DNA REPLICATION IN THE THREE DOMAINS OF LIFE

INTRODUCTION

In vivo replication of a double-stranded DNA molecule (chromosome) is a very complex process that requires the coordinated action of a conglomerate of enzyme activities involved in initiation, elongation and termination of DNA replication. Once DNA replication is initiated, it must progress to completion. Thus, cells have dedicated periods (S for synthesis) within the cell cycle to devote energy and metabolites toward replication, which is a highly costly task. The topology of DNA is also an important aspect of DNA replication. Furthermore, replicated daughter chromosomes must be partitioned equitably in coordination with cytokinesis to ensure the faithful transmission of chromosomes to the daughter cells.

DNA replication is semi-conservative

The self-complementary structure of the double-stranded DNA helix based on the formation of A-T and G-C pairs suggests that each strand might serve as a template to the formation of a complementary daughter strand. Hence, a DNA molecule (chromosome) might be copied with retention of the genetic information, thereby generating two identical daughter molecules starting with a single template molecule. This hypothesis of semi-conservative DNA replication was demonstrated by Matthew Meselson and Frank Stahl in 1958. In their experiment they separated daughter DNA

molecules and showed that both molecules consisted of one intact "old" or parental template strand and one intact newly synthesized complementary strand (daughter strand). To perform their experiment Meselson and Stahl used the heavy isotope 15N as a tag to differentially label the parental and daughter strands. Bacteria grown on a medium containing for instance (15NH4)2SO4 as a good source of nitrogen will incorporate 15N in their DNA and will, therefore, have denser DNA than bacteria grown under normal conditions, i.e. on (14NH4)2SO4 as a source of nitrogen. Light and dense molecules can be separated by ultracentrifugation through a CsCl gradient. When bacteria grown on 15N medium were transferred for one generation to 14N medium all the DNA molecules were found to have a density intermediate to

that of light and dense DNA (14N-15N hybrid molecules). After two generations 50% of the molecules had a density intermediate to that of light and heavy DNA (14N-15N hybrid molecules) and 50% were light. After three generations one would expect 25% hybrid and 75% light molecules. If DNA replication involves strand separation and synthesis of complementary new strands using both parental strands as a template, this is exactly what one would expect. Alternative models such as the conservative model (in which the parental strands remain together) or the distributive model (pieces of parental and newly synthesized DNA joined in one strand) can not account for the observations made by Meselson and Stahl. In the conservative model, 50% of the molecules are expected to be dense and 50% light in the first generation. In the dispersive model all the molecules will already have a variable density in the first generation. Therefore, the Meselson and Stahl experiment supports the semiconservative model for DNA replication. This model is valid in all kinds of organisms, Bacteria, Archaea and Eucarya.

A supplementary proof for the semi-conservative mode of DNA replication in eukaryotic cells was delivered by Herbert Taylor, Philip Woods and Walter Hughes(1957). They used root tips of *Vicia faba* (broad bean) and a cytologic technique. Taylor and his collegues were able to monitor the process of DNA replication by labeling the DNA with 3H-thymidine, a radioactive precursor of DNA that is detectable by autoradiography. The root tip cells were grown for about one generation in the presence of 3H-thymidine, then washed and transferred to unlabeled medium in which cell division continued. After each generation part of the cultures was arrested at the metaphase by adding colchicine, a chemical that poisens the spindle fiber formation such that sister chromatids do not separate. Taylor and collaborators found that the distribution of labeled thymidine is compatible with a model for DNA replication in which each strand is used as a template for the synthesis of a complementary strand, i.e. semi-conservative DNA replication.

Chemistry and enzymology of DNA synthesis

The enzymatic synthesis of DNA requires two essential substrates: the four nucleoside triphosphates (dATP, dGTP, dCTP, dTTP) and a primer:template junction. The nucleoside triphosphates have three phosphoryl groups (α = proximal position,β = middle position, γ = distal position) attached to the 2'-deoxyribose via the 5'-hydroxyl. The β and γ bonds are high energy bonds. The primer:template junction has already two essential components: (i) the single

