Evolution of Replication Machines

Nina Y. Yao[1] and **Mike E. O'Donnell**[1,2]

[1]DNA Replication Laboratory, The Rockefeller University, New York, NY, 10065, USA.

[2]Howard Hughes Medical Institute, The Rockefeller University, New York, NY, 10065, USA.

Abstract

The machines that decode and regulate genetic information require the translation, transcription and replication pathways essential to all living cells. Thus, it might be expected that all cells share the same basic machinery for these pathways that were inherited from the primordial ancestor cell from which they evolved. A clear example of this is found in the translation machinery that converts RNA sequence to protein. The translation process requires numerous structural and catalytic RNAs and proteins, the central factors of which are homologous in all three domains of life, bacteria, archaea and eukarya. Likewise, the central actor in transcription, RNA polymerase, shows homology among the catalytic subunits in bacteria, archaea and eukarya. In contrast, while some "gears" of the genome replication machinery are homologous in all domains of life, most components of the replication machine appear to be unrelated between bacteria and those of archaea and eukarya. This review will compare and contrast the central proteins of the "replisome" machines that duplicate DNA in bacteria, archaea and eukarya, with an eye to understanding the issues surrounding the evolution of the DNA replication apparatus.

Keywords

LUCA; evolution; DNA replication; sliding clamp; clamp loader; DNA helicase; DNA polymerase; primase; replisome

LUCA, the Last Universal Common Ancestor

The Last Universal Common Ancestor cell, referred to by the acronym "LUCA", is the cell from which the three domains of life evolved, bacteria, archaea and eukarya (Steel *et al.*, 2010; Theobald, 2010). This last universal common ancestor cell is thought to have evolved around 3.5 billion years ago (Doolittle, 2000; Glansdorff *et al.*, 2008). How did the first cell come about (i.e. either LUCA or its predecessors)? Cells that self-replicate couldn't have appeared all at once, in just one step. One prevailing theory is that pre-cellular life started in an aqueous environment, possibly in, or near deep thermal vents on the ocean floor where plenty of heat and organic chemicals abound and could freely float about. Complex organic polymers may have formed on inorganic catalytic surfaces and concentrated in nooks and

crannies within and around deep-sea vents. Polymers of many types probably formed, and just about as many probably fell apart. The polymer RNA has been proposed to be the basis for the first cellular life (Woese, 1968; Orgel 1968; Crick 1968), sometimes referred to as the "RNA world" hypothesis. This hypothesis gained momentum with the discovery that RNA can catalyze a reaction (Kruger *et al.*, 1982). The RNA world hypothesis posits that the RNA polymer had the properties required to catalyze a wide variety of reactions, more so than the inorganic catalysts in the rocks, and perhaps sufficient for life to occur (Copley *et al.*, 2007; Gilbert, 1986; Orgel, 2003; Robertson and Joyce, 2012). RNA can fold into very complex shapes, like proteins, and catalyze many types of chemical reactions. The RNA in modern cells is formed from only 4 types of nucleotide building blocks, and one might think this would limit the repertoire of chemistry that RNA can provide. But the repertoire was sufficient to form the very complicated and huge ribosome machinery (which is a ribozyme) that makes proteins (Steitz, 2008). The ribozyme, and all of the tRNAs that supply the individual amino acid building blocks of protein polymers, is present in all three domains of life, the bacteria, archaea and eukarya (Woese *et al.*, 1990). While RNA utilizes only 4 different monomeric nucleotide units (e.g. rC, rG, rA, rU), the current genetic code of triplet nucleotide units in modern-day cells enable use of 20 different amino acid units. The development of translation and the use of many amino acids in proteins resulted in the evolution of protein molecules that could catalyze many different types of reactions, and could bind ribozymes and enhance thier reaction rates. The first proteins probably simply helped RNA fold into tighter structures, by shielding the strong repelling forces of the phosphate backbone and (e.g. like the ribosomal proteins). Thus proteins would allow RNA to adopt more precise shapes and to develop into the precision ribozymes that we know of today (most ribozymes have a protein component). Eventually proteins would develop to become catalysts in their own right. For example, a ribozyme probably catalyzed the synthesis of other RNA polymers, but modern-day cells use a protein, RNA polymerase, to make RNA polymers. Eventually, the chemical efficiency afforded by 20 different amino acid side chains replaced most ribozymes.

Protein synthesis, the process of translation, is highly complex. It requires a large multi-protein/multi-rRNA ribosome, numerous tRNAs, 20 amino acids, a method to match each different tRNA with the exact amino acid that its anticodon decodes (i.e. catalyzed by aminoacyl tRNA synthetases), and a messenger RNA sequence that is "read" by the ribosome and uses the amino-acyl tRNAs to make protein. The process of translation is fascinating, but exactly how the translation machinery evolved is one of the mysteries of cellular evolution that remains a major question to this day. No protein machine has yet evolved to substitute for RNA in this complex process. In fact, despite the enormous genetic diversity in the viral world, no virus encodes its own translation system, yet many viruses have evolved to encode various methods of synthesizing RNA and DNA.

We don't know if the first cell used protein or RNA polymers for enzymatic catalysis, but we can be pretty certain that LUCA harbored the ribosome machinery, and thus synthesized protein molecules, because all cells today contain ribosomes composed of homologous rRNAs and use the same universal genetic code to make proteins. LUCA also probably contained DNA polymers because ribonucleotide reductase (RDR), which makes the dNMP building blocks of DNA, is homologous in all cells (Forterre *et al.*, 2004; Leipe *et al.*, 1999).

Another compelling argument for the existence of DNA in LUCA is the role of homologous recombination in evolution of cellular life (see next section), and the fact that all known recombinases (RecA/Rad52) use the DNA polymer as the substrate. However, whether LUCA used DNA as the genetic material in addition to using DNA as a recombination substrate, or used RNA for its genome instead of DNA, is less certain. There is substantial precedent in modern-day viruses for use of RNA as their genetic material. By way of example, the scheme in **Fig. 3** suggests one possible scheme (out of many) for information flow within LUCA in which RNA, DNA and protein exist, yet DNA is not used as the genome. In this scheme, the genome is RNA, and DNA is present for use as a substrate for homologous recombination required for exchange of genes among genomes (horizontal gene transfer), and thus sampling of combinations of genes that facilitated evolution of a self-replicating cell. In this hypothetical cell, RNA polymerase uses DNA to make RNA for protein synthesis. It also uses RNA polymerase to generate the RNA genome from the DNA molecule. The DNA molecule is made by a simple process, similar to the replication of modern-day retroviruses, which only require one enzyme, a reverse transcriptase. The reverse transcriptase initiates DNA synthesis using a tRNA for a primer, and when finished it makes the second strand of DNA while digesting away the original RNA.

Recombination promoted genetic diversity needed to evolve life

The exchange of a gene from one organism to another type of organism is referred to as horizontal gene transfer, and requires homologous recombination catalyzed by a recombinase. Before the advent of genome sequencing, horizontal gene transfer was thought to be a rare event. But sequencing of numerous cellular genomes indicates that horizontal gene transfer was a frequent process during evolution. A high rate of horizontal gene transfer could not last long, as the genomic instability would prevent stable species development. But DNA swapping by horizontal gene transfer appears to have been frequent early in evolution, and this was probably necessary to sample sufficient combinations of enzymes that catalyzed different types of reactions needed for evolution of a free living self-reproducing cell. Accordingly, the recombinases, bacterial RecA and eukaryotic Rad51, are homologous and their common ancestor enzyme is thought to have been present in LUCA.

A relatively new view is that viruses predated cellular life (Forterre *et al.*, 2004; Koonin *et al.*, 2015; Leipe *et al.*, 1999). In this scenario, viruses did not evolve from cells. It used to be thought that cells must have come first, and then viruses devolved from them by becoming a trimmed down selfish DNA held inside a protein capsid container. But this is inconsistent with the fact that genetic diversity in the viral world dwarfs the diversity of cellular genetic information. Thus the argument has been turned around to propose that the DNA sequences in cellular genomes derived from a sampling of information in viruses that predated cells. In this view, viruses were agents of information transfer between compartments, or nooks and crannies in deep-sea vents, that held pre-cellular assemblages. This viral mediated information transfer was performed by homologous recombination, taking chunks of nucleic acid from one compartment to another. In this view, viruses predated cellular life and were needed to explore the large variety of combinations needed to "pick the combination lock" of life. Even today, viruses are the predominant form of biological assemblage, being 10-100

fold more prevalent in number than all cells on earth combined (Forterre *et al.*, 2004; Koonin *et al.*, 2015; Leipe *et al.*, 1999).

Note that the recombinases act upon DNA, not RNA, and presumably acted on duplex DNA in LUCA. It can be reasoned that DNA only became a prevalent polymer after the emergence of proteins, as follows. The nucleotide precursors of RNA are made from "scratch", meaning that the metabolic pathways that piece together nucleotide bases and the ribose sugars synthesize the four rNMP precursors from numerous smaller molecules. The nucleotide precursors for DNA (dNMPs) are simply made in one step by removing the 2' OH from the ribose sugar of the rNMPs. The abstraction of the 2' hydroxyl moiety may not have been possible to achieve using the chemistry afforded by amino acid side chains, because all modern-day cells require a RDR enzyme that employs an iron radical to perform this reaction. Iron radicals destroy RNA and thus would have destroyed an RNA enzyme that catalyzes this process. This is one reason that proteins are thought to have predated the existence of the DNA polymer. DNA uses the base T instead of U, used in RNA. The likely reason for the evolution of T to replace U for DNA is because deamination of C is a frequent spontaneous process, due to hydrolysis, and U is the product of C deamination. Since U correctly base pairs with A, frequent deamination of C, followed by replication over the template U would have led to loss of GC base pairs over evolutionary time, replacing GC base pairs with AT base pairs. But the use of T in DNA enables a mechanism to detect and repair DNA containing U produced by deamination of C. All cells contain an enzyme called uracyl glycosylase, which detects U residues in DNA, and removes the base. Other enzymes then replace the U with C to restore the GC base pair.

Based on homology of genes and the universal genetic code, LUCA contained the ribosome, the tRNAs, tRNA syntheses, RDR, RNA polymerase, recombinase and a host of other molecules that made life possible. These molecules are homologous in every cell type, which is the hallmark definition of a molecule that was present in LUCA. One may think that by comparing many cellular genome sequences, one should be able to derive the minimal set of genes for a living cell, but it isn't that simple. The types and variety of the minimal genes needed for life is a very active area of investigation (Glass *et al.*, 2006; Mushegian and Koonin, 1996).

DNA polymerases and the chemistry of replication

The nucleotide units of DNA and RNA are identical except for the 2' OH on the ribose sugar of RNA. The loss of the 2' OH from the ribose sugar provides DNA with much greater stability than RNA, but DNA lacks the catalytic power of RNA. With the evolution of proteins, the catalytic properties of RNA were no longer needed, and eventually most ribozymes were replaced by more efficient protein counterparts. The synthesis of RNA and DNA is a comparatively simple process and only one enzyme is needed, RNA or DNA polymerase. The chemistry is the same, and is shown for DNA in **Fig. 4**. The 3' OH is activated by proton abstraction and the 3' oxy anion performs a nucleophilic attack at the 5' alpha phosphate of an incoming dNTP, with the beta-gamma phosphates leaving as a molecule of pyrophosphate (Kornberg and Baker, 1992). The energy of the reaction is not too far downhill, and the reverse reaction can occur provided a sufficient pyrophosphate

concentration. While use of dNDP substrates is just as energetically favorable as use of dNTPs, phosphate is plentiful in cells and would reverse the reaction. Therefore long chains of DNA and RNA would be difficult to synthesize using dNDP substrates. The use of dNTPs, and production of pyrophosphate during synthesis enables the cell to prevent the reverse reaction by splitting the pyrophosphate into two phosphate molecules (i.e. by pyrophosphatase) (Kornberg and Baker, 1992).

