Course Code- Zoo-22-CR   **3rd Sem**

DNA Biosynthesis In Prokaryotes

DNA synthesis is the natural or artificial creation of deoxyribonucleic acid (DNA) molecules. The term DNA synthesis can refer to any of the following in various contexts:

- DNA replication - DNA biosynthesis (*in vivo* DNA amplification)
- Polymerase chain reaction - enzymatic DNA synthesis (*in vitro* DNA amplification)
- Gene synthesis - physically creating artificial gene sequences

In molecular biology, DNA replication is the biological process of producing two identical replicas of DNA from one original DNA molecule. This process occurs in all living organisms and is the basis for biological inheritance. DNA is made up of a double helix of two complementary strands. During replication, these strands are separated. Each strand of the original DNA molecule then serves as a template for the production of its counterpart, a process referred to as semiconservative replication. Cellular proofreading and error-checking mechanisms ensure near perfect fidelity for DNA replication.

In a cell, DNA replication begins at specific locations, or origins of replication, in the genome. Unwinding of DNA at the origin and synthesis of new strands results in replication forks growing bi-directionally from the origin. A number of proteins are associated with the replication fork to help in the initiation and continuation of DNA synthesis. Most prominently, DNA polymerase synthesizes the new strands by adding nucleotides that complement each (template) strand. DNA replication occurs during the S-stage of interphase.

DNA replication can also be performed *in vitro* (artificially, outside a cell). DNA polymerases isolated from cells and artificial DNA primers can be used to initiate DNA synthesis at known sequences in a template DNA molecule. The polymerase chain reaction (PCR), a common laboratory technique, cyclically applies such artificial synthesis to amplify a specific target DNA fragment from a pool of DNA.

DNA   STRUCTURE

DNA usually exists as a double-stranded structure, with both strands coiled together to form the characteristic double-helix. Each single strand of DNA is a chain of four types ofnucleotides. Nucleotides in DNA contain a deoxyribose sugar, a phosphate, and a nucleobase. The four types of nucleotide correspond to the four nucleobases adenine,cytosine, guanine, and thymine, commonly abbreviated as A,C, G and T. Adenine and guanine are purine bases, while cytosine and thymine are pyrimidines. These nucleotides form phosphodiester bonds, creating the phosphate-deoxyribose backbone of the DNA double helix with the nuclei bases pointing inward (i.e., toward the opposing strand). Nucleotides (bases) are matched between strands through hydrogen bonds to form base pairs. Adenine pairs with thymine (two hydrogen bonds), and guanine pairs with cytosine (stronger: three hydrogen bonds).

DNA strands have a directionality, and the different ends of a single strand are called the "3' (three-prime) end" and the "5' (five-prime) end". By convention, if the base sequence of a single strand of DNA is given, the left end of the sequence is the 5' end, while the right end of the sequence is the 3' end. The strands of the double helix are anti-parallel with one being 5' to 3', and the opposite strand 3' to 5'. These terms refer to the carbon atom in deoxyribose to which the next phosphate in the chain attaches. Directionality has consequences in DNA synthesis, because DNA polymerase can synthesize DNA in only one direction by adding nucleotides to the 3' end of a DNA strand.

The pairing of complementary bases in DNA (through hydrogen bonding) means that the information contained within each strand is redundant. Phosphodiester (intra-strand) bonds are stronger than hydrogen (inter-strand) bonds. This allows the strands to be separated from one another. The nucleotides on a single strand can therefore be used to reconstruct nucleotides on a newly synthesized partner strand.

DNA POLYMERASE

DNA polymerases are a family of enzymes that carry out all forms of DNA replication. DNA polymerases in general cannot initiate synthesis of new strands, but can only extend an existing

DNA or RNA strand paired with a template strand. To begin synthesis, a short fragment of RNA, called a primer, must be created and paired with the template DNA strand.