stranded DNA template that will direct the addition of the complementary deoxynucleotides in the growing chain, and (ii) the short primer that is also complementary to the template but consists of RNA (not DNA). The primer must have a free 3'-OH adjacent (hybridized) to the single stranded region of the template. It is this free 3'-OH that will be extended by the addition of new nucleotides. All DNA synthesis proceeds in the direction 5' →3' (nobody has ever found a primase or a DNA polymerase synthesizing in the opposite direction). Since the two strands of a double stranded DNA molecule have an antiparallel orientation, the template strand must have an orientation (3' → 5') that is opposite to that of the growing DNA strand. Only the primer is chemically modified during DNA synthesis, the template just provides the information necessary to add the correct nucleotides. During the synthesis the phosphodiester bond is formed in an SN2 reaction(a type of reaction where one bond is broken and one bond is formed). The OH-group at the 3'-end of the primer strand attacks the α- phosphoryl group of the incoming nucleoside triphosphate. The leaving group in the reaction is pyrophosphate.

The whole process is catalyzed by an enzyme, DNA polymerase. DNA polymerase (and also RNA polymerase) is an unusual enzyme in that it uses a single active site to catalyze the addition of any of four dNTPs (most enzymes have an active site that catalyzes only one reaction). The DNA polymerase can do this by exploiting the nearly identical geometry of the A-T and G-C bps (see constant diameter of the α- helix). Therefore, the DNA polymerase monitors the ability of the nucleotide to form a correct pair with the template information, rather than detecting the exact nucleotide that enters the active site. Only when a correct pair is formed, the 3'-OH of the primer and the α-phosphate of the incoming nucleotide will be in the optimum position for catalysis to occur (formation of the phophodiester bond). Incorrect base pairing leads to dramatically lower rates (up to 10,000-fold) of nucleotide addition due to a catalytically unfavorable alignment of the substrates. DNA polymerases show a high discrimination between ribo- and deoxyribonucleoside triphosphates. Although rNTPs are present at about 10-fold higher concentrations in the cell than the dNTPs, they are incorporated at a rate that is more than 1,000-fold lower than the dNTPs. This discrimination is mediated by the steric exclusion of the rNTPs from the active site. The binding pocket is too small to accomodate the extra 2'-OH group on the sugar moieity of the incoming nucleotide. This space is occupied by two amino acids, the

discriminator amino acids, of the active site that make van der Waals contacts with the sugar ring. The presence of a 2'-OH group on the incoming nucleotide results in a steric clash with the discriminator amino acids and the α-phosphate of the nucleotide is then misaligned with the 3'-OH of the primer, resulting in a dramatically reduced rate of catalysis. Replacing these discriminator amino acids by smaller ones (glutamate to alanine substitutions) results in an enzyme with a strongly reduced discrimination between dNTPs and rNTPs.

DNA synthesis is semi-discontinuous

In the cell, the two strands of the DNA duplex are replicated at the same time. Therefore, the two strands must be separated to create two template DNA strands with a template: primer junction. The junction between the newly synthesized DNA and the unreplicated duplex is the replication fork. The replication fork moves continuously, thereby generating two single-stranded templates that direct the formation of two daughter DNA duplexes. The antiparallel nature of the DNA strands (innate chemical asymmetry of DNA) in a duplex molecule creates a supplementary complication for the simultaneous replication of the two template strands. Since DNA is only synthesized in the 5' → 3' direction by extending a 3'-OH end, only one of the two exposed templates can be replicated continuously as the replication fork moves. This continuously synthesized new strand is called the leading strand. Synthesis of the other strand is more problematic since this template directs the DNA polymerase to move in the direction
opposite to the direction of the replication fork. Therefore, this strand, the lagging strand, must be synthesized in a discontinuous manner. Synthesis of the lagging strand must wait for movement of the replication fork to expose a stretch of template before it can be replicated. Each time a stretch is exposed, DNA synthesis will be initiated and it will stop when it reaches the 5'-end of the previous newly synthesized stretch of lagging strand. These short fragments are called Okazaki fragments. They may vary in length from 1,000 to 2,000 nucleotides in bacteria and are generally much shorter in eukaryotes (100 to 400 nucleotides). Shortly after their synthesis Okazaki fragments are covalently joined to generate a continuous intact new strand. Leading and lagging strands often possess distinct sequence properties.

