structures an even more important subject for research. All of the above mentioned roles of telomeres depend upon their shortening and, therefore, telomere shortening mechanisms are among the key aspects of telomere biology, because the loss of chromosome functions, cell aging and degeneration are associated with the telomeric regions of chromosomes' length shortening. In this respect, the study of these mechanisms as well as the factors involved in their protection and elongation are of primary importance as long as our cumulative knowledge can help in the future in the struggle against aging, tumours and many other diseases, the treatment of which requires a high replicative capacity in the cells. Based on the great significance of the telomere shortening process, the nub of various telomere-shortening mechanisms will be considered in detail in this chapter, namely the end replication and telomere repair problems.

## 2. The end replication problem

### 2.1. Early views of the end replication problem

*2.1.1. The end replication problem as a cause of telomere length shortening, which determines the replicative potential of cells*

For the first time, the problem of eukaryotic linear chromosome ends' replication was proposed and described in detail in the form of the theory of marginotomy by A. M. Olovnikov in Russian in 1971 (Olovnikov, 1971). One year later, in 1972, the problem was also described by J. D. Watson, independently of Olovnikov (Watson, 1972). In 1973, the problem proposed by Olovnikov was represented in its English version (Olovnikov, 1973). At the heart of this problem, as was suggested, lies the inability of the usual DNA replication system to fully complete the replication of linear DNA ends in the process of cell division. As a result of this,

it was assumed that the telomeric regions of chromosomes are shortened by roughly tens of base pairs (bps) at each cell doubling (Olovnikov, 1973). This state of affairs should explain why normal somatic cells, having divided a number of times, stop their further reproduction, start ageing and undergo apoptosis (i.e. the causes of the Hayflick cell division limit) (Hayflick, 1965). It was therefore suggested that in all of the non-transformed somatic cells of the organism, the telomere replication mechanism is absent and as a result of which they are gradually shortened on their division. When telomeres shorten to the definite minimal length needed for their normal functioning, the cells stop their division, age and then die. In other words, it was suggested that the telomere shortening process is a kind of "counter" which determines the replicative potential of cells (Olovnikov, 1973). These suppositions have been confirmed experimentally in several studies (Allsopp et al., 1992; Allsopp et al., 1995; Levy, et al., 1992). Thus, in one such experiment it was found that cells with shortened telomeres could perform far fewer divisions than cells with long telomeres (Allsopp et al., 1992).

### 2.1.2. The old theoretical model of the end replication problem

We now consider the actual mechanisms of the incomplete replication of the ends of linear DNA. As is known, every human chromosome consists of two anti-parallel DNA strands, which together form a single linear double-stranded DNA molecule with two ends. When the end replication problem was described for the first time, it was still considered that according to the generally accepted DNA model its ends would also have a double-stranded structure. Proceeding from this understanding of DNA, has been formulated the old theoretical model of the end replication problem, which was based on two possible mechanisms by means of which DNA ends could not be completely replicated.

The first mechanism suggests that DNA polymerase implementing DNA synthesis only in the 5' → 3' direction should have besides the catalytic centre also the DNA binding site, which should be located in front of catalytic one and be responsible for attachment of enzyme to a parent DNA strand. As such, and during DNA replication, when a DNA polymerase approaches the very end of the template by moving in front of the DNA binding site it cannot continue synthesis and so dissociates from the DNA because it will have nothing more to bind to. As a result, the end portion of a template which is equal in length to a DNA binding centre cannot be replicated, since a DNA polymerase cannot simply bring its catalytic centre to the last nucleotides of a parent strand without being dissociated from the DNA. Thus, after an incomplete replication process of such a kind, the 3'-end of a new DNA strand should become shortened by several nucleotides when compared with the parental one (Olovnikov, 1973).

The second mechanism of the incomplete replication of DNA ends was based on the fact that a DNA polymerase is not able to begin new DNA synthesis itself but is capable only of elongating already existing oligonucleotides. Therefore, at the very beginning of replication, primase synthesises a short RNA primer of around 9-12 nucleotides long, which is subsequently elongated by a leading strand DNA polymerase. When the DNA polymerase has already synthesised a long enough DNA strand, the RNA primer is removed by RNase after that the gap is filled by the polymerase and the ends between the DNA fragments are connected by a ligase. The DNA end regions there do not form an exception and are replicated according

to the same principle. However, the problem arises with the RNA primer, which is attached to the 3'-end of the DNA and defines the 5'-end of a new strand. The end gap arising after removal of the RNA primer cannot be filled later by the DNA polymerase, as there is no free 3'-end, which it could elongate. As a result, such incomplete end replication mechanism leads, this time, to the shortening of the 5'-end of a new DNA strand (Olovnikov, 1973).

Thus, according to the old theoretical model of the end replication problem, the 5' overhang at one end and the 3' overhang at the other can be formed at both daughter DNAs arising due to the peculiarities in the functioning of the DNA polymerase system. Such single-stranded protrusions should be cut by nucleases further in order to achieve double-stranded DNA ends' structure - which, as it was supposed earlier, they had initially. Consequently, the daughter chromosome telomere ends upon completion of the replication process will have the same configuration as they had before doubling. However, at the same time they will be shortened by a certain number of nucleotides (Olovnikov, 1973). The old theoretical model of the end replication problem is depicted in Figure 1.

### 2.1.3. Experimental confirmation of the old theoretical model of the end replication problem

As is shown in Fig. 1, the old theoretical model of the end replication problem assumes that incomplete replication can result from two mechanisms at both DNA ends - both on leading and lagging strands, resulting in each daughter chromosome being shortened at each of it's telomeres simultaneously. Further experimental verification of these circumstances has demonstrated quite different results for the leading strand and has completely confirmed the suppositions concerning the lagging DNA strand. In the course of one piece of research into the end replication problem using the artificially-created linear DNA replication system of the SV40 virus *in vitro*, it has been clearly determined that leading strand is synthesised entirely up to the very 5'-end of the template (Ohki et al., 2001). The explanation for this is that the DNA helicase unwinds its parent DNA to the very end and thus allows the completion of the synthesis of a new DNA strand. This discovery could abruptly undermine the possibility of the existence of the first proposed incomplete DNA replication mechanism, which assumes that due to the peculiarities of the functioning of DNA polymerase, the leading DNA strand synthesis cannot be fully completed up to the very end of the template. However, it is perhaps too early to judge this.

The results of another study of the end replication problem have shown the absolutely opposite situation. In the course of experiments looking into G-rich and C-rich single-stranded DNA in human fibroblasts, researchers unexpectedly revealed that the 5'-end of the DNA leading strand template is not replicated completely in the proliferating cells. Therefore, the 5'-overhangs appears in these cells predominantly during S-phase. This information provides grounds to suppose that the replication fork can terminate before reaching the chromosome's end. The authors of this study explain this in such a way that if the last RNA primer of the lagging strand is to be created as closely as possible to the 3'-end of the template then, in this case, the complete synthesis of the leading strand up to the very end of the 5'-end of a template is possible. If priming occurs more centromerically, then incomplete DNA leading strand

replication and related to it, enhanced telomere shortening may be observed (Cimino-Reale et al., 2003).

Similar results have also been observed in one more study, although it was not carried out with nuclear DNA but rather with the linear mitochondrial DNA end regions of the yeasts (in humans, mitochondrial DNA is ring-shaped). It was found during this work that the DNA polymerase stops at a distance of approximately 110 nucleotides from the 5'-end of a template and does not continue further leading strand synthesis, thereby again leaving the 5'-overhangs. However, the authors of the research could not explain why this happens (Nosek et al., 1995).

Consequently, the results of the researches just reviewed are quite conflicting. Unlike the first investigation described, where DNA leading strand synthesis continues up to the very end of the 5'-end of the parent DNA, creating "blunt" end, in the second and in the third studies we observed the incomplete replication of the leading strand with the creation of a "sharp" end of the DNA molecule. Moreover, none of the studies describe the possibility that complete or incomplete replication of the parent DNA 5'-ends was due to the peculiarities in the function of DNA polymerase as had been suggested by Olovnikov in its first incomplete DNA ends replication mechanism. Instead they consider quite other reasons for - in one case of complete leading strand synthesis to the very end of a template, and in another instance of incomplete leading strand replication.

Therefore, this data cannot fully support or refute the possibility of the inability of DNA polymerase to complete the replication of the 5'-end through a failure to bring its catalytic site to the last nucleotides of a template. It is also very important to note that the above mentioned studies, which describe the incomplete synthesis of the DNA leading strand with the creation of a 5' overhang, in practice are almost unique in their kind. The prevailing majority of studies show that leading strand is synthesised completely up to the very end of a template (Chai et al., 2006; Lingner et al., 1995, Wright et al., 1997).

It was also shown in the experimental research reviewed by us initially that lagging strand synthesis stops within the area located at a distance of approximately 500 bps from the end of a parent strand leaving 3' overhangs there (Ohki et al., 2001). This in its turn fully supports assumptions concerning existence of the second mechanism of incomplete DNA replication described above. As too long extension of incomplete replication was found here, in this work the authors have reviewed somewhat in a new way this mechanism, performed on the lagging strand. It is known that the length of RNA primers range from 9 to 12 nucleotides, which has been described in most of the studies that we have analysed (Griep, 1995; Hao & Tan, 2002; Sfeir et al., 2005). In some cases, primers of 20-30 nucleotides in length are also mentioned (Bouche et al., 1978; Dai et al., 2009). Nonetheless, the length of an incomplete replication of a DNA lagging strand is much longer, and has been discovered to be as long as 500 nucleotides. Therefore, the authors of this research propose the following mechanisms for the incomplete replication of the DNA lagging strand. It should be noted that the first mechanism completely corresponds to that proposed in the old theoretical model of the end replication problem for DNA. As has already been noted, it is based on the removal of the end RNA primer and the further failure to fill the resulting gap with deoxyribonucleotides. As lagging strand's incomplete replication reached approximately 500 nucleotides and the RNA primer length ranges
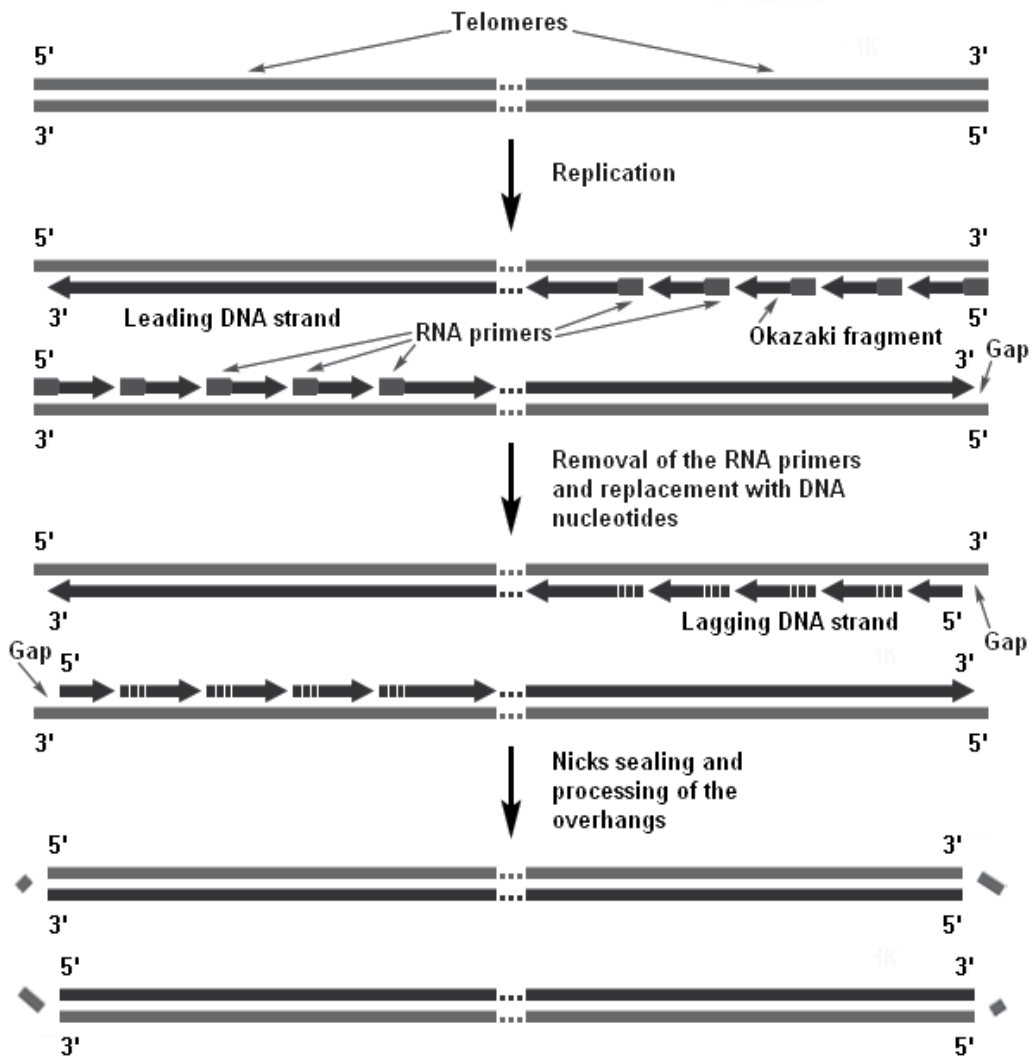
**Figure 1.** Schematic representation of the old theoretical model of the end replication problem, described by A. M. Olovnikov in 1971.

from 9 to 12 nucleotides, the authors of the work considered it very unlikely that this mechanism was the only one for the shortening of the 5'-end of daughter DNA. The second assumed mechanism consists of the inability of the DNA polymerase $\alpha$-primase to begin lagging strand synthesis at the very end of a linear DNA molecule that can be the main cause of the end replication problem of the parent DNA's 3'-end (Ohki et al., 2001). It is also interesting that the length of Okazaki fragments, which represent the short DNA fragments with RNA primers at the 5'-end and are the key feature of the lagging strand, can range from between 100 to 500 nucleotides (Burgers, 2009; Mackenney et al., 1997; Zheng & Shen, 2011). Therefore, if such a fragment cannot be formed at the very end of the parent strand then the DNA daughter strand

after the replication process will appear to be shortened in its length. Besides this research, many other studies showing similar results with incomplete lagging strand synthesis are also known.

Thus, most of the conditions of the old theoretical model of the end replication problem initially proposed by Olovnikov were confirmed absolutely in the course of the experiments carried out. In particular, it has been confirmed that telomeres are shortened at every cell division and that, specifically, these circumstances define the replicative potential of the cells and appear to be the cause of their aging and subsequent programmed cell death. Nevertheless, it was a failure to acquire any information absolutely confirming that the specifically incomplete replication of the DNA strands and subsequent cleavage of the resulting single-stranded overhangs of the molecule by nucleases results in telomeres' ends shortening, as was supposed by the old theoretical model. Unfortunately, we also failed to find any experimental data which fully confirms the first mechanism for the incomplete replication of DNA ends, assuming that DNA polymerase is not able to completely copy the 5'-end of the DNA leading strand template since it is incapable of bringing its catalytic site to the last nucleotides of a parent strand. At the same time, the results of the studies show that the 5'-end of a template remains in some cases not completely replicated, but other reasons for this, which are not directly related to DNA polymerase are specified in these cases. Therefore, the assumption regarding incomplete replication of the DNA leading strand is basically confirmed, but it is still unclear whether DNA polymerase directly plays a key role here or whether some other factors are involved (such as the ones that have been mentioned by the authors of the studies already discussed). Given all this, the assumption concerning the second mechanism of the incomplete replication of DNA ends is completely confirmed. Thus, the 3'-ends of the parent DNA, as was confirmed by the results of the experiments and initially stated in the theory, cannot be completely replicated during the lagging strand synthesis. As long as the extension of the incomplete replication of a lagging strand was much longer than the RNA primer length, it was supposed that the reason for the incomplete formation of a lagging strand along with the end primer removal might be due to the inability to prime and create the whole Okazaki fragment at the very DNA end. Given that the incomplete synthesis of the DNA leading strand is described only in some studies, but in the overwhelming majority of works it is shown that leading strand synthesis is performed completely up to the very end of a template, and that the 3'-end of the DNA template cannot for sure be replicated completely, the old theoretical model of end replication problem was also named a problem of incomplete lagging strand synthesis.

