MOLECULAR BIOLOGY

DNA replication (Including mutation and repair)

Saleem Javed
Reader
Department of Biochemistry
Faculty of Science
Jamia Hamdard
New Delhi – 110062

27-Sep-2006

CONTENTS

Keywords

Replication, DNA repair, Mutagenesis, Mutation

DNA Replication in Prokaryotes

DNA molecules are synthesized through a process called replication. Replication begins with the unwinding of the double helix by an enzyme called helicase. The unwinding can start anywhere along the strand, and once begun, enzymes create two "replication forks" that continue to unzip the helix in both directions. After the DNA has started to unwind and straighten out, another enzyme called DNA polymerase goes to work. Its job is to match up the exposed nitrogenous bases with new nucleotides, which it finds in the surrounding nuclear fluid. These nucleotides hydrogen bond to the bases on two separated polymers, one at a time, according to the normal Watson-Crick pairing rules, the hydrogen bonded bases then get linked to the previous (3') base through a 3'-5' linkage with the help of DNA polymerases. When the entire DNA molecule has been separated and re-matched in this manner the result is two perfect copies of the original. Based on the Watson-Crick model of DNA, three methods of DNA replication were suggested. They are called conservative, dispersive and semi-conservative replication (Figure 1).

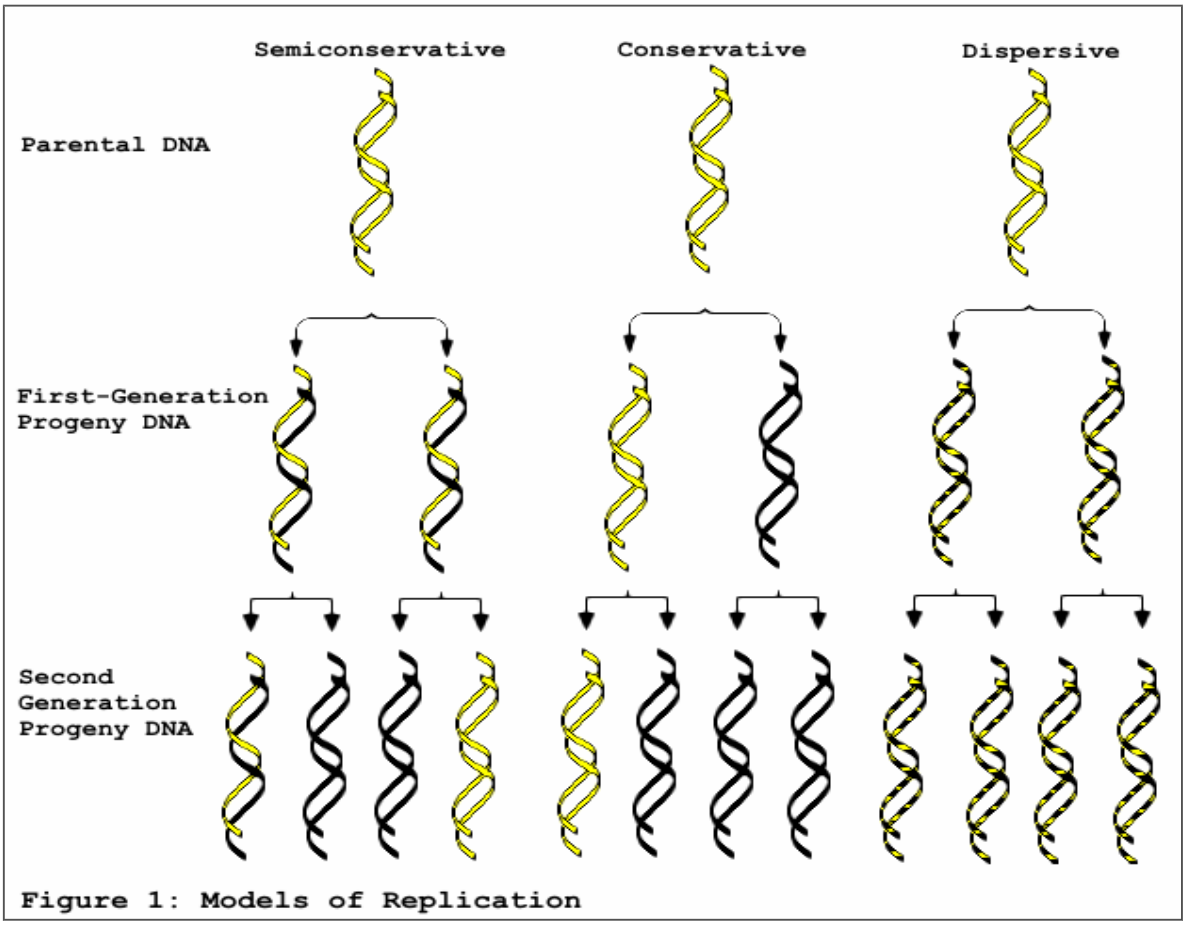

Figure 1: Models of Replication

Conservative replication

In conservative replication, the double helix remains completely intact during the replication process, and an entirely new double helix is formed without destroying the original copy.

Dispersive replication

The least likely candidate for DNA replication is dispersive replication. In this hypothesis, the DNA molecule is broken up into many small segments. Alongside each segment forms an appropriate complementary segment, and then all of the segments are joined back together into two molecules of DNA. The final product is two strands of DNA with small pieces from the original DNA and small pieces from the new DNA.

Semi-conservative replication

A second hypothesis is known as semi-conservative replication, whereby the hydrogen bonds between the complementary bases are broken and the DNA molecule "unzips" into two strands. Alongside each of the two strands forms a new strand with the appropriate base pairs. Thus the final copies are half original DNA and half new DNA, in contrast to conservative replication in which the copies are either completely original or completely new.

Experimental evidence for semi-conservative replication:

Watson and Crick proposed that when the time came for DNA to be replicated, the two strands of the molecule separated from each other but remained intact as each served as the template for the synthesis of a complementary strand. Matthew Meselson and Franklin Stahl did an experiment in 1957, to confirm which model truly represents the DNA replication.

Messelson and Stahl cultured bacteria in a medium containing heavy nitrogen *isotope (*^{15}N). They allowed the bacteria to grow in this culture media for several generations so that all of the nitrogen incorporated into the bases of DNA contained this form of nitrogen. The DNA, thus obtained, was subjected to equilibrium density centrifugation and bonded at a certain density. This DNA was “heavier” than the DNA obtained from bacteria grown in normal nitrogen. When the bacteria grown in heavy nitrogen was grown for one generation in medium containing normal nitrogen, the DNA obtained bonded midway between ‘heavy’ and normal DNA.

It was thus inferred from this experiment that the DNA replication is semi-conservative in nature i.e. one strand comes from the parental strands while the other is synthesized *ab inito* (Figure 2).

DNA Polymerases, other enzymes and proteins factors involved in replication

As mentioned earlier, DNA is the genetic material of the cell. It is the store house of all the information necessary for the normal functioning of cell; therefore it should be passed on to the next generation. The DNA thus has to replicate and the fidelity of the DNA sequence has to be maintained during replication. Replication of DNA is a complicated process. There are several reasons as to why replication is said to be a complicated problem, the first being the fact that the two DNA strands of the double stranded DNA molecule run in opposite directions (one strand runs in 5'- 3' while the other in 3'-5' direction.), therefore this polarity is to be maintained during replication process. Replication of a double stranded DNA molecule requires the double helix to unwind, and the unwinding requires energy. Moreover the single stranded structure formed after unwinding is likely to reanneal to the complementary bases either in the same strand or in the opposite strand. The process of replication requires several reactions therefore a number of enzymes are involved in the whole process. A strong and very efficient mechanism to repair the

error (misincorporation of bases) is also needed to replicate the enormous length of the genetic material.

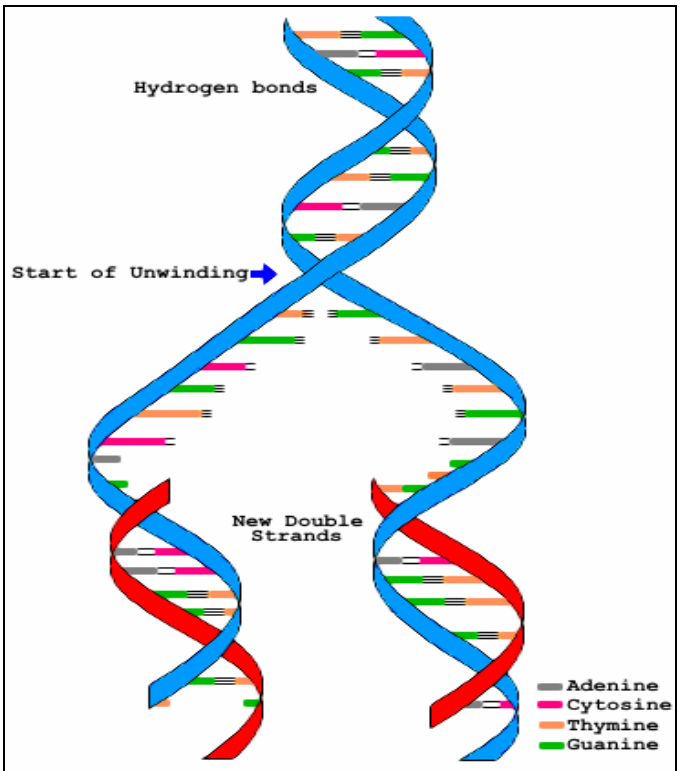

Figure 2: Semi-Consevative Model of DNA Replication

DNA Polymerases

In the 1950s, Arthur Kornberg found out that the energetics of the biochemical synthesis of a polynucleotide chain require deoxynucleoside triphosphates (dNTPs). He also worked out a very sensitive assay for DNA synthesis using radio labelled dNTPs. He chose *E. coli* for his studies and prepared large quantities of cells to obtain large amounts of purified enzyme. This led to the discovery of an enzyme in *E. coli* that could catalyse the synthesis of DNA, he called it DNA polymerase. We now know it as DNA polymerase I. We also now know that there are four other DNA polymerase enzymes in *E. coli.* Kornberg was awarded the 1959 noble prize in physiology jointly with Severo Ochoa, who had discovered an enzyme that catalyzed the synthesis of RNA. However, although both Kornberg's and Ochoa's enzymes are important, neither is the principal enzyme of either DNA or RNA synthesis.