Crystal structures of RNA and DNA polymerases, along with biochemical experiments, reveal that the catalysis of nucleotide addition does not directly involve any amino acid side chains. Instead, the chemistry of activating the 3' OH of the primer strand, and stabilizing the pyrophosphate leaving group of the incoming dNTP, is mainly accomplished by two metal ions (i.e. Mg++) that are bound to the polymerase active site by three conserved acidic residues (Beese and Steitz, 1991; Steitz, 2006). Interestingly, RNA is fully capable of chelating Mg++, and thus the use of two metals to catalyze this reaction may be a reflection of the RNA world. However, the most important aspect of RNA and DNA synthesis is accuracy of synthesis, and this is possibly more efficiently achieved by an enzyme than by a ribozyme. Base pairing efficiency of a NTP substrate to the template strand will strongly bias the choice of inserting the correct nucleotide, and an RNA or protein catalyst would be no different in this regard. But the AT and GC base pairs only differ by one hydrogen bond and therefore base pairing alone cannot provide the high accuracy (fidelity) required of DNA synthesis in making exact copies of entire genomes for cell duplication. Instead of base pairing energy, it is the nearly identical geometry of the AT and CG base pairs that enable polymerases to be highly accurate. The nucleotide selection step of DNA polymerases involves a conformation change of the enzyme that completely buries the base pair prior to chemistry. Only if the base pair adopts the correct geometry does chemistry occur (e.g. **Fig. 5a**) (Doublie *et al.*, 1999; Johnson *et al.*, 2003; Joyce *et al.*, 2008; Li *et al.*, 1998; Luo *et al.*, 2007; Santoso *et al.*, 2010).

DNA polymerases assort into different "families" based on non homologous sequences (families: A, B, C, D, X, Y and RT (reverse transcriptase) (Steitz, 1999; Yang and Woodgate, 2007). Crystal structure analysis of representatives of each family show they all have the shape of a right hand, with subdomains referred to as palm, fingers and thumb (e.g. see **Fig. 5b**) (Johansson and Macneill, 2010; Steitz, 1999). Hence, the evolutionary relationship of polymerases is not completely clear, because even though they have non homologous sequences they adopt the same general shape. Each subdomain carries a specific function. The fingers domain binds the incoming dNTP, the thumb domain helps grip duplex DNA, and the palm domain contains the amino acids that bind the two catalytic metal ions. The catalytic palm domain is the most conserved of the domains, and the topological chain folding pattern of this domain is similar in the A, B, and Y families, indicating an ancestral relationship among these polymerases (**Fig. 5c**) (Yang and Woodgate, 2007). Likewise, the C and X family polymerases have similar palm folding patterns (**Fig. 5c**) (Bailey *et al.*, 2006; Lamers *et al.*, 2006; Wing *et al.*, 2008). Perhaps the palm domain evolved twice independently. It is interesting that all bacteria to date use the C family polymerase for chromosome replication, and this family is not represented in a eukaryotic cell. Eukarya use three B family polymerases for replication. Archaea also use a B family polymerase, or a D

family polymerase for replication (Kelman and Kelman, 2014), and genome sequence analysis of archaea indicate that this domain of life evolved along the same line as eukaryotes, separate from bacteria (Xie *et al.*, 2012; Yutin *et al.*, 2008). Hence, the use of C- and B-family polymerases to replicate bacterial and eukaryotic/archaeal genomes, respectively, suggests the replication machinery evolved twice independently after LUCA (Leipe *et al.*, 1999).

The fingers domain binds each of the 4 dNTPs. As described above, the highly accurate dNTP selection step of DNA polymerases involves a conformation change in which the fingers domain closes over the palm, burying the dNTP-template base pair in a confined chamber within which only a correct base pair can fit (e.g. **Fig. 5a**). Base pairs with the correct shape are joined, while those without the correct shape prevent the conformation change needed for catalysis, and dissociate from the polymerase, followed by binding of another dNTP to try again. The entire replication process is highly accurate, with only 1 mistake per 10^{-9} base pairs (Echols and Goodman, 1991; Kunkel, 2004; Modrich *et al.*, 1996). This high fidelity of synthesis involves several enzymes, but the replicative DNA polymerase active site alone is quite accurate, making only about one mistake every 10,000 base pairs. Cellular genomes are much larger than this, and therefore this level of accuracy would not be sufficient to avoid mutations. Replicative DNA polymerases have an associated 3'-5' exonuclease active site that proofreads the product of the DNA polymerase. Thus when a DNA polymerase makes a mistake and incorporates an incorrect dNTP, the polymerase stalls because it is very slow at extending a mismatch. This gives time for the 3' mismatch primer terminus to relocate from the palm domain to the 3'-5' exonuclease active site which readily removes mismatched 3' nucleotides. The exonuclease proofreader gives an extra 100-fold accuracy, increasing the fidelity of synthesis to only one mistake every 1 million base pairs. Despite this impressive accuracy, most genomes are longer than one million base pairs, so even this level of accuracy is not sufficient for accurate genome duplication. But cells have repair systems that act after replication and detect and "fix" mistakes made by the polymerase. For example, the post-replication mismatch repair system recognizes mismatches (mistakes) made during replication and repairs them, increasing fidelity about 1000-fold. This brings the accuracy of genome replication to one mistake every 1 billion base pairs, well above the size of many (but not all) genomes.

While bacteria use a C-family DNA polymerase for replication, eukaryotes utilize three different B family DNA polymerases (Pols) for chromosome replication, Pols alpha, delta and epsilon (Garg and Burgers , 2005; Johansson and Macneill , 2010; Stillman, 2008). Pol alpha was the first eukaryotic replicative DNA polymerase identified and, surprisingly, it was found to synthesize primers (Conaway and Lehman, 1982). For many years it was thought to be the only replicative DNA polymerase in eukaryotes. Pol alpha consists of 4 subunits, the catalytic Pol1 subunit is the largest, the Pol12 subunit is second largest (unknown function), and two smaller subunits perform the RNA priming function. The smallest of the two priming subunits contains the primase catalytic site, but the larger priming subunit is required to form the first dinucleotide bond. After about 7 nucleotides, the RNA primer switches to the DNA polymerase subunit that extends RNA primers another 20-25 residues to produce a hybrid RNA/DNA primer (Garg and Burgers, 2005). A second

replicative polymerase, Pol delta, was discovered in biochemical studies of SV40 DNA replication (Waga *et al.*, 1998). Human Pol delta contains 4 subunits, the catalytic Pol3 is the largest, then Pol31 and Pol32 subunits in budding yeast, and a fourth very small subunit (p12) in human (Garg and Burgers, 2005). Characterization of SV40 replication showed that Pol delta synthesizes both the leading and lagging strands, and that Pol alpha serves as the primase (Tsurimoto and Stillman, 1991b; Tsurimoto and Stillman, 1991a; Waga *et al.*, 1998). It came as quite a surprise when biochemical and genetic studies revealed a third DNA polymerase, Pol epsilon, that is essential to cellular replication (Ohya *et al.*, 2002; Syvaoja *et al.*, 1990). Pol epsilon consists of 4 subunits, the large catalytic Pol2 subunit, the second largest subunit is Dpb2, and two small histone fold subunits, Dpb3 and Dpb4. The sequence of the Pol2 subunit predicts that it consists of two B family DNA polymerase sequences linked end to end (Tahirov *et al.*, 2009). The N-terminal half of Pol2 is the active polymerase and the C-terminal half of Pol2 encodes an inactive polymerase. Consistent with a role in DNA synthesis, point mutants in the active site of the catalytic N-terminal polymerase of Pol2 are not viable (Dua *et al.*, 1999). However, cells are viable upon deletion of the entire N-terminal half of Pol2, suggesting that a back-up polymerase rescues cells when the catalytic region of Pol epsilon is absent (Dua *et al.*, 1999; Kesti *et al.*, 1999). Surprisingly, the C-terminal inactive polymerase region of Pol2 is essential to cell viability, and is currently believed to serve an essential structural role (Dua *et al.*, 1999; Kesti *et al.*, 1999).

How did LUCA replicate its DNA?

Bacteria use distinctive enzymes to replicate their DNA compared to eukaryotes, and thus the replication process of modern-day cells may not have been utilized by LUCA. One of the differences between bacteria and eukaryotes, mentioned above, is that the DNA polymerases are widely diverged; bacteria use C family DNA polymerases while eukaryotes use polymerases of the B family. This has been taken as evidence that DNA replication evolved twice, independently for bacteria and eukarya (Leipe *et al.*, 1999). As explained above, DNA polymerases all have right hand shapes and use a 2 metal ion mechanism of catalysis. Thus it may be arguable as to whether DNA polymerases of different families are truly a case of independent evolution. But even if there are distinctly evolved polymerases, recent studies have shown that many different types of DNA polymerases can trade places with the main chromosomal polymerase during replication, and these different DNA polymerases are functional with the sliding clamp and helicase of the replicating machinery (Geertsema and van Oijen, 2013; Indiani *et al.*, 2005; Johansson and Dixon, 2013; Stukenberg *et al.*, 1991; Yang *et al.*, 2004). This plasticity is probably important to the replication process because replication, unlike transcription and translation, must succeed at all costs in the task of duplicating an entire genome. Replication only happens once in a cell's life, and the cell cannot afford failure. Transcription and translation are processes that can just start over whenever they fail. Consider replication of an entire genome. DNA polymerases differ in their capability to traverse various types of DNA sequences or DNA lesions. Thus, plasticity in the replicating machinery, through use of exchangeable polymerases, enables the replication machinery to advance over innumerable types of sequences and sites of DNA damage. Given this perspective, the use of distinct B and C family DNA polymerases in

bacteria versus eukaryotes may not signify distinct evolutionary lineage of the replicating machinery. For example, maybe bacteria started out using a B family polymerase, but switched to a C family polymerase.

However, one can make more supportive arguments for the hypothesis of distinct evolution of replication machinery in bacteria versus eukaryotes by looking at the other enzymes required for replication, and asking whether LUCA even needed them. The "trick", or most difficult aspect of cellular replication, is simultaneous replication of both strands of duplex DNA. LUCA may have replicated one strand at a time instead of simultaneous replication of both strands. Simultaneous synthesis of antiparallel strands requires that the lagging strand is made in the opposite direction of fork progression because DNA polymerases can only synthesize DNA in the 5'-3' direction (i.e. the nucleotide building blocks are only activated at the 5' position). Therefore replication of the lagging strand requires it to be synthesized in sections (i.e. Okazaki fragments). This required evolution of a robust priming activity, as well as enzymes to remove primers and ligate the sections of DNA together. However, if the two strands of dsDNA are not replicated simultaneously, then one strand can be synthesized continuously 5'-3', and after it is completed the second strand can also be extended continuously, 5'-3'. Indeed, many viruses and phage replicate this way, and often with a strand displacing polymerase instead of requiring a helicase. Bacterial and eukaryotic cells use very distinct, non homologous helicase and primases (as will be discussed further in a later section). The non homologous primases, helicases, and DNA polymerases, combined, make a compelling argument that replication machinery evolved twice, independently, and that LUCA may not have performed simultaneous duplication of both strands of DNA.