## 2.2. Modern views of the end replication problem

### 2.2.1. The establishment of the fact that telomere ends have a single-stranded structure

In the early 1980s, the data began to appear suggesting that both ends of each chromosome need not necessarily have a double-stranded structure but that they have a single-stranded structure (i.e. they are represented by 3'-overhangs). In 1981, it was noted for the first time that the ends of the linear minichromosomes, which are present in macronuclei of such ciliates as *Oxytricha*, *Stylonychia* and *Euplotes*, possess G-rich 3'-overhangs between 12 and 16 nucleotides

long (Klobutcher et al., 1981). Later on, in 1989, similar results were also acquired for the linear extrachromosomal ribosomal DNA of ciliate *Tetrahymena* and - evolutionarily distant from it – the slim mould *Didymium* (Henderson & Blackburn, 1989). Later, in 1993, it was found that the telomeres of the yeast *Saccharomyces cerevisiae* also gain 3'-end overhangs in the late S phase of the cell cycle and which differ a little in their dimensions from the ones which were described in previous works, being formed by more than 30 nucleotides (Wellinger et al., 1993). Unlike the above-mentioned organisms, which have a constant G-overhang length, the telomeric overhangs of higher eukaryotes display variability, even among the different cells studied in one group. As has been demonstrated by the results of a great number of studies, human telomeres possess very heterogeneous 3' overhangs, ranging from very short ones 35 nucleotides long and even less, to very long ones with an extension of 500 nucleotides or more. Furthermore, such varying in their dimensions G-overhangs are observed in all types of examined cells including the telomerase-positive transformed cells, telomerase-negative normal mitotic cells and post-mitotic cells (Cimino-Reale et al., 2001; Makarov et al., 1997; McElligott & Wellinger, 1997; Stewart et al., 2003; Wright et al., 1997). All these observations allow the supposition that G-overhangs are a general feature of eukaryotic chromosome telomeres.

### 2.2.2. A new theoretical model of the end replication problem

Based on numerous experimental observations showing that telomere ends' structure is not double-stranded but single-stranded, J. Lingner et al. have shown that this situation considerably changes established views as to the end replication problem. In particular, they demonstrated that the second mechanism of incomplete DNA replication, based on last RNA primer removal, no longer necessarily appears to be a problem for DNA replication machinery and the cause of telomere shortening. As primer cutting all the same leads to the creation of a 3' overhang, which also existed prior to replication and which is a normal structural feature of chromosome ends, so no genetic informational loss occurs in this case. In this respect, the incomplete synthesis of the lagging strand up to the end of a template can be considered now to be the mechanism of normal single-stranded 3' overhang telomere ends' structural formation. At the same time, a problem arises with the leading strand synthesis. This is caused by the fact that in the course of replication on such telomeres, daughter chromosomes lose the 3' overhang which was present in the parent chromosome and in the absence of telomerase this will accordingly result in their shortening. Moreover, if it is not restored by this enzyme to its previous state then, and only in this case, in the next round of replication might be observed the problem of incomplete lagging strand synthesis and already associated to it DNA shortening (Lingner et al., 1995) (Fig. 2). Thus, the result of replication with the new theoretical model proposed by Lingner et al. is a formation of two daughter DNAs which have one "sharp" end with a 3' overhang forming due to lagging strand synthesis and the other "blunt" end (or a "sharp" one with a 5' overhang if we take into consideration the possibility of performing the first mechanism of the old model of the end replication problem proposed by Olovnikov) forming on the leading strand. In contrast to the earlier proposed theoretical model of the end replication problem where overhangs should be cut, now single-stranded 3'-end protrusions remain intact, forming the natural eukaryotic chromosomes ends. Given all this, if according

to a new theoretical model leading strand synthesis results in the loss of 3' overhangs and the formation of "blunt" DNA ends, but the results of many experiments show that both chromosome ends have G-overhangs, and given that incomplete lagging strand replication assumes its formation only on one end, then there should also be a mechanism creating such an overhang on leading telomeres (Fig. 2).

Lingner et al. have also proposed two possible mechanisms for previous 3' overhang formation which also guarantee that DNA shortening, due to a problem of incomplete lagging strand synthesis, can never occur. The first mechanism presupposes that after the DNA replication process the end of the newly synthesised leading strand in the "blunt" DNA end can be elongated by telomerase and as a result a "sharp" DNA end, with a previous 3' overhang, will be restored. The possible caveat of this variant is that the DNA molecule with the "blunt" end acts here as a substrate for telomerase but telomerase is able to elongate only single-stranded ends rather than double-stranded ends as was found earlier *in vitro*. Nevertheless, the possibility that telomerase access could be provided in this case by helicases, nucleases or proteins binding single-stranded DNA has been considered. The second mechanism assumes that telomerase acts before the replication process, elongating the 3' overhang. It creates a template for the gap-filling synthesis of the complementary C-strand. As a result of the elongation of the 5'-end by conventional DNA replication machinery and the subsequent removal of the RNA primer, a telomere end region acquires a 3' overhang structure again but becomes much longer. Now, when the replication process approaches its end, the overhang on the leading strand is also lost but the genetic material no longer decreases, since before replication the parent strand of the leading strand was elongated. Further, the so-formed DNA "blunt" end and, in particular, its 5'-end region are exposed to processing by nucleases resulting in the formation of a "sharp" end with a previous 3' overhang that existed prior to the elongation by the telomerase (Lingner et al., 1995).

*2.2.3. Experimental evidence for the new theoretical model of the end replication problem*

These mechanisms describe the different possibilities for the creation of a previous 3' overhang and opposition to telomere shortening due to an inability during leading DNA strand synthesis to create such a single-stranded protrusion. However, both of these mechanisms are based on the possibility of telomerase action. At the same time telomerase activity is either very low in most normal human somatic cells or else is not observed at all (Rhyu, 1995). In spite of this, 3' overhangs are observed at both chromosome ends in normal cells (Makarov et al., 1997). There are also the results of focused experimental studies, showing that the removal of the genes coding for telomerase components does not affect the G-overhang structure considerably and this in turn also shows that the formation of such overhangs occurs irrespective of telomerase activity (Dionn & Wellinger, 1996; Hemann & Greider, 1999; Yuan et al., 1999). Besides, it was found during another study that these overhangs are exposed to cell cycle-regulated changes independent of telomerase activity (Dai et al., 2010). At the same time, telomerase in the cells where it is present is capable of elongating the 3' overhang after it is formed and thus make it like in the previous parental telomere. On this basis, it might be supposed in principle that previous 3' overhang in cells where there is no telomerase activity cannot be restored, but at the same time in its place a new overhang, by means of a special mechanism which will be reviewed

later, is formed which results in the telomere shortening. It is should also be mentioned that if it were forever restored to a previous state with the telomerase participation that was assumed in the above described mechanisms, it would lead to telomeres not be shortened during the course of cell doubling. This is equivalent to the acquisition of the unlimited replication potential which is observed mainly in the transformed cells. At the same time, the second mechanism reviewed presupposes that after preliminary elongation by polymerases, a DNA "blunt" end formed during replication due to the impossibility of creating a 3' overhang through leading strand synthesis is exposed to treatment by nucleases which process its C-rich strand and thus create an overhang of a specific length. Recent studies suggest that such post-replication treatment of a parent strand by nucleases, independently of whether there is telomerase in the cells or not, seems likely to be the main mechanism of 3' overhang formation in the leading telomeres (Lenain et al., 2006; van Overbeek & de Lange, 2006; Wu et al., 2010). If the parent 3' overhang before replication were to be elongated by telomerase, then the nucleases activity would further lead to previous 3' overhang formation, i.e. telomere end length does not decrease upon that and even increases, and if not, then these enzymes will create a new 3' overhang resulting in the shortening of the telomere's length. It is important to note that there are studies, showing that telomerase elongates the 3' overhangs of the leading daughter telomeres (Chai et al., 2006). Therefore, the first of the above reviewed mechanisms can be considered more realistic for previous 3' overhang restoration, especially taking into account that the DNA's "blunt" end after replication is necessarily exposed to the nuclease's influence and only after this does it become accessible for telomerase. It is also important to note that if in a case of accomplishment of the first mechanism of the old model of the end replication problem a DNA " sharp" end with a 5' overhang will be formed, the telomere ends shortening in that case would be even greater, as the incomplete synthesis of the DNA leading strand up to the end of a template and - related to this - excessive post-replication processing will take place. The latter is caused by the situation that nucleases now, in order to create a 3' overhang, will not only remove a certain number of C-strand nucleotides as a part of the double-stranded DNA, but also its single-stranded overhang. The schematic representation of the new theoretical model of the end replication problem is presented in Fig. 2.

It is clearly shown in Fig. 2 that under the new theoretical model of the end replication problem the incomplete DNA lagging strand synthesis, as a consequence of the impossibility of creating an Okazaki fragment and the removal of an end RNA primer, no longer leads to the daughter telomere's shortening but appears instead to be a kind of mechanism of their normal 3' overhanging structures' renewal. At the same time, during the synthesis of the leading strand, the DNA replication machinery is not able to recreate such an overhang on other chromosome ends as for its synthesis there is simply no template. Therefore, the leading telomeres of daughter chromosomes with respect to a parent chromosome lose their 3' overhang, which can be the cause of their further shortening. Experimental evidence for the claim that it is 3' overhang loss, which really leads to telomere shortening is derived from one study where it was found that the length of this overhang completely agrees with the chromosome end regions' shortening rate (Huffman et al., 2000). Nevertheless, there is also the data from another study showing that the G-overhang length does not correlate with the telomere-shortening rate (Keys et al., 2004). The authors of the research suggest that besides the 3' overhang loss in the course of DNA replication; the telomere-shortening rate is also influenced by damage from oxygen free radicals (Keys et al., 2004). As a result of such replication, there occurs the
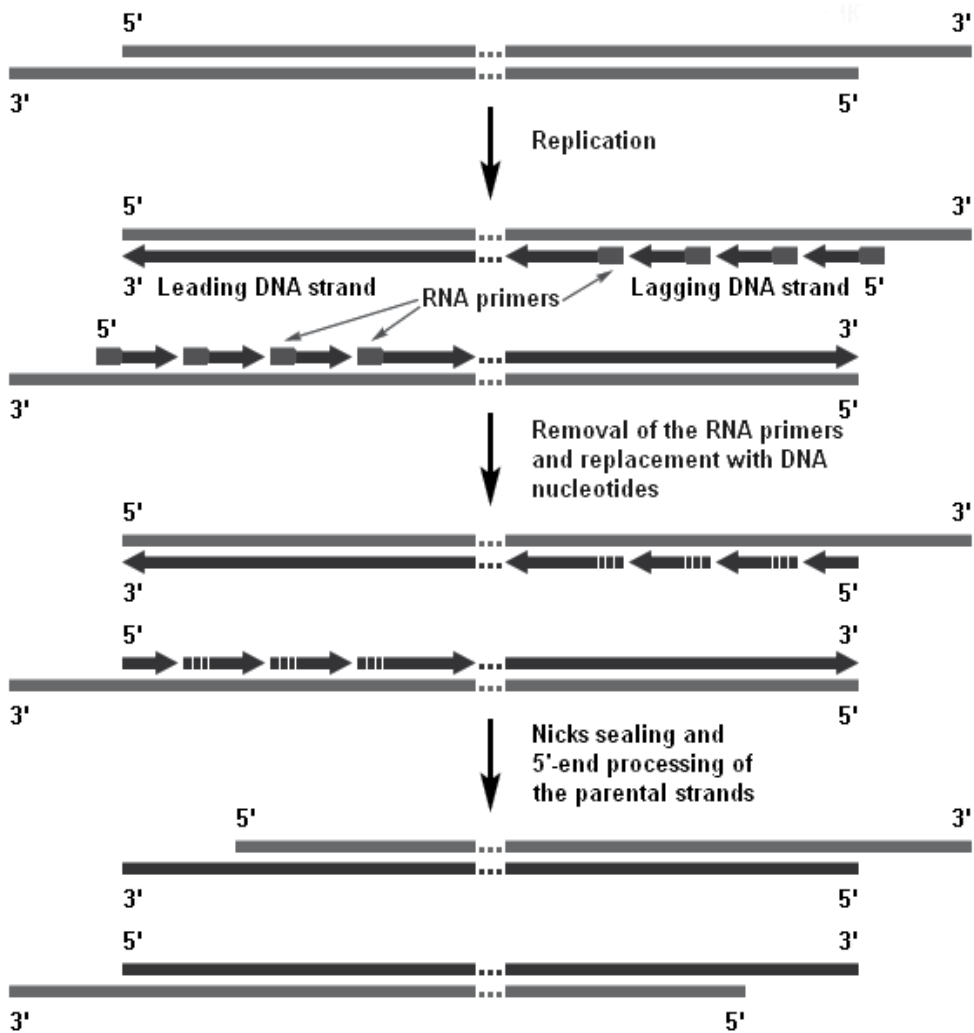
**Figure 2.** New theoretical model of the end replication problem.

formation of DNA daughter molecules that have one "sharp" end with a 3' overhang and the other "blunt" end. Taking into account that both chromosome ends have G-overhangs the "blunt" end on the leading telomere is further exposed to treatment by nucleases which cut its C-strand and thereby create the new 3' overhang, the length of which, will determine the rate of telomere shortening due to the end replication problem in the next replication cycle. The processing of the 5'-end of a parent strand of DNA can occur with the participation of such factors as the MRN protein complex (MRE11-RAD50-NBS1) as well as EXO1 and Apollo nucleases (Dewar & Lydall, 2010; Larrivee et al., 2004; Lenain et al., 2006; Maringele & Lydall, 2002; Tran et al., 2004; van Overbeek & de Lange, 2006; Wu et al., 2010; Zubko et al., 2004). Given this, the likely leading role is assigned to the Apollo nuclease, as RNA interference

mediated repression of the gene encoding Apollo nuclease, leads to the loss of 3' overhangs, subsequent cell cycle arrest and programmed death (van Overbeek & de Lange, 2006). The dominant role of such overhangs on the ends of chromosomes, as has long been established in the course of experiments, consists of the formation of special structures called telomeric loops (t-loops, see below), which protect DNA ends from being recognised as double-stranded breaks by the repair system proteins and other enzymatic influences (Grach, 2009; Griffith et al., 1999; Stansel et al., 2001). This is why it is so important that the leading telomere form a new 3' overhang, even taking into account some DNA parent strand shortening. It should also be noted that if we assume - hypothetically - the possibility that chromosomes could function normally if a 3' overhang was present on only one DNA end, at the same time if in this case the parent 5'-end was never cut by nucleases, it would lead to the impossibility of telomere shortening in a considerable number of primary cells and their immortalisation without telomerase. Fortunately, it is not possible because then there would be a high probability of such cells transforming. As is shown in Fig. 2, telomere shortening as a result of one round of DNA replication happens due to the impossibility of recreating a 3' overhang on a leading daughter telomere and a subsequent undercut of a 5'-end by nucleases in creating such a new overhang. If, after these events, a telomerase does not become active and does not elongate this new 3' overhang, thereby providing the possibility of recreating a previous overhang, then the shortened 3'-ends, having passed to the next round of replication will lead to a situation such that now, on their template, even shorter 5'-ends will be created as a result of incomplete lagging strand synthesis. Therefore, it is possible to say that in such cases telomere shortening can be performed by means of incomplete lagging strand synthesis, which, it should be especially emphasised, is possible only after the previous loss of the 3' overhang or, as some have noted, after incomplete DNA leading strand synthesis, and so cannot proceed on the initial chromosomes. It is interesting to notice here that in certain scientific works, which describe the new model of the end replication problem, the 3' overhang loss is designated as a problem of incomplete leading strand synthesis. This description - as it appears to us - does not fully correspond to the actuality because, in this case, a template is replicated to the very 5'-end, unlike the variant proposed by the first mechanism of the old model of the end replication problem, where its incomplete replication with C-overhang formation can be carried out. Therefore, with regard to a single-stranded template, the leading strand synthesis here is performed completely, however the truth is that in relation to the parent double-stranded DNA with a 3' overhang on both sides it does not do so completely. Thus, irrespective of these descriptions, but in the case of absence of telomerase activity, it seems to be possible that the following chain of events under the new theoretical model of the end replication problem lead to a daughter telomeres' shortening: a 3' overhang loss on the DNA daughter strand, the processing of the 5'-end of the DNA parent strand within one round of replication, and then the incomplete replication of a shortened 3'-end of the previous DNA daughter strand in the next one.

