Taylor & Francis
Taylor & Francis Group

Critical Review

Historical View of DNA Replication Studies, with Special Reference to Japan

Akio Sugino[1] and Hiroyuki Araki[2]
[1]*Graduate School of Frontier Biosciences, Osaka University, Osaka, Japan*
[2]*National Institute of Genetics, Shizuoka, Japan*

Summary

Almost forty years after the key contributions to the field by Okazaki and coworkers that gave rise to the concept of leading and the lagging strand, we are still at the state of uncertainty about the proteins that replicate each strand. Perhaps, one main conclusion that should be drawn from the data currently available is that the protein architecture at the fork is more plastic than originally thought.

IUBMB *Life*, 58: 323–327, 2006

Keywords Okazaki fragments; leading strand synthesis; lagging strand synthesis; eukaryotic chromosomal DNA replication forks.

Contemporary molecular biology is concerned principally with understanding the mechanisms responsible for transmission and expression of the genetic information that ultimately governs cell structure and function. All cells share a number of basic properties, and this underlying unity of cell biology is particularly apparent at the molecular level. Such unity has allowed scientists to choose simple organisms (such as bacteria and viruses) as models for many fundamental experiments, with the expectation that similar molecular mechanisms are operative in organisms as diverse as *E. coli* and humans. Numerous experiments have established the validity of this assumption, and it is now clear that the molecular biology of cells provides a unifying theme to understanding diverse aspects of cell function and behavior.

Initial advances in molecular biology were made by taking advantage of the rapid growth and readily manipulable genetics of simple bacteria, such as *E. coli*, and their viruses. More recently, not only the fundamental principles but also many of the experimental approaches first developed in prokaryotes have been successfully applied to eukaryotic cells. Furthermore, the development of recombinant DNA has had a tremendous impact, allowing individual eukaryotic genes to be isolated and characterized in detail. Current advances in recombinant DNA technology have allowed the human genome to be sequenced.

The most fundamental property of all livings is ability to reproduce. All organisms inherit from their parents the genetic information specifying their structure and function. Likewise, all cells arise from pre-existing cells, so the genetic material must be replicated and passed from parent to progeny cell at each cell division. How genetic information is replicated and transmitted from cell and organism to organism represents a question that is central to all of biology. Consequently, elucidation of the mechanisms of genetic transmission and identification of the genetic material as DNA were discoveries that formed the foundation of our current understanding of biology at the molecular level.

The molecular biology golden age began with the first important evidence that DNA is the genetic material: the discovery, reported in 1944, that DNA prepared from one strain of *Pneumococcus* could transform another strain. The purified DNA carried genetic information that could be absorbed and expressed by cells of another strain. DNA was further recognized to be a molecule far larger and more complex than four different repeating nucleotides, varying in composition from organism to organism.

Two influential discoveries were then made. The first was the demonstration in 1952 by Hershey and Chase that infection of *E. coli* by T2 phages involved injection of the DNA of the virus into the host cell. A second, remarkable event was the discovery by Watson and Crick, in 1953, the complementary, double-stranded structure of DNA and with it the recognition of how the molecule can be replicated. Complementary pairing of the nucleotide constituents of one strand to those of the second strand was postulated to explain in a simple way how one DNA duplex is able to direct the

Received 22 March 2006; accepted 30 March 2006
Address correspondence to: Dr Akio Sugino, Laboratories for Biomolecular Networks, Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamada-oka, Suita, Osaka 565-0871, Japan. Tel: +81 6 6879 4661. Fax: +81 6 6879 4663.
E-mail: asugino@fbs.osaka-u.ac.jp