DNA Polymerase I

DNA polymerase I is a single polypeptide chain with a molecular weight of 103 KD, having three distinct catalytic activities –

- **5′→3′ polymerase activity:** This is the predominant function of the DNA polymerase I involving formation of phosphordiester bond in a growing polynucleotide chain. This activity leads to the addition of a new nucleotide at the 3′−OH of an existing

oligonucleotide. This activity requires a single strand of DNA that can serve as a template, a primer to extend and form the complementary copy of the template DNA as DNA synthesis can not occur de-novo (DNA replication has an obligatory requirement for the primer and can not occur in its absence) The polymerization reaction can be summarized as follows-

[Polynucleotide]n + dNTP → [Polynucleotide]n+1 + PPi

Where PPi represents pyrophosphate released from dNTP.

The enzyme uses the four deoxynucleoside triphosphates (dNTPs) and Mg^{2+} to catalyze the template directed polymerization of nucleotides. There are certain important points that should be learned thoroughly, which are as under:

i. Polymerization occurs in only 5' to 3' direction.
ii. Polymerization requires a template to copy the complementary strand.
iii. Polymerization requires 4 dNTPs: dATP, dGTP, dCTP, dTTP
iv. Polymerization requires a pre-existing primer from which to extend.

- **5′→3′ exonuclease activity:** This activity is responsible for the removal of nucleotides from the 5′ end of the DNA chain. This activity is utilized for the removal RNA primer from the newly synthesized chain.

- **3′→5′ proofreading exonuclease activity:** The purpose of this activity is proofreading, that is used to remove any mismatched nucleotide. This activity is essential for maintaining the accuracy of replication. This proofreading activity is just opposite to the polymerase activity.

DNA polymerase I can be cleaved in to one smaller and one larger subunit by tryptic digestion. The larger subunit is called as Klenow fragment and it bears 5′→3′ polymerase activity and 3′→5′ proofreading exonuclease activity while the smaller subunit has 5′→3′ exonuclease activity.

Why don't DNA polymerases elongate chains in the 3' to 5' direction?

If a DNA polymerase could synthesize DNA in the 3' to 5' direction, then nucleotides would add to the primer terminus which has a 5'-triphosphate: (a) The 3'-OH of each incoming deoxyribonucleoside triphosphate would attack the 5'-triphosphate of the growing chain. (b) The editing removal of an incorrect 5'-terminal nucleoside triphosphate would prevent the DNA chain from being further extended because the primer terminus would now be a 5'-monophosphate, not a 5'-triphosphate.

Interestingly, E. coli temperature-sensitive (conditional lethal) mutants lacking the 5'-to-3' exonuclease activity of DNA polymerase I at non-permissive temperatures (40 degrees C) are not viable (remember, the polymerizing activity of DNA polymerase I is dispensable!). This enzymatic activity plays an essential role by removing RNA primers during replication. This process is similar to nick translation (here translation means "movement" not "protein synthesis"), the simultaneous polymerization in the 5'-to-3' direction from a nick with concurrent removal of nucleotides ahead by the 5'-to-3' exonuclease activity.

DNA Polymerase III

DNA polymerase III in contrast to DNA polymerase I consists of several subunits.

1. **α Subunit:** 130 KD subunit is responsible for DNA synthesis
2. **β Subunit:** 40 KD subunit has a probable role in template association.
3. **ε Subunit:** 25 KD subunit is responsible for proofreading activity.
4. **θ Subunit:** 10 KD subunit helps ε subunit.
5. **δ Subunit:** 32 KD subunit is responsible for the processivity of the enzyme molecule.
6. **γ Subunit:** 52 KD subunit helps δ subunit
7. **τ Subunit:** 71 KD subunit is responsible for dimmer formation.

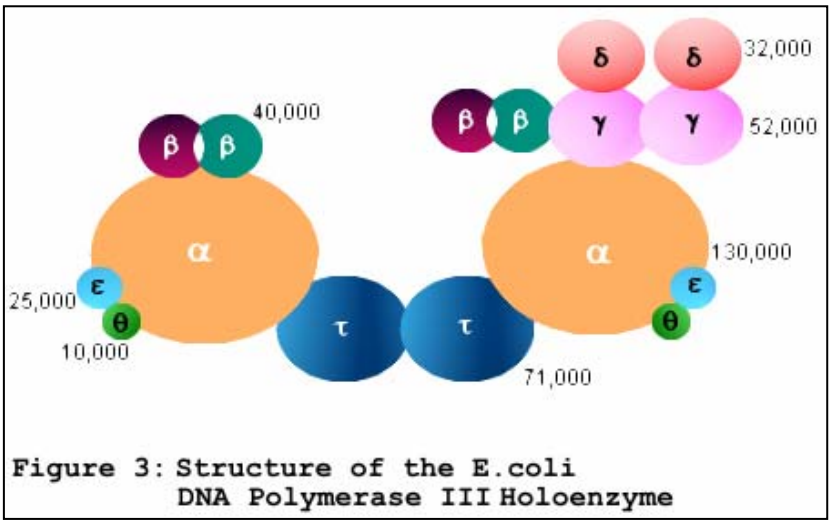

Figure 3: Structure of the E.coli DNA Polymerase III Holoenzyme

The catalytic core of DNA Polymerase III consists of α, ε and θ subunit (one each) and has a molecular mass of 130 KD. Two catalytic cores combines with the help of two τ subunits to give Pol III* which has a mass of 470 KD. Now two γ and two δ subunits combines to one of the catalytic core of the pole III* converting it to Pol III′ having a molecular mass of 750 KD. The joining of two β subunits to each of the catalytic cores converts the Pol III′ top holoenzyme with a molecular mass of 900 KD (Figure 3).

DNA Helicases

These proteins bind to the double stranded DNA and stimulate the separation of the two strands by unwinding DNA in front of opening replication fork. Helicases utilize ATP, makes single stranded cut and allows one strand to swivel freely around the other.

DNA single strand binding proteins

These proteins bind to the DNA as a tetramer and stabilize the single-stranded structure that is generated by the action of the helicases. Replication is everal times faster when these proteins are attached to the single-stranded DNA.

Primase

As DNA polymerases cannot start a growing chain from scratch rather they require a free 3' hydroxyl group. This requirement is fulfilled by the DNA-dependent RNA primase which synthesizes RNA primers at the initiation sites.

DNA Ligase

Enzymes that join DNA molecules together are called DNA ligases. DNA ligases reform phosphodiester linkages between adjacent 5'-phosphates and 3'-hydroxyls using an energy cofactor.Nicks occur in the DNA molecule being synthesized because the RNA primer is removed and synthesis proceeds in a discontinuous manner on the lagging strand. The final replication product does not have any nicks because DNA ligase forms a covalent phosphodiester linkage between 3'-hydroxyl and 5'-phosphate groups.

Mechanism of DNA Replication

The ability of a cell to replicate its DNA in a timely and faithful manner is fundamental for survival, but, despite decades of study, the structural and molecular basis for initiating DNA replication, and the degree to which these mechanisms have been conserved by evolution have been ill defined and hotly debated. DNA synthesis begins at a specific base sequence, termed the origin of replication. DNA replication is bidirectional, starting at the origin of replication and proceeding in both directions from that point. An eukaryotic chromosome may have multiple origins of replication and may replicate at many points along its length.

Individual nucleotides within DNA strands are covalently linked; the 3' carbon of one sugar is linked to the 5' phosphate of the adjacent sugar to form a 3'-5' phosphodiester linkage—a sugar-phosphate backbone. Therefore, each strand of DNA has a 5' carbon attached to a phosphate on one end (the 5' end) and a 3' carbon attached to a hydroxyl group on the other end (the 3' end). Though DNA replication is bidirectional with respect to the DNA helix, replication on a single strand always proceeds in a 5' to 3' direction (Figure 4).