Examples abound of relatively simple mechanisms of replication in eukaryotic viruses and bacteriophage that do not require a helicase or primase. For example, some viruses use a linear dsDNA genome and a protein-nucleotide conjugate to create the initial primer site, after which the DNA polymerase catalyzes synthesis by strand displacement to the end of the genome, circumventing the need for primase and helicase (Kornberg and Baker, 1992). Other viral genomes have covalently closed ends, and the polymerase simply generates a duplicate genome from a single nick and then extension around the entire genome without interruption. Again no primase or helicase are required. In another example, mentioned earlier, retroviruses use a reverse transcriptase to convert the ssRNA genome to a dsDNA genome, by initiating synthesis with a tRNA (i.e. not primase), and using RNase to remove the RNA thus circumventing need for a helicase. New viral RNA genomes are generated from the dsDNA by the host RNA polymerase.

Evolution of the DNA sliding clamp and clamp loader

The first protein that was discovered to encircle DNA was the *E. coli* beta subunit sliding clamp (Kong *et al.*, 1992; Kuriyan and O'Donnell, 1993; Stukenberg *et al.*, 1991). Since then, many different types of proteins have been found to encircle DNA for their function. The beta sliding clamp is shown in **Fig. 6a**; it is a dimer of a subunit that consists of three domains. Each of the three domains has the same folding pattern, giving the dimer a pseudo six-fold symmetry. Turned on its side, the clamp is relatively thin, about 1 turn of DNA, and has two distinctive faces (the C-face is labeled in **Fig. 6a**). The beta clamp is assembled onto

DNA by a 5-subunit clamp loader that uses ATP to open and close the clamp around DNA (Kuriyan and O'Donnell, 1993). After being loaded onto DNA, the beta clamp can diffuse along dsDNA. The clamp binds to Pol III, the chromosomal replicase of *E. coli*, and once bound to beta Pol III can synthesize DNA for thousands of nucleotides without coming off DNA at a rate of 500-1000 nucleotides each second (Kuriyan and O'Donnell, 1993). The main role of the clamp in chromosome replication is to impart high processivity onto the DNA polymerase. It does so by binding directly to the polymerase and sliding along behind it, acting as a mobile tether that holds polymerase to DNA during numerous rounds of dNTP incorporation.

In eukarya and archaea the sliding clamp is the proliferating cell nuclear antigen (PCNA). The structure of PCNA is essentially superimposable with that of beta (**Fig. 6b**) (Gulbis *et al.*, 1996; Krishna *et al.*, 1994). An example of an archaeal PCNA is shown in **Fig. 6c** (Matsumiya *et al.*, 2001). The major difference between PCNA and beta is that each PCNA subunit is composed of only 2 domains, instead of three, and PCNA trimerizes to form a 6-domain ring like beta. Numerous studies demonstrate that nearly all the different DNA polymerases of a cell interact with the sliding clamp, both in bacteria and eukarya/archaea (Yao and O'Donnell, 2015). The striking similarity in the structures of the bacterial and eukarya/archaea sliding clamps reveals that they are related by a common ancestor and thus LUCA contained the sliding clamp. Clamps require multi-subunit ATPase driven clamp loaders that open and close the clamp around a primed site. Both the bacterial and eukaryotic clamp loading machines are circular pentamers and the subunits have both sequence homology (Bunz *et al.*, 1993; O'Donnell *et al.*, 1993), and structural similarity (Bowman *et al.*, 2004; Jeruzalmi *et al.*, 2001; Kelch *et al.*, 2011). Thus LUCA contained the clamp and clamp loading system that was handed down during evolution to all three domains of life.

Considering that the sliding clamp provides processivity to DNA polymerases, it is natural to assume that in LUCA the clamp provided processivity to the polymerase that replicated the genome. But this does not need to be the case, nor the principle function for with the clamp evolved. For one, DNA polymerases can be processive without a clamp, as will be discussed later. But more importantly, the sliding clamp has many functions beyond its role in processive DNA replication. Take for example the PCNA clamp. PCNA binds a host of different enzymes, far too many to list here (see (Georgescu *et al.*, 2015; Maga *et al.*, 2003)). Therefore, the reason that clamps evolved in the first place could be for another purpose, and not replication. Many PCNA clamp binding proteins are involved in DNA repair. For example, PCNA is required for mismatch repair in eukaryotes, in which it dictates strand specificity to ensure that the misincorporated nucleotide of the new strand is excised, and not the correct parental nucleotide (Manosas *et al.*, 2012; Pluciennik *et al.*, 2010). This review will revisit the role of sliding clamps later (in the last section), in which we propose a new perspective on the modern-day function of the sliding clamp in cell physiology.

Bacterial and eukaryotic replication machines evolved independently

The structure of DNA is elegant, but the antiparallel double helix poses significant barriers to duplication by enzymes. Long chains of duplex DNA, must be unzipped and the helical

turns require enormous amounts of unwinding during replication, a fact pointed out by Watson and Crick in their classic model of DNA and its implications for cellular replication (Watson and Crick, 1953). We now know that the unwinding problem is solved by topoisomerases, and that unzipping of DNA is performed by ATP-driven helicases. As described earlier, the antiparallel strands require synthesis in opposite directions, yet the nucleotide precursors are only activated on the 5' end (the 5' triphosphate). Therefore, one strand is synthesized in segments extended in the opposite direction of unwinding, and each segment must be initiated *de novo*, by condensation of two nucleotides. In addition, all cells contain primase, a specialized RNA polymerase that makes a short RNA primer (Kornberg and Baker, 1992). Why don't DNA polymerases make the primer, and why is the primer made of RNA that requires other enzymes to repair and replace the RNA with DNA? One possible reason a RNA primase is required is that rNTPs are present in 10-100 fold higher concentration in cells than dNTPs, and a high concentration may be required for an enzyme to bind two rNTPs at the same time. There is another reason that RNA may be used to prime DNA synthesis. The initial condensations of a nucleotide chain are inaccurate, and in fact primases make a mistake about 1 % of the time (Sheaff and Kuchta, 1994; Zhang and Grosse, 1990). Hence, the use of RNA is a convenient "marker" of low fidelity nucleic acid synthesis, enabling it to be recognized and replaced with DNA. The RNA primer is distinguished from DNA by repair enzymes that excise the RNA and replace it with DNA using a high fidelity DNA polymerase, after which DNA ligase seals Okazaki fragments together.

Given the complexity required to coordinate the diverse enzymatic reactions required for simultaneous replication of both strands of duplex DNA, one may question whether this process was operative in LUCA. Previous sections have dealt with the DNA polymerase, and how bacteria and eukarya/archaea have distinctive replicative DNA polymerases, suggesting independent evolution of the replication machinery in these two branches of cellular life. In this section we compare and contrast the primase and helicase of bacteria and eukarya/archaea. It is these two components of bacterial and eukaryotic replication enzymes that differ the most.

Primase

Cells that simultaneously replicate both strands of duplex DNA must continually reinitiate synthesis on the antiparallel strand, the "lagging strand" that is synthesized in the opposite direction of fork movement. For this function, cells from all three domains of life have evolved a RNA primase activity (Kornberg and Baker, 1992). Primases generally use rNTPs to make a short RNA of a dozen nucleotides or less. The bacterial and eukaryotic primases appear the most divergent of all replication factors. The bacterial primase is a single-subunit protein and the active site displays homology to topoisomerases, sometimes referred to as a "toprim" fold, illustrated in **Fig. 5a** (Aravind *et al.*, 1998; Podobnik *et al.*, 2000). In bacteria the RNA primers are approximately 10-12 nucleotides. In contrast, the eukaryotic primase is the Pol alpha, and the RNA priming activity is contained within a heterodimer of the two smallest subunits. The eukaryotic priming subunits share no sequence or structural similarity to the toprim fold of bacterial primase, and instead share homology to the Pol X family (**Fig. 7b**) (Kilkenny *et al.*, 2013; Kirk and Kuchta, 1999). Priming activity requires both subunits;

the larger of the two is required to form the first dinucleotide bond, while the smaller subunit contains the active site for chain extension (Copeland *et al.*, 1993; Zerbe and Kuchta, 2002). The priming subunits make a RNA primer of about 7 rNMPs, and then the DNA polymerase subunit extends the RNA primer with about 20-25 nucleotides of DNA to make a hybrid RNA/DNA primer (Kilkenny *et al.*, 2013; Singh *et al.*, 1986). The archaeal primase lacks the polymerase and B subunits of Pol alpha and is a heterodimer like the two small subunits of Pol alpha (Lao-Sirieix *et al.*, 2005). Interestingly, archaeal primase can synthesize either DNA or RNA primers *in vitro* (Bocquier *et al.*, 2001; Liu *et al.*, 2001), but it binds rNTPs tighter than dNTPs (Lao-Sirieix and Bell, 2004). In archaeal cells the 5' ends of Okazaki fragments are in fact tipped with RNA (Matsunaga *et al.*, 2003). In summary, the bacterial and eukaryotic/archaeal primases are distinct in both sequence and structure, making a compelling argument that they share no common ancestor and must have evolved independently of one another.

Helicase

Helicases harness the power of nucleotide hydrolysis to translocate along a single-strand (ss) DNA, and are required for DNA replication in all cell types. Replicative helicases in all cells are hexamers that appear to encircle one strand and motor along it, while excluding the other strand from the internal channel (Ahnert and Patel, 1997; Hacker and Johnson, 1997; Lee *et al.*, 2014). This is referred to as the "steric exclusion" mechanism of DNA unwinding. While the common quaternary structure and steric exclusion mechanism might suggest that replicative helicases of all cells share a common ancestor, they share no sequence homology or subunit structure. The motor domains of the bacterial helicase (DnaB in *E. coli*) utilize a similar chain fold as the RecA recombinase (Forterre *et al.*, 2004; LeBowitz and McMacken, 1986; Leipe *et al.*, 1999). In contrast, the replicative helicase motors of the eukaryotic Mcm2-7 heterohexamer are based in the AAA+ fold (ATPases associated with a variety of cellular activities) (Chong *et al.*, 2000; Forterre *et al.*, 2004; Leipe *et al.*, 1999; Li *et al.*, 2015; Tye, 1999). The archaeal replicative Mcm homohexameric helicase is also based on the AAA+ fold (Chia *et al.*, 2010; Chong *et al.*, 2000). Furthermore, the bacterial helicase translocates 5'-3' on ssDNA (LeBowitz and McMacken, 1986), while archaeal Mcm and eukaryotic CMG translocate in the 3'-5' direction (Bochman and Schwacha, 2008; Bochman and Schwacha, 2009; Ilves *et al.*, 2012; Moyer *et al.*, 2006). Thus the bacterial and eukaryotic helicases evolved independently, and not from a common ancestor. Furthermore, the eukaryotic helicase consists of eleven distinct subunits needed for activity; six different AAA+ motor subunits comprise the Mcm2-7 heterohexamer (Li *et al.*, 2015), and 5 accessory factors bind the side of the Mcm2-7 ring (Ilves *et al.*, 2012; Moyer *et al.*, 2006). These accessory factors are Cdc45 protein, and the GINS heterotetramer; the 4 GINS subunits share homology to one another. The 11-subunit helicase complex is referred to as CMG (Cdc45, Mcm2-7, GINS). Unlike bacterial and archaeal helicase, 3D EM reconstruction reveals that eukaryotic CMG contains two channels; the Mcm2-7 subunits form the channel that encircles and translocates on ssDNA during unwinding, and the accessory factors form a second channel. The function, if any, of the second channel is not known (**Fig. 7b**) (Costa *et al.*, 2011). There are no bacterial homologues to the Cdc45 and

GINS subunits. The archaea contain homologues to the Cdc45/GINS accessory factors, but it is unknown whether they form a CMG complex (Makarova *et al.*, 2012).