In order to understand in more detail how incomplete DNA lagging strand synthesis and 3' overhang loss on the leading telomere is accomplished under the new theoretical model of the end replication problem, let us examine the structure of replication forks on both chromosome ends, as presented by Fig. 3.

Fig. 3 shows two replication bubbles on eukaryotic chromosome ends, each of which consists of a pair of replication forks moving in opposite directions. As is known, in most cases the initiation of replication is accomplished from a non-telomeric origin (Gilson & Geli, 2007). Later on, one fork of the replication bubble moves towards a centromere and another one towards a telomeric end. Here it is seen that helicase unwinds a double-stranded DNA molecule up to the very ends. It allows for DNA polymerases to finish leading strand synthesis completely, to the very end of a template. The double-stranded ends on the leading telomeres are formed upon that. As is clear from the figure, in such cases, the previous 3' overhang, which earlier was on the parent DNA ends, cannot be reproduced, in principle, in daughter molecules during leading strand formation as there is nothing for it to be synthesised on and, consequently, it is lost, resulting in telomere shortening. While the leading strand concerning the parent strand is synthesised completely, the lagging strand synthesis cannot be completed up to the end of the template. Earlier, it was thought that the removal of the end RNA primer is responsible for it. However, today many researchers are inclined to consider that this is not the only reason and it is also probably significantly complemented with the impossibility of creating the last Okazaki fragment. This situation is also well represented by Fig. 3. As is known, first of all the leading strand is synthesised in the motion of the unwinding of the parent DNA, and later on, after the DNA polymerase has synthesised a certain leading strand extension, it moves to a lagging strand and elongates it, thereby catching up with the first one. When such synthesis of both strands reaches the last point of unwinding - which can correspond to the 5'-end of the parent DNA - there remains a long stretch of single-stranded DNA in the form of a 3' overhang beyond its limits. Upon this, there is no more space for synthesising the leading strand in order that later on a DNA polymerase can move and fill such an overhang with a lagging strand. In this connection, the Okazaki fragment on the 3' overhang is not created and it remains non-replicated, and after the last primer removal its length increases a little more. However, as an overhang - which occurred prior to replication - is created anyway, the telomere's shortening does not happen in this case.
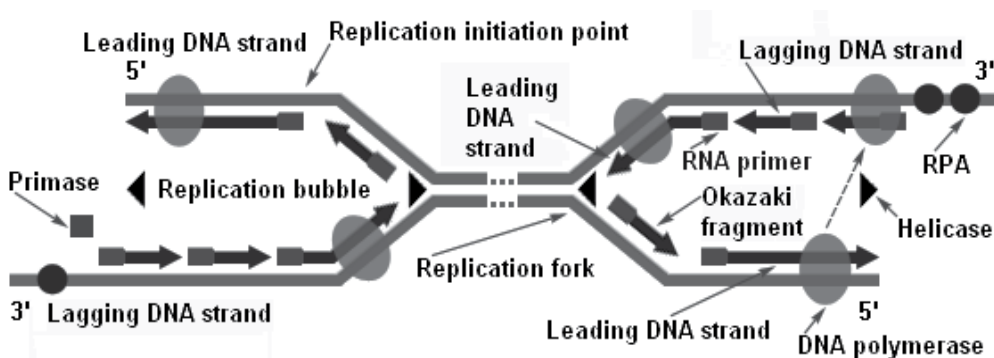


**Figure 3.** Telomeric replication forks.

Thereby, almost all conceptions of the new theoretical model of the end replication problem have been proved to be true in the course of the studies. It has been proved experimentally

that 3' overhang loss in particular leads to telomere shortening. Besides it has also been confirmed that a new overhang is formed due to leading telomere processing by nucleases. At the same time, the assumptions as to the point at which a telomerase itself directly restores a previous overhang were not confirmed. For all this, according to new views of the end replication problem and which have been confirmed by a number of studies, from now on it should be construed not as the lagging strand synthesis problem but rather as the DNA leading strand synthesis problem.

## 2.3. The conclusion of this section

Thus, summarising all of the aforesaid, the views of the end replication problem as the cause of telomere shortening have changed over a period of several decades. Initially, when everyone considered that the structure of chromosome ends was double-stranded, it was supposed that telomeres were shortened mainly due to incomplete DNA lagging strand synthesis, which leads to the formation of 3' overhangs in support of which there is much experimental data. Furthermore, some studies have demonstrated the possibility of the accomplishment of incomplete leading strand synthesis up to the very end of a template with 5' overhangs forming, but the truth is that their number is small. Such overhangs further should be undercut, which would lead to the single-step shortening of chromosomes from both ends. As most of these works nevertheless provide evidence in favour of the idea that the replication of 5'-ends was carried out completely, it was later considered that telomeres were shortened only due to the problem of incomplete lagging strand synthesis. Here, it would seem that if the telomere shortening mechanism acts only from one end of a chromosome then the its other end would never decrease in length. Actually, this is not precisely true. The matter is that, if we were to monitor two strands of any initially parent DNA then one of them - after a certain number of divisions and in case of the absence of telomerase - will be really shortened from one end and the other one would be from the opposite end. If we continuously monitor some formed daughter strands, then in the subsequent generations of the cells there will also appear chromosomes which are shortened at their own two ends. After the establishment of the important circumstance that the structure of telomere ends is not double-stranded but 3' overhanging single-stranded, the problem of incomplete lagging strand synthesis already actually ceased to be treated as being the problem, since it no longer led to telomere shortening now, and only restored a previous configuration of their ends, which is important for normal chromosome functioning. At the same time, the existence of 3' overhangs on the chromosomes' ends creates a significant problem for leading strand synthesis. It is caused by the fact that in the course of replication, the 3' overhang which is present in the parent DNA on two ends cannot be renewed in the daughter DNAs during leading strand formation because of the absence of a template for its synthesis; in this connection it will be absent at one end in one daughter molecule and on another end of another one. Such a 3' overhang loss, the further processing of the 5'-ends of leading telomeres resulting in the formation of new G-overhangs and also the subsequent incomplete lagging strand synthesis in the next generation on a template of an already shortened 3'-end, actually lead to telomere shortening. In this connection the end replication problem is inverted from the lagging strand to the leading strand. However, this does not mean that leading strand is necessarily synthesised incompletely up

to the very end of a template, and it can be reproduced fully in this case. Thus, all of the observations described above have shaped our current thinking about telomere shortening during cell division.

# 3. The telomere repair problem

### 3.1. Early ideas of the telomere repair problem — The incomplete repair of double-stranded DNA ends

The problem of the incomplete repair of the very ends of DNA was also described, first, by A. M. Olovnikov as early as 1995. As with the old model of the end replication problem it was based on the idea that telomere ends have a double-stranded structure. The two suggested mechanisms of incomplete DNA ends' repair that are actually similar to the mechanisms of incomplete replication described in the previous section were distinguished. The essence of the first one concluded that if a single-stranded break (SSB) or "nick" occurred close to a 3'-end of a DNA strand at a distance of just several nucleotides, then this damage could not in principle be repaired. It was presumably connected with the following situation. The short end oligonucleotide created by the nick could not remain hybridized to the rest of the DNA molecule resulting in the formation of a gap with a protruding 5'-end. Later on, a repair DNA polymerase should attach to a DNA molecule and, while moving along the undamaged C-rich strand, synthesizes the lost 3'-end region on its template. However, it could not be performed in this case as DNA polymerase again, as well as with respect to the replication of the very 5'-end, not able to bring its catalytic site to the last nucleotides of a template in order to reproduce them on a complementary strand. Therefore, the DNA molecule remains shortened at the 3'-end afterwards. The second suggested mechanism provided for the impossibility of damage repair if a nick happened near the 5'-end of the DNA strand. A gap formed after the separation of a DNA fragment that was too short to remain hybridized to the template, could not be filled in for another reason in this case. The chain which was subject to repair and shortened due to the single-stranded break has no 3'-end or primer to which a DNA polymerase can add nucleotides in the course of repair synthesis and, therefore, should also remain non-elongated. On this basis, in both cases of the incomplete repair of DNA ends, single-stranded overhangs are formed which should be cut further by nucleases that would subsequently lead to telomere shortening (Olovnikov, 1995a; Olovnikov, 1995b; Olovnikov, 1995c).

### 3.2. The end repair problem — The incomplete repair of DNA ends and direct damage-mediated telomere shortening

The discovery that telomere ends had a single-stranded 3' overhanging structure, besides changing views on the end replication problem also considerably changed conceptions of the repair of chromosome ends. Before considering in detail exactly what these changes are characterised by, it is necessary to clearly define what should be understood by an incomplete DNA ends' repair. It is suggested by A. M. Olovnikov that it can proceed in two cases. In the first case, damage resulting in the breaking off of a single-stranded DNA fragment happens

near the very end of the 3'-end of a double helix, and further DNA polymerase is not able to completely synthesise the lost region insofar as by moving along the opposite undamaged strand it cannot bring its catalytic site to the last nucleotides of the template. The second case assumes that damage with the separation of a DNA fragment happens near the 5'-end of a double-stranded molecule, and as a result DNA polymerase cannot restore the lost part again, since there is no available 3'-end as a primer to elongate. It is known that to begin polynucleotide synthesis at primer absence, as already noted, it is not able. Moreover there appears that the gap is so short that primase cannot even create an RNA primer. It is thus meant that in the first case a DNA polymerase binds to a DNA molecule and synthesizes several nucleotides, but the truth is that the last bases, which should lie opposite the very edge of the template, do not form (i.e. the damage repair process starts but does not come to its completion), and that in the second case the DNA polymerase, due to a RNA primer absence, cannot attach to a DNA double-stranded molecule and even to begin damage repair process. Based on these differences, specifically as to whether the repair process can start but does not come to its end or whether it does not start at all, we propose to look at the problem of incomplete DNA ends' repair and related to it telomere shortening somewhat in a new way. In particular, it is proposed that, under the problem of incomplete repair to consider that, the repair of damage at the end of double-stranded DNA starts but cannot finish completely because of the inability of most repair system proteins to correctly function at the very edge of a template that leads to chromosome ends shortening. It is necessary to emphasise that when noting that repair cannot be finished it is meant not only that DNA polymerase is unable to copy a template completely up to the very end in the course of repair DNA synthesis, but also that other enzymes can begin and even accomplish some stage of the repair process, however that later, due to certain reasons, repair cannot continue and so it finishes prematurely. As such, in the first case described, it is possible to say that what is actually accomplished is the incomplete repair of DNA ends. In addition, it is also proposed that if the repair of damage at a DNA end cannot start at all, again owing to the inability of repair proteins to work correctly at the edge of a template - and it will lead to chromosome telomere regions shortening, then this situation further should be designated as direct damage-mediated telomere shortening (DDMTS). The second described case can be related to this. Thus, we define two possible variants by means of which telomere shortening can be performed in a case where damage occurs at the very ends of a double-stranded DNA molecule, namely incomplete DNA ends' repair and direct damage-mediated telomere shortening.

Based on these new conceptions, let us consider once again Olovnikov's theoretical model of incomplete DNA ends' repair. As was mentioned, since it is described in the first variant that a DNA polymerase attaches to a DNA molecule, reproduces several nucleotides but subsequently cannot finish repair synthesis to the very end of a template, then in this case there occurs incomplete DNA repair. In the second variant, the enzyme cannot even attach itself to a template to begin repair - that was designated as DDMTS. It is necessary to note here that if, in the case of the first variant, the gap is very short then the DNA polymerase - even if it attaches itself to a template - will also not be able to begin repair because it will place its anchor region directly onto the very end of an undamaged DNA strand and, as a result, it will be immediately separated from the DNA molecule. This situation can already be regarded as DDMTS. At the

same time, if with the second variant the gap will be long enough for an RNA primer to be formed, and then it is possible that there will be two variants, in both cases of which there will now occur the incomplete repair of the DNA ends. If the gap will be long enough to fit just an RNA primer, then in this case it might be supposed that when a DNA polymerase attaches itself, it will not be able to synthesise nucleotides as well, but as far as primase synthesizes the primer, then it is possible to consider that repair has started and that one of its stages has finished, but also that another one is not able to begin. In the future, such a primer is removed by RNase and a gap of the same length as it was before the repair arises. In the case where the length of a gap is such that in spite of the RNA primer several nucleotides are able to fit there, then the DNA polymerase synthesises them. However, after primer removal there will be a gap anyway but which the truth is that will be smaller than before the repair. As a result, such situation should also be viewed as incomplete repair. Ultimately, it is important to note that at replicative and cell senescence stages it is known that repair systems act poorly. Therefore, if a long enough gap appears at one of the DNA ends in senescent cells, then it will probably not be even partially repaired, and as a result DDMTS will take place. Thus, if such gaps are repaired in young cells, even if incompletely, then in old ones they will be not be repaired. It is also necessary to emphasise that at one time is apparently possible to separate only terminal single-stranded DNA fragment that was less than nine nucleotides in length, since it is widely known that RNA primers 9-12 nucleotides in size are strongly hybridized to a template. For that matter, when it was mentioned that a gap can arise is longer than primer itself or else the same, it can proceed only according to several steps, i.e. a successive separation of several fragments 8 nucleotides in size or less. Thus, if a break occurs, e.g. at a distance of 9 or more nucleotides, then such a terminal oligonucleotide will not only be able to hang on a template but will also be reunited with the remaining proximal part by DNA ligase. In summary, it is also necessary to add that since it was experimentally discovered that 5'-end copying by a DNA polymerase in the course of replication is, in most cases, accomplished completely, then in such a case only the second variant of incomplete repair based on RNA primer removal so described could feasibly be carried out.