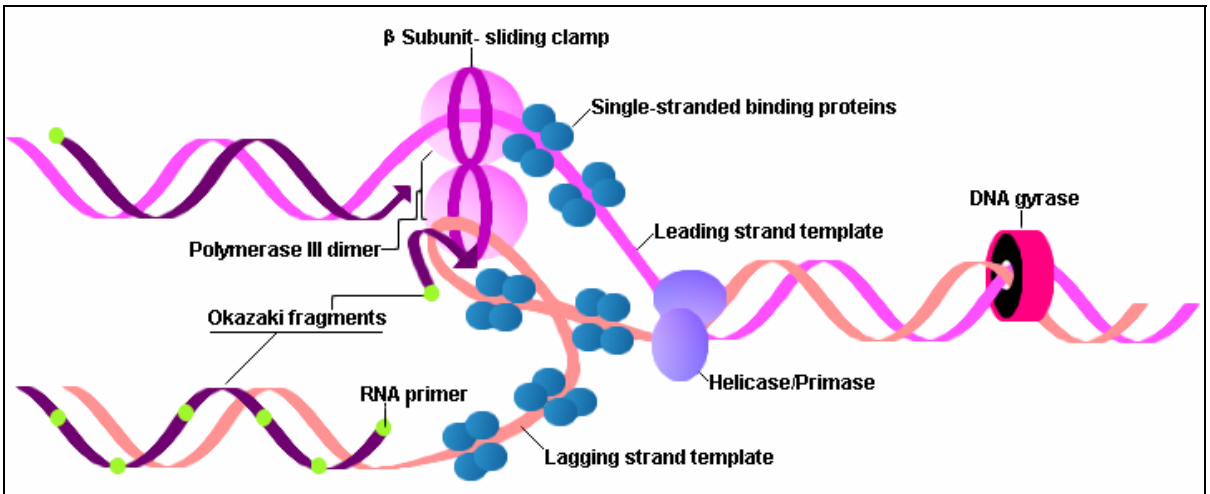

Figure:4 DNA Synthesis at a Single Replication Fork

The two DNA strands being synthesized at each replicating fork are extended in the same direction at the macro molecular level. The complementary strands of the DNA have reverse chemical polarity. One strand is extended in the 5' to 3' direction and the other in 3' to 5' direction. The DNA polymerase can only catalyze synthesis in the 5' to 3' direction. This posed a paradox. This crux was later solved by demonstrating that on one strand the synthesis is continuous, whereas on the other strand it is discontinuous. Both strands are extended in 5' to 3' direction. The synthesis on the strand that is continuous is called the leading strand, whereas the strand where extension is discontinuous, due to reverse polarity, is called the lagging strand. The lagging strand grows by the synthesis of small fragments known as the *Okazaki* or *nascent* fragments (Figure 5).

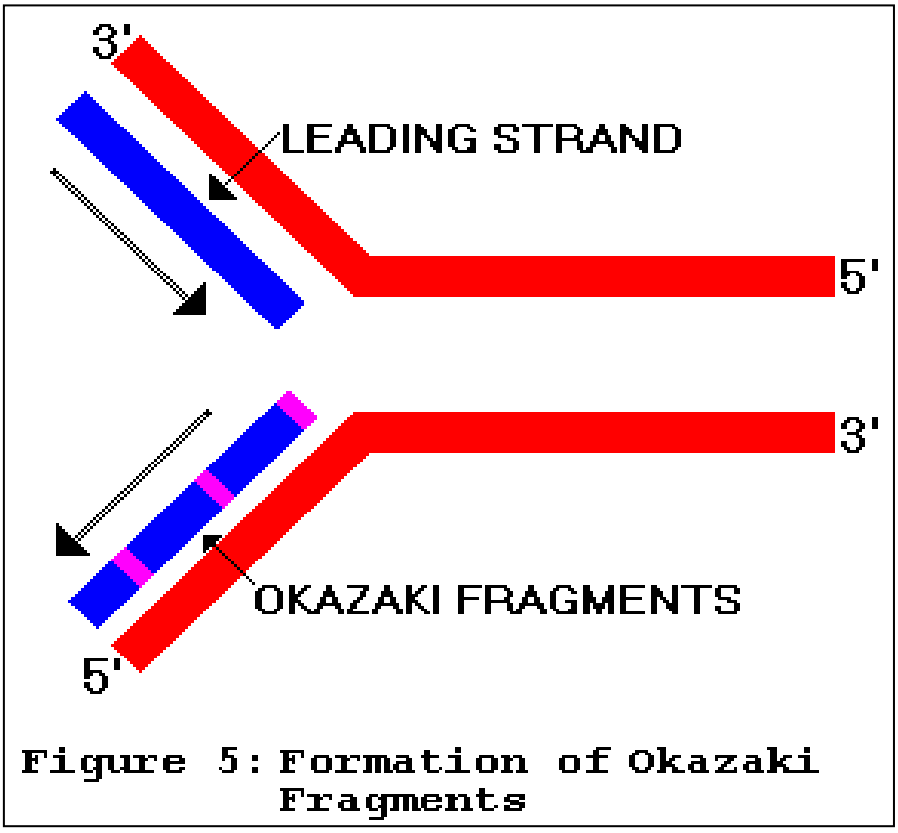

Figure 5: Formation of Okazaki Fragments

DNA replication begins with a partial unwinding of the double helix at an area known as the *replication fork***.** This unwinding is accomplished by an enzyme known as *DNA helicases* at the origin of replication (Figure 6).

These are enzymes that couple the unwinding of the double helical DNA to hydrolysis of ATP. E.coli has several different helicases and one in particular, *DnaB*, has the major role to unwind the DNA. DnaB moves along 5' to 3' template strand at the leading edge of the replication complex. Unwinding is facilitated by another protein known as SSB (*single stranded DNA binding protein*). SSB is called as *helix-destabilizing protein.* It binds to single stranded DNA and it lowers the melting temperature of the DNA. The enzyme *DNA gyrase* introduces the negative supercoiling. This unwound section appears under electron microscopes as a "bubble" and is thus known as a *replication bubble*.

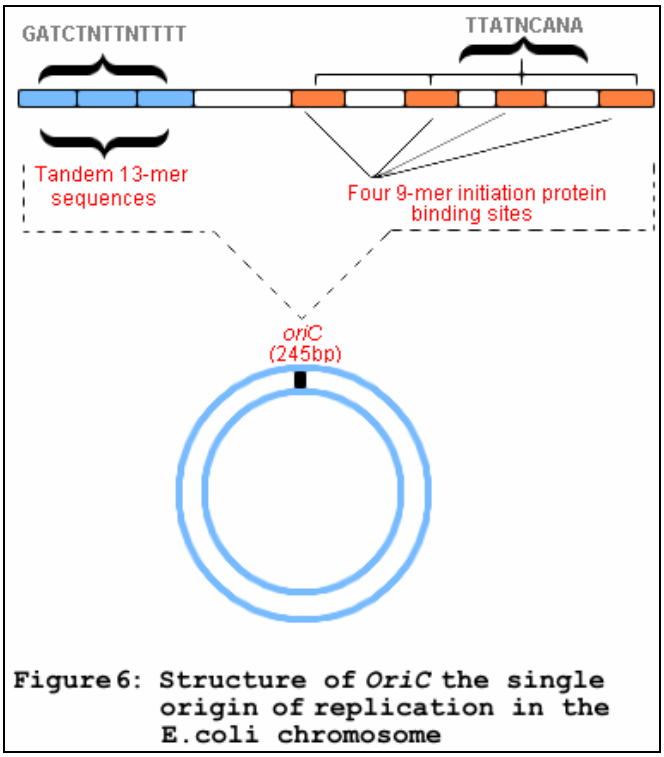

Figure 6: Structure of *OriC* the single origin of replication in the E.coli chromosome

The starting point for DNA polymerase is a short segment of RNA known as an *RNA primer*. The term "primer" is pinpointing of its role which is to "prime" or start DNA synthesis at certain points. The primer is "laid down" complementary to the DNA template by an enzyme known as *RNA polymerase* **or** *Primase***.**

The DNA polymerase (once it has reached its starting point as indicated by the primer) then adds nucleotides one by one in complementary manner, utilizing the principle of complementarity and reading from the template strand, A to T and G to C. DNA polymerase is described as being "template dependent" in that it will "read" the sequence of bases on the template strand and then "synthesize" the complementary strand. The template strand is always read in the 3' to 5' direction. The new DNA strand (since it is complementary) is synthesized in the 5' to 3' direction.

DNA polymerase catalyzes the reaction between the 5' phosphate on an incoming nucleotide and the free 3' OH on the growing polynucleotide. As a result, the new DNA strands *grow only in 5' to 3' direction.* Thus the orientation of the template strand is 3' to 5'.

Because the original DNA strands are complementary and run anti-parallel, only one new strand can begin at the 3' end of the template DNA and grow continuously as the point of replication (*the replication fork*) moves along the template DNA. The other strand must grow in the opposite direction because it is complementary, not identical to the template strand. The result of this side's *discontinuous* replication is the production of a series of short sections of new DNA called *Okazaki fragments* (after their discoverer, a Japanese researcher). When DNA polymerase III

reaches the primer, which has initiated an adjoining fragment, DNA pol I takes its place and, through the process of nick translation, removes the RNA and replaces it by DNA. The nick is sealed by DNA ligase. DNA ligase from E.coli requires NAD^+ as a co- reactant and the reaction mechanism involves the formation of an intermediate in which the adenyl group of NAD^+ is covalently attached to the enzyme. The 5'-phosphoryl terminus at the nick is activated by transfer of the adenyl group to form the DNA adenylate (Figure 7).

