COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH
UNIVERSITY GRANTS COMMISSION

LIFE SCIENCE CODE: 03

Unit – 3: FUNDAMENTAL PROCESSES

3.1.1 FUNDAMENTAL CONCEPT OF GENE:

Every cell contain nucleus. Nucleus is the place where maximum amount of DNA resides. Chromosome is the fully packaged structure of DNA. Each nucleus contains nuclear material, which condense to form chromatin fibre. Further condensation or packaging generates chromosome structure. Specific parts of a DNA that are responsible for any phenotypic character are known as gene. In more specifically those part of a DNA that can synthesize protein is known as gene. Also remember the part of DNA which helps in protein synthesis but cannot code any amino acid is also gene.

3.1.1 MOLECULAR BASIS OF GENE:

The whole nucleic acid sequence that is essential for the synthesis of a particular functional gene product (RNA or polypeptide) may referred as **gene.** A gene also includes the DNA sequences that required for synthesis of a particular RNA transcript. Most of the DNA sequences transcribed into mRNAs, that can produce proteins, some other DNA sequences are transcribed into RNAs other than mRNAs that do not encode for any proteins. For example tRNA, rRNA, snRNA, hnRNA, miRNA etc. The DNA parts which transcribed into these above mentioned RNAs can cause specific phenotypes when mutated that is these RNAs are indirectly involve in functional protein synthesis these can also be referred to as genes, even though the final products of these genes are not proteins. Functions various types of RNAs are given in the table mentioned below.

RNA	FUNCTION
mRNA	Protein synthesis
rRNA	Protein synthesis
tRNA	Protein synthesis
snRNA	mRNA splicing
snoRNAs	Pre-rRNA processing and rRNA modification
miRNA	Regulations of gene expression
U7 snRNA	Histone mRNA 3′ processing
Xist	X-chromosome inactivation
7SK	Transcription control
RNAse P	tRNA 5′ processing
7SL RNA	Protein secretion (component of signal recognition particle, SRP)
RNAse MRP	rRNA processing
Telomerase RNA	Templates for addition of telomeres
Vault RNA	Components of vault ribonucleoproteins
hY1, hY3, hY4, hY5	Components of Ro ribonucleoproteins
H19	Function unknown