DNA polymerase adds a new strand of DNA by extending the 3' end of an existing nucleotide chain, adding new nucleotides matched to the template strand one at a time via the creation of phosphodiester bonds. The energy for this process of DNA polymerization comes from hydrolysis of the high-energy phosphate (phosphoanhydride) bonds between the three phosphates attached to each unincorporated base. Free bases with their attached phosphate groups are called nucleotides; in particular, bases with three attached phosphate groups are called nucleoside triphosphates. When a nucleotide is being added to a growing DNA strand, the formation of a phosphodiester bond between the proximal phosphate of the nucleotide to the growing chain is accompanied by hydrolysis of a high-energy phosphate bond with release of the two distal phosphates as a pyrophosphate. Enzymatic hydrolysis of the resulting pyrophosphate into inorganic phosphate consumes a second high-energy phosphate bond and renders the reaction effectively irreversible.

In general, DNA polymerases are highly accurate, with an intrinsic error rate of less than one mistake for every 10^7 nucleotides added. In addition, some DNA polymerases also have proofreading ability; they can remove nucleotides from the end of a growing strand in order to correct mismatched bases. Finally, post-replication mismatch repair mechanisms monitor the DNA for errors, being capable of distinguishing mismatches in the newly synthesized DNA strand from the original strand sequence. Together, these three discrimination steps enable replication fidelity of less than one mistake for every 10^9 nucleotides added.

The rate of DNA replication in a living cell was first measured as the rate of phage T4 DNA elongation in phage-infected E. coli. During the period of exponential DNA increase at 37 °C, the rate was 749 nucleotides per second. The mutation rate per base pair per replication during phage T4 DNA synthesis is 1.7 per 10^8.
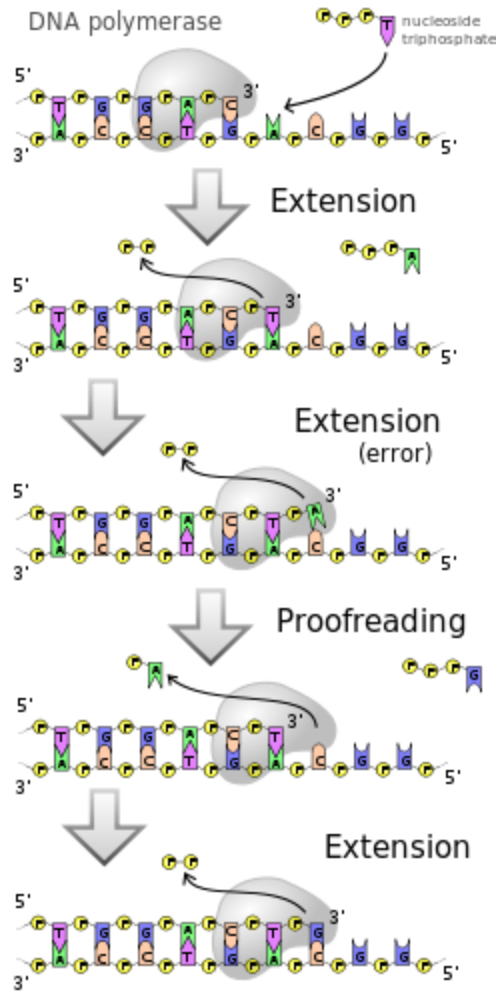
REPLICATION   PROCESS

DNA replication, like all biological polymerization processes, proceeds in three enzymatically catalyzed and coordinated steps: initiation, elongation and termination.

INITIATION

For a cell to divide, it must first replicate its DNA. This process is initiated at particular points in the DNA, known as "origins", which are targeted by initiator proteins. In *E. coli* this protein is DnaA; in yeast, this is the origin recognition complex. Sequences used by initiator proteins tend to be "AT-rich" (rich in adenine and thymine bases), because A-T base pairs have two hydrogen bonds (rather than the three formed in a C-G pair) and thus are easier to strand separate. Once the origin has been located, these initiators recruit other proteins and form the pre-replication complex, which unzips the double-stranded DNA.
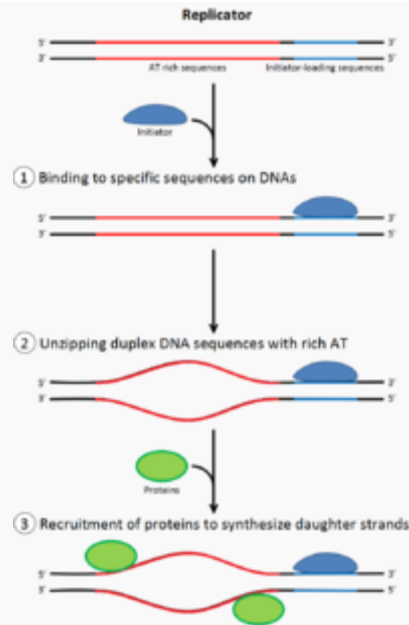
Role of initiators for initiation of DNA replication.
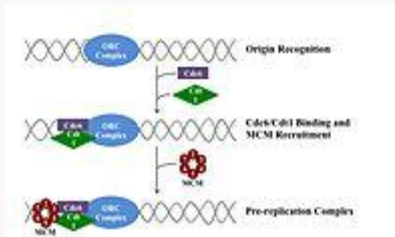
Formation of pre-replication complex.