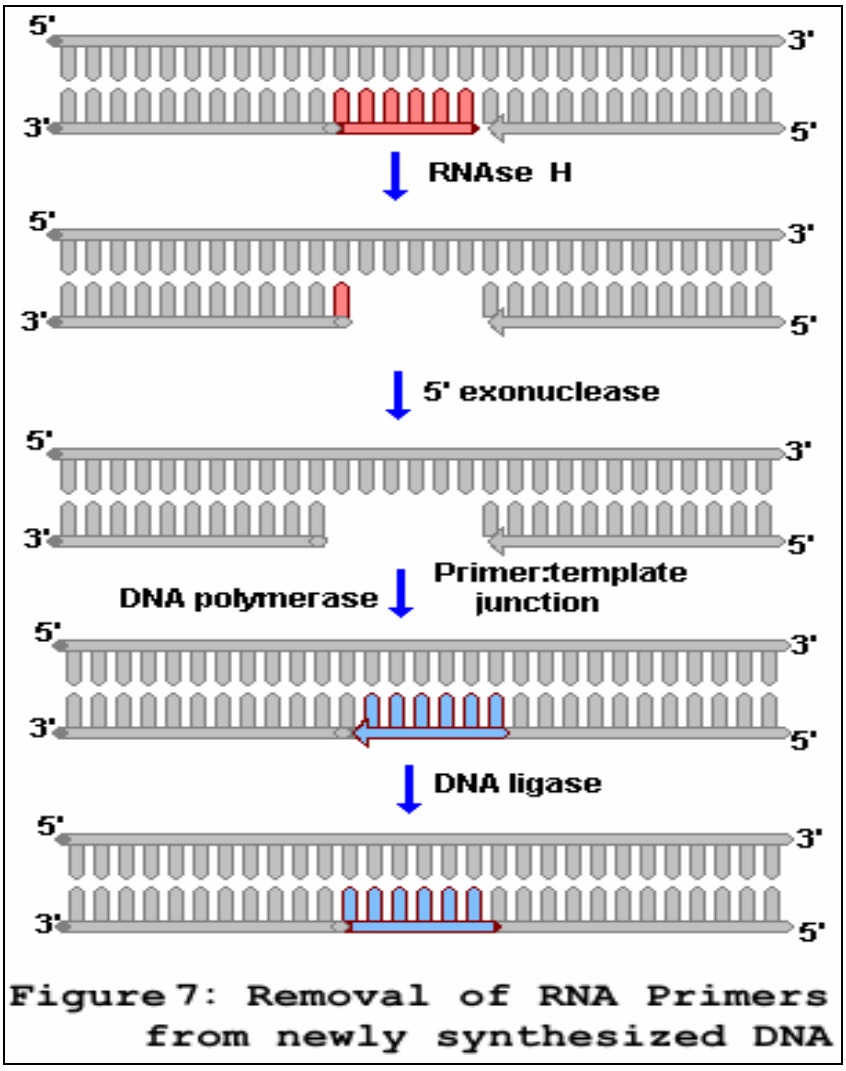

Figure 7: Removal of RNA Primers from newly synthesized DNA

Since each new strand is complementary to its old template strand, two identical new copies of the DNA double helix are produced during replication. In each new helix, one strand is the old template and the other is newly synthesized, a result described by saying that the replication is *semi-conservative*. Crick described the DNA replication process and the fitting together of two DNA strands as being like a hand in a glove. The hand and glove separate, a new hand forms inside the old glove, and a new glove forms around the old hand. As a result, two identical copies now exist.

There has been continuous efforts at the part of scientists throughout the world to uncover the mysteries behind the replication and recently researchers with the U.S. Department of Energy's

Lawrence Berkeley National Laboratory and the University of California at Berkeley have shown that the core machinery for initiating DNA replication is the same for all three domains of life - Archaea, Bacteria and Eukarya. The researchers have reported the identification of a helical substructure within a superfamily of proteins, called AAA+, as the molecular "initiator" of DNA replication in a bacteria, *Escherichia coli* (E. coli), and in a eukaryote, *Drosophila melanogaster,* the fruit fly. Taken with earlier research that identified AAA+ proteins at the heart of the DNA replication initiator in archaea organisms, these new findings indicate that DNA replication is an ancient event that evolved millions of years ago, prior to when Archae, Bacteria and Eukarya split into separate domains of life.

Inhibitors of DNA Replication

A number of compounds inhibits DNA replication process by acting at different targets.. They fall into various categories:

1. Inhibitors of Nucleotide Biosynthesis

These compounds interfere with the biosynthesis of nucleotides such as methotrexate.

2. Inhibitors that Interact with DNA Template

This class of inhibitors interferes with the replication process by binding to the DNA tempelate e.g. *Actinomycin D.* These compounds get intercalated between the two DNA strands and makes it difficult for the template to get replicated.

3. Nucleotide analogues

A number of nucleotide analogues function by blocking further chain growth at replication fork .The 2'3'-dideoxynucleotides, when incorporated on to the 3' hydroxyl end of a growing DNA chain causes termination of replication as they lack a 3' hydroxyl group due to which no further addition can occur.

4. Inhibitors that bind to replication proteins

Certain compounds inhibit the DNA replication by binding to proteins that are involved in replication and thus inhibit the process of DNA replication. Nalidixic Acid and Novobiocin bind to the A and B subunits, respectively, of E.coli DNA gyrase to inhibit its action and hence DNA replication.

Mutation

A mutation in a gene refers to a permanent change in the DNA sequence that makes up a gene. Mutations may range in size from a single DNA building block (DNA base) to a large segment of a chromosome. Most of these mutations are recognized as they lead to a change in phenotype of the organism. The process leading to the creation of a mutation is called mutagenesis and the agents causing mutation are called mutagens. Mutagenesis which does not involve a known mutagen is called Spontaneous mutagenesis and the mutations thus caused are called Spontaneous mutations. Similarly mutagenesis which involves a known mutagen is called induced mutagenesis and the mutations thus caused are called induced mutations. As the gene

consists of many individual units, and the specific changes in these units may lead to several mutant phenotypes. It is therefore pertinent to understand the nature of mutations for understanding the gene. Mutations are very important due to several reasons for example they may be deleterious or rarely advantageous to an organism, they are the major source of genetic variation which fuels evolutionary change. Mutations may also provide mutant phenotypes for genetic studies.

Types of Mutations

There are different basis used for the classification of mutations, as given below:

Transition and Transversion Mutations

These are the mutations where a single base is replaced by another base.

- **Transition:** If one purine (A or G) or pyrimidine (C or T) is replaced by the other purine or pyrimidine then this type of base substitution is called a transition**.**
- **Transversion:** If a purine is replaced by a pyrimidine or vice-versa, the substitution is called a transversion.

Deletion and Insertion

A deletion mutation refers to change in the number of DNA bases by removing a single base or a piece of DNA. Deletion in the DNA molecule may alter the functional behavior of the resulting protein.

An insertion mutation refers to change in the number of DNA bases by adding a single base or a piece of DNA. The protein product made by a mutant gene would be different from the original protein and may not function properly.

Frameshift mutation

Deletion or insertion of bases or a piece of DNA may lead to a change in the reading frame of genes**.** A reading frame consists of groups of 3 bases, a codon, that code for one amino acid. A frameshift mutation shifts the grouping of the bases and changes the code for amino acids. The resulting protein is usually nonfunctional.

Suppressor mutation

It is a mutation which counteracts the effects of another mutation. A suppressor is generally defined as a mutation that completely or partially restores the mutant phenotype of another mutation. A suppressor maps at a different site than the mutation it counteracts, it can be present either within the same gene or at a more distant locus. Different suppressors may act in different ways Suppressors can be broadly divided into two major groups, informational suppressors (either altered tRNAs or other components of the translational machinery) that act by misreading mRNAs and metabolic suppressors that act on genes common to the same pathway or to a single metabolic function.

Germinal and Somatic Mutations

There are two types of cells in eukaryotic organisms the germ cells and the somatic cells. When mutation alters a gene in a germ cell, then it is termed as germinal mutation, also called as hereditary mutations as they are passed on to the offspring. Germinal mutations may not affect the individuals in which they occur, but may result in genetic disorders in their offspring. Such mutations may be responsible for genetic disorders in which an affected child has a mutation in every cell, but has no family history of the disorder.

Somatic mutation is a mutation occurring in any cell that is not destined to become a germ cell; if the mutant cell continues to divide, the individual will come to contain a patch of tissue of genotype different from the cells of the rest of the body. These are the mutations that occur in the DNA of individual cells of non-germline tissues at some time during a person's life. There is another category of mutations that occur in a single cell within an early embryo. As all the cells divide during growth and development of an early embryo, the individual will have some cells with the mutation and some cells without the genetic change. This specific situation is called mosaicism.