DNA REPLICATION IN BACTERIA

The origin of bidirectional replication: *oriC*

In bacteria, replication is triggered when the cell mass increases past a threshold level. The initiation event occurs at a constant ratio of cell mass to the number of chromosome origins. Cells growing more rapidly are larger and possess a greater number of origins. The growth of *E. coli* can be described in terms of the unit cell, an entity of about 1.7 μm long. Replication of the *E. coli* chromosome starts at a unique, well defined site, the origin of replication (*oriC*) (located at min 83 of the *E. coli* map) and proceeds in both directions (bidirectional replication with two replication forks) till it meets another specific site, the terminator (*ter*). All the DNA replicated starting from one particular origin is called a replicon (the *E. coli* chromosome has only one replicon).

STEPS OF DNA REPLICATION

Initiation

Initiation of replication requires two components: the replicator and the initiator. The replicator is the entire set of *cis*-acting DNA sequences that is sufficient to direct the correct initiation of DNA replication (in contrast to the origin of replication which is the site where DNA is unwound and replication starts). Although the origin of replication is always part of the replicator, sometimes (and especially in eukaryotes) the origin is only a fraction of the replicator. The initiator protein specifically recognizes a DNA sequence element in the replicator and activates the initiation of replication. It is even the only sequence-specific DNA binding protein involved in the initiation of replication. The other proteins involved in the initiation and elongation of DNA replication do not bind DNA in a sequence-specific manner.

Instead, these proteins are recruited to the replicator by a combination of protein-protein interactions and affinity for specific DNA structures (such as ssDNA or template:primer junctions, but not for specific sequences). Initiation at *oriC* was found to require the formation of a complex that comprises six proteins: DnaA, DnaB, DnaC, Gyrase, SSB and HU (a histone-like protein). HU is not absolutely required *in vitro*, but it stimulates the reaction.

HU binds to DNA in a sequence-independent manner and has the capacity to bend DNA. HU is involved in building of the structure that leads to the unwound *oriC*.

Initiator proteins typically perform three functions:

-(i) they bind to the replicator

-(ii) once bound they distort or unwind the DNA

-(iii) they attract additional protein factors required for initiation of DNA replication.

Replicators in general contain initiator binding sites and easily unwound DNA sequences (AT-rich sequences). *E. coli oriC* is about 245 bp long and contains five binding sites of 9 bp for the initiator protein DnaA and three 13 bp long sites of AT rich DNA for the initial melting and formation of ssDNA. When bound to ATP, DnaA also interacts with DNA sequences in the vicinity of the 13 bp repeats of *oriC*. These additional interactions result in the separation of the two DNA strands over more than 20 bp. This unwound DNA provides a ssDNA template for additional replication proteins to begin the RNA and DNA synthesis steps of replication. The complex of DnaA and ssDNA recruits two copies (one for each of the two replication forks, bidirectional replication) of a complex of two proteins: the helicase DnaB (six monomers) and the homohexameric helicase loader, DnaC. Once bound to the ssDNA at the origin, the helicase loader directs the assembly of the associated ring-shaped helicase DnaB around the ssDNA (process analogous to the loading of the sliding clamp by the clamp loader). One helicase is loaded onto each of the two separated single strands at the origin. The orientation of the two helicase molecules is such that they will proceed toward each other as they move with a 5' → 3' polarity along their associated ssDNAs. Loading of the helicase leads to the dissociation with the helicase loader DnaC and activation of the helicase activity of DnaB. Then each of the two helicases will recruit a primase which will synthesize a primer.

Elongation

Initiation of DNA synthesis requires an RNA primer that must be removed subsequently All DNA polymerases (bacterial, archaeal, eukaryal enzymes) require a free 3'-OH primer to start. DNA polymerases (in contrast to RNA polymerases) cannot initiate a new strand *de novo*. To create this primer, cells use a special class of RNA polymerases. Primase (*dnaG* in *E. coli*, a monomeric protein) is a specialized RNA polymerase that synthesizes short RNA primers (5-12 nucleotides long) using the information content of a single stranded DNA template.