ISSN 1521-6543 print/ISSN 1521-6551 online © 2006 IUBMB
DOI: 10.1080/15216540600732070

assembly of two molecules identical to it. In this model, each strand of duplex serves as a template upon which the complementary strand is made. These discoveries and other important ones that followed led to the realization that DNA has two major and discrete functions. One is to carry the genetic information that brings about the specific phenotype of the cell. The other major function of DNA is its own replication. For duplicating the genotype of the cell, DNA serves as a template for converting one chromosome into two identical chromosomes. This process is called 'semi-conservative replication' because one strand of parental DNA is conserved in each progeny DNA molecule. Direct support for semi-conservative DNA replication was obtained in 1958 by Meselson and Stahl. An enzyme, named DNA polymerase, was discovered by Kornberg and his associates in 1956 to have the unprecedented property of taking instructions from this template and duplicating it by assembling activated nucleotides into long stretches of DNA. These discoveries happened in the era when Japanese economy had been completely destroyed due to the World War II, but led many Japanese scientists (we list here several of those; J. Tomizawa, Y. Hirota, H. Yoshikawa, N. Sueoka, K. Matsubara, M. Takanami, R. Okazaki, and T. Okazaki) to investigate DNA replication problems in bacteria and their viruses. Subsequently, they greatly contributed to our understanding of the mechanism of DNA replication in prokaryotes. Thus, they established their rich tradition of molecular biological studies on DNA replication in Japan and also inspired many young Japanese scientists, who inherited their DNA replication studies in both prokaryotes and eukaryotes.

The synthesis of new DNA strands complementary to both strands of the parental molecule posed an important problem to understand the biochemistry of DNA replication. Since the two strands of double-helical DNA run in opposite (anti-parallel) directions, continuous synthesis of two new strands at the replication forks would require that one strand be synthesized in the $5'-3'$ direction while the other is synthesized in the opposite ($3'-5'$) direction. But, DNA polymerase catalyzes the polymerization of dNTPs only in the $5'-3'$ direction. How, then, can the other progeny strand of DNA be synthesized?

This enigma or 'DNA replication dilemmas' was resolved by the discovery of so-called 'Okazaki fragments' in *E. coli,* by R. Okazaki and his associates in 1966 (*1*). In 1967, at the 7th International Congress of Biochemistry in Tokyo, R. Okazaki presented their experimental results and proposed his discontinuous DNA replication model (Fig. 1), showing that only one strand of DNA is synthesized in a continuous manner in the direction of overall DNA replication. The other is formed from small, discontinuous pieces of DNA that are synthesized backwards with respect to the direction of movement of the replication fork (*2*). These small pieces of newly synthesized DNA (called 'Okazaki fragments' after their discoverer) are joined by the action of DNA ligase, which was discovered in 1967 by Gellert (from *E. coli*), and by Lehman and Richardson (from T4 infected *E. coli*) forming an intact new DNA strand. The continuously synthesized strand is called the 'leading strand', since its elongation in the direction of replication fork movement exposes the template strand used for the synthesis of Okazaki fragments (the lagging strand).

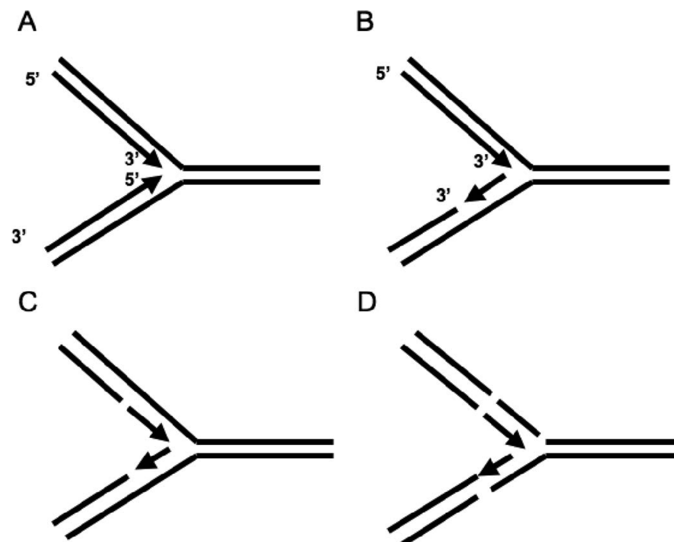

Figure 1. Original model for the possible structure and reaction in the replicating region of DNA (modified from Okazaki et al., 1968 (*2*)).