Point Mutations

Point mutations refer to change in a single DNA nucleotide (Figure 8).

- **Missense mutation**: A missense mutation leads to the substitution of a different amino acid for the original one by changing one DNA base pair. Change in the base pair results in the change of codon of one amino acid for another in the protein made by that gene.

- **Nonsense mutation**: A nonsense mutation is also a change in one DNA base pair. Instead of substituting one amino acid for another, it results in a stop codon (TAA, TAG, or TGA) being inserted at the place of mutation before the end of the gene. This may results in the formation of a shortened protein that may function improperly or not at all.

- **Silent mutation**: In this case the new nucleotide alters the codon but so as to produce the same amino acid in the protein product (degeneracy of codon). Therefore there is no net effect on the biological function. Such mutations are said to be silent because they cause no change in their product and cannot be detected without sequencing the gene (or it's mRNA).

Morphological and Biochemical Mutations

Mutations that affect the external appearance of an individual are called Morphological Mutations while mutations that influence one specific step of an enzymatic pathway are known as Biochemical mutations. In case of biochemical mutants of bacteria the mutants are grown on a media supplemented with a specific nutrient, such mutants are called auxotrophic mutants. In strict genetic sense morphological mutants are the direct result of a mutation in a biochemical pathway such as albinism results from a mutation in the pathway which converts the amino acid tyrosine to the skin pigment melanin.

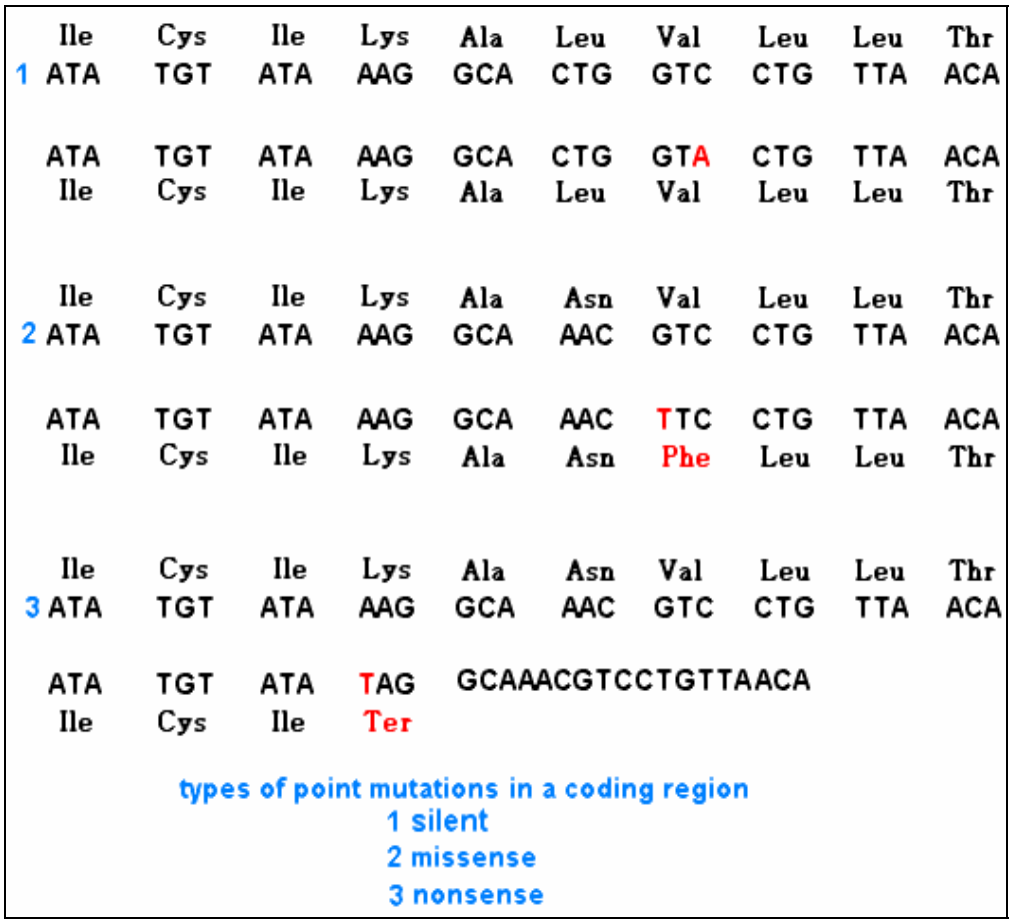

Figure 8: Point Mutations in DNA Sequences

Backward and Forward mutations

Forward mutation is the mutation from wild type allele to the detrimental allele. Backward mutations undo the forward mutation. Since there are many ways to destroy the function but fewer ways to undo that harm, backward mutations are normally rarer than forward mutations.

True Reversion

It is a type of suppression mutation that leads to the restoration of the natural genetic code

Dominant and Recessive Mutation

Alternate forms of the same gene are called alleles. Diploid organisms carry two copies of each gene, therefore either they may carry identical alleles, referred to as homozygous situation for a gene, or they may carry different alleles, referred to as heterozygous situation for that gene.

Dominant mutations have an altered gene product which acts antagonistically to the wild-type allele. Dominant mutations usually result in an altered molecular function (which is often inactive) and are characterized by a dominant phenotype. The phenotypic effects of a dominant mutation are observed in a heterozygous individual carrying one mutant and one normal allele

In contrast to this recessive mutation refers to a situation where both alleles must be mutant in order for the mutant phenotype to be observed; that is, the individual must be homozygous for the mutant allele to show the mutant phenotype.

Loss of Function and Gain of Function Mutations

As discussed above wild type alleles of a gene typically encode a product which is necessary for a specific biological function. If a mutation occurs in that allele, the specific biological function is lost, these mutations are called loss of function mutations. Recessive mutations inactivate the affected gene and lead to a loss of function. For instance, recessive mutations may remove part of or the entire gene from the chromosome, disrupt expression of the gene, or alter the structure of the encoded protein, thereby altering its function. In such mutations the extent to which the function is lost may vary. If the mutation results in the loss of function entirely, the mutation is called a null mutation, but sometimes some function still remains though it may not be at the level of the wild type allele; such mutations are called leaky mutations. Although it would be expected that most mutations would lead to a loss of function, it is possible that a new and important function could result from the mutation. In cases where mutation creates a new allele associated with a new function such mutations are called gain of function mutations. In such a case any heterozygous individual containing the new allele along with the original wild type allele will express the new allele, thus acting as a dominant mutation.

Conditional Mutations

In some cases mutations will show its effect only if the individual is placed in a specific environment, called as restrictive condition. Individual grown in the absence of restrictive conditions leads to the expression of the wild type phenotype. Such mutations are called Conditional mutations.

Lethal Mutations

Lethal mutations are the mutations which lead to the death of the individual. In such mutations death necessarily does not take place immediately, it may take some time even years, the only criterion is that if the longevity of an individual is significantly reduced in comparison to the expected longevity then the mutation is considered a lethal mutation.

Polar Mutations

Polar mutation is a mutation that influences the transcription or translation of part of the gene or Operon downstream of the mutant site. A number of mutations such as nonsense mutations and frameshift mutations can lead to polar mutations. This type of mutation may influence the expression of the other gene downstream to the affected gene, a good example of such a mutation is found in case of lac operon where a mutation in the beta galactosidase gene affects the expression of permease and acetylase genes also. This type of effect is called the polarity effect.

Spontaneous and Induced Mutations

Mutations arise spontaneously, though at low frequency owing to the chemical instability of purine and pyrimidine bases (Tautomerism, Depurination, Deamination) and to errors during DNA replication. Natural exposure of an organism to certain environmental factors, such as ultraviolet light and chemical mutagens, also can cause mutations. This class of mutation is termed spontaneous mutations. Spontaneous mutations are responsible for evolution. and they are also the source of diseases. The rate at which spontaneous mutation arise varies greatly, larger genes tend to mutate more frequently as they provide a larger target.

Mutations can also arise due to induction by exposure to ultraviolet rays and alpha, beta, gamma, and X radiation, by extreme changes in temperature, and by certain mutagenic chemicals such as nitrous acid, nitrogen mustard. Therefore it can be inferred that there are three general approaches used to generate mutations that is radiation, chemical and transposon insertion. These treatments can induce point mutations (changes in a single nucleotide) or deletions (loss of a chromosomal segment). The first induced mutations were created by treating Drosophila with X-rays.

Chemical mutagens work mostly by inducing point mutations. Two major classes of chemical mutagens are routinely used. These are alkylating agents and base analogs Ethylmethane sulfonate (EMS), a commonly used mutagen, alkylates guanine in DNA, forming O^6-ethylguanine. During subsequent DNA replication, O^6-ethylguanine directs incorporation of deoxythymidylate, not deoxycytidylate, resulting in formation of mutant cells in which a G·C base pair is replaced with an A·T base pair. In contrast to this, a base analog (Such as 5-bromouracil and 2-aminopurine) only mutate DNA when the analog is incorporated into replicating DNA.