ELONGATION

DNA polymerase has 5'-3' activity. All known DNA replication systems require a free 3' hydroxyl group before synthesis can be initiated (note: the DNA template is read in 3' to 5' direction whereas a new strand is synthesized in the 5' to 3' direction—this is often confused). Four distinct mechanisms for DNA synthesis are recognized:

1. All cellular life forms and many DNA viruses, phages and plasmids use a primase to synthesize a short RNA primer with a free 3' OH group which is subsequently elongated by a DNA polymerase.

2. The retroelements (including retroviruses) employ a transfer RNA that primes DNA replication by providing a free 3′ OH that is used for elongation by the reverse transcriptase.

3. In the adenoviruses and the φ29 family of bacteriophages, the 3' OH group is provided by the side chain of an amino acid of the genome attached protein (the terminal protein) to which nucleotides are added by the DNA polymerase to form a new strand.

4. In the single stranded DNA viruses — a group that includes the circoviruses, the geminiviruses, the parvoviruses and others — and also the many phages and plasmids that use the rolling circle replication (RCR) mechanism, the RCR endonuclease creates a nick in the genome strand (single stranded viruses) or one of the DNA strands (plasmids). The 5′ end of the nicked strand is transferred to atyrosine residue on the nuclease and the free 3′ OH group is then used by the DNA polymerase to synthesize the new strand.

The first is the best known of these mechanisms and is used by the cellular organisms. In this mechanism, once the two strands are separated, primase adds RNA primers to the template strands. The leading strand receives one RNA primer while the lagging strand receives several. The leading strand is continuously extended from the primer by a DNA polymerase with high processivity, while the lagging strand is extended discontinuously from each primer forming Okazaki fragments. RNase removes the primer RNA fragments, and a low processivity DNA polymerase distinct from the replicative polymerase enters to fill the gaps. When this is complete, a single nick on the leading strand and several nicks on the lagging strand can be found. Ligase works to fill these nicks in, thus completing the newly replicated DNA molecule.

The primase used in this process differs significantly between bacteria and archaea/eukaryotes. Bacteria use a primase belonging to the DnaG protein superfamily which contains a catalytic domain of the TOPRIM fold type. The TOPRIM fold contains an α/β core with four conserved strands in a Rossmann-like topology. This structure is also found in the catalytic domains

of topoisomerase Ia, topoisomerase II, the OLD-family nucleases and DNA repair proteins related to the RecR protein.

The primase used by archaea and eukaryotes, in contrast, contains a highly derived version of the RNA recognition motif (RRM). This primase is structurally similar to many viral RNA-dependent RNA polymerases, reverse transcriptases, cyclic nucleotide generating cyclases and DNA polymerases of the A/B/Y families that are involved in DNA replication and repair. In eukaryotic replication, the primase forms a complex with Pol α

Multiple DNA polymerases take on different roles in the DNA replication process. In *E. coli*, DNA Pol III is the polymerase enzyme primarily responsible for DNA replication. It assembles into a replication complex at the replication fork that exhibits extremely high processivity, remaining intact for the entire replication cycle. In contrast, DNA Pol I is the enzyme responsible for replacing RNA primers with DNA. DNA Pol I has a 5' to 3' exonuclease activity in addition to its polymerase activity, and uses its exonuclease activity to degrade the RNA primers ahead of it as it extends the DNA strand behind it, in a process called nick translation. Pol I is much less processive than Pol III because its primary function in DNA replication is to create many short DNA regions rather than a few very long regions.