Unlike other RNA polymerases that synthesize mRNA, RNA primase (which is also much smaller, 60 kDa) does not require specific DNA sequences (promoter sequences) to initiate RNA synthesis. Furthermore, in contrast to the mRNA synthesizing RNA polymerases, primases are only activated when they are associated with other proteins involved in DNA replication such as DNA helicase (*dnaB* in *E. coli*). Primers start with the sequence pppAG, positioned opposite the sequence 3'-GTC-5' in the template. Even though DNA polymerases only incorporate dNTPs into DNA they can initiate synthesis using either an RNA primer or a DNA primer annealed to a DNA template. RNA primers are used *in vivo* to initiate DNA synthesis of the leading and lagging strands. DNA primers are used *in vivo* to fill the gaps after digestion of the RNA primers (Okazaki fragments) and in repair of DNA damage, and *in vitro* (PCR amplifications). Both the leading and the lagging strand synthesis requires RNA primers to initiate DNA synthesis, but the number of primers is much higher for lagging strand synthesis since the synthesis of each Okazaki fragment requires a new primer. To complete the DNA synthesis these primers have to be removed and replaced with DNA to fill the gaps. To do so, the enzyme RNase H (H for hybrid DNA:RNA) recognizes and removes most of each RNA primer. RNase H does, however, not remove the ribonucleotide directly linked to the DNA end (RNase H can only cleave bonds between two ribonucleotides, not between a ribo- and a deoxyribonucleotide). The final ribonucleotide is removed by an exonuclease that degrades RNA or DNA starting from the 5'-end (Pol I can do it in *E. coli*). The digestion of the primer leaves a gap in the double stranded DNA that is an ideal substrate (primer-template junction) for a DNA polymerase that will fill the gap. The action of the polymerase will generate a molecule that is complete except for one single break in the backbone between the 3'-OH of the last incorporated nucleotide of the repaired gap and the 5'- phosphate (monophosphate, not a triphosphate) of the DNA part of the initial Okazaki

fragment. This nick can be repaired by DNA ligase, an enzyme that uses the energy of a high energy co-factor (ATP, mostly, or NAD+) to create a phosphodiester bond. After all primers are replaced and the associated nicks sealed the DNA synthesis is complete. Ligases use two cofactors : Mg2+ ions and either ATP or NAD+ (nicotinamide adenine dinucleotide). ATP-dependent ligases are present in eukaryotes, viruses and some bacteria and archaea. NAD+-dependent enzymes are found in most bacteria (including *E. coli*). NAD+ is commonly involved in oxidation reduction reactions, but its role in DNA ligation is quite different and is similar to the role of ATP. In step 1 of the ligation reaction an adenylate group (AMP) is transferred from either ATP or NAD+ to a lysine residue in the active site of the enzyme. In step 2 the enzyme binds to DNA, at the site of a strand break, and transfers the AMP to the 5'-phosphate of the DNA substrate. This activates the 5'-phosphate for nucleophilic attack by the 3'-OH group of the DNA. Finally, the nucleophilic attack leads to displacement of the AMP and the formation of a new phosphodiester bond that seals the nick.

Termination , Finishing Dna Replication And Disentangling Daughter Chromosomes

The completion of chromosome replication requires a specific set of events: arrest of replication and physical separation of the daughter molecules. For a circular chromosome, the normal replication fork machinery can replicate the entire molecule (this is not so for linear chromosomes). In *E. coli*, termination of the bidirectional replication occurs when the replication forks meet (about 40 min after initiation at *oriC*) in a region opposite to *oriC*. This region are located a series of 10 termination sites called *ter* sites (*terA* to *terJ*). A *ter* site consists of a 23 bp long asymmetric sequence that is the binding site for the 36 kDa monomeric Tus protein (Termination utilization substance). A Ter site will block replication forks moving in one direction but not in the other. This creates a "replication fork trap" that allows the fork to enter but not to leave the terminus region. DNA replication will stop because the *ter*/Tus complex inhibits the helicase activity of DnaB (that will eventually dissociate from the replisome) in an orientation dependent manner. Unwinding is therefore blocked, and the whole process of DNA replication will stop.

After the replication of a circular chromosome is complete, the two resulting daughter molecules remain linked together as catenates. A similar problem will arise each time that two homologous DNA molecules recombine an uneven number of times in the cell (a dimeric molecule will then be formed). The physical separation of these daughter molecules is a prerequisite for their segregation into separate daughter cells. This disentangling is accomplished by the action of Type II topoisomerases. In *E. coli* this is performed by the heterotetrameric XerC-XerD recombinase (a tyrosine type of recombinase or resolvase) that acts at a specific site, the *dif* site. Resolution also requires the FtsK protein (a large transmembrane protein) that is localized in the septum and is required for both chromosome segregation and cell division. If disentangling would not occur correctly, septum formation would release an anucleated cell and a cell containing both daughter chromosomes.