3.2 DNA REPLICATION:

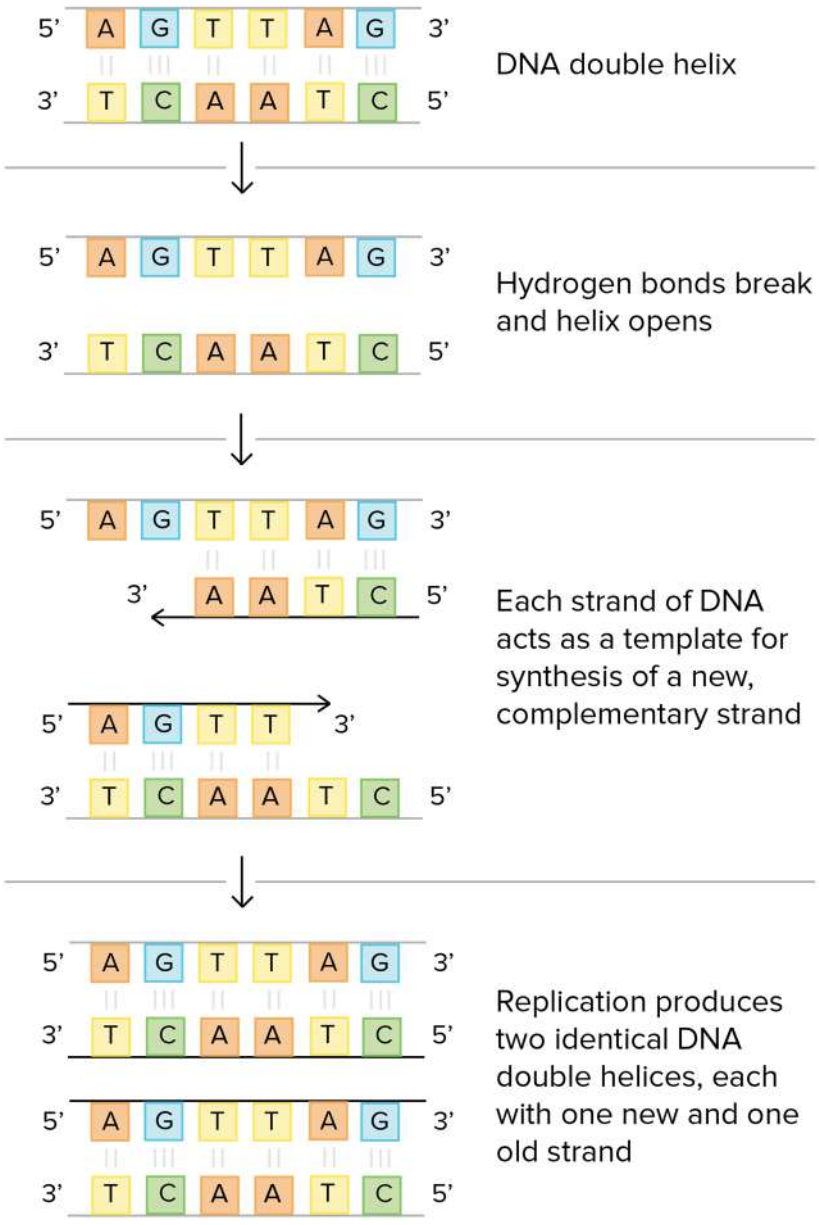
DNA replication is the process of duplication of the DNA content. During S-phase of cell cycle DNA replication takes place.

Eukaryotic DNA is a double helix structure and duplicate in semi-conservative manner, according to Watson and Crick.

3.2.1 The basic idea

DNA replication is **semiconservative**, meaning that each strand in the DNA double helix acts as a template for the synthesis of a new, complementary strand.

This process takes us from one starting molecule to two "daughter" molecules, with each newly formed double helix containing one new and one old strand.



Schematic of Watson and Crick's basic model of DNA replication.

1. DNA double helix.
2. Hydrogen bonds break and helix opens.
3. Each strand of DNA acts as a template for synthesis of a new, complementary strand.

4. Replication produces two identical DNA double helices, each with one new and one old strand.

In a sense, that's all there is to DNA replication! But what's actually most interesting about this process is how it's carried out in a cell.

Cells need to copy their DNA very quickly, and with very few errors (or risk problem such as cancer). To do so, they use a variety of enzymes and proteins, which work together to make sure DNA replication is performed smoothly and accurately.

Each steps of DNA replication **(Prokaryotic)** process is discussed below.

3.2.2 DNA replication fork formation:

During replication first the intertwined double helix strand unwound to make the bases available for base pairing with complement nucleotides. This unwinding of DNA double helix strand is mediated by **helicase** enzyme at a specific region of DNA strand at the very beginning of the replication process. This specific site is known as the origin of replication or simply the origin. The nucleotide sequences of origin are species specific. This process of unwinding is mediated by means of denaturation. The denatured area forms a special structure known as **replication bubble.** In case of E.coli bacteria the origine of replication is also known as **OriC** region.

A special initiator protein DnaA (encoded by dnaA gene)bind at the OriC region. **DNA helicase** enzyme (enocoded by dnaB gene) now recruited and loaded on DNA by DNA helicase loader protein DnaC (encoded by dnaC gene). **Helicase starts untwisting the DNA strand in both direction from OriC region** by breaking the H-bond between the Nitrogen bases with the help of energy released by breaking of ATP molecule. Another very important enzyme **Topoisomerase-I** helps to release the torsional stress generated due to unwinding of DNA strands.