Although the discovery of discontinuous synthesis of the lagging strand provided a mechanism for the elongation of both strands of DNA at the replication fork, it soon raised another question. Since DNA polymerase requires a primer and cannot initiate synthesis *de novo*, how is the synthesis of Okazaki fragments initiated? The answer is that short RNA fragments serve as primers for DNA replication. In contrast to DNA synthesis, the synthesis of RNA can initiate *de novo*; an enzyme called primase (the product of DnaG in *E. coli*) synthesizes short fragments of RNA (e.g., 3 to 10 nucleotides long) complementary to the lagging strand template at the replication fork. Okazaki fragments are then synthesized via extension of these RNA primers by DNA polymerase. An important consequence of such RNA priming is that newly synthesized Okazaki fragments contain an RNA-DNA join, the discovery of which provided critical evidence for the role of RNA primers in DNA replication.

To form a continuous lagging strand of DNA, the RNA primer must be eventually removed from the Okazaki fragments and replaced with DNA. In *E. coli*, RNA primers are removed by the combined action of RNase H, an enzyme that degrades the RNA strand of RNA-DNA hybrids, and DNA polymerase I. This is the aspect of *E. coli* DNA replication in which DNA polymerase I plays a critical role. In addition to its DNA polymerase activity, DNA polymerase I acts as an exonuclease that can hydrolyze DNA (or RNA) in either the $3'-5'$ or $5'-3'$ direction. The action of DNA

polymerase I as a $5'-3'$ exonuclease removes ribonucleotides from the $5'$-ends of Okazaki fragments, allowing them to be replaced with deoxyribonucleotides to yield fragments consisting entirely of DNA. The different DNA polymerases thus play distinct roles at the replication fork. In prokaryotic cells, DNA polymerase III is the major replicative polymerase, although to date, three additional DNA polymerases (II, IV and V) have been found beside DNA polymerases I and III, functioning in the synthesis of both the leading strand of DNA and Okazaki fragments by the extension of RNA primers. DNA polymerase I then removes RNA primers and fills the gaps between Okazaki fragments (*3*).

Each eukaryotic chromosome contains many replication origins (multi replicons) and its replication is initiated at these origins spaced approximately $30-100$ kb apart. Each one of the large number of replication origins in the cell directs the assembly of two divergently migrating replication forks that faithfully replicate their portion of the chromosome. While these origins vary widely in strength, DNA sequence environment and chromatin structure, and in the time of each origin firing in the S phase, it is generally assumed that the identical elongation protein apparatus is formed at each of these origins. In eukaryotic cells, however, multiple DNA polymerases α, δ and ε (Pol α, δ, and ε) are required to do what in *E. coli* is accomplished by DNA polymerase III alone. Since the replication of SV40 DNA *in vitro*, which has been reconstituted by B. Stillman and his associates (*4*), requires only two Pol α and δ, Pol α, which is found in a complex with primase, appears to function in conjunction with primase to synthesize short RNA-DNA fragments during lagging strand synthesis. Pol δ can then synthesize both the leading and lagging strands, acting to extend the RNA-DNA primers initially synthesized by Pol α-primase complex. In addition, DNA polymerase δ can take the place of *E. coli* DNA polymerase I in filling the gaps between Okazaki fragments following primer removal. However, Pol ε is required for cellular chromosomal DNA replication, thus, instead of Pol δ, Pol ε can synthesize the leading strand of DNA. As a consequence, a wide variety of fork models have been proposed in reviews of the literatures.