Spontaneous alteration in DNA & DNA damage

The major source of alteration in DNA sequence during normal DNA metabolism giving rise to mismatches is the mispairing of bases. Replication fidelity in E. coli is affected by several parameters. The difference in the free energy for the stable pairing of a complementary base pairing in comparison to a non-complementary base pairing during replication is only 2-3 Kcal/mol, provided no other factors are involved, this will translate into a high rate of error . However the actual error frequency in newly replicated DNA is much less than the predicted error frequency, which can be attributed to the components of replication machinery including the complex of DNA polymerase and accessory proteins constituting the replisome.

Deamination of bases

Spontaneous deamination of cytosine, adenine and guanine under pH and temperature dependent reaction of DNA leads to the conversion of the affected base to uracil, hypoxanthine and xanthine respectively. This conversion sometimes gives rise to mutations.

Deamination of cytosine has special significance as the uracil is normally confined to RNA. Therefore it can be inferred that the use of thymine (instead of uracil) in DNA enables the cell to identify the deamination product of cytosine and help in selecting the uracil as an inappropriate base in DNA. Biologically conversion of cytosine to uracil by deamination is very important

with respect to the rate of mutation, as in case of cell where the cell is not capable of removing uracil from DNA, the rate of spontaneous mutations is increased.

The conversion of cytosine to uracil can also be enhanced by a number of chemicals, UV radiation and intercalating agents. For example nitrous acid and sodium bisulfite can promote the deamination of cytosine. Nitrous acid acts rather nonspecifically as it causes deamination of adenine and guanine in DNA, and also promotes the cross-linking of DNA strands. On the other hand sodium bisulfite promotes the deamination of cytosine exclusively in single stranded regions of DNA, sodium bisulfite for cytosine residues under highly specific experimental conditions.

Tautomeric shifts

The four bases in the DNA can spontaneously undergo transient rearrangement of bonding that is termed as tautomeric shift, forming a structural isomer (Figure 9). This change results in the alteration of base pairing properties of the base. The base can change from their normal amino (NH_2) form to imino (NH) form or from usual keto (C=0) to enol (C-OH) form. When either cytosine or adenine is in the imino form, it can form a mispair with the other two available hydrogen bonds, similarly if the guanine or thymine shift from the keto to the enol form, they can lead to the base pairing by forming three hydrogen bonds.

Mutagenecity testing: Correlation of mutagenecity and carcinogenicity, Ames testing

Rapid industrial development has resulted in the advent of new products as well as the pollutants as their byproducts. It becomes essential to determine the hazardous effect of these xenobiotics as potent mutagens or carcinogens. The available evidence have indicated that most of the mutagen known are also carcinogens. The use of the Ames test is based on the assumption that any substance that is mutagenic may also turn out to be a carcinogen. Although, in fact, some substances that cause cancer in laboratory animals do not give a positive Ames test (and vice-versa), the ease and low cost of the test make it invaluable for screening substances in our environment for possible carcinogenicity.

Directly assaying potential carcinogens by testing for their ability to form tumors in animals is difficult and expensive. However, in addition to causing tumors in animal cells, most carcinogens are mutagens. Based upon this insight, Bruce Ames and colleagues developed an indirect assay for potential carcinogens. Ames test is a simple and inexpensive procedure that uses a bacterial test organism to screen for mutagens (Figure 10).

The test organism is a histidine negative (his-) and biotin negative (bio-) auxotrophic strain of *Salmonella typhimurium* that will not grow on a medium deficient in histidine unless a back mutation to histidine positive (his+) has occurred. In this test the *his* mutants plated on minimal medium with a very small amount of histidine are exposed to the potential mutagen. Sometimes the compound in question itself may not be mutagenic but once it gets metabolized in the body the metabolite formed may be mutagenic. Therefore in such cases the compound is first incubated with liver enzymes and then the *his* mutants plated on minimal medium with a very small amount of histidine are exposed to the potential mutagen. The concentration of histidine used is limiting, so after the cells go through several cell divisions the histidine is used up and

the auxotrophs stop growing. However, if the potential mutagen induces His^+ revertants during the initial few cell divisions, then each of the resulting revertants will continue to divide and form a colony. The number of colonies produced is proportional to how efficiently the mutagen reverts the original mutation. By definition in the Ames test, a mutagen is any chemical agent which induces twice the number of mutants as occurred spontaneously, and thus is potentially carcinogenic for humans.

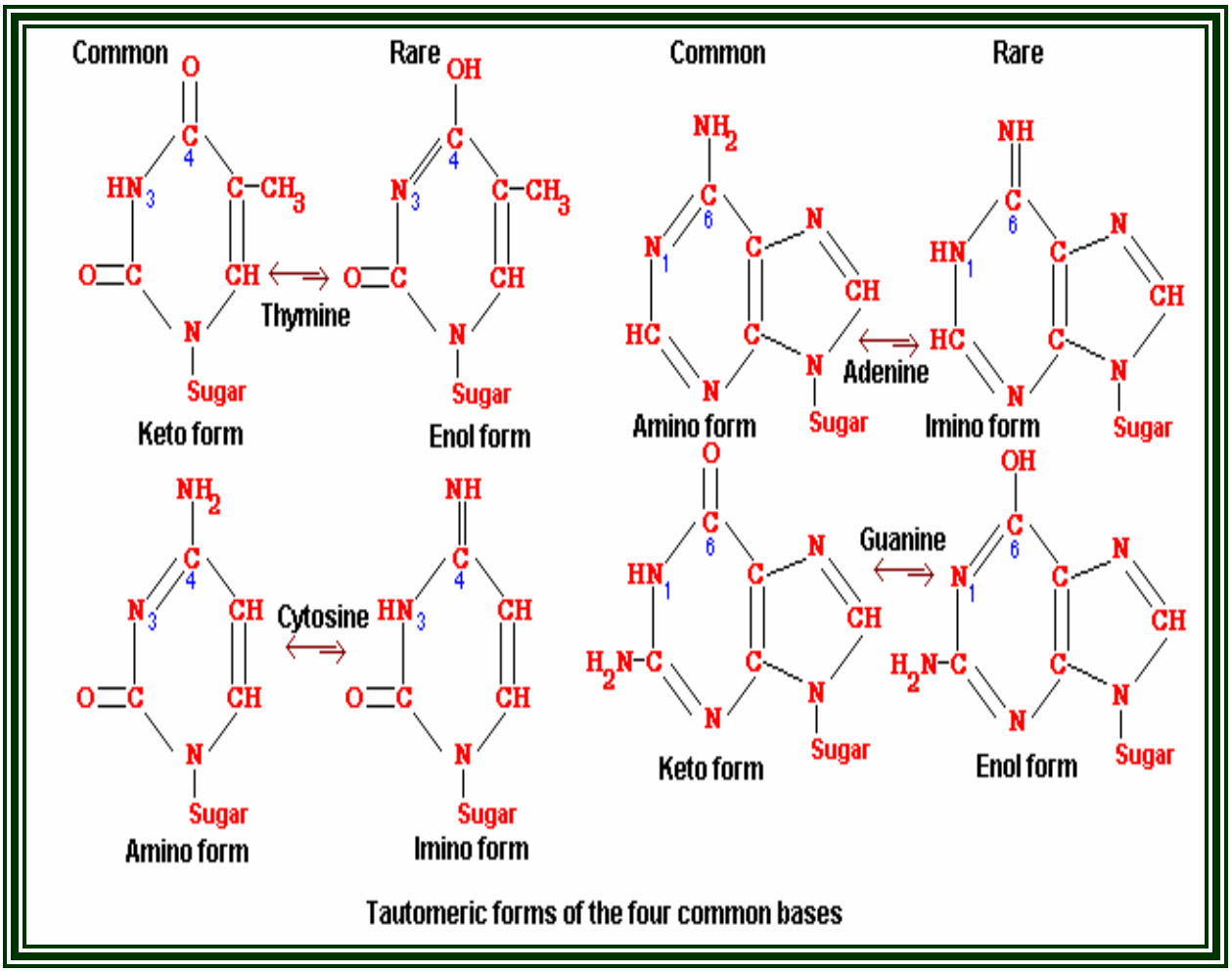

Figure 9: Tautomeric forms of bases

Replica Plating Technique

Fluctuation test carried out by Luria and Delbruck in 1943 proved statistically that bacteria acquire resistance against phages through mutation. Joshua and Esther Lederberg in 1952 provided direct proof of pre-existing mutant. They introduced replica plating technique and provided a direct method for demonstrating the undirected spontaneous origin of bacterial mutant; i.e., the mutants occurred independently of any selective agent or environment. The procedure given by them made it practical to examine large numbers of clones for a particular characteristic. This technique for example can be used to demonstrate the occurrence of phage-resistant mutants in a culture which was known to be phage- sensitive by using sufficiently large samples. Similarly the spontaneous appearance of antibiotic-resistant strains could also be demonstrated without previous exposure of the culture to the antibiotic. The replica plating method can also be used for isolating nutritional mutants. In essence the technique provides a

practical means for finding the one cell in a million (more or less) which has mutated. Thus the experiment conducted by Lederbergs demonstrated that mutation against the phage had its origin in spontaneous mutation. The growth of resistant colonies on replica plates arose from cells that were already present and were already resistant on the original non-selective plate prior to exposure to the selective agent, such as a lytic phage. Since then many other types of mutations have been found in bacteria, and it is now firmly established that bacteria have a hereditary system just like higher organisms.