Each DNA helicase enzyme activate a special enzyme **Primase** (encoded by dnaG gene). **DNA polymerase** (the enzyme which add the complement base to template strand) **cannot initiate** the synthesis of DNA molecules, it requires a pre-existing strand. **DNA Primase or Primase** is a modified form of RNA polymerase enzyme, add a short strand of nucleotides **(5-10 nucleotides)** or RNA strand. DNA polymerase enzyme can add nucleotides to the **free hydroxyl** group of the **3′ end.** This RNA primer removed later and replaced with DNA at the end of DNA replication.

Template- A template strand is the one on which new strand is synthesized by means of complementary base pairing.

Primer- Primer is the short strand or segment of nucleotides bound to the template strand so that DNA polymerase can add DNA nucleotides.

Concept about DNA Polymerase: Scientist Kornberg recognize and isolated an enzyme that was capable ton synthesis DNA in E.coli. This enzyme originally was named as Kornberg enzyme, but is now known as DNA Polymerse-I or DNA Pol-I. Later studies show that all type of DNA Polymerase enzymes from prokaryotes to eukaryotes can catalyse the polymerisation of nucleotide precursor or deoxyribo nucleotide phosphate (dNTPs) into a DNA chain. DNA Polymerase enzyme can form a phosphodiester bond between the free 3′-OH group of the deoxyribose on the last nucleotide and the 5′-phosphate group of the of the dNTP molecule. The energy require to add nucleotides is generated from the release of two of three phosphate molecule from dNTP precursor.

Nucleotides are added very rapidly. For example in case of E.coli the rate is 850 nucleotides per second whereas 60-90 nucleotides per second in case of human. The direction of DNA polymerase of adding nucleotides or the direction of the synthesis of DNA is from 5′ to 3′ end.

Till now there are 5 types of DNA polymerase have been discovered, DNA Pol-I to DNA Pol-V. DNA polymerase-I and DNA polymerase-III are necessary for DNA replication. DNA polymerase-I, DNA Pol-II, DNA Pol-IV, DNA Pol-V are necessary for DNA repair mechanism.