Each subunit within both bacterial and eukaryotic helicase motor rings have large C-terminal and N-terminal domains, and this bilobed structure gives the helicase the appearance of two rings stacked on top of one another (**Fig. 7a**). The N-terminal domains may form one ring and the C-terminal motor domains form another. The eukaryotic Mcm2-7 and archaeal Mcm hexamer both encircle the leading strand with the motor domains pointing toward the forked junction (Costa *et al.*, 2014; Rothenberg *et al.*, 2007). Structure analysis of bacterial and eukaryotic viral replicative helicases propose that ATP hydrolysis drives rotary movements of DNA binding elements in the central channel for DNA translocation (**Fig. 7c**) (Enemark and Joshua-Tor, 2006; Itsathitphaisarn *et al.*, 2012).

Different architectures of bacterial and eukaryotic replisome machines

Replication proteins work together as a replisome machine that replicate both strands of DNA at the same time (Kornberg and Baker, 1992; Benkovic *et al.*, 2001). The current view of the bacterial replisome, from studies of *E. coli*, is illustrated in **Fig. 6a** (McHenry, 2011; Yao and O'Donnell, 2015). The clamp loader organizes the replication enzymes at the fork. Three of the five clamp loading subunits are identical (called tau) and contain a 24 kda C-terminal extension that bind three molecules of Pol III core for every clamp loader. The C-region of tau subunits also connect to one molecule of the hexameric helicase that encircles the lagging strand. Studies of *E. coli* and bacteriophage T4 and T7, show that a polymerase-helicase contact is required for leading strand synthesis, and the energy of dNTP incorporation gives an extra push to the helicase (Benkovic *et al.*, 2001; Donmez and Patel, 2008; Kim *et al.*, 1996; Kulczyk *et al.*, 2012). While it would seem a waste to use 3 Pol molecules at the *E. coli* replication fork, when there are only two DNA strands to replicate, the lagging strand is made as multiple Okazaki fragments and thus two polymerases could be used on this strand. Indeed, experimental evidence indicates that two of the three Pol III's function on the lagging strand (Georgescu *et al.*, 2011; Reyes-Lamothe *et al.*, 2010). RNA primers of about 12 nucleotides are synthesized by primase, which acts stochastically coming in and out of the replisome (i.e. is not an integral member of the replisome) (Wu *et al.*, 1992). As each Okazaki fragment is extended, a DNA loop forms on the lagging strand to accommodate the opposite direction of extension of the lagging strand relative to the leading strand. Okazaki fragments in bacteria are 1-2 kb, and upon finishing a section of DNA Pol III rapidly ejects from the beta clamp, leaving it behind on DNA (O'Donnell, 1987; Studwell *et al.*, 1989; Stukenberg *et al.*, 1994). The beta clamp is then targeted by Pol I and ligase (Lopez de Saro and O'Donnell, 2001), which remove the RNA primer, fill in the resulting gap with DNA and seal the Okazaki fragments together (Kornberg and Baker, 1992).

The formation of the lagging strand DNA loops, one for each Okazaki fragment, is due to the attachment of the lagging strand polymerase(s) to the helicase. The DNA loops serve no known function and my simply be a consequence of identical leading and lagging strand DNA polymerases in bacterial and the T4 and T7 bacteriophage systems (Hamdan *et al.*, 2009; Nossal *et al.*, 2007; Park *et al.*, 1998). The leading strand Pol-helicase contact is

important to activity, and the same contact, between the lagging strand Pol-helicase, result in DNA loops. DNA loops take little energy to form, and in the absence of a negative effect, evolution would not select to prevent their formation. The ssDNA generated in the DNA loop is tightly wrapped by the single-strand DNA binding protein, SSB (Chastain *et al.*, 2003; Georgescu *et al.*, 2014b). SSB is ubiquitous in all cell types and not only protects ssDNA against nucleases but also removes secondary structure blocks to the lagging strand polymerase.

The eukaryotic replisome architecture is not yet clear, although some details have emerged. Several studies indicate that Pol epsilon is the leading strand polymerase while Pol delta replicates the lagging strand (Clausen *et al.*, 2015; Kunkel and Burgers, 2008; Miyabe *et al.*, 2011; McElhinny *et al.*, 2008; Pursell *et al.*, 2007). Biochemical studies of replication forks driven by CMG also support these polymerase assignments (Georgescu *et al.*, 2014a; Georgescu *et al.*, 2015; Langston *et al.*, 2014). However this assignment is still in contention (Johnson *et al.*, 2015) and further studies will be required to sort this out. CMG forms a central organizing unit of the eukaryotic replisome, and antibody pull-outs of CMG from cell extracts identify a large complex of numerous proteins referred to as the RPC (Replication Promoting Complex (Gambus *et al.*, 2006; Gambus *et al.*, 2009). Among these proteins are CMG, Ctf4, Pol alpha, Pol epsilon, Mcm10, Tof1, Csm3, Mrc1 and FACT. While the function of CMG helicase, and the DNA polymerase/primases are understood, the exact function of most of the other proteins of the RPC are largely unknown. Nor is it known whether the lagging strand Pol delta connects stably to the replisome. Genetic studies indicate that Mcm10 is essential to replication, and it binds Pol alpha, although very little is known about its function (Warren *et al.*, 2009). Ctf4 mutants display a chromosome segregation phenotype, and Ctf4 is known to bind both CMG and Pol alpha, linking them together (Miles and Formosa, 1992; Simon *et al.*, 2014). Tof1, Csm3, Mrc1 are involved in a replication checkpoint pathway, although how they function at the fork remains largely unknown. Recent biochemical studies have reconstituted leading and lagging strand replication *in vitro* in the absence of Ctf4, Mcm10, Tof1, Mrc1 and Csm3, and therefore these particular proteins do not appear to be required for the central actions of leading and lagging strand synthesis (Georgescu *et al.*, 2014a; Georgescu *et al.*, 2015).

The current view of the architecture of the eukaryotic replisome is shown in **Fig. 8b**. The CMG encircles the leading strand as discussed above (Bochman and Schwacha, 2008; Bochman and Schwacha, 2009; Ilves *et al.*, 2012, Moyer *et al.*, 2006). Studies with pure proteins have shown a direct connection between CMG and Pol epsilon, consistent with Pol epsilon acting on the leading strand (Langston *et al.*, 2014). Indeed, *in vitro*, Pol epsilon is much more active with CMG on the leading strand compared to Pol delta, while Pol delta is more active on the lagging strand with CMG than Pol epsilon (Georgescu *et al.*, 2014a). Unlike the bacterial replisome, where the clamp loader organizes the replisome, there is no evidence that the eukaryotic RFC clamp loader (Replication Factor C) travels with the replisome (e.g. RFC is not pulled-down with the RPC). CMG links the lagging strand Pol alpha to the replisome through the Ctf4 trimer, and Pol alpha-primase is a component of the RPC, indicating that it travels with the replisome, unlike the primase of *E. coli* (Gambus *et al.*, 2009; Simon *et al.*, 2014). The lagging strand Pol delta is not reported to bind CMG, nor

is it part of the RPC, and therefore lagging strand looping may not occur. Okazaki fragments in eukaryotic cells are only 100-200 nucleotides in length, much shorter than in bacteria. The PCNA clamp helps Pol delta to perform strand displacement synthesis of the RNA portion of Okazaki fragments, and PCNA recruits the endonuclease (Fen1) that excises the ssRNA, and recruits the ligase that seals Okazaki fragments (Burgers, 2009). Each of these proteins bind PCNA, and thus a PCNA trimer might bind all three proteins simultaneously (Burgers, 2009). The lagging ssDNA is coated by RPA (Replication Protein A), a ssDNA binding protein that serves an analogous function to bacterial SSB. Both RPA and SSB bind DNA using OB folds, four within the SSB tetramer (one per subunit) and four within the RPA heterotrimer.

Replicative polymerases evolved an "addiction" to DNA sliding clamps

Among the core replication proteins (Table I), only the clamp and clamp loader are clearly conserved in the three domains of life, and thus these components were likely present in LUCA before the divergence of bacteria, archaea and eukaryotes. An obvious possible use of the sliding clamp in LUCA was to provide processivity for the replicative polymerase. However, a much wider role for the sliding clamp may be proposed (as in Kelman and Hurwitz, 1988), based on the fact that the clamp interacts with numerous factors in modern day cells. The reasoning is as follows: any protein that functions with DNA and interacts with a clamp that encircles DNA will be tethered to DNA in a mobile fashion, and can scan along the DNA surface to locate its proper site of action. Scanning along a linear DNA would, in effect, be equivalent to an increase in concentration of the protein in the cell. In other words, in the absence of a clamp, the protein would need to be present in much higher amounts to locate its site of action in 3D space at the same rate compared to linear 1D scanning along DNA. Thus the evolution of a clamp loader and a promiscuous clamp that binds different DNA metabolic proteins would, with one protein (and the clamp loader), increase the effective concentration of a host of proteins.