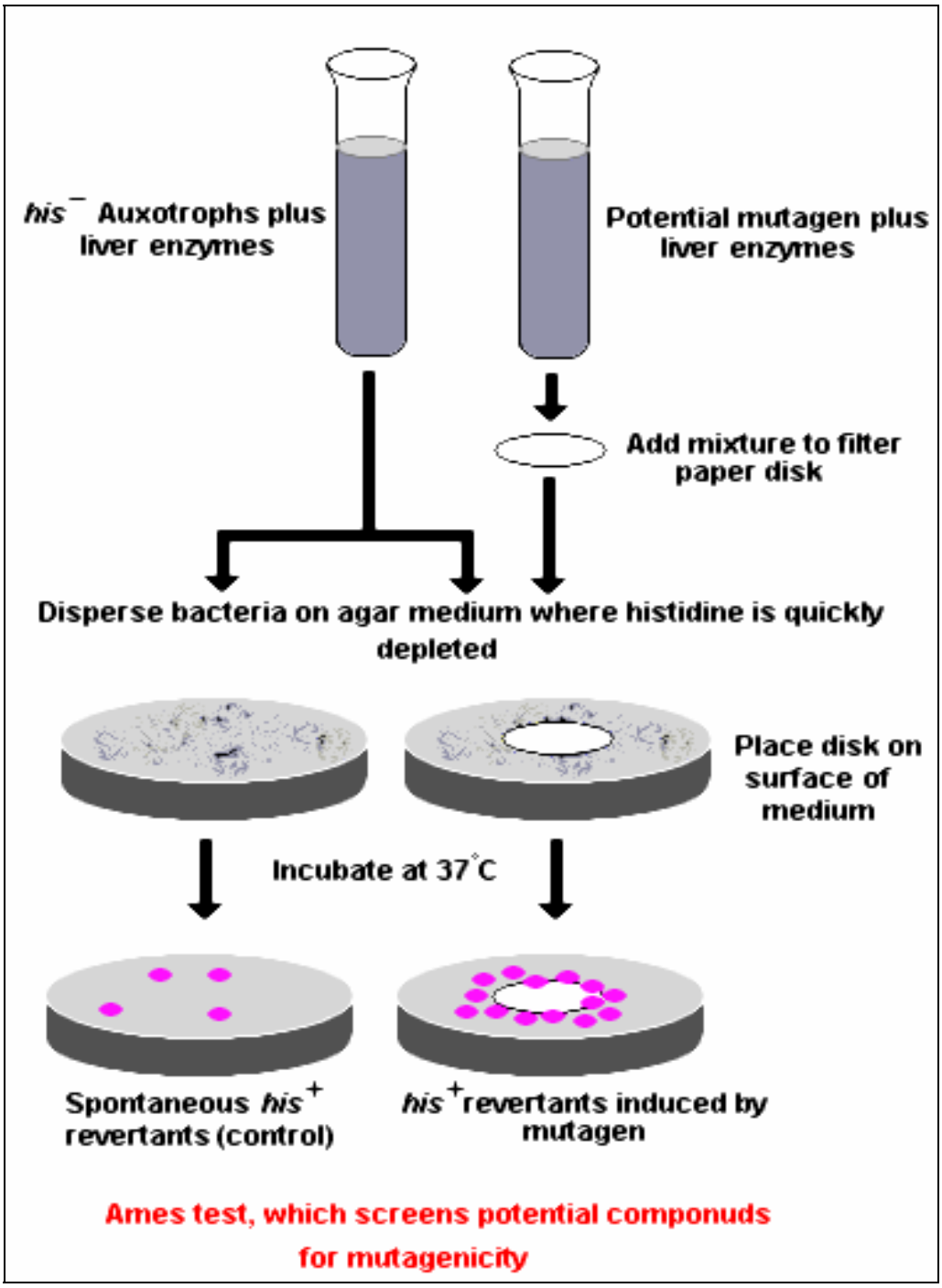

Figure 10: Ames test for screening of mutagenic compounds

In Vitro Mutagenesis (Random and Site Directed Mutagenesis)

In vitro mutagenesis has become a standard method for the functional analysis of the gene. With this technique we can create one mutant at a time or we can also make hundreds of mutants simultaneously using different methodology. Different methodologies used in case of *In Vitro* mutagenesis can be broadly classified as Random and site directed mutagenesis. Site directed mutagenesis technique is used to produce a specified mutation at a predetermined position in a DNA molecule, while the random mutagenesis can create mutations anywhere in the DNA molecule. Random mutagenesis can be helpful in the identification of the location and the boundaries of a particular function within a cloned gene. Random mutagenesis is the initial step and it provides only the simple identification of the functional region but does not give any idea about how things work at molecular level. This is useful as a quick strategy to narrow down the focus of attention from a larger region of DNA molecule to a smaller region to help in the detailed study of that region subsequently. There are a number of approaches that can be used for random mutagenesis such as inserting a linker into the plasmid randomly, damaging DNA with chemicals, altering the sequences within the sites for restriction enzymes etc.

Site Directed mutations on the other hand are introduced precisely where they are required to define the role of specific sequence. Site directed mutagenesis is of special significance in the field of protein engineering as it helps the scientists to introduce specific amino acid residues at selected sites. There are again several approaches for site directed mutagenesis but the best one is based on the use of a synthetic oligonucleotide. In this case the desired mutant sequence is simply build by designing an oligonucleotide that carries a mutation flanked by 10 to 15 nucleotides of wild type sequence. The mutation can be introduced by enzymatic extension of the of a mutagenic oligonucleotide, which is carried out by hybridizing the mutagenic oligonucleotide to its complementary sequence in a single stranded wild type DNA prepared from a plasmid thus forming a heteroduplex with mismatched nucleotides at the site of the mutation (Figure 11). The oligonucleotide serves as a primer and converts the single stranded DNA into double stranded form, once the primer has been extended completely a double stranded circular DNA is formed by ligating the ends of newly synthesized strand. In this heteroduplex one of the strand has wild type sequence while the other has the mutant sequence, it is then introduced into the *E coli* by transformation where either strand can be replicated. The bacterial colonies formed usually may either contain wild type or mutant.

DNA Repair: UV repair system in E. coli

As a major defense against environmental damage to cells DNA most organisms possess some capacity to repair their DNA. There are a number of systems in the cell that perform repair of damaged DNA inside the cell. Some of these are specific for a particular type of damage while other can handle a range of mutation types. These systems also differ in the degree to which they are able to restore the wild-type sequence. We can broadly classify the DNA repair mechanisms into 3 categories; one that can reverse the damage involves the simple enzymatic action which restores normal structure without breaking backbone (Photo reactivation, Ligation of single strand breaks). Second class of DNA repair mechanism removes the damage by cutting out and replacing a damaged or inappropriate base or section of nucleotides (Base excision repair, Mismatch repair, Nucleotide excision repair). The third category of DNA repair mechanism does not involve repair in real rather it is based on damage tolerance therefore it is a way of coping

with damage so that life can go on. Some of the important of these mechanisms are discussed below:

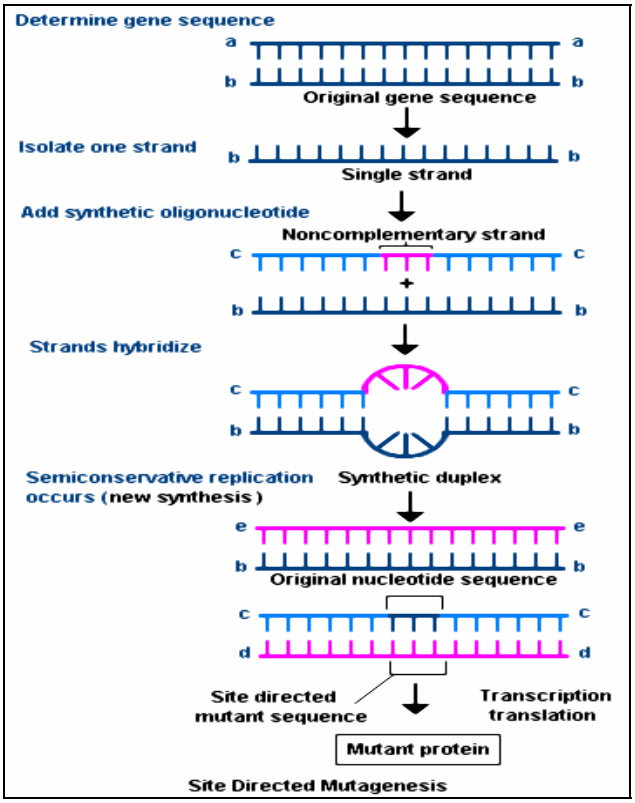

Figure 11: Oligonucleotide Directed Mutagenesis

Base excision repair

In base excision the damaged or inappropriate base is removed from its sugar linkage and is replaced. The excision is initiated by the action of specific class of DNA repair enzymes called glycosylase. These enzymes catalyze the hydrolysis of the N-glycosidic bond resulting in creation of apurinic or apyrimidinic site. An example base excision repair enzyme uracil glycosylase--enzyme which removes uracil from DNA. Uracil can occur in DNA if RNA primers are not removed in DNA replication or (more likely) if cytosine is deaminated. The enzyme recognizes uracil and cuts the glyscosyl linkage to deoxyribose. The sugar is then cleaved and a new base put in by DNA polymerase using the other strand as a template. Mutants lacking uracil glycosylase have elevated spontaneous mutation levels and are hyper-sensitive to killing and mutation by nitrous acid (Figure 12).