3.2.3 SEMICONSERVATIVE MODE OF DNA REPLICATION:

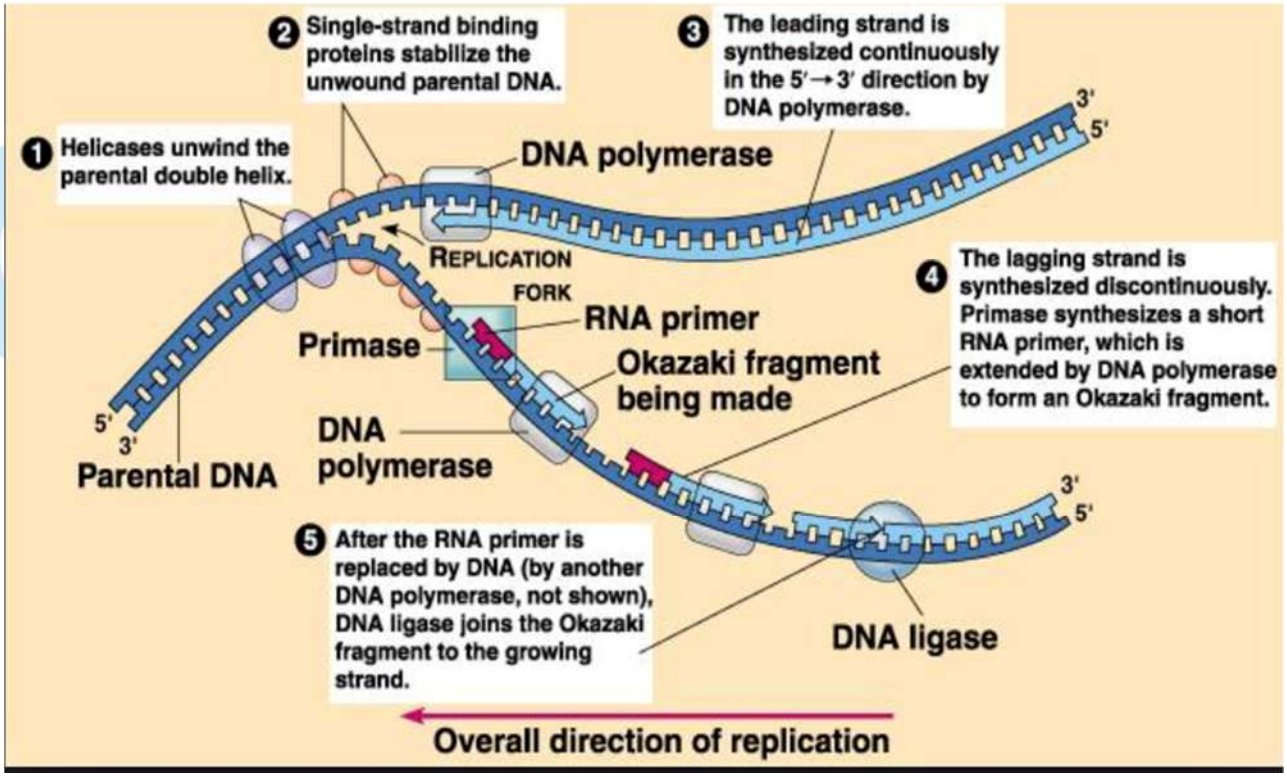
After the untwisting of DNA strand a special protein **Single strand DNA binding (SSB) proteins** bind to each single stranded DNA to make them stabilize and prevent further reforming of DNA double strand by complementary base pairing (reannealing). Each RNA primer synthesized earlier by the Primase enzyme is extended by addition of DNA nucleotides by the help of DNA polymerase-III. During this process DNA pol enzyme displace SSB proteins as they move along the template strands. The addition of nucleotides on template strand to generate new DNA strand follow the complementary base pairing rule that is A binds to T and G binds with C.

DNA polymerase-III enzyme can synthesize DNA in the direction of 5′-3′ direction but the two DNA strand direction are opposite. To maintain the 5′-3′ direction of DNA synthesis on each template DNA is made on opposite direction on each strand. The new strand which synthesized on the same direction of the movement of replication fork is the **Leading strand** and the strand which synthesize on the opposite direction of the movement of replication fork is the **Lagging strand.** In leading strand only a single RNA primer is needed as the DNA synthesis in this strand is continuous because DNA synthesis continues in the direction of the replication fork movement. Whereas in lagging strand needs a series of RNA primer as synthesis of DNA is in the opposite direction of the movement of replication fork. DNA synthesis in lagging strand stops after a certain time due to unavailability of free template. To start again lagging strand need RNA primer, and that is how in lagging strand DNA synthesis take place. So the synthesis of DNA in leading strand is continuous and DNA synthesis in