The sliding clamp was discovered in the field of DNA replication, and it provides a dramatic increase in processivity of the replicative DNA polymerase of *E. coli* (Stukenberg *et al.*, 1991). Thus it is natural to assume that the clamp is required for a processive polymerase. However, we now know of several viral and phage DNA polymerases that do not require a clamp and clamp loader for their dramatically high processivity. In fact, to these authors' knowledge the only examples of phages/viruses that encode their own clamp and clamp loader are provided by the T-even family bacteriophages. There are viral/phage polymerases that are highly processive by themselves, or have a single accessory factor that aids processivity, but is not a ring. An excellent example of this is the phi29 bacteriophage DNA polymerase, which is highly processive as a single protein (Salas, 2009). The phi29 DNA polymerase can undergo rapid and processive strand displacement synthesis with no help of an accessory factor, and it works so well that industry has employed it for whole genome amplification (Dean *et al.*, 2002). The herpes simplex virus 1, vaccinia poxvirus, and T7 phage polymerases are also highly processive (they bind one accessory factor but do not use a ring or clamp loader) (Kornberg and Baker, 1992). As another example, RNA polymerase has long been known to be highly processive (Kornberg and Baker, 1992). Hence, sliding clamps are not inherently required for a polymerase to be highly processive in synthesis. In

this line of reasoning, we have proposed that the replicative polymerases of modern-day cells primarily use the clamp for another purpose, specifically to mark newly replicated DNA to distinguish it from old parental DNA (Georgescu *et al.*, 2015). For example, it is of utmost importance that mismatch repair distinguish the parental strand having the correct nucleotide from the newly replicated strand that contains the incorrect mismatch. The front and back surfaces of sliding clamps are structurally distinct (see **Fig. 6**), and they are loaded onto a 3' terminus with a distinctive polarity. When the clamp is loaded onto newly replicated DNA for replisome function, the clamp provides directional information to mismatch repair proteins that must distinguish the new strand from the old strand in the daughter duplexes. Although *E. coli* uses DNA methylation to direct mismatch repair to the new strand, most bacteria do not have this methylation system, and they use the clamp to direct mismatch repair to the new strand, as do eukaryotic cells (Kunkel and Erie, 2005; Lenhart *et al.*, 2015; Modrich, 2006). Another important function of clamps that mark newly replicated DNA is the assembly of nucleosomes. The Caf1 factor that assembles nucleosomes on daughter strands requires PCNA for function (Sharp *et al.*, 2001; Shibahara and Stillman, 1999).

One way to ensure that clamps are placed on newly replicated DNA, and thus mark newly replicated DNA, is to evolve DNA polymerases that simply can not function without a clamp. In this view, the "clamp dependent" DNA polymerase must also periodically leave the clamp, let it diffuse away along DNA, and then associate with a new clamp (after the clamp loader assembled a new clamp on the primed site), thereby populating the product DNA with clamps. This is an observed characteristic of both *E. coli* Pol III (Stukenberg et. al., 1994) and eukaryotic Pol delta (Langston and O'Donnell., 2008). Thus we propose that replicative DNA polymerases evolved to become "addicted" to sliding clamps as a mechanism that ensures that newly replicated DNA is marked by sliding clamps.

Summary and perspective

Viruses may have predated cells (Forterre *et al.*, 2004; Koonin *et al.*, 2015; Leipe *et al.*, 1999), and thereby provided a large pool of genetic diversity that homologous recombination could plumb for genes and ultimately splice them together into a genome that encodes sufficient chemistry to enable a growing, self-reproducing cell. The translation and transcription pathways of all cells are performed by homologous machineries, indicating that LUCA (the last universal common ancestor cell) had settled on a solution to these processes and handed this solution down to all modern-day cells. Evolution of the replication process appears far different, as the core enzymes of replication are not homologous in bacteria and archaea/ eukarya. The helicase and primase components of the replication machinery may not have been in LUCA, as they are only required when both strands of duplex DNA are replicated simultaneously. Instead, LUCA may have had a simple replication system similar to numerous modern-day viruses and phage that replicate one strand at a time, and require no primase or helicase. Interestingly, the sliding clamps and clamp loader are homologous in all cells and thus were present in LUCA. It is possible that sliding clamps had other functions in LUCA beyond replication, and that replicative polymerases of modern-day cells use sliding clamps as a method to mark newly replicated DNA for DNA repair and packaging.

Acknowledgment

REFERENCES

Ahnert P, Patel SS. Asymmetric interactions of hexameric bacteriophage T7 DNA helicase with the 5′- and 3′-tails of the forked DNA substrate. J Biol Chem. 1997; 272:32267–73. [PubMed: 9405431]

Aravind L, Leipe DD, Koonin EV. Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Nucleic Acids Res. 1998; 26:4205–13. [PubMed: 9722641]

Bailey S, Wing RA, Steitz TA. The structure of T. aquaticus DNA polymerase III is distinct from eukaryotic replicative DNA polymerases. Cell. 2006; 126:893–904. [PubMed: 16959569]

Beese LS, Steitz TA. Structural basis for the 3′-5′ exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. EMBO J. 1991; 10:25–33. [PubMed: 1989886]

Benkovic SJ, Valentine AM, Salinas F. Replisome-mediated DNA replication. Annu Rev Biochem. 2001; 70:181–208. [PubMed: 11395406]

Bochman ML, Schwacha A. The Mcm2-7 complex has in vitro helicase activity. Mol Cell. 2008; 31:287–93. [PubMed: 18657510]

Bochman ML, Schwacha A. The Mcm complex: unwinding the mechanism of a replicative helicase. Microbiol Mol Biol Rev. 2009; 73:652–83. [PubMed: 19946136]

Bocquier AA, Liu L, Cann IK, et al. Archaeal primase: bridging the gap between RNA and DNA polymerases. Curr Biol. 2001; 11:452–6. [PubMed: 11301257]

Bowman GD, O'Donnell M, Kuriyan J. Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. Nature. 2004; 429:724–30. [PubMed: 15201901]

Bunz F, Kobayashi R, Stillman B. cDNAs encoding the large subunit of human replication factor C. Proc Natl Acad Sci U S A. 1993; 90:11014–8. [PubMed: 8248204]

Burgers PM. Polymerase dynamics at the eukaryotic DNA replication fork. J Biol Chem. 2009; 284:4041–5. [PubMed: 18835809]

Chastain PD 2nd, Makhov AM, Nossal NG, et al. Architecture of the replication complex and DNA loops at the fork generated by the bacteriophage t4 proteins. J Biol Chem. 2003; 278:21276–85. [PubMed: 12649286]

Chia N, Cann I, Olsen GJ. Evolution of DNA replication protein complexes in eukaryotes and Archaea. PLoS One. 2010; 5:e10866. [PubMed: 20532250]

Chong JP, Hayashi MK, Simon MN, et al. A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. Proc Natl Acad Sci U S A. 2000; 97:1530–5. [PubMed: 10677495]

Clausen AR, Lujan SA, Burkholder AB, et al. Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. Nat Struct Mol Biol. 2015; 22:185–91. [PubMed: 25622295]

Conaway RC, Lehman IR. A DNA primase activity associated with DNA polymerase alpha from Drosophila melanogaster embryos. Proc Natl Acad Sci U S A. 1982; 79:2523–7. [PubMed: 6806812]

Copeland WC, Wang TS. Enzymatic characterization of the individual mammalian primase subunits reveals a biphasic mechanism for initiation of DNA replication. J Biol Chem. 1993; 268:26179–89. [PubMed: 8253737]

Copley SD, Smith E, Morowitz HJ. The origin of the RNA world: co-evolution of genes and metabolism. Bioorg Chem. 2007; 35:430–43. [PubMed: 17897696]

Costa A, Ilves I, Tamberg N, et al. The structural basis for MCM2-7 helicase activation by GINS and Cdc45. Nat Struct Mol Biol. 2011; 18:471–7. [PubMed: 21378962]

Costa A, Renault L, Swuec P, et al. DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. Elife. 2014:e03273. [PubMed: 25117490]

Crick FHC. The origin of the genetic code. J. Mol. Biol. 1968; 38:367–379. [PubMed: 4887876]

Dean FB, Hosono S, Fang L, et al. Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A. 2002; 99:5261–6. [PubMed: 11959976]

Donmez I, Patel SS. Coupling of DNA unwinding to nucleotide hydrolysis in a ring-shaped helicase. EMBO J. 2008; 27:1718–26. [PubMed: 18497749]

Doolittle WF. Uprooting the tree of life. Sci Am. 2000; 282:90–5. [PubMed: 10710791]

Doublie S, Sawaya MR, Ellenberger T. An open and closed case for all polymerases. Structure. 1999; 7:R31–5. [PubMed: 10368292]

Dua R, Levy DL, Campbell JL. 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. 1999; 274:22283–8. [PubMed: 10428796]

Echols H, Goodman MF. Fidelity mechanisms in DNA replication. Annu Rev Biochem. 1991; 60:477–511. [PubMed: 1883202]

Enemark EJ, Joshua-Tor L. Mechanism of DNA translocation in a replicative hexameric helicase. Nature. 2006; 442:270–5. [PubMed: 16855583]

Forterre, P.; Filee, J.; Myllykallio, H. Origin and Evolution of DNA and DNA Replication Machineries. Landes Bioscience; Oct 4. 2004 2004

Gambus A, Jones RC, Sanchez-Diaz A, et al. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat Cell Biol. 2006; 8:358–66. [PubMed: 16531994]

Gambus A, van Deursen F, Polychronopoulos D, et al. A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome. EMBO J. 2009; 28:2992–3004. [PubMed: 19661920]

Garg P, Burgers PM. DNA polymerases that propagate the eukaryotic DNA replication fork. Crit Rev Biochem Mol Biol. 2005; 40:115–28. [PubMed: 15814431]

Geertsema HJ, van Oijen AM. A single-molecule view of DNA replication: the dynamic nature of multi-protein complexes revealed. Curr Opin Struct Biol. 2013; 23:788–93. [PubMed: 23890728]

Georgescu R, Langston L, O'Donnell M. A proposal: Evolution of PCNA's role as a marker of newly replicated DNA. DNA Repair (Amst). 2015; 29:4–15. [PubMed: 25704660]

Georgescu RE, Kurth I, O'Donnell ME. Single-molecule studies reveal the function of a third polymerase in the replisome. Nat Struct Mol Biol. 2011; 19:113–6. [PubMed: 22157955]

Georgescu RE, Langston LD, Yao NY, et al. Mechanism of Asymmetric Polymerase Assembly at the Eukaryotic Replication Fork. Nat Struct Mol Biol. 2014a

Georgescu RE, Schauer GD, Yao NY, et al. Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. Elife. 2015; 4:e04988. [PubMed: 25871847]

Georgescu RE, Yao N, Indiani C, et al. Replisome mechanics: lagging strand events that influence speed and processivity. Nucleic Acids Res. 2014b; 42:6497–510. [PubMed: 24829446]

Gilbert W. The RNA world. Naure. 1986; 319:618.

Glansdorff N, Xu Y, Labedan B. The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. Biol Direct. 2008; 3:29. [PubMed: 18613974]

Glass JI, Assad-Garcia N, Alperovich N, et al. Essential genes of a minimal bacterium. Proc Natl Acad Sci U S A. 2006; 103:425–30. [PubMed: 16407165]

Gulbis JM, Kelman Z, Hurwitz J, et al. Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. Cell. 1996; 87:297–306. [PubMed: 8861913]

Hacker KJ, Johnson KA. A hexameric helicase encircles one DNA strand and excludes the other during DNA unwinding. Biochemistry. 1997; 36:14080–7. [PubMed: 9369480]

Hamdan SM, Loparo JJ, Takahashi M, et al. Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. Nature. 2009; 457:336–9. [PubMed: 19029884]

Ilves I, Tamberg N, Botchan MR. Checkpoint kinase 2 (Chk2) inhibits the activity of the Cdc45/ MCM2-7/GINS (CMG) replicative helicase complex. Proc Natl Acad Sci U S A. 2012; 109:13163–70. [PubMed: 22853956]

Indiani C, McInerney P, Georgescu R, et al. A sliding-clamp toolbelt binds high- and low-fidelity DNA polymerases simultaneously. Mol Cell. 2005; 19:805–15. [PubMed: 16168375]

Itsathitphaisarn O, Wing RA, Eliason WK, et al. The hexameric helicase DnaB adopts a nonplanar conformation during translocation. Cell. 2012; 151:267–77. [PubMed: 23022319]

Jeruzalmi D, O'Donnell M, Kuriyan J. Crystal structure of the processivity clamp loader gamma (gamma) complex of *E. coli* DNA polymerase III. Cell. 2001; 106:429–41. [PubMed: 11525729]

Johansson E, Dixon N. Replicative DNA polymerases. Cold Spring Harb Perspect Biol. 2013:5.