Regulation of initiation of DNA replication by DnaA, ATP levels and SeqA

In all organisms initiation of chromosome replication must be tightly controlled to ensure that the chromosome number and the cell number remain balanced and that every daughter cell gets an intact copy of the genetic content (genome) of the cell. This problem is more acute in eukaryotic cells with several chromosomes (see below) that in prokaryotes with one single chromosome (some prokaryotes such as the bacterium *Agrobacterium tumefaciens* have more than one chromosome). Yet, even organisms with only one chromosome have to prevent runaway chromosome duplication. *E. coli* does so by inhibiting recently initiated origins from reinitiating. This is done by a combination of different mechanisms that act to prevent rapid replication reinitiation from *oriC*. One strategy exploits the changes in the methylation state of the *E. coli* DNA before and after replication. The enzyme Dam methylase (DNA adenosine methyl transferase) adds a methyl group to every adenine residue (N6 position) in the palindromic sequence GATC. Typically the *E. coli* genome is fully methylated but this situation is different immediately after replication. The A residues in the newly synthesized strand have not yet been methylated and, therefore, the recently replicated DNA is hemimethylated (one strand methylated only, the template strand). This hemimethylated state of the newly replicated *oriC* site is detected by the SeqA protein. SeqA binds tightly to the GATC sequence, but only when it is hemimethylated. There are many GATC sites in the immediate vicinity of the *oriC* site and

SeqA binds these sites before they can be fully methylated by the Dam enzyme. Binding of SeqA has two consequences: (i) it dramatically reduces the rate at which the GATC sequences are methylated, (ii) when bound to these sites SeqA inhibits the binding of DnaA to *oriC* and prevents the initiation of a new round of replication from the two newly synthesized copies of *oriC*. A second control mechanism exploits the fact that only ATP-bound DnaA can direct initiation of replication at *oriC*. However, during the initiation process, the DnaA bound ATP is converted to ADP and Pi. This DnaA-bound ADP must then be replaced by ATP, but this is a slow process. There are more than 300 DnaA 9-bp long binding sites outside of the *oriC* region, and as they are replicated their number increases to > 600. This important increase in DnaA binding sites helps to reduce the levels of available DnaA for initiation. It is the combination of these two strategies that rapidly and dramatically reduces the ability of *E. coli* to initiate replication from newly synthesized copies of *oriC*.

REPLICATION OF LINEAR EUKARYOTIC CHROMOSOMES

Multiple origins for initiation of replication

In eukaryotic cells, DNA replication is restricted to part of the cell cycle, the S phase, which usually lasts a few hours in higher eukaryotes. During this time all the DNA in the cell must have been duplicated exactly once. Incomplete replication of any part would cause inappropriate links between daughter chromosomes. The segregation of such linked chromosomes will cause chromosome breakage or loss. Also re-replication of chromosomes would have severe consequences. Increasing the number of copies of particular genes (proto oncogenes) can lead to catastrophic defects in gene regulation, cell division, or the response to environmental signals. Therefore, every bp in each chromosome is replicated once and only once each time a eukaryotic cell divides. Notice that not every replicator has to be activated, some replicons may be passively replicated by a growing fork derived from an adjacent origin. *E. coli* requires about 40 min to replicate its chromosome, while Drosophila with 40- times more DNA, is known to accomplish the same task in about 3 min only (during embryogenic cell division). Replication of the huge amount of DNA within one single eukaryotic chromosome is accomplished by dividing it into many individual replicons. Eukaryotic cells also have larger amounts of the DNA polymerases,