## 3.3. Modern views of the end repair problem

Now let us consider what exactly are characterized by the changing of conceptions of repair at the ends of chromosomes, if the telomeres have single-stranded 3' overhanging structure. They are characterised by the following circumstances. First, given such telomere ends organisation, the problem of incomplete repair can arise, as it seems to be possible, only if a single-stranded break occurs at a distance of up to approximately eight nucleotides towards the centromere from a place where the 3' overhang begins and the 5'-end of complementary strand is situated (Fig. 4). In such a case, if a DNA polymerase even manages to copy a template up to the very 5'-end in the course of the repair synthesis of the lost single-stranded DNA fragment, the previous 3' overhang would still not be able to renew, and so it can be seen that in such a situation repair has begun but cannot be finished, insofar as the damage could not be fully repaired. If we take into account that a DNA polymerase might not be able to copy the last nucleotides of a template, then in such a case if a break with a subsequent separation of a DNA fragment occurs at a very short distance (e.g. of one nucleotide) from the above

mentioned place, then repair will not begin and it will be designated as DDMTS. Secondly, if a break occurs somewhere at the 3' overhang or near its base (i.e. at a region where the opposite C-rich strand ends), then the distal part of the overhang or its entirely will separate from the DNA molecule and will be lost, as a result of which DDMTS will be observed - as far as repair in that case cannot even begin in principle due to single-stranded DNA fragment loss and the absence of a template for synthesizing the new one (Figure 5). With both variants, the new overhang will form in the future through 5'-end processing by the nucleases. Thirdly, an incomplete repair problem of the 5'-end, which should arise due to RNA primer removal on telomeres with "blunt" ends, is no longer a problem, and hence the reason for telomere shortening in instances with "sharp" ends, because as is the case with lagging strand synthesis at DNA replication, its cutting out leads only to the formation of the previous 3' overhanging configuration of telomere ends (Figure 6). Thus, single-stranded end breaks, at the 3' over-hanging telomere structure, can lead to incomplete telomere repair and further telomere shortening only if they will occur on a G-rich strand at a distance of several nucleotides in front of a place where a complementary C-strand ends. If the breaks affect the 3' overhang itself, then this will lead to DDMTS. Finally, the breaks of a C-rich strand occurring near the very 5'-ends will repair completely.

As is known, apart from single-stranded breaks, there are also such basic types of DNA damage as various nucleotide modifications, double-stranded breaks and cross-links (Sancar et al., 2004). Various nucleotide modifications (of a single one, a pair or else several) arising at DNA ends in most cases cannot lead to single-stranded breaks of the molecule in themselves. Further, they are exposed to various repair pathways, such as a direct repair (DR), a base excision repair (BER) and a nucleotide excision repair (NER) (Sancar et al., 2004). Since a direct repair is accomplished without any breakage of the phosphodiester backbone it cannot, in principle, lead to telomere shortening and, therefore, in our case, is of no particular interest. This type of reactions includes the photoreactivation of ultraviolet-induced pyrimidine dimers by a DNA photolyase enzyme, the removal of the $O^6$-methyl group from $O^6$-methylguanine ($O^6$MeGua) in DNA by the DNA methyltransferase enzyme, and the repair of apurinic/apyrimidinic sites through the direct insertion of bases by the insertase enzyme (Sancar et al., 2004). In addition, the repair of single-stranded DNA breaks by DNA ligase enzymes also belongs to this type but only if they do not arise at the very ends of DNA and do not lead to gap formation. Base excision repair consists of the cutting out of damaged nucleotide from a DNA strand by means of several reactions with the participation of DNA glycosylases, AP-endonuclease and phosphodiesterase, resulting in the formation of a very short gap (Fromme & Verdine, 2005; Krokan et al., 1997; Seeberg et al., 1995). This gap should be filled further by a DNA polymerase on a template of an undamaged complementary strand, after which the free ends are sealed by a ligase. If such a form of repair is carried out somewhere in the middle of the DNA molecule or near its 5'-end, then no problem will arise. However, if it proceeds at a distance of, say, 9 nucleotides from the place where the 3' overhang begins, i.e. the 9th nucleotide will be removed at that, then the end DNA fragment 8 nucleotides long up to the 5'-end of an opposite strand, together with the 3' overhang, would be lost. It will lead to gap formation, which can be filled further to form a "blunt" DNA end, but upon this, as well as in the case of a single-stranded break forming at a distance of up to 8 nucleotides and the

subsequent formation of the same gap as discussed earlier, the previous 3' overhang it will not be able to restore itself and the telomere will shorten. Thus, base excision repair imposed on the end regions of a G-strand may be lead to an incomplete DNA repair. It is also interesting to discuss the situation where such a damaged nucleotide arises within the 3' overhang itself. BER system enzymes are apparently incapable of acting on a single-stranded DNA. As a result, such damage will not be repaired, and where further the base modification can similarly lead to DNA strand breaking and 3' overhang distal part loss, then this situation should be considered as DDMTS. However, if repair enzymes all the same could cut out a damaged nucleotide, which again will lead to the separation and loss of the 3' overhang terminal fragment, and so there will be an incomplete repair in this case. As for the nucleotide excision repair, it is very similar to BER but is accomplished by other enzymes, and in this case not only one damaged nucleotide is removed but up to thirty (de Laat et al., 1999; Reardon & Sancar, 2005; Sancar et al., 2004). With NER as well as with BER, if the gap is formed on the G-strand of telomeric DNA, in such a way that 8 or less nucleotides remain up to the end of a double-stranded DNA structure in the G-strand, then again a short end fragment of a G-strand together with the 3' overhang will be separated and lost and as a result an incomplete repair and related to it telomere shortening will be observed subsequently again. In order not to repeat this, it may be said that all other situations involving NER at the end of the telomere, including whether several nucleotides on the 3' overhang will be damaged, are similar to those that have been reviewed in relation to BER.

Double-stranded breaks (DSBs) of DNA can be repaired by three mechanisms: non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ) and homology-directed repair (HDR) (Chu, 1997; Liang et al., 2008; Lieber et al., 2003). The first mechanism consists of the direct joining of broken ends by a specialised enzyme DNA ligase IV with the participation of the protein Ku and DNA-PK, which is carried out within G0/G1 and the early S phases of the cell cycle (Lieber et al., 2003). The second mechanism does not depend on these proteins and also differs from NHEJ in that this mechanism of DSBs repair uses 5-25 base pair microhomologous sequences to align the broken strands before joining, and it is carried out within the S phase of the cell cycle (Liang et al., 2008). The third mechanism is based on homologous recombination of a damaged chromosome with a sister chromatid or homologous undamaged chromosome, and therefore the damaged chromosome is repaired on their template that is carried out within the late S and G2 phases of the cell cycle (Chu, 1997). Nevertheless, NHEJ and HDR seem to be the main mechanisms for DSB's repair. If DSB occurs near the very end of a chromosome at a distance of only several nucleotides from the place where the 3' overhang begins, then such damage will probably not be repaired since the distal double-stranded fragment of DNA would be too short for the repair enzymes to bind to it, and, in the case of NHEJ and MMEJ, connect it to the rest of a molecule. Additionally, HDR also would be ineffective in repairing such damage since the homology tract would be again too short to effectively engage the enzymes that catalyze homologous recombination. As a result, the repair of such damage will not begin and there will be observed DDMTS. In addition, it is also necessary to note that if telomere damage occurs at a great enough distance from the very end of a DNA, but still within the telomere region organized by the telosome, it is unlikely that it could be repaired by homologous recombination. This is caused by the ability of

telomeric proteins to block recombination events for the preventing of the elongation of chromosome telomere regions through ALT (Grach, 2011b).
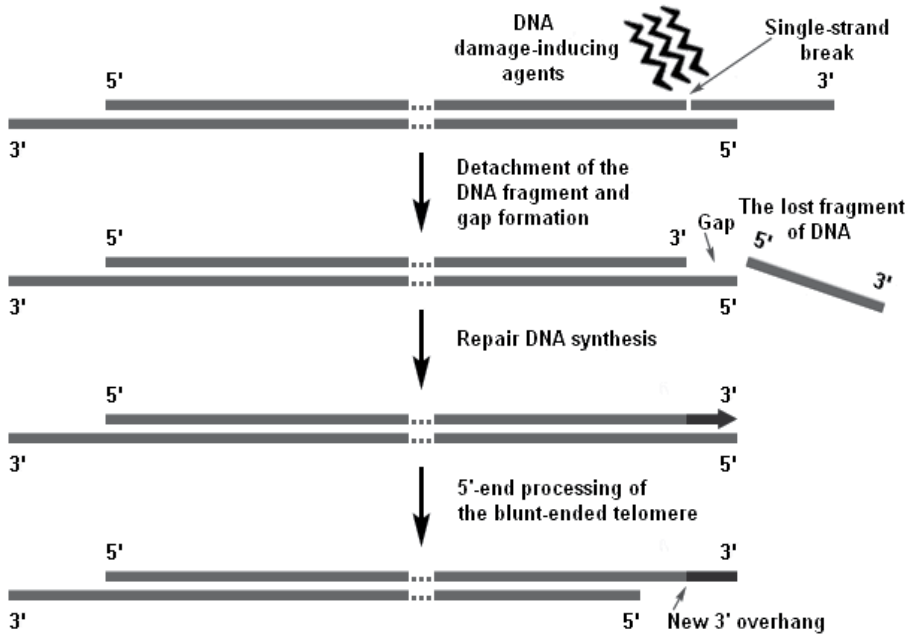


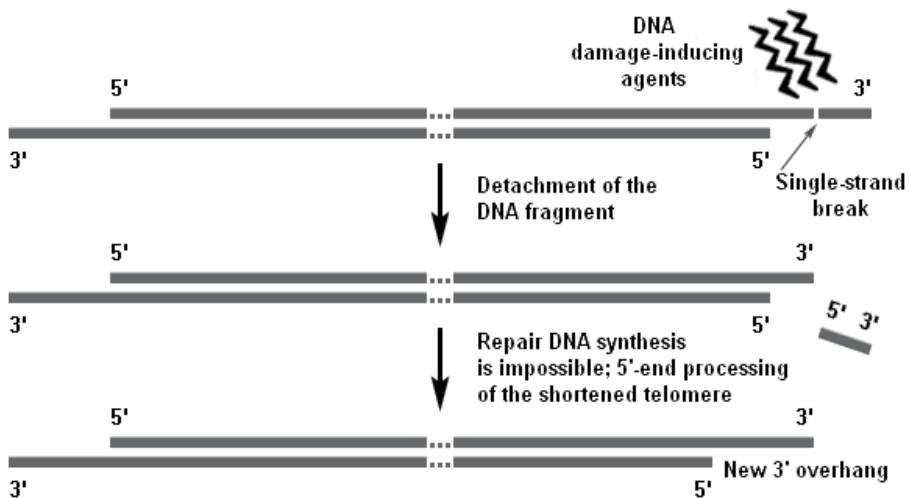**Figure 4.** Incomplete double-stranded DNA end repair.



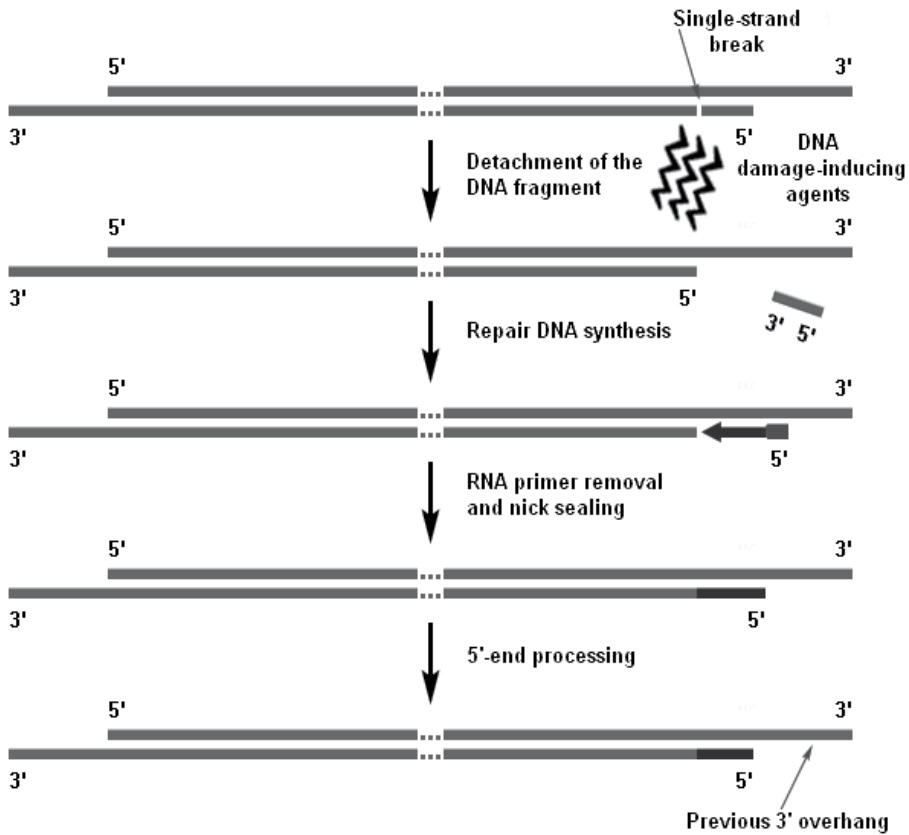**Figure 5.** Direct damage-mediated telomere shortening.

**Figure 6.** Complete double-stranded DNA end repair.

DNA cross-links represent connections between the nucleotide bases by covalent bonds (normally they are bound by hydrogen bonds) which can be of two types: intrastrand and interstrand (Dronkert & Kanaar, 2001). The latter ones represent a very serious problem for DNA replication machinery insofar as before a DNA polymerase replicates a parent DNA, it is at first unwound by helicase through the breaking of the hydrogen bonds between two strands. Therefore, if two complementary nucleotides are covalently linked, helicase will not be able to separate them and this will lead to the stalling of the replication fork and potential DSB formation. Two mechanisms for the repair of such damage are known which can be carried out during different phases of a cell cycle. Both of these mechanisms begin identically by forming cuts on both sides of the cross-link on one DNA strand by NER system proteins, after which there occurs a twisting of the oligonucleotide carrying the damage and gap formation. Later, such a gap will be filled; one mechanism for this is by TLS (Translesion synthesis), which makes use of DNA polymerases that are able to replicate DNA despite template damage arising before them. This is followed by another round of NER during which the second DNA strand is cleaved, and adduct removal is carried out. The second gap that is formed can than be filled by a conventional DNA polymerase on a complementary strand template and the ends are connected by DNA ligase. In the case of the other mechanism, the

filling of the first gap is carried out during the course of recombination on a homologous chromosome template within a G2 phase of a cell cycle, upon completion of which NER proceeds again. The subsequent stages of repair are the same as those for the first mechanism. If cross-links occur somewhere on the non-telomeric chromosome regions, then these two repair mechanisms can act without any problems arising. If, however, they arise too close to an end of a double-stranded DNA, in such a way that there are 8 or less nucleotides left towards the place where the 3' overhang begins on a G-strand during gap formation, then NER system activity will lead subsequently to such overhang loss and incomplete repair with telomere shortening. It could probably be restored to a previous state only in the case of the recombination-mediated DNA synthesis, which on the very ends of chromosomes, is very tightly blocked, as has already been mentioned. Every other event here is similar to those that have been described in relation to BER and NER.

Thus, we have reviewed the possible cases of the occurrence of incomplete repair and of DDMTS for various types of the damage of chromosome ends with 3' overhangs, which in their turn, should lead to the telomere shortening. At the same, not all such possible cases (as well as variations of damage and mechanisms for their repair) have been analysed but only those that seem to be the most important. Also, it should be emphasized that every possible case of incomplete repair and DDMTS, which has been assumed, can arise on uncapped linear telomere ends. If telomere ends are in a capped condition (i.e. in the form of a t-loop) then already other such cases will probably be observed, which also will lead to their shortening. Nevertheless, it seems for us that if telomere ends are organised into t-loops then the cases of incomplete repair and DDMTS characterised by telomere shortening will occur much less frequently than with linear telomeres. It should be noted that experimental data fully and directly confirming the appearance of incomplete repair or DDMTS for different described variants of damage could not be found. At the same time, there are many studies providing general information, demonstrating that various kinds of damage can occur on telomeres, which are repaired much less efficiently than those which are formed on the non-telomeric chromosome regions, and that they lead to telomere shortening (Passos et al., 2007).