The functions of Pol α-primase and Pol δ at the fork were initially established through biochemical studies of SV40 viral DNA replication as mentioned above. While using its large T antigen as both initiator and DNA helicase, the virus appropriates cellular enzymes for all other replication functions. Biochemical studies have revealed that the primase component of Pol α-primase synthesizes an approximately 10 nt long RNA which is then elongated by the polymerase subunit to yield a $\sim$30 nt primer (called as pre-Okazaki fragment), which is in turn elongated by Pol δ. The switch from Pol α to Pol δ is mediated through loading of the processivity clamp PCNA (proliferating cell nuclear antigen). On the leading strand, Pol δ continues elongation until all viral DNA is replicated while on the lagging strand a reiterative switch from Pol α to Pol δ ensures initiation (pre-Okazaki fragment) and elongation of pre-Okazaki fragments to form a mature Okazaki fragment, respectively. However, this viral DNA replication mechanism has provided us with one view of the cellular replication fork (namely the 'Two DNA Polymerases Model') (*4, 5*). This became apparent when in 1990 yeast Pol ε was identified as another essential DNA polymerase required for proper chromosomal DNA replication (*6*).

Of the large number of DNA polymerases that exist in the nucleus of a eukaryotic cell, only two enzymes, Pol δ and Pol ε, have the necessary high-fidelity required for accurate chromosome duplication, and are also capable of proofreading their errors through the $3'-5'$-exonuclease activity. Both Pol δ and Pol ε are multi-subunit enzymes. But unlike *E. coli* DNA polymerase III, these enzymes are monomeric with regard to their catalytic cores. Pol α-primase synthesizes the initial primer for leading strand DNA synthesis, and this enzyme also functions repeatedly at the lagging strand where it initiates Okazaki fragments. Pol α-primase lacks a proofreading exonuclease activity. Based on these results, it made sense to position both Pol δ and Pol ε at the replication fork, but how? Complicating and confusing the issue were studies showing that while the gene for Pol ε is essential, partial deletions that lack the polymerase domain retain viability in yeasts, *S. cerevisiae* and *S. pombe* (*7–9*). Does this mean that the SV40 model is correct after all? We do not think so. But we would also suggest that it is no longer is appropriate to look at the replication fork as a single fixed structure. Methodologies to place the DNA polymerases at specific strands by physical means remain yet to be developed. Therefore, studies addressing this problem have been genetic in nature so far. In particular, replication fidelity has been exploited as an experimental tool for polymerase function. This experimental approach is possible because proofreading-deficient forms of Pol δ (*pol3exo$^-$*) and of Pol ε (*pol2exo$^-$*) can be tolerated in the cell, instead of temperature-sensitive each polymerase mutants, such as *pol2-9*, *pol2-18*, or *cdc2*. As expected, such mutants show an increase in spontaneous mutation rates. The distribution of the mutations accumulating in a selectable target has been used as the genetic read-out in strand determination studies. In two independent studies (*10, 11*), it was concluded that the exonuclease activities of Pol δ and Pol ε proofread opposite strands of the replication fork. However, since proofreading by a DNA polymerase is expected to couple to its polymerization activity, this conclusion can reasonably be interpreted that Pol δ and Pol ε replicate opposite strands of the fork. Furthermore, replication forks assembled at an origin appear to retain their distinctive architecture during their entire progression. The same unique mutational signature of a target is observed regardless whether the target is very close to the origin, or moved as far as 40 kb away from the origin, provided the same direction of replication through this target is maintained (*12*). These