In eukaryotes, the low-processivity enzyme, Pol α, helps to initiate replication because it forms a complex with primase. In eukaryotes, leading strand synthesis is thought to be conducted by Pol ε; however, this view has recently been challenged, suggesting a role for Pol δ. Primer removal is completed Pol δ[while repair of DNA during replication is completed by Pol ε.

As DNA synthesis continues, the original DNA strands continue to unwind on each side of the bubble, forming a replication fork with two prongs. In bacteria, which have a single origin of replication on their circular chromosome, this process creates a "theta structure" (resembling the Greek letter theta: θ). In contrast, eukaryotes have longer linear chromosomes and initiate replication at multiple origins within these.

Replication fork

The replication fork is a structure that forms within the nucleus during DNA replication. It is created by helicases, which break the hydrogen bonds holding the two DNA strands together. The resulting structure has two branching "prongs", each one made up of a single strand of DNA.

These two strands serve as the template for the leading and lagging strands, which will be created as DNA polymerase matches complementary nucleotides to the templates; the templates may be properly referred to as the leading strand template and the lagging strand template.

DNA is always synthesized in the 5' to 3' direction. Since the leading and lagging strand templates are oriented in opposite directions at the replication fork, a major issue is how to achieve synthesis of nascent (new) lagging strand DNA, whose direction of synthesis is opposite to the direction of the growing replication fork.

Leading strand

The leading strand is the strand of nascent DNA which is being synthesized in the same direction as the growing replication fork. A polymerase "reads" the leading strand *template* and adds complementary nucleotides to the nascent leading strand on a continuous basis.

Lagging strand

The lagging strand is the strand of nascent DNA whose direction of synthesis is opposite to the direction of the growing replication fork. Because of its orientation, replication of the lagging strand is more complicated as compared to that of the leading strand. As a consequence, the DNA polymerase on this strand is seen to "lag behind" the other strand.

The lagging strand is synthesized in short, separated segments. On the lagging strand *template*, a primase "reads" the template DNA and initiates synthesis of a short complementary RNA primer. A DNA polymerase extends the primed segments, forming Okazaki fragments. The RNA primers are then removed and replaced with DNA, and the fragments of DNA are joined together by DNA ligases.

Dynamics at the replication fork

As helicase unwinds DNA at the replication fork, the DNA ahead is forced to rotate. This process results in a build-up of twists in the DNA ahead.[19] This build-up forms a torsional resistance that would eventually halt the progress of the replication fork. Topoisomerases are enzymes that temporarily break the strands of DNA, relieving the tension caused by unwinding the two strands of the DNA helix; topoisomerases (including DNA gyrase) achieve this by adding negative supercoils to the DNA helix. Bare single-stranded DNA tends to fold back on itself forming secondary structures; these structures can interfere with the movement of DNA

polymerase. To prevent this, single-strand binding proteins bind to the DNA until a second strand is synthesized, preventing secondary structure formation.

Clamp proteins form a sliding clamp around DNA, helping the DNA polymerase maintain contact with its template, thereby assisting with processivity. The inner face of the clamp enables DNA to be threaded through it. Once the polymerase reaches the end of the template or detects double-stranded DNA, the sliding clamp undergoes a conformational change that releases the DNA polymerase. Clamp-loading proteins are used to initially load the clamp, recognizing the junction between template and RNA primers.

Replication machinery

Replication machineries consist of factors involved in DNA replication and appearing on template ssDNAs. Replication machineries include primosotors are replication enzymes; DNA polymerase, DNA helicases, DNA clamps and DNA topoisomerases, and replication proteins; e.g. single-stranded DNA binding proteins (SSB). In the replication machineries these components coordinate. In most of the bacteria, all of the factors involved in DNA replication are located on replication forks and the complexes stay on the forks during DNA replication. These replication machineries are called **replisomes** or **DNA replicase systems**. These terms are generic terms for proteins located on replication forks. In eukaryotic and some bacterial cells the replisomes are not formed.