up to 50,000 copies in animal cells (compared to 15 copies for *E. coli* Pol III). Individual replicons are relatively small, typically about 40 kb in the yeast *Saccharomyces cerevisiae* which has about 400 replicons for the ensemble of 16 chromosomes. The human genome (46 chromosomes) has about 10,000 growing forks, with an average replicon length of about 100 kb. The rate of replication is about 2,000 bp/min, which is much slower (about 30-fold) than the movement of the bacterial replication fork (1,000 bp/sec). The mammalian genome could be replicated in about 1h if all the replicons were functioning simultaneously. However, this is not the case and the S phase lasts about 6h in a typical somatic cell, which indicates that only about 15% of the replicons are active at any given moment. Eukaryotic chromosomal replicons do not have termini at which the replication fork stops. It appears that a growing replication fork continues until it meets and fuses with another fork proceeding towards it, initiated from the adjacent replicon. Evidently, initiation at the various replicons and replication of the whole set of linear chromosomes of an eukaryotic cells will have to be strictly controlled and coordinated. The origin in each replicon is activated once and only once in a single division cycle. The start of the S phase is signaled by the activation of the first replicons. Over time, initiation events occur at other replicons, in an ordered manner. Quite large regions of the chromosome can be characterized as "early replicating" and others as "late replicating". Initiation of replication requires two steps that occur in distinct times in the cell cycle: replicator selection and origin activation (in bacteria these two events are intimately linked). Replicator selection occurs in the G1 phase, prior to the S phase. This process leads to the assembly of a multiprotein complex at each replicator in the genome. Origin activation occurs later, in the S phase. It triggers the replicator associated complex to initiate unwinding of the DNA and recruitment of the polymerase.

Replicator selection starts with the formation of the pre-replicative complexes (pre- RCs). This pre-RC is composed of 4 proteins (complexes) that assemble in a specific order at each replicator. Pre-RCs are activated to initiate replication by the action of two protein kinases, Cdk (cyclin dependent kinase) and Ddk, which covalently attach phosphate groups to their target proteins, pre-RC components and other replication proteins. Cdk and Ddk are inactive in the G1 phase, but are activated upon entry in the S phase. Phosphorylation results in the assembly of

additional replication proteins at the origin and the initiation of replication. Among these are the three eukaryotic DNA polymerases.

The first step in this process is the recognition of the replicator by ORC (origin recognition complex), the eukaryotic initiator. ORC contains six separate proteins: Orc1-6. Once ORC is bound it recruits two helicase loading proteins: Cdc6 (cell division cycle protein 6 ; ATPase) and Cdt1 (cdc1- dependent transcript 1). In this way the replisome can be assembled only before the S phase. Together ORC and the loader proteins recruit the eukaryotic replication fork helicase, the Mcm2-7 complex (Mcm = mini chromosome maintenance). So far, only *Saccharomyces cerevisiae* has been shown to have clearly defined replication origins, known as the autonomously replicating sequences (ARS). The replicator of yeast is a 100 bp long sequence that contains three important regions: A, B1 and B2. ORC recognizes the conserved sequence of the A-element which contains an 11 bp A+T rich sequence, and the less conserved element B-1. Like DnaA, ORC binds and hydrolyzes ATP. The binding of ATP to ORC is required for sequence specific binding to the origin of replication. Unlike DnaA, ORC does not itself direct strand separation. Instead, ORC recruits all the remaining replication proteins to the replicator. The B-2 element facilitates the melting (unwinding) of the DNA and is involved in the binding of the replication factors. The structure of replicators from higher, multicellular eukaryotes is still not yet well understood, but they appear to be much larger than the yeast replicator, generally more than 1,000 bp.

Regulation of eukaryotic replication initiation

Initiation at the various replicators has to be strictly controlled. Cdks (cyclin dependent kinases) play two apparently contradictory roles in regulating pre-RC function: (i) they are required to activate pre-RCs to initiate (see above), but (ii) Cdk activity also inhibits the formation of new pre-RCs. It is the tight connection between pre-RC function, Cdk levels, and the cell cycle that ensures that the eukaryotic genome is replicated only once per cell cycle. Active Cdk is absent during the G1 phase. In contrast, high levels of Cdk are present during the remainder of the cell cycle (S, G2, M phases). Thus, during each cell cycle there is only one opportunity for pre-Rcs to

form (during G1) and only one opportunity of the pre-RC to be activated (in theory during S, G2, and M, but in practice all pre-RCs are activated, or disrupted by another moving replication fork, in the S phase). Pre-RC are disassembled after they are activated or after the DNA to which they are bound is replicated.