### 3.4. The shelterin-mediated telomere repair problem

Many investigations have been performed focussing on the influence of reactive oxygen species (ROS) on the occurrence of telomere damage and associated telomere shortening. In one of the earliest studies it was found that mild hyperoxia leads to accelerated telomere shortening and inhibits the proliferation of fibroblasts which, as it was supposed at that time, could happen due to the accumulation of single-stranded DNA breaks on chromosome end regions (von Zglinicki et al., 1995). In another study, it was found that oxidative stress really leads to the accumulation of single-stranded breaks on telomeres whereupon they actually become rapidly shortened (Petersen et al., 1998). In yet another study, it was found that ultraviolet radiation combined with riboflavin induces the formation of 8-oxo-7, 8-dihydro-2'-deoxyguanosine (8-oxodG) in DNA fragments with telomere sequences that further leads to the appearance of breaks in the area of the central guanine of GGG sequences. It was also shown that under the influence of hydrogen dioxide ($H_2O_2$) together with Cu (II) on these

fragments DNA damage also occurred, which included the formation of 8-oxodG at the GGG sequence in the telomere sequence (5'-TTAGGG-3'), and which also led to breaks. Therefore, it was concluded that the formation of 8-oxodG in a GGG telomere sequence triplet induced by oxidative stress could play an important role in the acceleration of telomere shortening (Kawanishi & Oikawa, 2004). Along with these studies, many others showing that ROS leads to telomeric DNA damage formation and to their subsequent shortening are known (Passos & Von Zglinicki, 2006; Richter & von Zglinicki, 2007; Saretzki et al., 1999; Tchirkov & Lansdorp, 2003; Toussaint et al., 2000; von Zglinicki, 2000; von Zglinicki et al., 2000; von Zglinicki, 2002). It is also possible that besides the occurrence of single-stranded breaks on telomeres under the influence of ROS, there could also be oxidative modifications of nucleotides, which should be subject to repair by BER and NER systems, as well as double-stranded breaks (Passos et al., 2007). ROS has special importance concerning telomere damage, because unlike other mutagenic factors such as ionizing radiation, ultraviolet emanation, different chemical agents etc, ROS are constantly formed by mitochondria in a human organism during its normal metabolic activity. This is what triggered a strong interest to their study. Based on this, it may be that these other damaging agents can affect an organism and damage its telomeres in only very rare cases, while ROS continuously damages chromosomes' end regions, leading to their shortening. This situation actually should explain the fact that under normal conditions in the course of a cell's division, the telomere shortening rate considerably exceeds that which is expected only as a result of 3' overhang loss under the end replication problem (Keys et al., 2004). There are also studies which demonstrate that ROS can directly damage mitochondria themselves, and their mitochondrial DNA (mtDNA) in particular, thereby leading to their dysfunction which in turn can lead to the more intensive production of free radicals and, as a consequence, can result in even more intensive telomere damage and their shortening (Liu et al., 2002; Passos et al., 2006; Passos et al., 2007).

As was already noted, the damage occurring on telomeres is repaired less efficiently than that which originates in other genome regions (Kruk et al., 1995; Petersen et al., 1998; von Zglinicki, 2002). The reasons for such deficiencies in telomere-specific repair have not yet been completely established. At the same time, it is supposed that a basic role in the significant lessening of damage repair efficacy on telomeres belongs to the formation at their ends of the t-loops (capped telomere condition) (Passos et al., 2007). It was mentioned earlier that t-loops protect telomere ends from the activity of repair system proteins and another enzymatic influences (Grach, 2009; Griffith et al., 1999; Stansel et al., 2001). The example of the latter, incidentally, could be the telomerase attaching to the 3' overhang and its elongation. How does such repair suppression on telomeres by t-loops proceed? In order to answer this question let us first consider in detail what these t-loops represent and what actually characterises the response to DNA damage.

A t-loop represents a structure on eukaryotic chromosome ends which is formed at the bending back and subsequent insertion of a single-stranded telomeric DNA (3' overhang) into a double-stranded one (Grach, 2009). Upon this, the G-overhang forms a heteroduplex with the complementary C-strand region in double-stranded telomeric DNA, which is accompanied by the local untwisting of the latter and which leads to the formation of the so-called displacement

loop (D-loop). The latter represents a triple-stranded structure which consists of a double-stranded DNA, formed by a 3' overhang and a C-strand of the previous double-stranded DNA, and single-stranded DNA, corresponding to the G-strand region of the previous double-stranded DNA (Fig. 7).
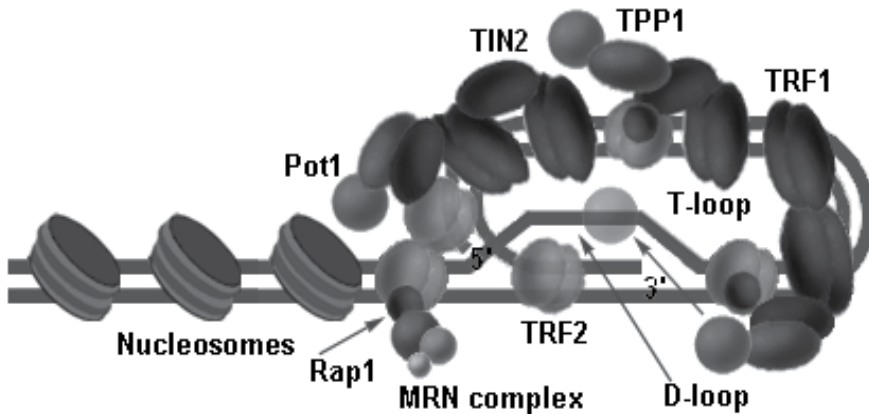


**Figure 7.** Structural organisation of a mammalian t-loop (Grach, 2009).

In t-loop formation, the primary role belongs to specific telomeric proteins, which are collectively referred to as the shelterin complex or telosome. Telomeric proteins differ slightly with different groups of organisms, but along with this they carry out similar functions. In mammals, shelterin includes six basic proteins, namely TRF1, TRF2, Rap1, TIN2, Pot1 and TPP1. These in turn can be divided into three groups: 1). double-stranded telomeric DNA binding proteins (TRF1 and TRF2); 2). single-stranded telomeric DNA binding proteins (Pot1); 3). proteins necessary for higher-order nucleoprotein complex formation (Rap1, TIN1 and TPP1) (Grach, 2009). The role of TRF1 function in t-loop formation it is to promote the bending back, twisting and linking of double-stranded telomeric DNA regions (Bianchi et al., 1997; Bianchi et al., 1999; Griffith et al., 1998; Griffith et al., 1999). Subsequently, TRF2 carries out the self-introduction of single-stranded telomeric DNA into a double-stranded one (Greider et al., 1999; Griffith et al., 1999; Stansel et al., 2001). Pot1 binds to single-stranded regions of telomeric DNA, which are represented only by a G-strand, and stabilises them (Baumann & Cech, 2001; Bunch et al., 2005; Churikov et al., 2006). Moreover, if a telomere end turns up in an uncapped condition, then Pot1 will cover the 3' overhang. If it turns up in a capped condition, then this protein will bind a single-stranded G-strand region in a D-loop structure. The Rap1 protein interacts with telomeres through binding with TRF2 (Li et al., 2000). It is suggested that the main functions of Rap1 are connected with its ability to recruit various repair system proteins, including Mre11/Rad50/Nbs1, Ku70/80 and PARP-1, to the telomeres (O'Connor et al., 2004). TIN2's role consists of both the binding of TRF1 proteins among themselves (Kim et al., 1999) and the binding of TRF1 with TRF2 (Ye et al., 2004). TPP1 is also necessary for the binding of TRF1 and TRF2 sub-complexes. It assists in the stabilising of TRF1-TIN2-TRF2 interaction. Besides this, TPP1 also directly binds Pot1 and regulates its activity (O'Connor et

al., 2006). Thus, TIN2 and TPP1 play a key role in the association of different telomeric proteins in a single functional complex, which participates in the formation of t-loops and the capping of telomere ends, and this provides telomeres protection from different kinds of enzymatic action.

The DNA damage response is characterised by the following circumstances. It is possible to distinguish such basic key points as DNA damage detection from repair itself. Earlier, we considered the most important repair mechanisms, and therefore we will analyse the damage detection pathways. As is known, there are two distinct DNA damage detection pathways, which can potentially be activated by a chromosome's natural ends, namely the ATM kinase pathway and the ATR kinase pathway (de Lange, 2010). The ATM kinase pathway is triggered in response to double-stranded DNA breaks (Lee & Paull, 2007). Upon this, the process by which the ATM kinase response is accomplished is still not completely clear (Lee & Paull, 2007). It is known that the Mre11/Rad50/Nbs1 complex (sensor proteins which take part in double-stranded breaks' detection) binds to DNA ends upon this and activates the ATM kinase in a combination with Tip60 HAT (Carson et al., 2003; de Lange, 2010; Williams et al., 2010). Later on, the ATM kinase phosphorylates some key proteins involved in the damage response to DNA double-strand breaks and initiates the activation of the DNA damage checkpoint, which pauses the cell cycle and allows time for a cell to repair damage before continuing its division (Lee & Paull, 2007). Upon this, Tip60 HAT, through histone acetylation, modulates repair proteins' loading and repair of double-stranded DNA breaks (Murr et al., 2006). In that case, if the damage is not resolved, the p53 protein is then activated, which triggers an apoptosis program (Polyak et al., 1997). The ATR kinase pathway is activated in response to the single-stranded DNA, and is based on the point that the abundant single-strand DNA binding protein RPA recognises and associates with single-stranded DNA, resulting in an RPA-ssDNA complex (Cimprich & Cortez, 2008; de Lange, 2010; Nam & Cortez, 2011). Further, the ATR kinase together with the ATRIP protein recognises such a single-stranded DNA coated with RPA and attached to the DNA's damage site (Cimprich & Cortez, 2008). At the same, the ATR kinase's recruitment to the RPA-ssDNA complex strongly depends on the ATRIP protein, which itself directly attaches to RPA-ssDNA, and thus binds this complex with the ATR (Zou & Elledge, 2003). The checkpoint clamp complex containing RAD9-HUS1-RAD1 (9-1-1) proteins, which take part in checkpoint activation, cell cycle arrest, and recruitment of specific DNA polymerases and other repair proteins to damaged DNA is also independently recruited to the DNA damage site by RAD17 protein (checkpoint clamp loader) (Bermudez et al., 2003; Sohn & Cho, 2009). Besides, the ATR activator TOPB1 is recruited to the DNA damage site (Choi et al., 2009). After these events, the activation of ATR by TOPBP1 and the phosphorylation of downstream targets in a signal transduction cascade proceeds, which eventually leads to checkpoint activation, cell cycle arrest and subsequent damage repair (Cimprich & Cortez, 2008). Later on, all of the events are similar to those which were considered concerning the ATM kinase pathway, i.e. if damage is completely repaired then the cell cycle is resumed and the cell will continue its division, and if not then there will occur a trigger of the apoptosis program and subsequent cell death.

Now that we have considered what t-loops represent in themselves and what the DNA damage response is characterised by, it is possible to answer the question – how the repair of damage on telomeres is so strongly repressed. As such, it was proposed that shelterin hides the chromosome end from the ATM kinase pathway of DNA damage detection by remodelling telomeres into a closed structure – the t-loop. In a t-loop, Mre11/Rad50/Nbs1 is unlikely to recognise the telomere end as a double-stranded DNA end, which thus prevents ATM kinase activation, with subsequent cell cycle arrest and initiation of DNA damage repair (de Lange, 2009; de Lange, 2010; Griffith et al., 1999). This situation can be implemented as after the DNA replication, as well as after the double-stranded breaks occurred at telomeres. On the other hand, ATR signalling on telomeres is blocked by the shelterin Pot1 protein. It was noted earlier that, telomeres, on their own ends, contain a single-stranded DNA. This DNA at the uncapped condition of telomere ends is represented by 3' overhangs, and at the capped condition by a single-stranded region of a G-strand as a part of a D-loop. Furthermore, such single-stranded DNA can arise after a single-stranded damage and double-stranded breaks occurred at telomeres. It is potentially capable of activating the ATR kinase; however it was suggested that Pot1 binds a single-stranded telomere DNA and excludes the RPA protein from it. Later on, in the absence of this protein, such single-stranded DNA can no longer be distinguished by the ATR-ATRIP complex as damage, which prevents ATR kinase pathway activation and all subsequent events, including repair (de Lange, 2009; de Lange, 2010; Denchi & de Lange, 2007).

Along with the blocking of ATM and ATR kinase DNA damage detection pathways, shelterin can also block the DNA repair reactions by the direct blocking of repair proteins. It was mentioned earlier that the repair of double-stranded breaks can basically be performed by two pathways - NHEJ and HDR. These two pathways in turn are triggered in a manner similar to the ATM and ATR signalling pathways in various ways (de Lange, 2010). NHEJ first employs the ring-shaped Ku70/80 protein complex, which loads onto DNA ends and facilitates their further synapsis and ligation by DNA ligase IV. As such, there is the suggestion, that a t-loop in addition to the repression of the ATM signalling pathway, also - probably - effectively blocks Ku70/80 joining and thus could thwart NHEJ in its earliest steps (de Lange, 2009; de Lange, 2010; Palm & de Lange, 2008). Besides this, the possibility was discussed that additional mechanisms not involving the t-loop can be used for telomere ends' protection from NHEJ (de Lange, 2010). It is suggested that POT1 contributes in NHEJ repression, especially after DNA replication when the t-loop is not yet formed (de Lange, 2009). HDR is initiated when Rad51 (the protein playing one of the most important roles in homologous recombination, since it organizes the proteinaceous complex which is necessary for chromosomes pairing and subsequent DNA strands exchange (Babynin, 2007)) replaces RPA on a single-stranded DNA (de Lange, 2010). In this connection, for blocking such a pathway of DSB's repair at telomeres, it is enough simply to repress RPA binding. Such repression on the telomere ends of mammals is carried out with the help of the POT1 protein, the binding of which to a single-stranded telomeric DNA, as was already noted, excludes RPA (de Lange, 2010). Therefore, POT1 is probably capable not only of blocking the ATR signalling pathway, and NHEJ after DNA replication, but also HDR on telomeres (de Lange, 2010). Besides this, there is data suggesting that Ku70/80 is also capable of repressing HDR in the absence of POT1 (Celli et al., 2006). It is necessary to also note that there is experimental data showing that TRF2 overexpression

weakens the repair of single-stranded breaks on telomeres, resulting in their accelerated shortening. This suggests the possibility that the repair of such damage on telomeres, as well as the other types of single-stranded damage, is again hindered by t-loops and shelterin, in the formation of which TRF2 participates. In this connection, at increased TRF2 production, the t-loops are probably formed more intensively in this case, and damages are repaired less effectively, leading to accelerated telomere shortening (Richter et al., 2007).

Thus, shelterin can inhibit repair on telomeres by the repression of various pathways of DNA damage detection, as well as of their repair itself. T-loop formation, in particular, leads to the blocking of the access of Mre11/Rad50/Nbs1 and Ku70/80 to double-stranded DNA ends, which prevents the activation of ATM signalling pathway and NHEJ that in its turn protects DNA natural ends, but blocks the repair of double-stranded breaks at telomeres. Besides, it is possible that the t-loop sterically blocks the repair of single-stranded damage at the telomeres by its three-dimensional structure. POT1 binding to single-stranded telomeric DNA excludes RAP from it and, therefore, prevents the recognition of damage by ATR in a complex with ATRIP, and which in turn prevents ATR kinase pathway activation. POT1 also blocks the binding of the Rad51 protein to single-stranded telomeric DNA, which prevents repair through HDR. Besides this, it is supposed that POT1 - after DNA replication when the t-loop yet is not formed - takes part in NHEJ repression, which is carried out, apparently, by Ku70/80 hetero-dimer blocking. There is also data suggesting that telomeric proteins themselves directly repress the pathways of single-stranded damage repair as well.