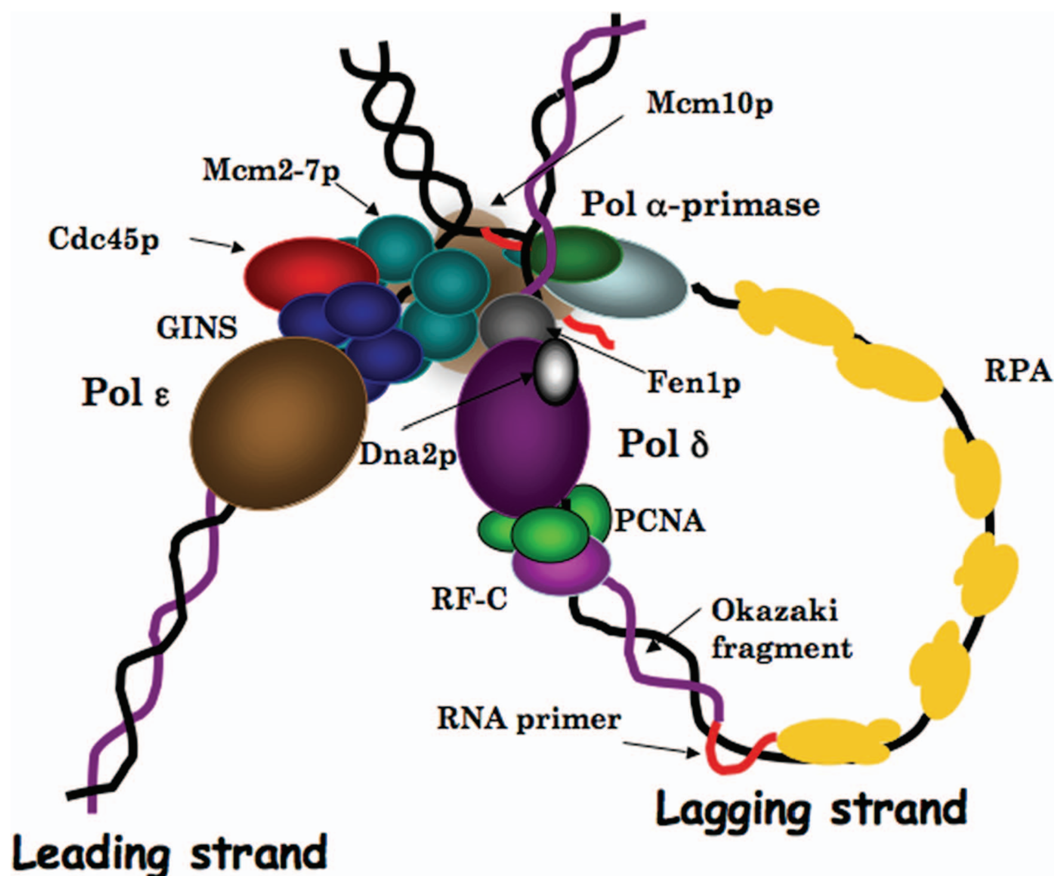

Figure 2. Current model for molecular dynamics at eukaryotic chromosomal DNA replication forks (*19*).

results indicate that forks started in early S phase in yeast still look the same when they arrive at their target in mid S phase.

The complex nature of Okazaki fragment synthesis on the lagging strand has provided two solid targets for such studies: Pol α-primase initiates Okazaki fragments and the flap endonuclease FEN1 degrades the initiator RNA during their maturation (*13*). In a very recent study (*14*), an epistatic relationship was established between Pol α and Pol δ. Some mutations in the polymerase domain of Pol α confer a mutator phenotype on yeast. When such a Pol α polymerase mutation was combined with exonuclease-deficient mutant of Pol δ, the resulting double mutant showed severe hypermutability. These data are consistent with the model that the exonuclease activity of Pol δ can proofread most errors introduced by the imperfect Pol α, but that these errors persist when proofreading is inactivated in the double mutant. Therefore, both Pol α and Pol δ synthesize the same strand, the lagging strand of the fork. On the other hand, no hypermutability was observed when the mutant Pol α was combined with exonuclease-deficient mutation of Pol ε, indicating that Pol ε cannot proofread errors made by Pol α. This is the expected result if Pol ε travels along the leading strand, instead of the lagging strand where Pol α-primase and Pol δ work during elongation of DNA synthesis.