Nucleotide excision repair

Nucleotide excision repair also works on base dimers. It involves an endonucleolytic cleavage near the dimer, followed by a polymerase which cuts out the thymine dimer with a 5'-3' exonucleolytic activity. This system works on DNA damage which is "bulky" and creates a block to DNA replication and transcription. This polymerase (the polA gene product) simultaneously synthesizes an appropriate matching strand. This system and photo-reactivation share the property that they both function effectively when they are able to repair the dimer

before a replication fork comes to that dimer. The mechanism consists of cleavage of the DNA strand containing the damage by endonucleases on either side of damage followed by exonuclease removal of a short segment containing the damaged region. DNA polymerase can fill in the gap that result (Figure 13).

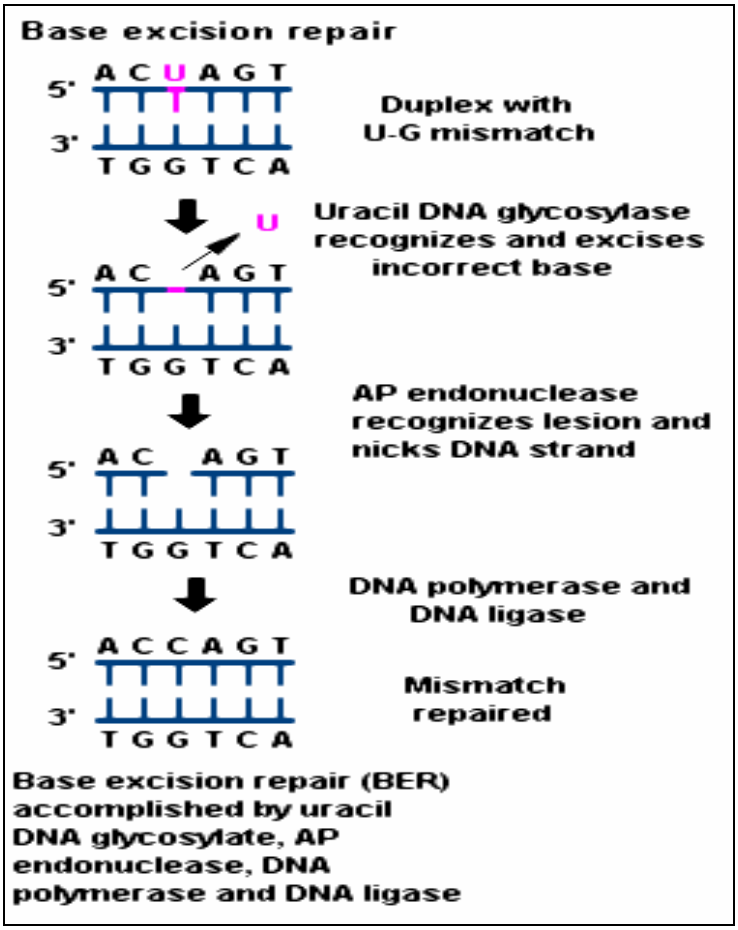

Figure 12: Base excision repair system

Mismatch repair

Both eukaryotic and prokaryotic cells are capable of repairing mismatched base pairs in their DNA. This process occurs after DNA replication as a last "spellcheck" on its accuracy. In E. coli, it adds another 100-1000-fold accuracy to replication. To repair mismatched bases, the system has to know which base is the correct one. In *E. coli*, this is achieved by a special methylase called the "Dam methylase", which can methylate all adenines that occur within (5') GATC sequences. Immediately after DNA replication, the template strand has been methylated, but the newly synthesized strand remains unmethylated thus helping to distinguish the template strand and the new strand. The incorrect nucleotide is removed as part of a short stretch and then the DNA polymerase gets a second try to put in the right sequence.

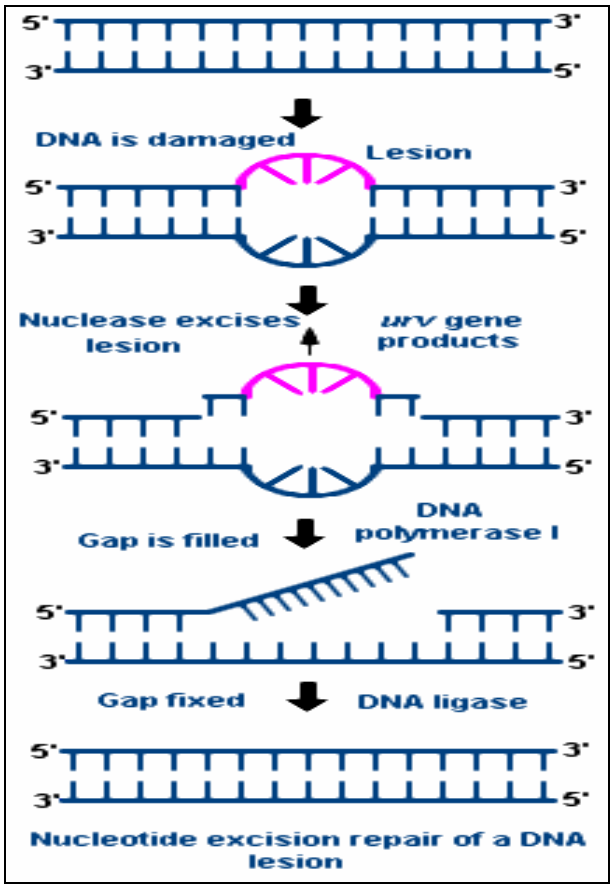

Figure 13: Nucleotide excision repair system

Photo reactivation

Photoreactivation is one of the simplest repair systems. It consists of a single enzyme which can split pyrimidine dimers (break the covalent bond) in presence of light. The reaction is catalyzed by photolyase enzyme. It is found in many bacteria, lower eukaryotes, insects, and plants (Figure 14).

Suggested Readings

1. Essential Genes (International Edition) by Benjamin Lewin. Pearson Education Inc. NJ.
2. DNA repair and Mutagenesis by Errol C. Friedberg, Graham C. Walker and Wolframe Siede. ASM Press, Washington D.C.
3. Genes VII by B. Lewin. Oxford University Press, New York.
4. Freifelder's Essentials of Molecular Biology (Fourth Edition) by George M. Malacinski. Narosa Publishing House Pvt Ltd, New Delhi
5. Microbiology: Concepts and Applications by M.J. Pelczar. McGraw Hill.

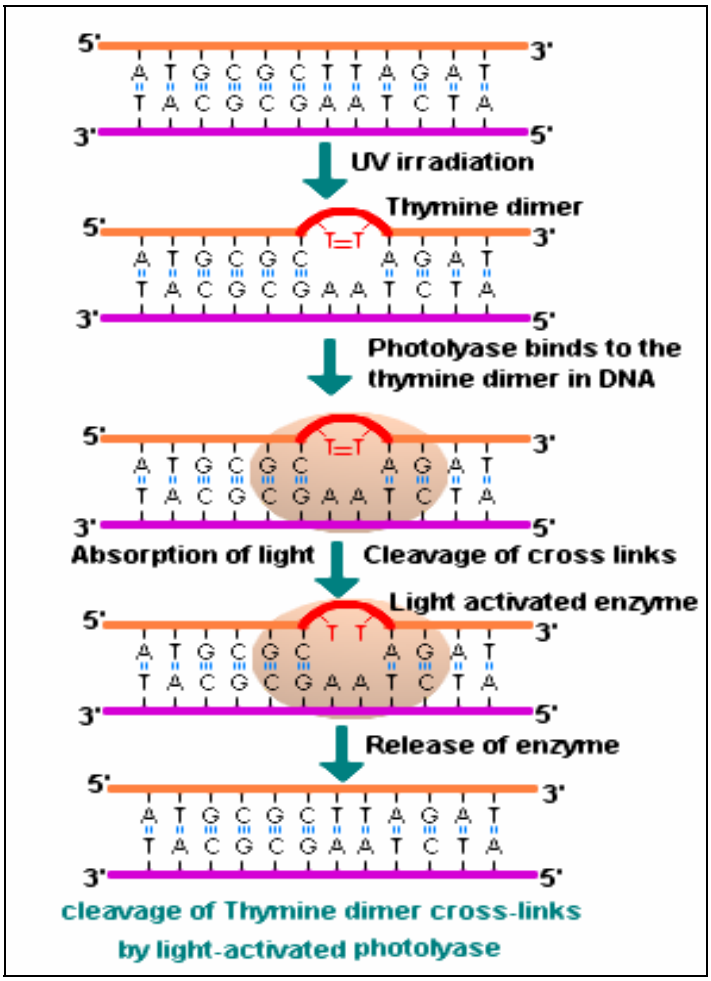

Figure 14: Photo reactivation repair system