All of these mechanisms for repair repression on telomeres would seem, at first sight, to be the enemies for them, since repair deficiency leads to the circumstance where damage, occurring at chromosomes' ends is badly repaired and this leads to their shortening and, ultimately leads to cell cycle arrest and apoptosis. Nevertheless, a certain amount of time is needed for this purpose, which in some cases can be a quite considerable. On the other hand, if telomeric proteins did not repress the DNA damage response at chromosomes' ends, it would result in apoptosis being triggered immediately rather than after telomeres had been shortened to a critical degree, which is caused by the following circumstances. If shelterin were be absent on distal telomere regions, or in other words if nucleosome organisation would be observed rather than telosome organisation, that probably would not distinguish them from other chromosome areas – it would lead to such a case whereby DNA natural ends would be recognised by the damage detection system as double-stranded breaks. In this connection, Mre11/Rad50/Nbs1 would activate subsequently the ATM kinase pathway. Besides this, it is known that MRN, attaching to double-stranded DNA ends and recognising them as DSBs, is also capable of performing the 5'-3' resection of such ends, thus creating 3' single-stranded tails or overhangs (Mimitou & Symington, 2009). With the absence of telosome and, in this case, of the POT1 protein in particular, this would now lead to the activation of the ATR kinase pathway. The activation of any of the DNA damage detection pathways would lead to cell cycle arrest and damage repair. The repair of such, let us say, false DSBs in the case of ATM signalling pathway would probably be carried out by the NHEJ way. Insofar as during NHEJ broken chromosome ends are directly joined, but in our hypothetical variant the natural ends of different chromo-somes would be recognized as broken ends, then NHEJ in this case could lead to various

chromosomes' fusion with one another. This situation will result in genome instability, which leads to the initiation of the apoptosis program. The repair of damage by the ATR kinase pathway due to the presence of recombinogenic 3' single-stranded protrusions would possibly occur through HDR. In such a case, the fusion of different chromosomes and, later on, cell destruction through the initiation of the apoptosis program, will also take place. There is experimental data supporting all of this and showing that repression of the shelterin proteins leads to chromosomes' fusion and subsequent apoptosis. Besides this, such chromosome ends' vulnerability from homologous recombination could lead to continuous telomere elongation through the ALT which would essentially increase the probability of cell transformation. Thus, the repair suppression mechanisms on telomeres, although they lead to the accumulation of damage and shortening, at the same time protect chromosomes' ends from fusion among themselves and ALT activation. In the case of protection against chromosome fusion, it appreciably extends cells' lifespan. As for protection from ALT, thus shelterin prevents inappropriate telomere elongation and importantly the probability of cells' transformation. Therefore, the repair suppressing mechanisms on telomeres appear as friends for them.

It follows from the discussion above that the DNA damage response on telomeres, as well as damage detection and their repair pathways is potentially detrimental. This view was held for a long period of time. However, opinions changed when it was discovered that the proteins involved in the DNA damage response were present on the functional telomeres of mammals and interact with shelterin components, as previously in some way mentioned in reference to telomeric Rap1 protein, which recruits various proteins of the repair system to telomeres (Boulton & Jackson, 1996; Francia et al., 2006; Nugent et al., 1998). Moreover, further experiments have shown that they are also involved in telomere maintenance. This discovery was absolutely unexpected and somewhat paradoxical, as earlier it was thought that these proteins were the enemies for telomeres. However, it has now been discovered that this is not exactly the case, which has radically changed established views. Moreover, from now on DNA damage response proteins should be considered at the same time as equally the enemies and friends of telomeres. For the first time, such a role for DNA damage response proteins was found in yeast, where the NHEJ factor Ku is required for the maintenance of telomeres (Boulton & Jackson, 1996; Nugent et al., 1998; Polotnianka et al., 1998). In particular, in one such study it was found that in cells lacking telomerase but with functional Ku, telomere shortening slightly decelerates, i.e. it is less than in cells with repressed Ku function (Nugent et al., 1998). Further studies have shown that in addition to Ku, other proteins, involved in various DNA damage detection and their repair pathways, widely interact with telomeric proteins (Francia et al., 2006; Hsu et al., 1999; Hsu et al., 2000; Lenain et al., 2006; Palm & de Lange, 2008; Tarsounas et al., 2004; van Overbeek & de Lange, 2006; Zhu et al., 2000). Such factors in the mammals were called "shelterin accessory factors". Although they are present on telomeres transiently, at the same time they are very important for the maintenance of their normal structural organisation and functioning (Palm & de Lange, 2008).

Thus we emphasise, once again, that the main cause of poor damage repair on telomeres in comparison with other chromosome regions is the formation by their ends of t-loops, a process in which shelterin directly participates. Shelterin protects telomeres from the influence of both

different sensor proteins that carry out DNA damage detection and the repair proteins themselves. Although this in turn leads to an accumulation of damage on telomeres and their shortening, at the same time it protects chromosomes from joining with each other, which ensures the maintenance of the integrity and normal efficiency of the cell genome. While shelterin protects telomeres from the action of DNA damage response proteins, it effectively interacts with them, which is also very important for telomere maintenance. It is necessary to add to all this also that in our opinion the cause of an inefficient damage repair on telomeres - specifically on their very ends - and their subsequent shortening, besides formation by telomeric proteins of the t-loops as well as their direct inhibiting influence on the DNA repair proteins is the fact that DNA repair proteins themselves cannot act correctly on the very DNA molecule end, which was discussed in the beginning of the section and can lead to incomplete repair or DDMTS.

### 3.5. The conclusion of this section

In summary, Olovnikov first described the problem of repairing the damage occurring on telomeres in the early 1990s. As at that time, it was considered that telomere ends had a double-stranded structure and so the model of telomere repair problem was described according to that conception. At that time, this problem was named the incomplete DNA repair problem, under which was considered the possibility of whether damage (DNA single-stranded break) occurred at a distance of only several nucleotides from the very end of a DNA molecule, then it could not be completely repaired in the course of copying of the complementary DNA strand because of DNA polymerases' functional peculiarities. As a result of this, telomeres should be shortened. In addition, based on that model, as well as on the point that in certain cases repair can start but does not come to completion or else does not begin at all, we considered it necessary to propose – an incomplete DNA repair and damage-mediated telomere shortening, or DDMTS. Upon this, incomplete repair and DDMTS - which in several different ways can lead to telomere shortening - can be considered to be two variants of the end repair problem. Considering these proposed new conceptions, the old model of the end repair problem was revised. As the time, it was found that the structure of telomere ends is not double-stranded, as was supposed earlier, but 3' overhanging single-stranded, we considered the eventualities of incomplete repair and the occurrence of DDMTS on the very linear telomere ends in instances of single-stranded breaks and other types of damage in accordance with this model. Theoretically, it is expected that incomplete repair and DDMTS for different variants of damage should uniquely lead to telomere shortening. However, it is not known whether it can actually occur, as unfortunately we were unable to find the experimental data confirming these assumptions. At the same time, there is much general data showing that damage on telomeres occurs and that this damage is repaired less effectively than damage in other regions of the chromosome. It was demonstrated that researchers' main attention in this has been given to telomere damage by reactive oxygen species, which are constantly formed in mitochondria at normal cell vital activity. It should explain why the genuine telomere-shortening rate exceeds the one that is expected as being only as a result of end replication problem. The circumstances by which the damage on telomeres is repaired more poorly has been explained well by many researchers in terms of t-loop formation. Therefore, in order to demonstrate exactly how t-

loops repress repair, we considered what they represent in themselves and also what DNA damage response is characterised by. As a result, it was shown that t-loops, formed with the participation of the shelterin protein complex, as well as the telomeric proteins themselves, block various damage detection mechanisms and their repair pathways directly, and this actually causes telomere repair deficiency. This can be designated as the "shelterin-mediated telomere repair problem". It should be distinguished from the end repair problem, which can be carried out not along the whole telomere length but only on their ends, because of the inability of repair proteins to act on a template end. Thus, in principle, it is possible to identify two problems of repair on telomeres – the end repair problem, which is carried out near the very DNA molecule ends, and the shelterin-mediated telomere repair problem, which can affect all telomere regions where there is telosome organisation. At the same time, both of these problems can be referred to, in general, as the telomere repair problem.

## 4. Conclusion

In summarising all the data, it is necessary to emphasise that there exist two basic telomere-shortening mechanisms – the end replication problem and the telomere repair problem. The end replication problem, which is based on the 3' overhang loss during the course of DNA leading strand synthesis while the genetic material is doubling, has been studied in depth. The study of this problem allowed for the discovery of the telomerase enzyme and finding of the connection between telomere shortening and ageing, as well as carcinogenesis and various degenerative diseases. At the same time, many aspects of the end replication problem are still not absolutely clear. Therefore, further detailed studies of this process are necessary. The problem of telomere repair has been studied much less. Thus, there is no experimental data fully confirming that the end repair problem - which includes incomplete repair and DDMTS - can really be carried out on telomere ends and so lead to their shortening. At the same time, there is general data showing that damage on telomeres is frequently formed and repaired much less efficiently than on other chromosome regions leading to telomeric shortening. The researchers' main focus has been given to the influence of ROS on telomere damage. Given that they are constantly formed in cells by mitochondria, this is quite justifiable. At the same time, it is also necessary to study other negative factors concerning telomere damage and their shortening. Faint damage repair on telomeres is explained mainly by the formation at their ends of the t-loops, which are created with the participation of the telomeric protein complex shelterin and block different proteins involved in DNA damage response in order to prevent chromosomes' fusion with each other. This situation was referred to as the shelterin-mediated telomere repair problem. In this connection, it is necessary to study in more detail the structure of the proteins included in shelterin and their functions as well as mechanisms for t-loop formation. It is also important to study the proteins, which take part in DNA damage detection and the repair process itself. The latter should be studied in relation to both DNA damage response and their role in telomere maintenance. Thus, telomere-shortening mechanisms remain quite poorly understood and require further research.

## Author details

Andrey  Grach[*]

Khmelnitsky Regional Hospital, Ukraine

## References

[1] Allsopp, R. C.; Vaziri, H., Patterson, C., Goldstein, S., Younglai, E. V., Futcher, A. B., Greider, C. W. & Harley, C. B. (1992). Telomere Length Predicts Replicative Capacity of Human Fibroblasts. *Proceedings of the National Academy of Sciences of the United States of America*, Vol.89, No.21, (November 1992), pp. 10114-10118.

[2] Allsopp, R. C.; Chang, E., Kashefi-Aazam, M., Rogaev, E. I., Piatyszek, M. A., Shay, J. W. & Harley C. B. (1995). Telomere Shortening is Associated with Cell Division *in vitro* and *in vivo*. *Experimental Cell Research*, Vol.220, No.1, (September 1995), pp. 194-200.

[3] Aubert, G. & Lansdorp, P. M. (2008). Telomeres and Aging. *Physiological Reviews*, Vol. 88, No.2, (April 2008), pp. 557-579.

[4] Azzalin, C. M.; Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner J. (2007). Telomeric Repeat Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. *Science*, Vol.318, No.5851, (November 2007), pp. 798-801.

[5] Babynin, E. V. (2007). Molecular Mechanism of Homologous Recombination in Meiosis: Origin and Biological Significance. *Tsitologiia*, Vol.49, No.3, (March 2007), pp. 182-193.

[6] Baumann, P. & Cech, T. R. (2001). Pot1, the Putative Telomere End-Binding Protein in Fission Yeast and Humans. *Science*, Vol.292, No.5519, (May 2001), pp. 1171-1175.

[7] Baur, J. A.; Zou, Y., Shay, J. W. & Wright, W. E. (2001). Telomere Position Effect in Human Cells. *Science*, Vol.292, No.5524, (June 2001), pp. 2075-2077.

[8] Bermudez, V. P.; Lindsey-Boltz, L. A., Cesare, A. J., Maniwa, Y., Griffith, J. D., Hurwitz, J. & Sancar, A. (2003). Loading of the Human 9-1-1 Checkpoint Complex onto DNA by the Checkpoint Clamp Loader hRad17-Replication Factor C Complex *in vitro*. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 100, No.4, (February 2003), pp. 1633-1638.

[9] Bianchi, A.; Smith, S., Chong, L., Elias, P. & de Lange, T. (1997). TRF1 is a Dimer and Bends Telomeric DNA. *The EMBO Journal*, Vol.16, No.7, (April 1997), pp. 1785-1794.

[10] Bianchi, A.; Stansel, R. M., Fairall, L., Griffith, J. D., Rhodes, D. & de Lange, T. (1999). TRF1 Binds a Bipartite Telomeric Site With Extreme Spatial Flexibility. *The EMBO Journal*, Vol.18, No.20, (October 1999), pp. 5735-5744.

[11] Blackburn, E. H. (2001). Switching and Signaling at the Telomere. *Cell*, Vol.106, No.6, (September 2001), pp. 661-673.

[12] Blackburn, E. H. & Collins, K. (2011). Telomerase: an RNP Enzyme Synthesizes DNA. *Cold Spring Harbor Perspectives in Biology*, Vol.3, No.5, (May 2011), pii: a003558. doi: 10.1101/cshperspect. a003558.

[13] Bouché, J. P.; Rowen, L. & Kornberg, A. (1978). The RNA Primer Synthesized by Primase to Initiate Phage G4 DNA Replication. *The Journal of Biological Chemistry*, Vol. 253, No.3, (February 1978), pp. 765-769.

[14] Boulton, S. J. & Jackson, S. P. (1996). Identification of a Saccharomyces cerevisiae Ku80 Homologue: Roles in DNA Double Strand Break Rejoining and in Telomeric Maintenance. *Nucleic Acids Research*, Vol.24, No.23, (December 1996), pp. 4639-4648.

[15] Bunch, J. T.; Bae, N. S., Leonardi, J. & Baumann, P. (2005). Distinct Requirements for Pot1 in Limiting Telomere Length and Maintaining Chromosome Stability. *Molecular and Cellular Biology*, Vol.25, No.13, (July 2005), pp. 5567-5578.

[16] Burgers, P. M. (2009). Polymerase Dynamics at the Eukaryotic DNA Replication Fork. *The Journal of Biological Chemistry*, Vol.284, No.7, (February 2009), pp. 4041-4045.

[17] Carson, C. T.; Schwartz, R. A., Stracker, T. H., Lilley, C. E., Lee, D. V. & Weitzman, M. D. (2003). The Mre11 Complex is required for ATM Activation and the G2/M Checkpoint. *The EMBO Journal*, Vol.22, No.24, (December 2003), pp. 6610-6620.

[18] Celli, G. B.; Denchi, E. L. & de Lange, T. (2006). Ku70 Stimulates Fusion of Dysfunctional Telomeres Yet Protects Chromosome Ends from Homologous Recombination. *Nature Cell Biology*, Vol.8, No.8, (August 2006), pp. 885-890.

[19] Chai, W.; Du, Q., Shay, J. W. & Wright, W. E. (2006). Human Telomeres Have Different Overhang Sizes at Leading Versus Lagging Strands. *Molecular Cell*, Vol.21, No.3, (February 2006), pp. 427-435.

[20] Choi, J. H.; Lindsey-Boltz, L. A. & Sancar, A. (2009). Cooperative Activation of the ATR Checkpoint Kinase by TopBP1 and Damaged DNA. *Nucleic Acids Research*, Vol. 37, No.5, (April 2009), pp. 1501-1509.