During each eukaryotic cell cycle, many nicks are generated during the maturation of Okazaki fragments on the lagging strand. Therefore, maturation of Okazaki fragments needs to be carried out with extraordinary efficiency and fidelity. Any nicks and gaps, which are not ligated, result in the formation of double-stranded breaks during the next cell cycle, and the cell has only a limited capacity to repair double-stranded breaks. Therefore, a few double-stranded breaks lead to cell death. Biochemical studies have shown that efficient and faithful nick processing requires the coordinated action of Pol δ and the flap endonuclease FEN1. In this process, the 3′-exonuclease activity of Pol δ is important for maintaining a ligatable nick. Mutational studies show strong genetic interactions between mutations in the exonuclease domain of Pol δ and mutations in *RAD27*, the gene for FEN1. Therefore, Pol δ functions in the maturation of Okazaki fragments *in vivo*. Together with its known genetic interactions with Pol α, a coherent view is presented in which Pol δ is the enzyme responsible for both elongation and maturation of Okazaki fragments on the lagging strand (*13*). Based upon our previous assertion that Pol δ and Pol ε replicate opposite strands of the fork, the logical conclusion appears to be that Pol ε is likely the leading strand polymerase. But if this is the case, then how can mutants lacking the polymerase domain of Pol ε still be

viable? A first clue is that such mutants are far from healthy; they show severe phenotypic defects in the progression of DNA replication (*15*). A second clue comes from the observation that point mutations in the active site of the polymerase confer lethality (*7*). Thus, when Pol ε is actually incorporated into the replisome, a catalytically active polymerase domain is essential.

Additional evidence that Pol ε travels with the replication fork comes from chromatin immunoprecipitation studies in yeast. Furthermore, DNA replication in *Xenopus* extracts depleted for Pol δ or for Pol ε results in a marked decrease in DNA synthesis, as shown by us (*15, 16*). The products formed in the absence of Pol δ are most consistent with a defect in lagging strand DNA synthesis (*17*), suggesting that Pol ε may be the leading strand enzyme.

The essential function of the C-terminal domain of the catalytic subunit of Pol ε has not been fully addressed. This domain is required for S phase checkpoint regulation, which is another eukaryotic specific process to maintain stably its chromosome from generation to generation. However, since Pol ε is known to be an essential factor for replisome assembly, its C-terminal domain may fulfill that essential function. The truncated domain of Pol ε is present at early origin, although it is released from the fork after firing. The DNA synthetic function on the leading strand must be assumed by Pol δ because this crippled fork can still replicate DNA with reasonable fidelity. On the other hand, the realization that the cell can get along with Pol δ replicating both strands in the Pol ε partial deletion mutation makes one wonder whether this mutational abnormality actually highlights a specialized form of the replication fork that can form in a wild-type strain under certain circumstances? What these conditions are, and whether assembly would occur at origins or through remodeling of preexisting normal forks, remains to be determined.

It is remarkable that we were able to isolate a new type of DNA replication mutants using temperature-sensitive Pol ε and its subunit Dpb2 mutants, all of which are the initiation protein mutants, indicating that Pol ε and its associated proteins play an important role during chromosomal DNA replication (*18*). Those new initiation proteins including GINS give us a new look at the chromosomal DNA replication forks.

More remarkably, almost 40 years after the key contributions to the field by Okazaki and coworkers that gave rise to the concept of the leading and the lagging strand, we are still in a state of uncertainty about the proteins that replicate each strand. Perhaps one main conclusion that should be drawn from the data currently available is that the protein architecture at the fork is more capable of undergoing to continuous deformation than originally thought.