REPLICATING THE ENDS OF LINEAR EUKARYOTIC CHROMOSOMES: TELOMERASE

Replication of the ends of a linear DNA chromosome poses a specific problem to the synthesis of the lagging strand: the end replication problem. The problem results from the requirement for an RNA primer to initiate all new DNA synthesis. There is no particular problem for the synthesis of the leading strand, which can be initiated from an internal replicator with a single primer that can be extended till the extreme 5'-terminus of its template strand. However, the situation is different for the lagging strand, whose synthesis requires multiple primers. Here, a complete copy cannot be made because even if the end of the last RNA primer for Okazaki fragment synthesis anneals to the final base of the lagging strand template, degradation of this RNA primer will result in a short region of unreplicated DNA (DNA filling synthesis cannot start because there is no 3'-OH group available at a template: primer junction).

As a result, each round of DNA replication would result in shortening of one of the two daughter DNA molecules. Obviously this should not happen because this would impair the propagation of the intact genetic material from generation to generation. Specific sequences at the end of eukaryotic chromosomes, the telomeres and a special enzyme, telomerase, are used to solve the problem. Telomeres are usually composed of many head-to-tail repeats of a TG-rich sequence (200-400 copies). The human telomeres consist of repeats of the sequence 5'-TTAGGG-3' (*Tetrahymena* has the sequence 5'-TTGGGG). Although most of these repeats are double stranded the 3'-end of each chromosome extends beyond the 5'-end as single-stranded DNA. This end recruits the telomerase. Telomerase is a special enzyme that is composed of protein and RNA (ribonucleoprotein). Telomerase, like all other DNA polymerases, acts to extend the 3'-end of its DNA substrate, but unlike the other DNA polymerases, telomerase does not need an exogenous DNA template to direct the addition of dNTPs. It is the RNA component of the telomerase that serves as a template for adding the telomeric repeat sequence to the 3'-terminus

at the end of the lagging strand. The key to the telomerase function is the composition of its RNA component that contains 1.5 copies of the complement of the telomere sequence (human: 5'-TAACCCTAA-3'). Therefore this region can anneal to the single-stranded DNA at the end of the telomere and thus create a primer:template junction. Repeated elongation, dissociation and reassociation-elongation result in the addition of many copies of the telomeric repeat sequence to the 3'-end of the molecule. Specific telomere-binding proteins further regulate telomerase activity and telomere length. Telomerase specifically elongates the 3'-OH of particular ssDNA sequences using its own RNA as a template. Therefore, telomerase belong to the class of Reverse transcriptases (synthesize DNA with an RNA as primer). The newly synthesized DNA is single-stranded. By providing an extended 3'-end the telomerase provides an additional template for the action of the lagging strand replication machinery which can then extend the 5'-end of the DNA. It is important to notice that there will still be a single stranded region at the end of the chromosome. The combined action of the telomerase and of the lagging strand replication machinery can however ensure that the telomere is maintained at sufficient length to protect the end of the chromosome from becoming too short (risk of deleting important sequences).

DNA REPLICATION IN ARCHAEA:

Archaea are true prokaryotes (they have no nucleus). Yet, their information processing machineries (replication, transcription, translation) are more closely related to the eukaryotic homologues than to the bacterial ones. The fact that the archaeal replication machinery is generally percieved as a simplified version of the eukaryal machinery has generated considerable interest in the Archaea as an experimentally tractable model for the fundamentally related yet massively more complex eukaryotic replication machinery.

The origin binding proteins in archaea are homologues of the related eukaryotic Orc1 and Cdc6 proteins. All archaeal genomes that have been sequenced to date contain at least one gene with homology to both Orc1 and Cdc6 (gene is generally close to the origin(s) of replication), some carry several copies (up to nine). They are members of the AAA+ protein family (ATPases associated with various cellular activities). The single Orc1/Cdc6 protein from *P. furiosus* was shown to bind the single origin of replication region. Therefore, it appears to act as the origin

recognition and binding protein in archaeal replication. All characterized archaeal origins of replication possess a number of repeated sequence elements that are bound by the Orc1/Cdc6 homologs and have stretches that are highly rich in A and T bases indicative of readily meltable DNA.