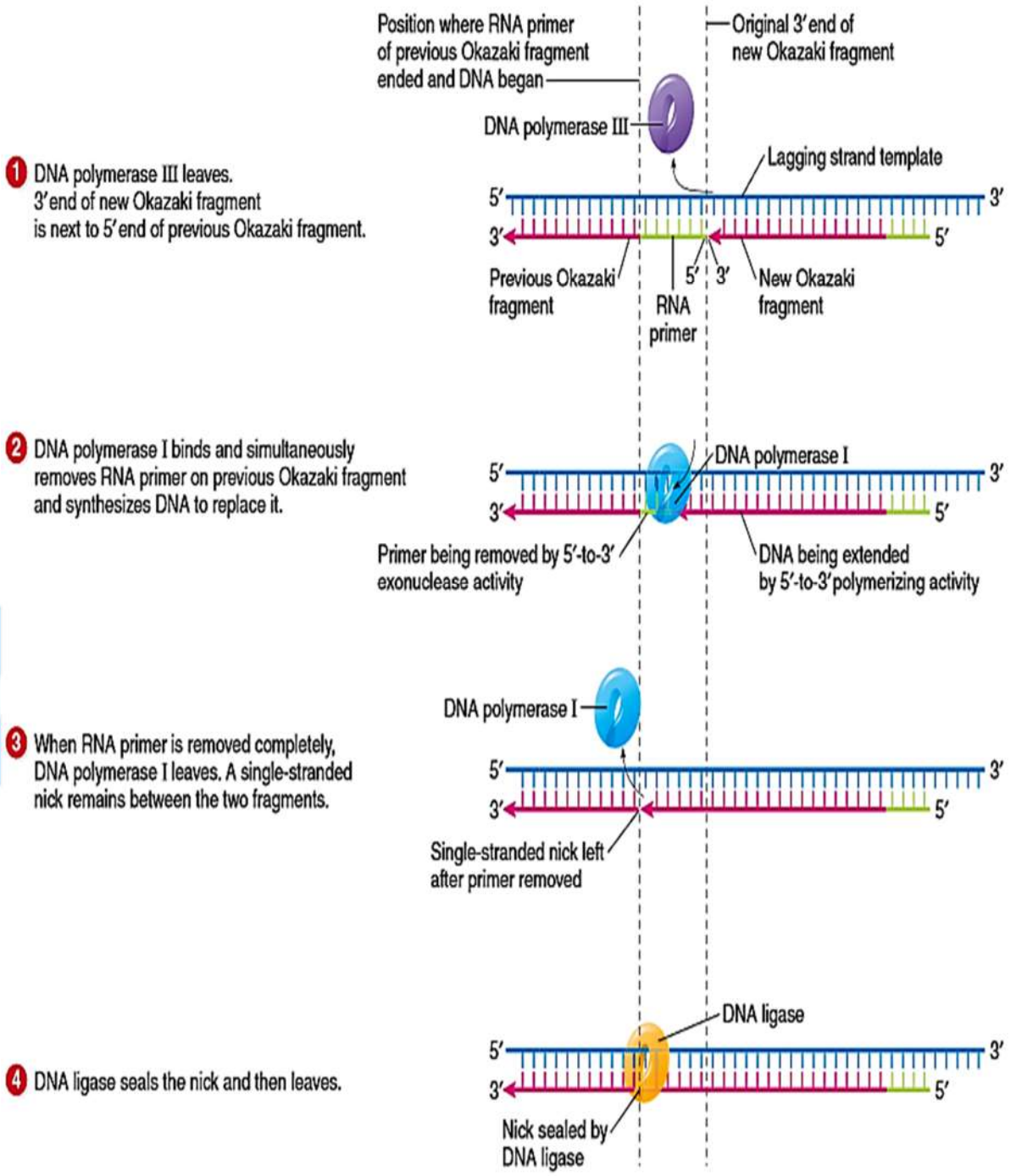
lagging strand is discontinuous. That is why DNA replication as a whole is a semi discontinuous process.

3.2.3.1 Helicase enzyme untwist more DNA portion along the length of chromosome that leads to the movement of replication fork along the length of chromosome. DNA gyrase another form of topoisomerase enzyme release the tension produced in the DNA ahead of the replication fork.