Since replication machineries do not move relatively to template DNAs such as factories, they are called a **replication factory**. In an alternative figure, DNA factories are similar to projectors and DNAs are like as cinematic films passing constantly into the projectors. In the replication factory model, after both DNA helicases for leading stands and lagging strands are loaded on the template DNAs, the helicases run along the DNAs into each other. The helicases remain associated for the remainder of replication process. Peter Meister et al. observed directly replication sites in budding yeast by monitoring green fluorescent protein(GFP)-tagged DNA polymerases α. They detected DNA replication of pairs of the tagged loci spaced apart symmetrically from a replication origin and found that the distance between the pairs decreased markedly by time. This finding suggests that the mechanism of DNA replication goes with DNA factories. That is, couples of replication factories are loaded on replication origins and the factories associated with each other. Also, template DNAs move into the factories, which bring

extrusion of the template ssDNAs and nascent DNAs. Meister's finding is the first direct evidence of replication factory model. Subsequent research has shown that DNA helicases form dimers in many eukaryotic cells and bacterial replication machineries stay in single intranuclear location during DNA synthesis.

The replication factories perform disentanglement of sister chromatids. The disentanglement is essential for distributing the chromatids into daughter cells after DNA replication. Because sister chromatids after DNA replication hold each other by **Cohesin** rings, there is the only chance for the disentanglement in DNA replication. Fixing of replication machineries as replication factories can improve the success rate of DNA replication. If replication forks move freely in chromosomes, catenation of nuclei is aggravated and impedes mitotic segregation.

TERMINATION

Eukaryotes initiate DNA replication at multiple points in the chromosome, so replication forks meet and terminate at many points in the chromosome; these are not known to be regulated in any particular way. Because eukaryotes have linear chromosomes, DNA replication is unable to reach the very end of the chromosomes, but ends at the telomereregion of repetitive DNA close to the ends. This shortens the telomere of the daughter DNA strand. Shortening of the telomeres is a normal process in somatic cells. As a result, cells can only divide a certain number of times before the DNA loss prevents further division. (This is known as the Hayflick limit.) Within the germ cell line, which passes DNA to the next generation, telomerase extends the repetitive sequences of the telomere region to prevent degradation. Telomerase can become mistakenly active in somatic cells, sometimes leading to cancer formation. Increased telomerase activity is one of the hallmarks of cancer.

Termination requires that the progress of the DNA replication fork must stop or be blocked. Termination at a specific locus, when it occurs, involves the interaction between two components: (1) a termination site sequence in the DNA, and (2) a protein which binds to this sequence to physically stop DNA replication. In various bacterial species, this is named the DNA replication terminus site-binding protein, or Ter protein.

Because bacteria have circular chromosomes, termination of replication occurs when the two replication forks meet each other on the opposite end of the parental chromosome.*E. coli* regulates this process through the use of termination sequences that, when bound by the Tus

protein, enable only one direction of replication fork to pass through. As a result, the replication forks are constrained to always meet within the termination region of the chromosome.

Regulation
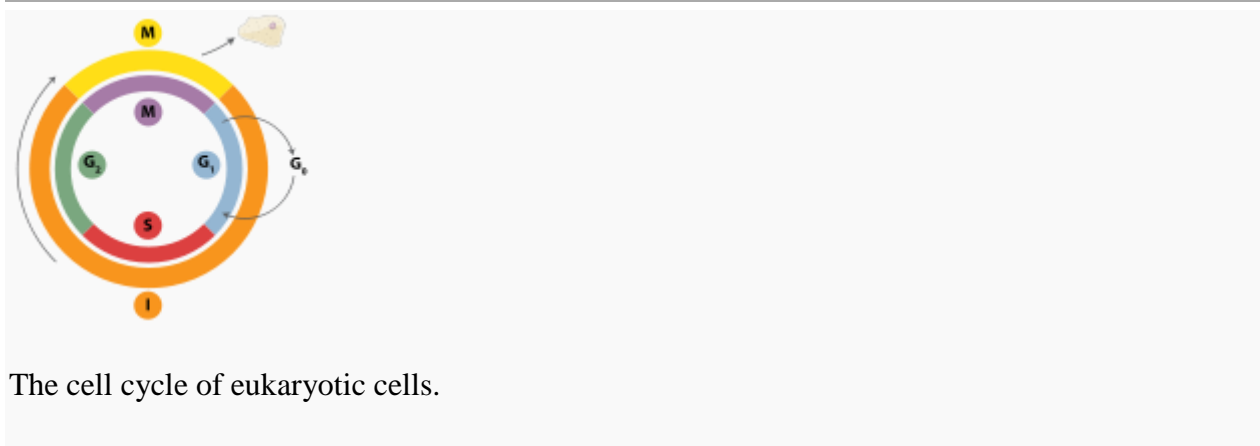
The cell cycle of eukaryotic cells.