Johansson E, Macneill SA. The eukaryotic replicative DNA polymerases take shape. Trends Biochem Sci. 2010; 35:339–47. [PubMed: 20163964]

Johnson RE, Klassen R, Prakash L, et al. A Major Role of DNA Polymerase delta in Replication of Both the Leading and Lagging DNA Strands. Mol Cell. 2015; 59:163–75. [PubMed: 26145172]

Johnson SJ, Taylor JS, Beese LS. Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. Proc Natl Acad Sci U S A. 2003; 100:3895–900. [PubMed: 12649320]

Joyce CM, Potapova O, Delucia AM, et al. Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. Biochemistry. 2008; 47:6103–16. [PubMed: 18473481]

Kelch BA, Makino DL, O'Donnell M, et al. How a DNA polymerase clamp loader opens a sliding clamp. Science. 2011; 334:1675–80. [PubMed: 22194570]

Kelman Z, Hurwitz J. Protein-PCNA interactions: a DNA scanning mechanism? TIBS. 1998; 23:236–8. [PubMed: 9697409]

Kelman LM, Kelman Z. Archael DNA replication. Annu Rev Genet. 2014; 48:71–97. [PubMed: 25421597]

Kesti T, Flick K, Keranen S, et al. DNA polymerase epsilon catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. Mol Cell. 1999; 3:679–85. [PubMed: 10360184]

Kilkenny ML, Longo MA, Perera RL, et al. Structures of human primase reveal design of nucleotide elongation site and mode of Pol alpha tethering. Proc Natl Acad Sci U S A. 2013; 110:15961–6. [PubMed: 24043831]

Kim S, Dallmann HG, McHenry CS, et al. Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement. Cell. 1996; 84:643–50. [PubMed: 8598050]

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; 38:7727–36. [PubMed: 10387012]

Kong XP, Onrust R, O'Donnell M, et al. Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. Cell. 1992; 69:425–37. [PubMed: 1349852]

Koonin EV, Dolja VV, Krupovic M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology. 2015; 479-480:2–25. [PubMed: 25771806]

Kornberg, A.; Baker, TA. DNA Replication. W.H. Freeman; New York: 1992. p. 106p. 275p. 227p. 471p. 553p. 689

Krishna TS, Kong XP, Gary S, et al. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. Cell. 1994; 79:1233–43. [PubMed: 8001157]

Kruger K, Grabowski PJ, Zaug AJ, et al. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell. 1982; 31:147–57. [PubMed: 6297745]

Kulczyk AW, Akabayov B, Lee SJ, et al. An interaction between DNA polymerase and helicase is essential for the high processivity of the bacteriophage T7 replisome. J Biol Chem. 2012; 287:39050–60. [PubMed: 22977246]

Kunkel TA. DNA replication fidelity. J Biol Chem. 2004; 279:16895–8. [PubMed: 14988392]

Kunkel TA, Burgers PM. Dividing the workload at a eukaryotic replication fork. Trends Cell Biol. 2008; 18:521–7. [PubMed: 18824354]

Kunkel TA, Erie DA. DNA mismatch repair. Annu Rev Biochem. 2005; 74:681–710. [PubMed: 15952900]

Kuriyan J, O'Donnell M. Sliding clamps of DNA polymerases. J Mol Biol. 1993; 234:915–25. [PubMed: 7903401]

Lamers MH, Georgescu RE, Lee SG, et al. Crystal structure of the catalytic alpha subunit of E. coli replicative DNA polymerase III. Cell. 2006; 126:881–92. [PubMed: 16959568]

Langston LD, O'Donnell. DNA polymerase delta is highly processive with proliferating cell nuclear antigen and undergoes collision release upon completing DNA. J Biol Chem. 2008; 283:29522–31. [PubMed: 18635534]

Langston LD, Zhang D, Yurieva O, et al. CMG helicase and DNA polymerase epsilon form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. Proc Natl Acad Sci U S A. 2014; 111:15390–5. [PubMed: 25313033]

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; 344:1251–63. [PubMed: 15561142]

Lao-Sirieix SH, Pellegrini L, Bell SD. The promiscuous primase. Trends Genet. 2005; 21:568–72. [PubMed: 16095750]

LeBowitz JH, McMacken R. The Escherichia coli dnaB replication protein is a DNA helicase. J Biol Chem. 1986; 261:4738–48. [PubMed: 3007474]

Lee SJ, Syed S, Enemark EJ, et al. Dynamic look at DNA unwinding by a replicative helicase. Proc Natl Acad Sci U S A. 2014; 111:E827–35. [PubMed: 24550505]

Leipe DD, Aravind L, Koonin EV. Did DNA replication evolve twice independently? Nucleic Acids Res. 1999; 27:3389–401. [PubMed: 10446225]

Lenhart JS, Pillon MC, Guarne A, et al. Mismatch repair in Gram-positive bacteria. Res Microbiol. 2015

Li N, Zhai Y, Zhang Y, et al. Structure of the eukaryotic MCM complex at 3.8 A. Nature. 2015; 524:186–91. [PubMed: 26222030]

Li Y, Korolev S, Waksman G. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. EMBO J. 1998; 17:7514–25. [PubMed: 9857206]

Liu L, Komori K, Ishino S, et al. The archaeal DNA primase: biochemical characterization of the p41-p46 complex from Pyrococcus furiosus. J Biol Chem. 2001; 276:45484–90. [PubMed: 11584001]

Lopez de Saro FJ, O'Donnell M. Interaction of the beta sliding clamp with MutS, ligase, and DNA polymerase I. Proc Natl Acad Sci U S A. 2001; 98:8376–80. [PubMed: 11459978]

Luo G, Wang M, Konigsberg WH, et al. Single-molecule and ensemble fluorescence assays for a functionally important conformational change in T7 DNA polymerase. Proc Natl Acad Sci U S A. 2007; 104:12610–5. [PubMed: 17640918]

Maga G, Hubscher U. Proliferating cell nuclear antigen (PCNA): a dancer with many partners. J Cell Sci. 2003; 116:3051–60. [PubMed: 12829735]

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. [PubMed: 22329974]

Manosas M, Spiering MM, Ding F, et al. Mechanism of strand displacement synthesis by DNA replicative polymerases. Nucleic Acids Res. 2012; 40:6174–86. [PubMed: 22434889]

Matsumiya S, Ishino Y, Morikawa K. Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from Pyrococcus furiosus. Protein Sci. 2001; 10:17–23. [PubMed: 11266590]

Matsunaga F, Norais C, Forterre P, et al. Identification of short 'eukaryotic' Okazaki fragments synthesized from a bacterial replication origin. EMBO Rep. 2003; 4:154–8. [PubMed: 12612604]

McHenry CS. Bacterial replicases and related polymerases. Curr Opin Chem Biol. 2011; 15:587–94. [PubMed: 21855395]

Miles J, Formosa T. Protein affinity chromatography with purified yeast DNA polymerase alpha detects proteins that bind to DNA polymerase. Proc Natl Acad Sci U S A. 1992; 89:1276–80. [PubMed: 1741381]

Miyabe I, Kunkel TA, Carr AM. The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved. PLoS Genet. 2011; 7:e1002407. [PubMed: 22144917]

Modrich P. Mechanisms in eukaryotic mismatch repair. J Biol Chem. 2006; 281:30305–9. [PubMed: 16905530]

Modrich P, Lahue R. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu Rev Biochem. 1996; 65:101–33. [PubMed: 8811176]

Moyer SE, Lewis PW, Botchan MR. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. Proc Natl Acad Sci U S A. 2006; 103:10236–41. [PubMed: 16798881]

Mushegian AR, Koonin EV. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci U S A. 1996; 93:10268–73. [PubMed: 8816789]

Nick McElhinny SA, Gordenin DA, Stith CM, et al. Division of labor at the eukaryotic replication fork. Mol Cell. 2008; 30:137–44. [PubMed: 18439893]

Nossal NG, Makhov AM, Chastain PD 2nd, et al. Architecture of the bacteriophage T4 replication complex revealed with nanoscale biopointers. J Biol Chem. 2007; 282:1098–108. [PubMed: 17105722]

O'Donnell, M.; Langston, LD.; Stillman, B. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. Cold Spring Harbor Laboratory Press; New York: 2013.

O'Donnell M, Onrust R, Dean FB, et al. Homology in accessory proteins of replicative polymerases--*E. coli* to humans. Nucleic Acids Res. 1993; 21:1–3. [PubMed: 8441605]

O'Donnell ME. Accessory proteins bind a primed template and mediate rapid cycling of DNA polymerase III holoenzyme from Escherichia coli. J Biol Chem. 1987; 262:16558–65. [PubMed: 3316222]

Ohya T, Kawasaki Y, Hiraga S, et al. The DNA polymerase domain of pol(epsilon) is required for rapid, efficient, and highly accurate chromosomal DNA replication, telomere length maintenance, and normal cell senescence in Saccharomyces cerevisiae. J Biol Chem. 2002; 277:28099–108. [PubMed: 12015307]

Orgel LE. Evolution of the genetic apparatus. J. Mol. Biol. 1968; 38:381–393. [PubMed: 5718557]

Orgel LE. Some consequences of the RNA world hypothesis. Orig Life Evol Biosph. 2003; 33:211–8. [PubMed: 12967268]

Park K, Debyser Z, Tabor S, et al. Formation of a DNA loop at the replication fork generated by bacteriophage T7 replication proteins. J Biol Chem. 1998; 273:5260–70. [PubMed: 9478983]

Pluciennik A, Dzantiev L, Iyer RR, et al. PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair. Proc Natl Acad Sci U S A. 2010; 107:16066–71. [PubMed: 20713735]

Podobnik M, McInerney P, O'Donnell M, et al. A TOPRIM domain in the crystal structure of the catalytic core of Escherichia coli primase confirms a structural link to DNA topoisomerases. J Mol Biol. 2000; 300:353–62. [PubMed: 10873470]

Pursell ZF, Isoz I, Lundstrom EB, et al. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science. 2007; 317:127–30. [PubMed: 17615360]

Reyes-Lamothe R, Sherratt DJ, Leake MC. Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. Science. 2010; 328:498–501. [PubMed: 20413500]

Robertson MP, Joyce GF. The origins of the RNA world. Cold Spring Harb Perspect Biol. 2012:4.