REFERENCES

1. Sakabe, K., and Okazaki, R. (1966) A unique property of the replicating region of chromosomal DNA. *Biochim. Biophys. Acta* **129**, 255–260.
2. Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., and Sugino, A. (1968) A mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc. Natl. Acad. Sci. USA* **59**, 598–605.
3. Johnson, A., and O'Donnell, M. (2005) Cellular DNA replicases: components and dynamics at the replication fork. *Annu. Rev. Biochem.* **74**, 283–315.
4. Waga, S., and Stillman, B. (1998) The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* **67**, 721–751.
5. Waga, S., and Stillman, B. (1994) Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication *in vitro*. *Nature* **369**, 207–212.
6. Morrison, A., Araki, H., Clark, A. B., Hamatake, R. K., and Sugino, A. (1990) A third essential DNA polymerase in *S. cerevisiae*. *Cell* **62**, 1143–1151.
7. Kesti, T., Flick, K., Keranen, S., Syvaoja, J. E., and Wittenberg, C. (1999) DNA polymerase epsilon catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. *Mol. Cell* **3**, 679–685.
8. Dua, R., Levy, D. L., and Campbell, J. L. (1999) Analysis of the essential functions of the C-terminal protein/protein interaction domain of *Saccharomyces cerevisiae* pol ε and its unexpected ability to support growth in the absence of the DNA polymerase domain. *J. Biol. Chem.* **274**, 22283–22288.
9. Feng, W., and D'Urso, G. (2001) *Schizosaccharomyces pombe* cells lacking the amino-terminal catalytic domains of DNA polymerase epsilon are viable but require the DNA damage checkpoint control. *Mol. Cell. Biol.* **21**, 4495–4504.
10. Shcherbakova, P. V., and Pavlov, Y. I. (1996) 3′–5′ exonucleases of DNA polymerases ε and δ correct base analog induced DNA replication errors on opposite DNA strands in *Saccharomyces cerevisiae*. *Genetics* **142**, 717–726.
11. Karthikeyan, R., Vonarx, E. J., Straffon, A. F., Simon, M., Faye, G., and Kunz, B. A. (2000) Evidence from mutational specificity studies that yeast DNA polymerases δ and ε replicate different DNA strands at an intracellular replication fork. *J. Mol. Biol.* **299**, 405–419.
12. Pavlov, Y. I., Newlon, C. S., and Kunkel, T. A. (2002) Yeast origins establish a strand bias for replicational mutagenesis. *Mol. Cell* **10**, 207–213.
13. Garg, P., and Burgers, P. (2005) DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit. Rev. Biochem. Mol. Biol.* **40**, 115–128.
14. Pavlov, Y. I., Frahm, C., McElhinny, S. A., Niimi, A., Suzuki, M., and Kunkel, T. A. (2006) Evidence that errors made by DNA polymerase α are corrected by DNA polymerase δ. *Curr. Biol.* **16**, 202–207.
15. Ohya, T., Kawasaki, Y., Hiraga, S., Kanbara, S., Nakajo, K., Nakashima, N., Suzuki, A., and Sugino, A. (2002) The DNA polymerase domain of pol ε is required for rapid, efficient, and highly accurate chromosomal DNA replication, telomere length maintenance, and normal cell senescence in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **277**, 28099–28108.
16. Waga, S., Masuda, T., Takisawa, H., and Sugino, A. (2001) DNA polymerase ε is required for coordinated and efficient chromosomal DNA replication in Xenopus egg extracts. *Proc. Natl. Acad. Sci. USA* **98**, 4978–4983.
17. Fukui, T., Yamauchi, K., Muroya, T., Akiyama, M., Maki, H., Sugino, A., and Waga, S. (2004) Distinct roles of DNA polymerases δ and ε at the replication fork in Xenopus egg extracts. *Genes Cell* **9**, 179–191.
18. Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A., and Araki, H. (2003) GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev.* **17**, 1153–1165.
19. Seki, T., Akita, M., Kamimura, Y., Muramatsu, S., Araki, H., and Sugino, A. (2006) GINS is a DNA polymerase epsilon accessory factor during chromosomal DNA replication in budding yeast. *J. Biol. Chem.* **281**, in press.