Eukaryotes

Within eukaryotes, DNA replication is controlled within the context of the cell cycle. As the cell grows and divides, it progresses through stages in the cell cycle; DNA replication takes place during the S phase (synthesis phase). The progress of the eukaryotic cell through the cycle is controlled by cell cycle checkpoints. Progression through checkpoints is controlled through complex interactions between various proteins, including cyclins and cyclin-dependent kinases. Unlike bacteria, eukaryotic DNA replicates in the confines of the nucleus.

The G1/S checkpoint (or restriction checkpoint) regulates whether eukaryotic cells enter the process of DNA replication and subsequent division. Cells that do not proceed through this checkpoint remain in the G0 stage and do not replicate their DNA.

Replication of chloroplast and mitochondrial genomes occurs independently of the cell cycle, through the process of D-loop replication.
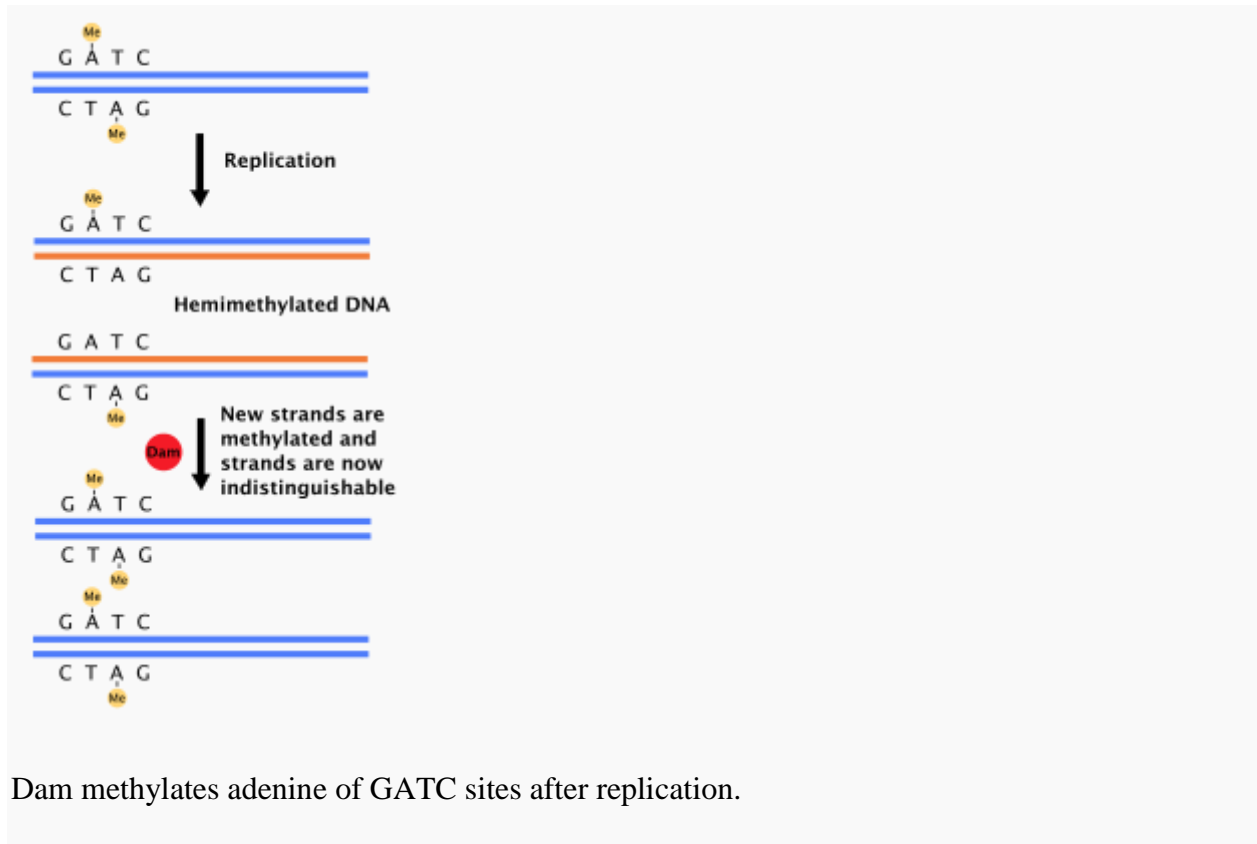
Replication focus