[21] Chu G. (1997). Double Strand Break Repair. *The Journal of Biological Chemistry*, Vol. 272, No.39, (September 1997), pp. 24097-24100.

[22] Churikov, D.; Wei, C. & Price, C. M. (2006). Vertebrate POT1 Restricts G-Overhang Length and Prevents Activation of a Telomeric DNA Damage Checkpoint but is Dispensable for Overhang Protection. *Molecular and Cellular Biology*, Vol.26, No.18, (September 2006), pp. 6971-6982.

[23] Cimino-Reale, G.; Pascale, E., Battiloro, E., Starace, G., Verna, R. & D'Ambrosio, E. (2001). The Length of Telomeric G-rich Strand 3'-Overhang Measured by Oligonucleotide Ligation Assay. *Nucleic Acids Research*, Vol.29, No.7, (April 2001), E35.

[24] Cimino-Reale, G.; Pascale, E., Alvino, E., Starace, G. & D'Ambrosio, E. (2003). Long Telomeric C-Rich 5'-Tails in Human Replicating Cells. *The Journal of Biological Chemistry*, Vol.278, No.4, (January 2003), pp. 2136-2140.

[25] Cimprich, K. A. & Cortez, D. (2008). ATR: an Essential Regulator of Genome Integrity. *Nature Reviews Molecular Cell Biology*, Vol.9, No.8, (August 2008), pp. 616-627.

[26] Conrad, M. N.; Dominguez, A. M. & Dresser, M. E. (1997). Ndj1p, a Meiotic Telomere Protein Required for Normal Chromosome Synapsis and Segregation in Yeast. *Science*, Vol.276, No.5316, (May 1997), pp. 1252-1255.

[27] Dai, H.; Liu, J., Malkas, L. H. & Hickey, R. J. (2009). Characterization of RNA Primers Synthesized by the Human Breast Cancer Cell DNA Synthesome. *Journal of Cellular Biochemistry*, Vol.106, No.5, (April 2009), pp. 798-811.

[28] Dai, X.; Huang, C., Bhusari, A., Sampathi, S., Schubert, K. & Chai, W. (2010). Molecular Steps of G-Overhang Generation at Human Telomeres and its Function in Chromosome End Protection. *The EMBO Journal*, Vol.29, No.16, (August 2010), pp. 2788-2801.

[29] de Laat, L.; Jaspers, N. G. & Hoeijmakers, J. H. (1999). Molecular Mechanism of Nucleotide Excision Repair. *Genes and Development*, Vol.13, No.7, (April 1999), pp. 768-785.

[30] de Lange, T. (2009). How Telomeres solve the End-Protection Problem. *Science*, Vol. 326, No.5955, (November 2009), pp. 948-952.

[31] de Lange, T. (2010). Telomere Biology and DNA Repair: Enemies with Benefits. *FEBS Letters*, Vol.584, No.17, (September 2010), pp. 3673-3674.

[32] Denchi, E. L. & de Lange, T. (2007). Protection of Telomeres through Independent Control of ATM and ATR by TRF2 and POT1. *Nature*, Vol.448, No.7157, (August 2007), pp. 1068-1071.

[33] Desmaze, C.; Soria, J. C., Freulet-Marrière, M. A., Mathieu, N. & Sabatier, L. (2003). Telomere-Driven Genomic Instability in Cancer Cells. *Cancer Letters*, Vol.194, No.2, (May 2003), pp. 173-182.

[34] Dewar, J. M. & Lydall, D. (2010). Telomere Replication: Mre11 Leads the Way. *Molecular Cell*, Vol.38, No.6, (June 2010), pp. 777-779.

[35] Dionne, I. & Wellinger, R. J. (1996). Cell Cycle-Regulated Generation of Single-Stranded G-Rich DNA in the Absence of Telomerase. *Proceedings of the National Academy of Sciences of the United States of America*, Vol.93, No.24, (November 1996), pp. 13902-13907.

[36] Dong, C. K.; Masutomi, K. & Hahn, W. C. (2005). Telomerase: Regulation, Function and Transformation. *Critical Reviews in Oncology Hematology*, Vol.54, No.2, (May 2005), pp. 85-93.

[37] Dronkert, M. L. & Kanaar, R. (2001). Repair of DNA Interstrand Cross-Links. *Mutation Research*, Vol.486, No.4, (September 2001), pp. 217-247.

[38] Dynek, J. N. & Smith, S. (2004). Resolution of Sister Telomere Association is Required for Progression Through Mitosis. *Science*, Vol.304, No.5667, (April 2004), pp. 97-100.

[39] Francia, S.; Weiss, R. S., Hande, M. P., Freire, R. & d'Adda di Fagagna, F. (2006). Telomere and Telomerase Modulation by the Mammalian Rad9/Rad1/Hus1 DNA-Damage-Checkpoint Complex. *Current Biology*, Vol.16, No.15, (August 2006), pp. 1551-1558.

[40] Fromme, J. C. & Verdine, G. L. (2004). Base Excision Repair. *Advances in Protein Chemistry*, Vol.69, (n.d.) pp. 1-41.

[41] Gilson, E. & Geli, V. (2007). How Telomeres are Replicated. *Nature Reviews Molecular Cell Biology*, Vol.8, No.10, (October 2007), pp. 825-838.

[42] Grach, A. A. (2009). Structural Organization of Telomeres in Various Kinds of Organisms. *Tsitologiia*, Vol.51, No.11, (November 2009), pp. 869-879.

[43] Grach, A. A. (2011a). Alternative Telomere-lengthening Mechanisms. *Cytology and Genetics*, Vol.45, No.2, (March-April 2011), pp. 121-130.

[44] Grach, A. A. (2011b). The Role of Alternative Lengthening of Telomeres Mechanisms in Carcinogenesis and Prospects for Using of Anti-Telomerase Drugs in Malignant Tumors Treatment. *Tsitologiia*, Vol.53, No.10, (October 2011), pp. 3-15.

[45] Greider, C. W. (1999). Telomeres do D-Loop-T-Loop. *Cell*, Vol.97, No.4, (May 1999), pp. 419-422.

[46] Griep, M. A. (1995). Primase Structure and Function. *Indian Journal of Biochemistry and Biophysics*, Vol.32, No.4, (August 1995), pp. 171-178.

[47] Griffith, J.; Bianchi, A. & de Lange, T. (1998). TRF1 Promotes Parallel Pairing of Telomeric Tracts *in vitro*. *Journal of Molecular Biology*, Vol.278, No.1, (April 1998), pp. 79-88.

[48] Griffith, J. D.; Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H. & de Lange, T. (1999). Mammalian Telomeres End in a Large Duplex Loop. *Cell*, Vol.97, No.4, (May 1999), pp. 503-514.

[49] Hao, Y. H. & Tan, Z (2002). The Generation of Long Telomere Overhangs in Human Cells: a Model and its Implication. *Bioinformatics*, Vol.18, No.5, (May 2002), pp. 666-671.

[50] Hayflick, L. (1965). The Limited *in vitro* Lifetime of Human Diploid Cell Strains. *Experimental Cell Research*, Vol.37, (March 1965), pp. 614-636.

[51] Hediger, F.; Neumann, F. R., Van Houwe, G., Dubrana, K. & Gasser, S. M. (2002). Live Imaging of Telomeres: yKu and Sir Proteins Define Redundant Telomere-Anchoring Pathways in Yeast. *Current Biology*, Vol.12, No.24, (December 2002), pp. 2076-2089.

[52] Hemann, M. T. & Greider, C. W. (1999). G-Strand Overhangs on Telomeres in Telomerase-Deficient Mouse Cells. *Nucleic Acids Research*, Vol.27, No.20, (October 1999), pp. 3964-3969.

[53] Henderson, E. R. & Blackburn, E. H. (1989). An Overhanging 3' Terminus is a Conserved Feature of Telomeres. *Molecular and Cellular Biology*, Vol.9, No.1, (January 1989), pp. 345-348.

[54] Henson, J. D.; Neumann, A. A., Yeager, T. R. & Reddel, R. R. (2002). Alternative Lengthening of Telomeres in Mammalian Cells. *Oncogene*, Vol.21, No.4, (January 2002), pp. 598-610.

[55] Hsu, H. L.; Gilley, D., Blackburn, E. H. & Chen, D. J. (1999). Ku is Associated with the Telomere in Mammals. *Proceedings of the National Academy of Sciences of the United States of America*, Vol.96, No.22, (October 1999), pp. 12454-12458.

[56] Hsu, H. L.; Gilley, D., Galande, S. A., Hande, M. P., Allen, B., Kim, S. H., Li, G. C., Campisi, J., Kohwi-Shigematsu, T. & Chen, D. J. (2000). Ku Acts in a Unique Way at the Mammalian Telomere to Prevent End Joining. *Genes and Development*, Vol.14, No. 22, (November 2000), pp. 2807-2812.

[57] Huffman, K. E.; Levene, S. D., Tesmer, V. M., Shay, J. W. & Wright, W. E. (2000). Telomere Shortening is Proportional to the Size of the G-Rich Telomeric 3'-Overhang. *The Journal of Biological Chemistry*, Vol.275, No.26, (June 2000), pp. 19719-19722.

[58] Kawanishi, S. & Oikawa, S. (2004). Mechanism of Telomere Shortening by Oxidative Stress. *Annals of the New York Academy of Sciences*, Vol.1019, (June 2004), pp. 278-284.

[59] Keys, B.; Serra, V., Saretzki, G. & Von Zglinicki, T. (2004). Telomere Shortening In Human Fibroblasts is not Dependent on the Size of the Telomeric-3'-Overhang. *Aging Cell*, Vol.3, No.3, (June 2004), pp. 103-109.

[60] Kim, S. H.; Kaminker, P. & Campisi, J. (1999). TIN2, a New Regulator of Telomere Length in Human Cells. *Nature Genetics*, Vol.23, No.4, (December 1999), pp. 405-412.

[61] Kirk, K. E.; Harmon, B. P., Reichardt, I. K., Sedat, J. W. & Blackburn E. H. (1997). Block in Anaphase Chromosome Separation Caused by a Telomerase Template Mutation. *Science*, Vol.275, No. 5305, (March 1997), pp. 1478-1481.

[62] Klobutcher, L. A.; Swanton, M. T., Donini, P. & Prescott, D. M. (1981). All Gene-Sized DNA Molecules in Four Species of Hypotrichs have the Same Terminal Sequence

and an unusual 3' Terminus. *Proceedings of the National Academy of Sciences of the United States of America*, Vol.78, No.5, (May 1981), pp. 3015-3019.

[63] Krokan, H. E.; Standal, R., & Slupphaug, G. (1997). DNA Glycosylases in the Base Excision Repair of DNA. *Biochemical Journal*, Vol.325, No.Pt 1, (July 1997), pp. 1-16.

[64] Kruk, P. A.; Rampino, N. J. & Bohr, V. A. (1995). DNA Damage and Repair in Telomeres: Relation to Aging. *Proceedings of the National Academy of Sciences of the United States of America*, Vol.92, No.1, (January 1995), pp. 258-262.

[65] Kurenova, E. V. & Mason, J. M. (1997). Telomere Functions. A Review. *Biochemistry (Moscow)*, Vol.62, No.11, (November 1997), pp. 1242-1253.

[66] Larrivee, M.; LeBel, C. & Wellinger, R. J. (2004). The Generation of Proper Constitutive G-Tails on Yeast Telomeres is Dependent on the MRX Complex. *Genes and Development*, Vol.18, No.12, (June 2004), pp. 1391-1396.

[67] Lee, J. H. & Paull, T. T. (2007). Activation and Regulation of ATM Kinase Activity in Response to DNA Double-Strand Breaks. *Oncogene*, Vol.26, No.56, (December 2007), pp. 7741-7748.

[68] Lenain, C.; Bauwens, S., Amiard, S., Brunori, M., Giraud-Panis, M. J. & Gilson, E. (2006). The Apollo 5' Exonuclease Functions Together with TRF2 to Protect Telomeres from DNA Repair. *Current Biology*, Vol.16, No.13, (July 2006), pp. 1303-1310.

[69] Levy, M. Z.; Allsopp, R. C., Futcher, A. B., Greider, C. W. & Harley, C. B. (1992). Telomere End-Replication Problem and Cell Aging. *Journal of Molecular Biology*, Vol.225, No.4, (June 1992), pp. 951-960.

[70] Li, B.; Oestreich, S. & de Lange, T. (2000). Identification of Human Rap1: Implications for Telomere Evolution. *Cell*, Vol.101, No.5, (May 2000), pp. 471-483.

[71] Liang, L.; Deng, L., Nguyen, S. C., Zhao, X., Maulion, C. D., Shao, C. & Tischfield, J. A. (2008). Human DNA Ligases I and III, but not Ligase IV, are required for Microhomology-Mediated End Joining of DNA Double-Strand Breaks. *Nucleic Acids Research*, Vol.36, No.10, (June 2008), pp. 3297-3310.

[72] Lieber, M. R.; Ma, Y., Pannicke, U. & Schwarz, K. (2003). Mechanism and Regulation of Human Non-Homologous DNA End-Joining. *Nature Reviews Molecular Cell Biology*, Vol.4, No.9, (September 2003), 712-720.

[73] Lingner, J.; Cooper, J. P. & Cech, T. R. (1995). Telomerase and DNA End Replication: No Longer a Lagging Strand Problem? *Science*, Vol.269, No.5230, (September 1995), pp. 1533-1534.

[74] Liu, L.; Trimarchi, J. R., Smith, P. J. & Keefe, D. L. (2002). Mitochondrial Dysfunction Leads to Telomere Attrition and Genomic Instability. *Aging Cell*, Vol.1, No.1, (October 2002), pp. 40-46.

[75] Londoño-Vallejo, J. A. (2008). Telomere Instability and Cancer. *Biochimie*, Vol.90, No. 1, (January 2008), pp. 73-82.

[76] Mackenney, V. J.; Barnes, D. E. & 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. *The Journal of Biological Chemistry*, Vol. 272, No.17, (April 1997), pp. 11550-11556.

[77] Makarov, V. L.; Hirose, Y. & Langmore, J. P. (1997). Long G Tails at Both Ends of Human Chromosomes Suggest a C Strand Degradation Mechanism for Telomere Shortening. *Cell*, Vol.88, No.5, (March 1997), pp. 657-666.

[78] Maringele, L. & Lydall, D. (2002). EXO1-Dependent Single-Stranded DNA at Telomeres Activates Subsets of DNA Damage and Spindle Checkpoint Pathways in Budding Yeast yku70Delta Mutants. *Genes and Development*, Vol.16, No.15, (August 2002), pp. 1919-1933.

[79] McElligott, R. & Wellinger, R. J. (1997). The Terminal DNA Structure of Mammalian Chromosomes. *The EMBO Journal*, Vol.16, No.12, (June 1997), pp. 3705-3714.

[80] Meeker, A. K. & Coffey, D. S. (1997). Telomerase: A Promising Marker of Biological Immortality of Germ, Stem, and Cancer Cells. A Review. *Biochemistry (Moscow)*, Vol. 62, No.11, (November 1997), pp. 1323-1331.

[81] Mimitou, E. P. & Symington, L. S. (2009). DNA End Resection: Many Nucleases Make Light Work. *DNA Repair (Amsterdam)*, Vol.8, No.9, (September 2009), pp. 983-995.

[82] Muntoni, A. & Reddel R. R. (2005). The First Molecular Details of ALT in Human Tumor Cells. *Human Molecular Genetics*, Vol. 14, No.2, (October 2005), pp. 191-196.