Rothenberg E, Trakselis MA, Bell SD, et al. MCM forked substrate specificity involves dynamic interaction with the 5′-tail. J Biol Chem. 2007; 282:34229–34. [PubMed: 17884823]

Salas M. My life with bacteriophage phi29. J Biol Chem. 2009; 287:44568–79. [PubMed: 23124207]

Santoso Y, Joyce CM, Potapova O, et al. Conformational transitions in DNA polymerase I revealed by single-molecule FRET. Proc Natl Acad Sci U S A. 2010; 107:715–20. [PubMed: 20080740]

Sharp JA, Fouts ET, Krawitz DC, et al. Yeast histone deposition protein Asf1p requires Hir proteins and PCNA for heterochromatic silencing. Curr Biol. 2001; 11:463–73. [PubMed: 11412995]

Sheaff RJ, Kuchta RD. Misincorporation of nucleotides by calf thymus DNA primase and elongation of primers containing multiple noncognate nucleotides by DNA polymerase alpha. J Biol Chem. 1994; 269:19225–31. [PubMed: 8034683]

Shibahara K, Stillman B. Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. Cell. 1999; 96:575–85. [PubMed: 10052459]

Simon AC, Zhou JC, Perera RL, et al. A Ctf4 trimer couples the CMG helicase to DNA polymerase alpha in the eukaryotic replisome. Nature. 2014

Singh H, Brooke RG, Pausch MH, et al. Yeast DNA primase and DNA polymerase activities. An analysis of RNA priming and its coupling to DNA synthesis. J Biol Chem. 1986; 261:8564–9. [PubMed: 2424899]

Steel M, Penny D. Origins of life: Common ancestry put to the test. Nature. 2010; 465:168–9. [PubMed: 20463725]

Steitz TA. DNA polymerases: structural diversity and common mechanisms. J Biol Chem. 1999; 274:17395–8. [PubMed: 10364165]

Steitz TA. Visualizing polynucleotide polymerase machines at work. EMBO J. 2006; 25:3458–68. [PubMed: 16900098]

Steitz TA. A structural understanding of the dynamic ribosome machine. Nat Rev Mol Cell Biol. 2008; 9:242–53. [PubMed: 18292779]

Stillman B. DNA polymerases at the replication fork in eukaryotes. Mol Cell. 2008; 30:259–60. [PubMed: 18471969]

Studwell, PS.; Stukenberg, PT.; Onrust, R., et al. Replication of the lagging strand by DNA polymerase III holoenzyme.. In: Richardson, Cc; Lehman, Ir, editors. Molecular Mechanisms in DNA Replication and Recombination. John Wiley and Sons, Inc; New York: 1989.

Stukenberg PT, Studwell-Vaughan PS, O'Donnell M. Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. J Biol Chem. 1991; 266:11328–34. [PubMed: 2040637]

Stukenberg PT, Turner J, O'Donnell M. An explanation for lagging strand replication: polymerase hopping among DNA sliding clamps. Cell. 1994; 78:877–87. [PubMed: 8087854]

Syvaoja J, Suomensaari S, Nishida C, et al. DNA polymerases alpha, delta, and epsilon: three distinct enzymes from HeLa cells. Proc Natl Acad Sci U S A. 1990; 87:6664–8. [PubMed: 1975694]

Tahirov TH, Makarova KS, Rogozin IB, et al. Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. Biol Direct. 2009; 4:11. [PubMed: 19296856]

Theobald DL. A formal test of the theory of universal common ancestry. Nature. 2010; 465:219–22. [PubMed: 20463738]

Tsurimoto T, Stillman B. Replication factors required for SV40 DNA replication *in vitro*. I. DNA structure-specific recognition of a primer-template junction by eukaryotic DNA polymerases and their accessory proteins. J Biol Chem. 1991a; 266:1950–60. [PubMed: 1671045]

Tsurimoto T, Stillman B. Replication factors required for SV40 DNA replication *in vitro*. II. Switching of DNA polymerase alpha and delta during initiation of leading and lagging strand synthesis. J Biol Chem. 1991b; 266:1961–8. [PubMed: 1671046]

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

Waga S, Stillman B. The DNA replication fork in eukaryotic cells. Annu Rev Biochem. 1998; 67:721–51. [PubMed: 9759502]

Warren EM, Huang H, Fanning E, et al. Physical interactions between Mcm10, DNA, and DNA polymerase alpha. J Biol Chem. 2009; 284:24662–72. [PubMed: 19608746]

Watson JD, Crick FH. Genetical implications of the structure of deoxyribonucleic acid. Nature. 1953; 171:964–7. [PubMed: 13063483]

Wing RA, Bailey S, Steitz TA. Insights into the replisome from the structure of a ternary complex of the DNA polymerase III alpha-subunit. J Mol Biol. 2008; 382:859–69. [PubMed: 18691598]

Woese, CR. Genetic Code. New Edition edit. Joanna Cotler Books; 1968.

Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990; 87:4576–9. [PubMed: 2112744]

Wu CA, Zechner EL, Marians KJ. Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. I. Multiple effectors act to modulate Okazaki fragment size. J Biol Chem. 1992; 267:4030–44. [PubMed: 1740451]

Xie Q, Wang Y, Lin J, et al. Potential key bases of ribosomal RNA to kingdom-specific spectra of antibiotic susceptibility and the possible archaeal origin of eukaryotes. PLoS One. 2012; 7:e29468. [PubMed: 22247777]

Yang J, Zhuang Z, Roccasecca RM, et al. The dynamic processivity of the T4 DNA polymerase during replication. Proc Natl Acad Sci U S A. 2004; 101:8289–94. [PubMed: 15148377]

Yang W, Woodgate R. What a difference a decade makes: insights into translesion DNA synthesis. Proc Natl Acad Sci U S A. 2007; 104:15591–8. [PubMed: 17898175]

Yao, NY.; O'Donnell, ME. Comparison of bacterial and eukaryotic replisome components.. In: Bradshaw, Ra; Stahl, Pd, editors. Encyclopedia of Cell Biology. Academic Press; 2015.

Yutin N, Makarova KS, Mekhedov SL, et al. The deep archaeal roots of eukaryotes. Mol Biol Evol. 2008; 25:1619–30. [PubMed: 18463089]

Zerbe LK, Kuchta RD. The p58 subunit of human DNA primase is important for primer initiation, elongation, and counting. Biochemistry. 2002; 41:4891–900. [PubMed: 11939784]

Zhang SS, Grosse F. Accuracy of DNA primase. J Mol Biol. 1990; 216:475–9. [PubMed: 2258921]
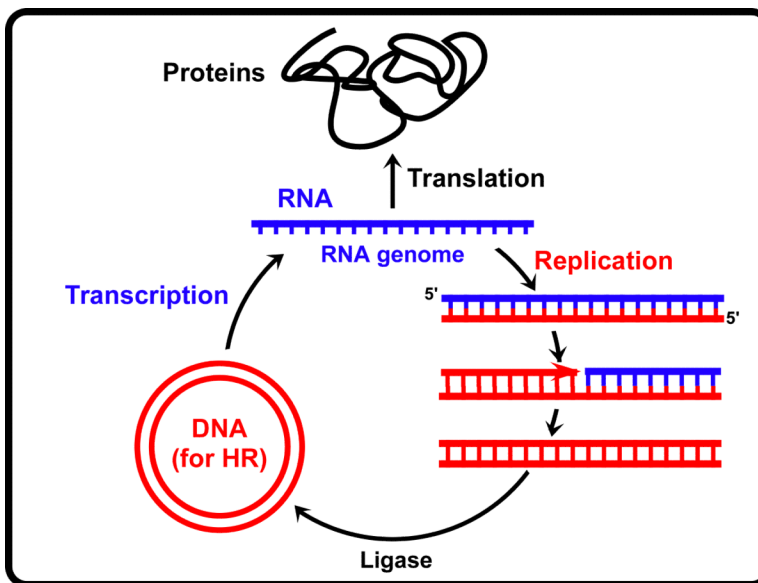
Figure 1.

Proposal for information flow in LUCA. LUCA contained proteins and the translation apparatus. Both RNA and DNA were probably present in LUCA, as inferred by homologous genes in all cell types that encode RNA polymerase, DNA ligase, ribonucleotide reductase and the RecA/Rad51 recombinase. In this diagram, RNA is the carrier of genomic information. The DNA is used as a substrate for homologous recombination (HR), RNA transcription, and to produce new copies of the RNA genome. RNA transcripts are decoded by the translation machinery that evolved the genetic code used in all cell types today. The RNA genome is replicated in simple fashion, requiring only a reverse transcriptase that first copies one strand, and then copies the second strand; no primase or helicase is required. A similar replication process is found in modern-day retroviruses (see color version of this figure at www.informahealthcare.com/bmg).
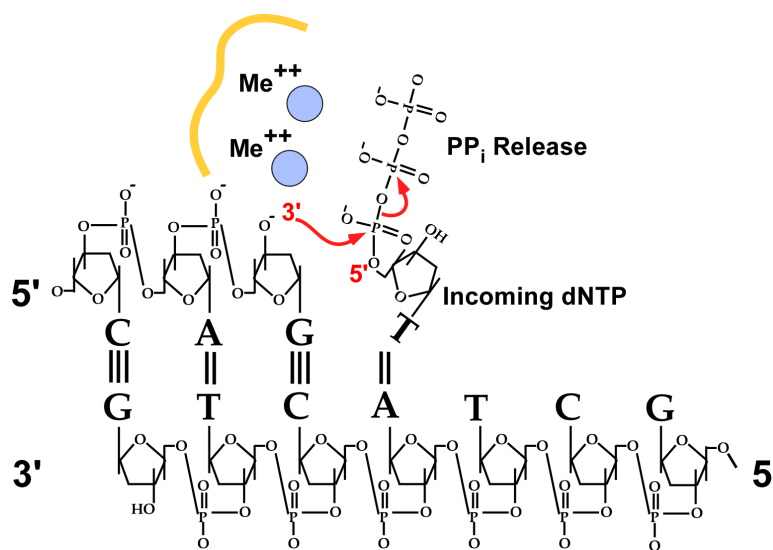
Figure 2.
Chemistry of DNA synthesis. The 3' OH terminus of the primer strand (top strand) is activated for nucleophilic attack on the alpha phosphate of an incoming dNTP, with release of PPi. The reaction involves two metal atoms (blue) that are bound within the active site of the polymerase (yellow). The metal atoms facilitate proton removal from the 3' OH, and stabilize the pyrophosphate leaving group (see color version of this figure at www.informahealthcare.com/bmg).
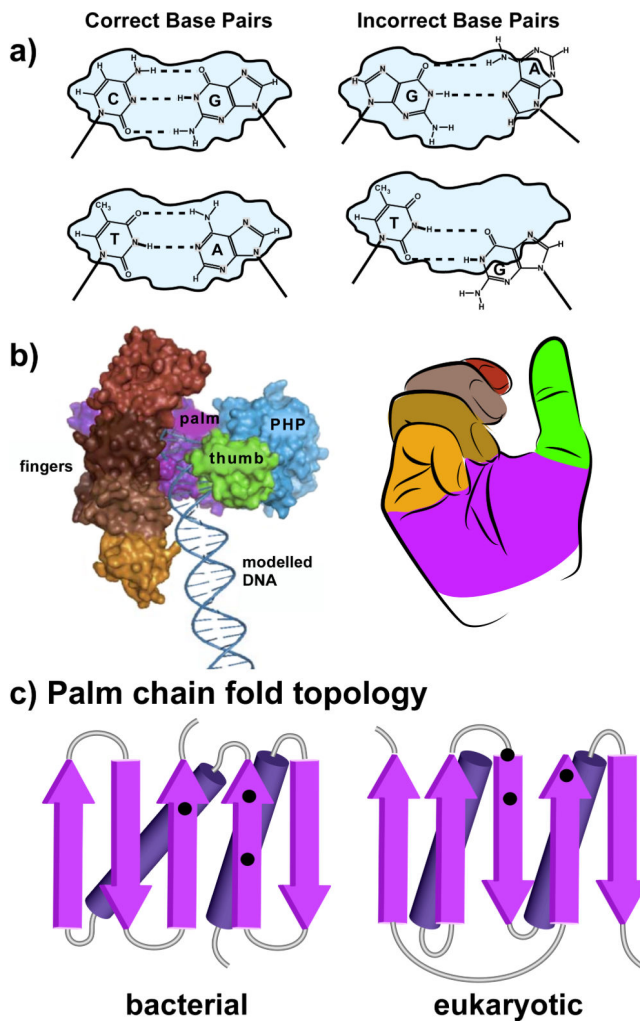
Figure 3.