In vertebrate cells, replication sites concentrate into positions called **replication foci**. Replication sites can be detected by immunostaining daughter strands and replication enzymes and monitoring GFP-tagged replication factors. By these methods it is found that replication foci of varying size and positions appear in S phase of cell division and their number per nucleus is far smaller than the number of genomic replication forks.

P. Heun et al.(2001) tracked GFP-tagged replication foci in budding yeast cells and revealed that replication origins move constantly in G1 and S phase and the dynamicsdecreased significantly in S phase. Traditionally, replication sites were fixed on spatial structure of chromosomes by nuclear matrix or lamins. The Heun's results denied the traditional concepts, budding yeasts don't have lamins, and support that replication origins self-assemble and form replication foci.

By firing of replication origins, controlled spatially and temporally, the formation of replication foci is regulated. D. A. Jackson et al.(1998) revealed that neighboring origins fire simultaneously in mammalian cells. Spatial juxtaposition of replication sites brings **clustering** of replication forks. The clustering do **rescue of stalled replication forks** and favors normal progress of replication forks. Progress of replication forks is inhibited by many factors; collision with proteins or with complexes binding strongly on DNA, deficiency of dNTPs, nicks on template DNAs and so on. If replication forks stall and the remaining sequences from the stalled forks are not replicated, the daughter strands have nick obtained un-replicated sites. The un-replicated sites on one parent's strand hold the other strand together but not daughter strands. Therefore, the resulting sister chromatids cannot separate from each other and cannot divide into 2 daughter cells. When neighboring origins fire and a fork from one origin is stalled, fork from other origin access on an opposite direction of the stalled fork and duplicate the un-replicated sites. As other mechanism of the rescue there is application of **dormant replication origins** that excess origins don't fire in normal DNA replication.

Bacteria

Dam methylates adenine of GATC sites after replication.

Most bacteria do not go through a well-defined cell cycle but instead continuously copy their DNA; during rapid growth, this can result in the concurrent occurrence of multiple rounds of replication. In *E. coli*, the best-characterized bacteria, DNA replication is regulated through several mechanisms, including: the hemimethylation and sequestering of the origin sequence, the ratio of adenosine triphosphate (ATP) toadenosine diphosphate (ADP), and the levels of protein DnaA. All these control the binding of initiator proteins to the origin sequences.

Because *E. coli* methylates GATC DNA sequences, DNA synthesis results in hemimethylated sequences. This hemimethylated DNA is recognized by the protein SeqA, which binds and sequesters the origin sequence; in addition, DnaA (required for initiation of replication) binds less well to hemimethylated DNA. As a result, newly replicated origins are prevented from immediately initiating another round of DNA replication.

ATP builds up when the cell is in a rich medium, triggering DNA replication once the cell has reached a specific size. ATP competes with ADP to bind to DnaA, and the DnaA-ATP complex is able to initiate replication. A certain number of DnaA proteins are also required for DNA

replication — each time the origin is copied, the number of binding sites for DnaA doubles, requiring the synthesis of more DnaA to enable another initiation of replication.

POLYMERASE CHAIN REACTION

Researchers commonly replicate DNA *in vitro* using the polymerase chain reaction (PCR). PCR uses a pair of primers to span a target region in template DNA, and then polymerizes partner strands in each direction from these primers using a thermostable DNA polymerase. Repeating this process through multiple cycles amplifies the targeted DNA region. At the start of each cycle, the mixture of template and primers is heated, separating the newly synthesized molecule and template. Then, as the mixture cools, both of these become templates for annealing of new primers, and the polymerase extends from these. As a result, the number of copies of the target region doubles each round, increasing exponentially.