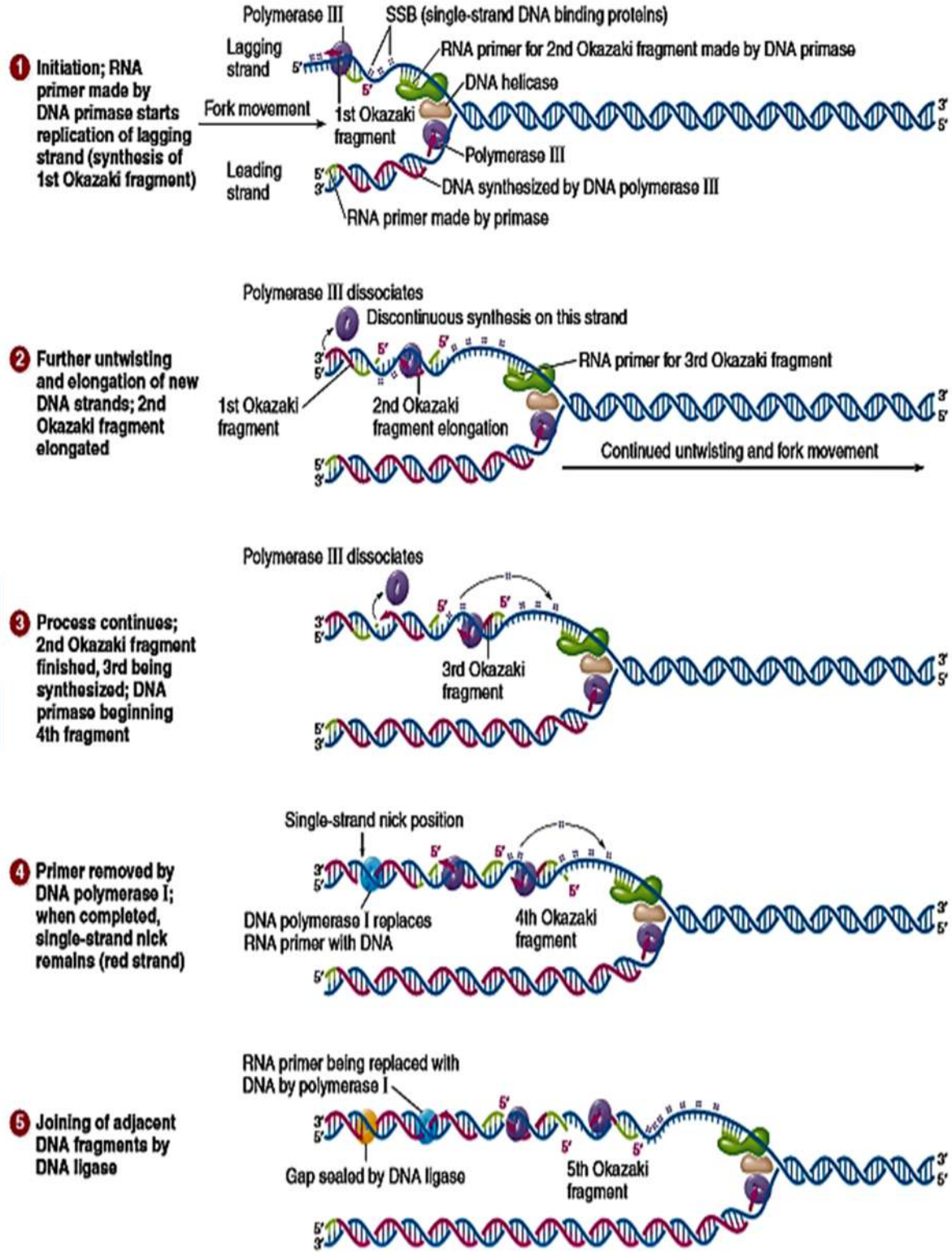
The so produced DNA fragments of lagging strand produced are collectively called **Okazaki Fragments.**