[83] Murr, R.; Loizou, J. I., Yang, Y. G., Cuenin, C., Li, H., Wang, Z. Q. & Herceg, Z. (2006). Histone Acetylation by Trrap-Tip60 Modulates Loading of Repair Proteins and Repair of DNA Double-Strand Breaks. *Nature Cell Biology*, Vol.8, No.1, (January 2006), pp. 91-99.

[84] Nam, E. A. & Cortez, D. (2011). ATR Signalling: More than Meeting at the Fork. *Biochemical Journal*, Vol.436, No.3, (June 2011), pp. 527-536.

[85] Nosek, J.; Dinouël, N., Kovac, L. & Fukuhara, H. (1995). Linear Mitochondrial DNAs from Yeasts: Telomeres with Large Tandem Repetitions. *Molecular and General Genetics*, Vol.247, No.1, (April 1995), pp. 61-72.

[86] Nugent, C. I., Bosco, G., Ross, L. O., Evans, S. K., Salinger, A. P., Moore, J. K., Haber, J. E. & Lundblad, V. (1998). Telomere Maintenance is Dependent on Activities Required for End Repair of Double-Strand Breaks. *Current Biology*, Vol.8, No.11, (May 1998), p. 657-660.

[87]    O'Connor, M. S.; Safari, A., Liu, D., Qin, J. & Songyang, Z. (2004). The Human Rap1 Protein Complex and Modulation of Telomere Length. *The Journal of Biological Chemistry*, Vol.279, No.27, (July 2004), pp. 28585-28591.

[88]    O'Connor, M. S.; Safari, A., Xin, H., Liu, D. & Songyang, Z. (2006). A Critical Role for TPP1 and TIN2 Interaction in High-Order Telomeric Complex Assembly. *Proceedings of the National Academy of Sciences of the United States of America*, Vol.103, No.32, (August 2006), pp. 11874-11879.

[89]    Ohki, R.; Tsurimoto, T. & Ishikawa, F. (2001). In Vitro Reconstitution of the End Replication Problem. *Molecular and Cellular Biology*, Vol.21, No.17, (September 2001), pp. 5753-5766.

[90]    Olovnikov, A. M. (1971). Principle of Marginotomy in Template Synthesis of Polynucleotides. *Doklady Akademiii Nauk SSSR*, Vol.201, No.6, (n.d.), pp. 1496-1499.

[91]    Olovnikov, A. M. (1973). A Theory of Marginotomy: The Incomplete Copying of Template Margin in Enzymatic Synthesis of Polynucleotides and Biological Significance of the Phenomenon. *Journal of Theoretical Biology*, Vol.41, No.1, (September 1973), pp. 181-190.

[92]    Olovnikov, A. M. (1995a). The Effect of the Incomplete Terminal Repair of the Linear Double-Stranded DNA Molecule. *Izvestiia Akademii Nauk Seriia Biologicheskaia*, No.4, (July-August 1995), pp. 501-503.

[93]    Olovnikov, A. M. (1995b). The Role of Incomplete Terminal Repair of Chromosomal DNA in the Aging of Neurons and Postmitotic Organisms. *Izvestiia Akademii Nauk Seriia Biologicheskaia*, No.4, (July-August 1995), pp. 504-507.

[94]    Olovnikov, A. M. (1995c). The Possible Cellular Use of the Effect of Incomplete DNA Terminal Repair to Control the Correct Sequence of Events in Individual Development. *Ontogenez*, Vol.26, No.3, (May-June 1995), pp. 254-256.

[95]    Palm, W. & de Lange, T. (2008). How Shelterin Protects Mammalian Telomeres. *Annual Reviews of Genetics*, Vol.42, (December 2008), pp. 301-334.

[96]    Passos, J. F. & Von Zglinicki, T. (2006). Oxygen Free Radicals in Cell Senescence: Are they Signal Transducers? *Free Radical Research*, Vol.40, No.12, (December 2006), pp. 1277-1283.

[97]    Passos, J. F.; von Zglinicki, T. & Saretzki, G. (2006). Mitochondrial Dysfunction and Cell Senescence: Cause Or Consequence? *Rejuvenation Research*, Vol.9, No.1, (Spring 2006), pp. 64-68.

[98]    Passos, J. F.; Saretzki, G. & von Zglinicki, T. (2007). DNA Damage in Telomeres and Mitochondria during Cellular Senescence: is there a Connection? *Nucleic Acids Research*, Vol.35, No.22, (December 2007), pp. 7505-7513.

[99]   Pedram, M.; Sprung, C. N., Gao, Q., Lo, A. W. I., Reynolds, G. E. & Murnane, J. P. (2006). Telomere Position Effect and Silencing of Transgenes near Telomeres in the Mouse. *Molecular and Cellular Biology*, Vol.26, No.5, (March 2006), pp. 1865-1878.

[100]  Pennaneach, V.; Putnam, C. D. & Kolodner, R. D. (2006). Chromosome Healing by *de novo* Telomere Addition in *Saccharomyces cerevisiae*. *Molecular Microbiology*, Vol.59, No.5, (March 2006), pp. 1357-1368.

[101]  Petersen, S.; Saretzki, G. & von Zglinicki, T. (1998). Preferential Accumulation of Single-Stranded Regions in Telomeres of Human Fibroblasts. *Experimental Cell Research*, Vol.239, No.1, (February 1998), pp. 152-160.

[102]  Podgornaya, O. I.; Bugaeva, E. A., Voronin, A. P., Gilson, E. & Mitchell, A. R. (2000). Nuclear Envelope Associated Protein That Binds Telomeric DNAs. *Molecular Reproduction and Development*, Vol.57, No.1, (September 2000), pp. 16-25.

[103]  Polotnianka, R. M.; Li, J. & Lustig, A. J. (1998). The Yeast Ku Heterodimer is Essential for Protection of the Telomere Against Nucleolytic and Recombinational Activities. *Current Biology*, Vol.8, No.14, (July 1998), pp. 831-834.

[104]  Polyak, K.; Xia, Y., Zweier, J. L., Kinzler, K. W. & Vogelstein, B. (1997). A Model for p53-Induced Apoptosis. *Nature*, Vol.389, No.6648, (September 1997), pp. 300-305.

[105]  Reardon, J. T. & Sancar, A. (2005). Nucleotide Excision Repair. *Progress in Nucleic Acid Research and Molecular Biology*, Vol.79, (n.d.), pp. 183-235.

[106]  Rhyu, M. S. (1995). Telomeres, Telomerase, and Immortality. *Journal of the National Cancer Institute*, Vol.87, No.12, (June 1995), pp. 884-894.

[107]  Richter, T.; Saretzki, G., Nelson, G., Melcher, M., Olijslagers, S. & von Zglinicki, T. (2007). TRF2 Overexpression Diminishes Repair of Telomeric Single-Strand Breaks and Accelerates Telomere Shortening in Human Fibroblasts. *Mechanisms of Ageing and Development*, Vol.128, No.4, (April 2007), pp. 340-345.

[108]  Richter, T. & von Zglinicki, T. (2007). A Continuous Correlation between Oxidative Stress and Telomere Shortening in Fibroblasts. *Experimental Gerontology*, Vol.42, No. 11, (November 2007), pp. 1039-1042.

[109]  Sancar, A.; Lindsey-Boltz, L. A., Unsal-Kaçmaz, K. & Linn, S. (2004). Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints. *Annual Review of Biochemistry*, Vol.73, (July 2004), pp. 39-85.

[110]  Saretzki, G.; Sitte, N., Merkel, U., Wurm, R. E. & von Zglinicki, T. (1999). Telomere Shortening Triggers a p53-dependent Cell Cycle Arrest via Accumulation of G-rich Single Stranded DNA Fragments. *Oncogene*, Vol.18, No.37, (September 1999), pp. 5148-5158.

[111]  Seeberg, E.; Eide, L. & Bjørås, M. (1995). The Base Excision Repair Pathway. *Trends in Biochemical Sciences*, Vol.20, No.10, (October 1995), pp. 391-397.

[112] Sfeir, A. J.; Chai, W., Shay, J. W. & Wright, W. E. (2005). Telomere-End Processing: the Terminal Nucleotides of Human Chromosomes. *Molecular Cell*, Vol.18, No.1, (April 2005), pp. 131-138.

[113] Sohn, S. Y. & Cho, Y. (2009). Crystal Structure of the Human Rad9-Hus1-Rad1 Clamp. *Journal of Molecular Biology*, Vol.390, No.3, (July 2009), pp. 490-502.

[114] Stansel, R. M.; de Lange, T. & Griffith, J. D. (2001). T-Loop Assembly *in vitro* Involves Binding of TRF2 near the 3' Telomeric Overhang. *The EMBO Journal*, Vol.20, No.19, (October 2001), pp. 5532-5540.

[115] Stewart, S. A.; Ben-Porath, I., Carey, V. J., O'Connor, B. F., Hahn, W. C. & Weinberg, R. A. (2003). Erosion of the Telomeric Single-Strand Overhang at Replicative Senescence. *Nature Genetics*, Vol. 33, No.4, (April 2003), pp. 492-496.

[116] Stewart, S. A. & Weinberg, R. A. (2006). Telomeres: Cancer to Human Aging. *Annual Rewiev of Cell and Developmental Biology*, Vol.22, (November 2006), pp. 531-557.

[117] Stewart, S. A. (2005). Telomere Maintenance and Tumorigenesis: An "ALT"ernative Road. *Current Molecular Medicine*, Vol.5, No.2, (March 2005), pp. 253-257.

[118] Tarsounas, M.; Muñoz, P., Claas, A., Smiraldo, P. G., Pittman, D. L., Blasco, M. A. & West, S. C. (2004). Telomere Maintenance Requires the RAD51D Recombination/ Repair Protein. *Cell*, Vol.117, No.3, (April 2004), pp. 337-347.

[119] Tchirkov, A. & Lansdorp, P. M. (2003). Role of Oxidative Stress in Telomere Shortening in Cultured Fibroblasts from Normal Individuals and Patients with Ataxia-Telangiectasia. *Human Molecular Genetics*, Vol.12, No.3, (February 2003), pp. 227-232.

[120] Testorelli, C. (2003). Telomerase And Cancer. *Journal of Experimental and Clinical Cancer Research*, Vol.22, No.2, (June 2003), pp. 165-169.

[121] Toussaint, O.; Medrano, E. E. & von Zglinicki, T. (2000). Cellular and Molecular Mechanisms of Stress-Induced Premature Senescence (SIPS) of Human Diploid Fibroblasts and Melanocytes. *Experimental Gerontology*, Vol.35, No.8, (October 2000), pp. 927-945.

[122] Tran, P. T.; Erdeniz, N., Symington, L. S. & Liskay, R. M. (2004). EXO1 – A Multi-Tasking Eukaryotic Nuclease. *DNA Repair (Amsterdam)*, Vol.3, No.12, (December 2004), pp. 1549-1559.

[123] van Overbeek, M. & de Lange, T. (2006). Apollo, an Artemis-Related Nuclease, Interacts with TRF2 and Protects Human Telomeres in S Phase. *Current Biology*, Vol.16, No.13, (July 2006), pp. 1295-1302.

[124] von Zglinicki, T.; Saretzki, G., Döcke, W. & Lotze, C. (1995). Mild Hyperoxia Shortens Telomeres and Inhibits Proliferation of Fibroblasts: a Model for Senescence? *Experimental Cell Research*, Vol.220, No.1, (September 1995), pp. 186-193.

[125]  von Zglinicki, T. (2000). Role of Oxidative Stress in Telomere Length Regulation and Replicative Senescence. *Annals of the New York Academy of Sciences*, Vol.908, (June 2000), pp. 99-110.

[126]  von Zglinicki, T.; Pilger, R. & Sitte, N. (2000). Accumulation of Single-Strand Breaks is the Major Cause of Telomere Shortening in Human Fibroblasts. *Free Radical Biology and Medicine*, Vol.28, No.1, (January 2000), pp. 64-74.

[127]  von Zglinicki, T. (2002). Oxidative Stress Shortens Telomeres. *Trends in Biochemical Sciences*, Vol.27, No.7, (July 2002), pp. 339-344.

[128]  Watson, J. D. (1972). Origin of Concatemeric T7 DNA. *Nature New Biology*, Vol.239, No.94, (October 1972), pp. 197-201.

[129]  Wellinger, R. J.; Wolf, A. J. & Zakian, V. A. (1993). Saccharomyces Telomeres Acquire Single-Strand TG1-3 Tails Late in S Phase. *Cell*, Vol.72, No.1, (January 1993), pp. 51-60.

[130]  Williams, G. J.; Lees-Miller, S. P. & Tainer, J. A. (2010). Mre11-Rad50-Nbs1 Conformations and the Control of Sensing, Signaling, and Effector Responses at DNA Double-Strand Breaks. *DNA Repair (Amsterdam)*, Vol.9, No.12, (December 2010), pp. 1299-1306.

[131]  Wright, W. E.; Tesmer, V. M., Huffman, K. E., Levene, S. D. & Shay, J. W. (1997). Normal Human Chromosomes Have Long G-rich Telomeric Overhangs at One End. *Genes and Development*, Vol.11, No.21, (November 1997), pp. 2801-2809.

[132]  Wu, P.; van Overbeek, M., Rooney, S. & de Lange, T. (2010). Apollo Contributes to G Overhang Maintenance and Protects Leading-End Telomeres. *Molecular Cell*, Vol.39, No.4, (August 2010), pp. 606-617.

[133]  Ye, J. Z.; Donigian, J. R., van Overbeek, M., Loayza, D., Luo, Y., Krutchinsky, A. N., Chait, B. T. & de Lange T. (2004). TIN2 Binds TRF1 and TRF2 Simultaneously and Stabilizes the TRF2 Complex on Telomeres. *The Journal of Biological Chemistry*, Vol. 279, No.45, (November 2004), pp. 47264-47271.

[134]  Yuan, X.; Ishibashi, S., Hatakeyama, S., Saito, M., Nakayama, J., Nikaido, R., Haruyama, T., Watanabe, Y., Iwata, H., Iida, M., Sugimura, H., Yamada, N. & Ishikawa, F. (1999). Presence of Telomeric G-Strand Tails in the Telomerase Catalytic Subunit TERT Knockout Mice. *Genes to Cells*, Vol.4, No.10, (October 1999), pp. 563-572.

[135]  Zheng, L. & Shen, B. (2011). Okazaki Fragment Maturation: Nucleases Take Centre Stage. *Journal of Molecular Cell Biology*, Vol.3, No.1, (February 2011), pp. 23-30.

[136]  Zhu, X. D.; Küster, B., Mann, M., Petrini, J. H. & de Lange, T. (2000). Cell-Cycle-Regulated Association of RAD50/MRE11/NBS1 with TRF2 and Human Telomeres. *Nature Genetics*, Vol.25, No.3, (July 2000), pp. 347-352.

[137] Zou, L. & Elledge, S. J. (2003). Sensing DNA Damage through ATRIP Recognition of RPA-ssDNA Complexes. *Science*, Vol.300, No.5625, (June 2003), pp. 1542-1548.

[138] Zubko, M. K.; Guillard, S. & Lydall, D. (2004). Exo1 and Rad24 Differentially Regulate Generation of ssDNA at Telomeres of *Saccharomyces cerevisiae* cdc13-1 Mutants. *Genetics*, Vol.168, No.1, (September 2004), pp. 103-115

*Edited by David Stuart*

DNA replication is a fundamental part of the life cycle of all organisms. Not surprisingly many aspects of this process display profound conservation across organisms in all domains of life. The chapters in this volume outline and review the current state of knowledge on several key aspects of the DNA replication process. This is a critical process in both normal growth and development and in relation to a broad variety of pathological conditions including cancer. The reader will be provided with new insights into the initiation, regulation, and progression of DNA replication as well as a collection of thought provoking questions and summaries to direct future investigations.

IntechOpen