All archaeal genomes sequenced to date have at least one MCM homologue. In contrast to eukaryotes, however, many archaea contain only one MCM gene, and the proteins form homohexamers *in vitro*.Like the eukaryotic MCM and the bacterial DnaB proteins, archaeal MCM is an AAA+ protein. *In vivo* MCM was found to interact functionally with Cdc6, and, via GINS (a tetrameric complex), with the primase (PriS + PriL). However, since genetic systems for archaeal replication are still in their infancy, it is not yet known whether MCM is essential for replication. For archaea there is little known about MCM loading. There is apparently no homologue of DnaC or Cdt1 in archaea. Therefore, it is suggested that the Orc/Cdc6 proteins may perform the functions carried out by ORC, Cdc6 and Cdt1 in eukaryotes. It is however not excluded that there would be a homologue of Cdt1 that would show only very weak sequence homology and would therefore not be easily recognized in genome sequences (yeast and human Cdt1 proteins show only 10 % sequence similarity). ATP hydrolysis by Cdc6 may be important for MCM loading, similar to the situation in eukaryotes. Single-stranded DNA-binding proteins are present in all three domains of life. They protect the ssDNA from nuclease degradation and chemical modification during replication, recombination and many other processes which require DNA to be unwound.

In common with bacteria and eukarya, archaea possess multiple DNA polymerases. Euryarchaeota possess DNA polymerases that belong to two distinct families, the ubiquitous family B and a specific euryarchaeal family D type. In bacteria the sliding clamp is a homodimer, the β-clamp. In contrast, in archaea and in eucaryotes the sliding clamp (PCNA) is a trimer. The helicase loader RFC catalyzes the opening of the PCNA ring and the deposition of PCNA on the DNA. This process is dependent on ATP binding by RFC. Archaeal RFC is a pentamer containing four identical copies of a small subunit and a single large subunit.

REPLICATION OF EXTRACHROMOSOMAL ELEMENTS

Replication of linear viral DNA molecules

Not all linear DNA chromosomes use the telomerase strategy to be correctly copied. Some use the protein primer strategy instead of an RNA primer. This strategy is mainly used by bacterial and animal viruses such as the bacteriophage λ 29 and Adenovirus, and also by certain linear bacterial chromosomes (such as *Streptomyces lividans*). The priming protein (encoded by the viral genome) binds covalently to the lagging strand template (5'-ends of the viral dsDNA) and uses an amino acid (serine in the case of Adenovirus) to provide an OH-group that replaces the 3'-OH normally provided by the RNA primer at the primer: template junction. These priming proteins remain covalently bound to the newly synthesized DNA. Some viruses and bacteriophages (such as bacteriophage λ) have linear DNA genomes in the capsids, but they circumvent the problem by converting the linear replicon into a circular one. The injected linear molecule circularizes in the host cell and is used as a circular template in DNA replication. Bacteriophage λ has 12 bp long single-stranded 5'- sticky ends. These sequences are complementary and allow the molecule to circularize upon hybridization of the *cos* (cohesive) sites. This circular intermediate is then used as a template for replication via the rolling circle mechanism. The long concatemeric DNA is then cleaved at specific sites (12 bp staggered cuts at the *cos* site) to generate the unit molecules that are subsequently encapsulated (as linear molecules) in the viral capsids.

Rolling circle replication

Some circular molecules (such as the circularized λ genome) use only one strand to generate multiple copies. A nick opens one strand and then the free 3'-OH end (equivalent to primer: template junction) is used to initiate DNA synthesis by the DNA polymerase. The newly synthesized strand displaces the original parental strand, thus generating a structure that is called "rolling circle" (the growing point can be envisaged as rolling around the circular template). This reaction could in principle continue indefinitely. The newly synthesized material is covalently linked to the original DNA. As a consequence the original unit is followed by a large number of unit genomes covalently linked to each other. The linear form may be maintained as a single-

stranded molecule and transformed in circular ssDNA molecules of unit length (which can in some instances be further converted into circular dsDNA by complementary strand synthesis) and encapsulated, or be transformed by complementary strand synthesis into concatemeric dsDNA that is then further processed into dsDNA molecules of unit length. The rolling circle mechanism is also used to generate amplified ribosomal DNA (rDNA) in the *Xenopus* oocyte. The genes encoding ribosomal RNA (rRNA) are organized as a large number of continuous repeats in the genome. A single repeating unit from the genome is converted into a rolling circle. The displaced tail, which contains many units (concatemers) is converted into duplex DNA, later it is cleaved from the circle so that the two ends can be joined together to generate a large circle of amplified rDNA that consists of a large number of identical repeating units. This molecule can be used to generate large amounts of rRNA.