Overview of DNA polymerase structure. (a) DNA polymerases match the incoming dNTP to the template strand and bury the base pair within a confined active site (gray) prior to catalysis. Only correct base pairs, GC and AT, fit the confines of the active site, thereby enabling catalysis (left diagrams). Mispairs do not fit the active site and catalysis does not occur (right diagrams). (b) DNA polymerases have the shape of a right hand. At left is a space filling model of *E. coli* Pol III, the chromosomal replicase (adapted with permission from Figure 5A of Lamers et al., 2006). To the right is an idealized right hand. The coloring of fingers, palm and thumb domains correspond to the colors of the analogous domains in the Pol III structure at the left. (c) Chain topology diagrams of the polypeptide chain folding pattern of the palm domains within: left, bacterial replicative polymerase (C family); right, eukaryotic replicative polymerases (B family) (see color version of this figure at www.informahealthcare.com/bmg).
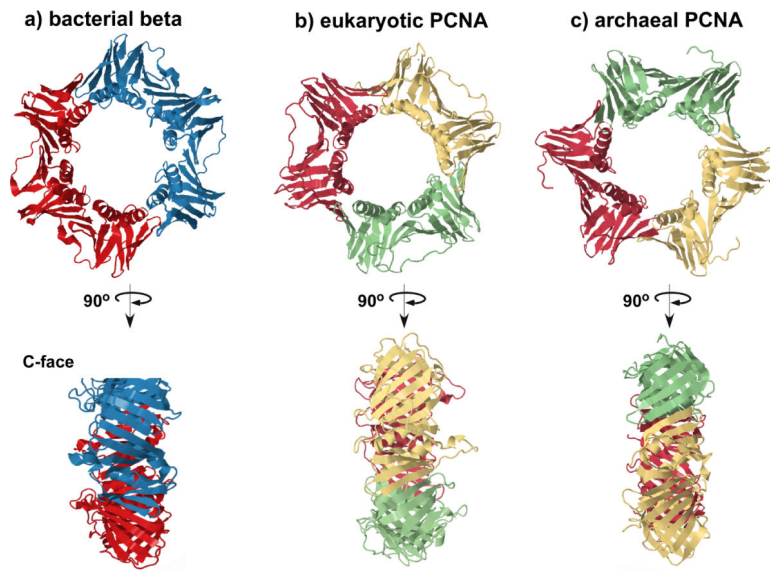
Figure 4.
Structures of sliding clamps from the three cellular domains of life. (a) The bacterial beta clamp from *E. coli* (2POL). (b) Eukaryotic PCNA clamp from budding yeast, *Saccharomyces cerevisiae* (1PLQ). (c) Archaeal PCNA clamp from *Pyrococcus furiosus* (1GE8). Different colors represent the identical protomers that comprise each ring (see color version of this figure at www.informahealthcare.com/bmg).
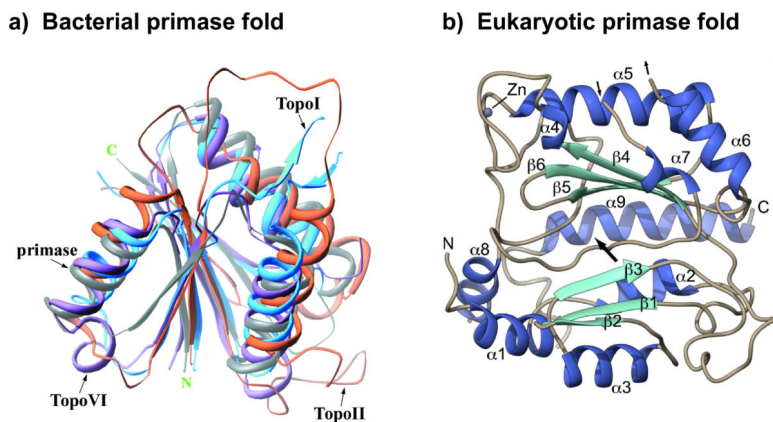
a) Bacterial primase fold

b) Eukaryotic primase fold

Figure 5.
Bacterial and eukaryotic primases adopt distinct chain folding patterns. (a) The structure of bacterial primases resembles the structure of topoisomerases, called a "Toprim" domain, and thus is thought to have shared a common ancestor with topoiosomerases. The figure shows a superimposition that includes the active site of *E. coli* primase, *E. coli* TopoI, *S. cerevisiae* TopoII and *M. jannaschii* TopoVI. Reprinted with permission of Elsevier from Figure 3(A) in Podobnik et al. (2000). (b) Human primase catalytic subunit adopts a folding pattern unrelated to topoisomerases, and instead has homology to the X-family of DNA polymerases. Figure adapted with permission from Figure 2 of Kilkenny et al. (2013) (see color version of this figure at www.informahealthcare.com/bmg).
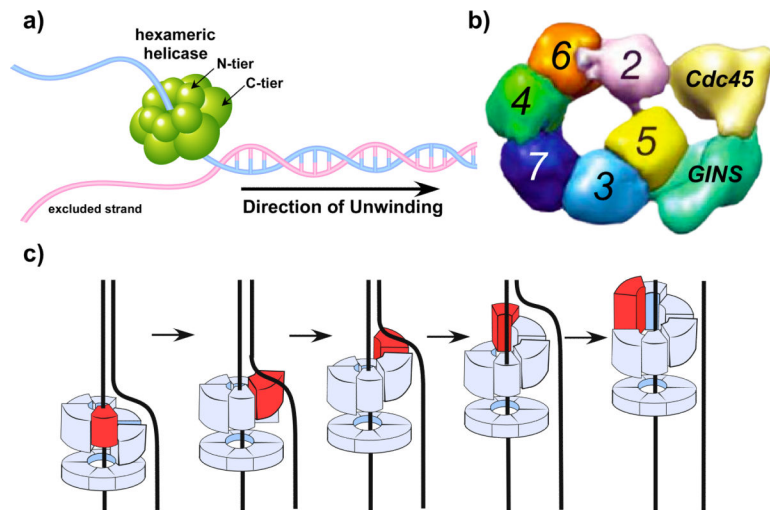
Figure 6.

Helicase architecture and proposed mechanism. (a) The bacterial replicative helicase (e.g. *E. coli* DnaB) consists of a homohexamer of motor proteins that encircle the lagging strand and hydrolyze ATP to move along the encircled strand, excluding the other strand and thereby unwinding DNA. This is referred to as the steric exclusion model of unwinding. (b) The eukaryotic helicase, CMG, contains a heterohexamer of Mcm2–7 motor subunits that are not homologous to the bacterial helicase, and also have five accessory factors, Cdc45 and the GINS tetramer, that form a second channel of unknown function. The figure is adapted with permission from Costa et al. (2011). (c) Rotary staircase model of DnaB hexameric helicase action in which each ATP hydrolysis step is proposed to cause one subunit (red) to translate forward along the ssDNA, and six ATP hydrolysis events would entail one ATP for each subunit around the hexameric ring. In this model, the "red" subunit is a different subunit of the hexamer in each drawing (see color version of this figure at www.informahealthcare.com/bmg).
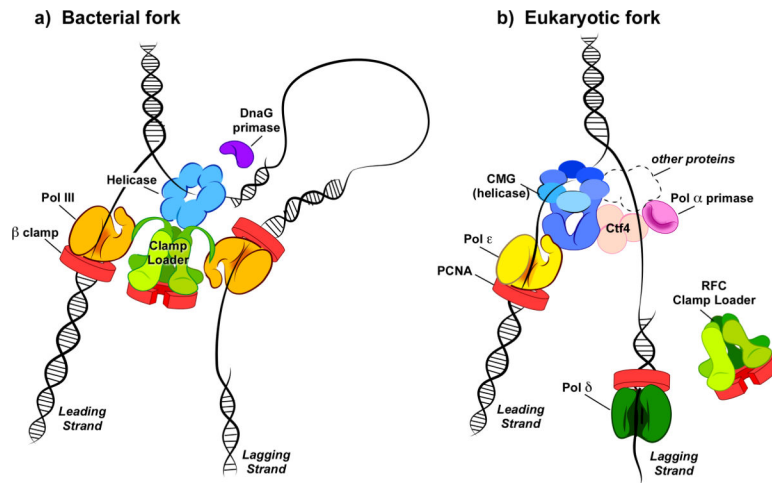
Figure 7.
Distinct replisome machines of bacterial and eukaryotic cells. (a) Replisome of *E. coli*. The clamp loader organizes the replisome using Cterminal domains that protrude from the three tau subunits of the clamp loader. Each tau subunit binds a separate copy of Pol III core (only two Pol Ill's are shown for clarity but the replisome contains three; see text for details) and each act with a beta clamp to synthesize the leading and lagging strands simultaneously. The tau arms also bind the helicase (DnaB), a homohexamer that encircles the lagging strand. The DnaG primase acts stochastically, binding the helicase periodically to initiate RNA primer synthesis. As RNA primers are extended into Okazaki fragments, DNA loops are formed. (b) Replisome of *S. cerevisiae*, a model eukaryote. Unlike the bacterial replisome, the organizing unit of the eukaryotic replisome is the CMG helicase, not the clamp loader (RFC). The six Mcm subunits of CMG helicase encircle the leading strand, and the five Cdc45 and GINS accessory factors of CMG form a second channel. CMG binds the Pol epsilon leading strand polymerase. The Ctf4 protein is a trimer that bridges CMG to Pol alpha (i.e. Ctf4 binds both CMG and Pol alpha), and Pol alpha acts as a primase to make hybrid RNA/DNA primers. The lagging strand Pol delta functions with PCNA to extend Okazaki fragments. Tight connection of Pol delta to the replisome is uncertain, and thus may not form DNA loops. Mcm10, Mrc1, Csm3 and Tof1 also travel with the replisome (these subunits are represented by the dashed perimeter) (see color version of this figure at www.informahealthcare.com/bmg).

Table I

Structural folds of core components of the bacterial and eukaryotic replisome

Component	Bacterial (fold)	Eukaryote (fold)	Conserved in LUCA
Sliding clamp	beta (6 domain ring)	PCNA (6 domain ring)	yes
Clamp loader	tau complex (AAA+)	RFC (AAA+)	yes
Helicase	DnaB (RecA)	CMG (AAA+)	no
Primase	DnaG (Toprim)	Pol alpha/primase (DnaX)	no
DNA Polymerase	Pol III (C-family)	Pols epsilon/delta (B-family)	no