3.2.3.2 Addition of Okazaki fragments: So now helicase continues to untwisting the DNA strand, DNA strand synthesized continuously on the leading strand template, and okazaki fragments continuously generating on lagging strand with the help of new RNA primer. Addition of okazaki fragments is a very interesting step of DNA replication. Two enzyme helps in this process, **DNA polymerase-I and DNA ligase** enzyme. After the completion of one Okazaki fragment synthesis DNA polymerase-III enzyme leaves the newer DNA fragment and DNA polymerase-I add to it. Let's consider two adjuscent okazaki fragments. We can see that 3′end of the newly synthesized DN fragment is adjuscent to the 5′end of the RNA primer of the previously synthesized DNA fragment. The **DNA polymerase-I** digest the RNA primer between the joining point of two Okazaki fragments. Now the RNA primer free 5′ end of the DNA fragment is ready to bind with 3′ end of the adjuscent DNA fragment. Now the DNA **ligase** enzyme binds these two ends of DNA fragment. In this way these two enzymes replace RNA primer with DNA strand throughout the length of lagging strand. DNA ligase catalyses the formation of a phosphodiester bond between a 3′-hydroxyl at the end of one DNA strand and a 5′ phosphate at the end of another strand. The phosphate must be activated by adenylylation. DNA ligases isolated from viruses and eukaryotes use ATP for this purpose. DNA ligases from bacteria are unusual in that they generally use NAD^+ a cofactor that normally functions in hydride transfer reactions



Mechanism of action of DNA polymerase-I and Ligase enzyme



DNA Replication in Schematic Way

3.2.3.3 ENZYMES INVOLVE IN DNA REPLICATION:

ENZYME	FUNCTION
Helicase	Untwist the DNA double helix
Topoisomerase or Gyrase	Release the tension in the DNA strand generated due to untwisting of double helix structure
Primase	Add RNA primer
DNA polymerase-III	Synthesize DNA
DNA polymerase-I	Remove RNA primer after completion of the DNA synthesis
Ligase	Add Okazaki fragments after deletion of RNA primers.

3.2.4 REPLICATION IN EUKARYOTE:

Replication process in eukaryotes is almost similar to replication in prokaryotic cell. In case of human genome there present 23 pairs of chromosome in single cell and each chromosome averagely consist about 10^8 base pairs, much much longer than E.coli chromosome. Replication movement is much slower in eukaryotic cells than prokaryotic cells. So if there was a single replication origin in eukaryotic DNA then it takes nearly few days to complete replication process. But eukaryotic chromosome replicate in very quick manner because DNA replication in eukaryotic chromosome shows many origin of replication rather than a single origin of replication like prokaryotic cell. As like prokaryotic replication at each origin of replication DNA unwinds to single strand DNA template and replication proceeds bidirectionally.

The length of DNA from the origin of replication to the two termini (two new DNA strand) point of replication where adjuscent replication forks fuse is known as the replication unit or **replicon.** The movement of replication fork in case of E.coli is about 1000 bp per second and complete the replication in about 42 minutes. Where as in human genome there are about 10000-100000 replicon for an average of 30-300kb, and the rate of replication fork movement is about 100 bp per second. Replication of the whole genome takes about 8 hours, but each replicon is replicating for a part of that time.

3.2.4.1 Enzymes of Eukaryotic Replication: Eukaryotic cells may contain about 15 or more DNA polymerase enzymes. Typically, replication of eukaryotic nuclear DNA requires three of these DNA polymerase enzymes like, **DNA pol α (alpha)/primase complex**, **DNA pol δ (delta), DNA pol ε (epsilon).** Pol α/primase complex initiates new strand in replication by adding RNA primer of about 10 nucleotides with the help of primase enzyme. Then poly α adds about 10-20 nucleotides of DNA. DNA pol ε seems to synthesize the **leading strand** whereas the **lagging strand** is synthesized by the involvement of DNA pol δ. Other eukaryotic DNA polymerases involved in specific DNA repair mechanism and extra nuclear DNA (mitochondria, chloroplast) replication.

As in prokaryotes, joining of Okazaki fragments on the lagging-strand template involves removing the primer on the older Okazaki fragment and replacing it with DNA by extension of the newer Okazaki. Primer removal does not involve the progressive removal of nucleotides, as is the case in prokaryotes. Rather, Pol continues extension of the newer Okazaki fragment; this activity displaces the RNA/DNA ahead of the enzyme, producing a flap. Nucleases remove the flap. The two Okazaki fragments are then joined by the eukaryotic DNA ligase.

3.2.5 DNA POLYMERASE PROOFREADING- REPLICATION FIDELITY:

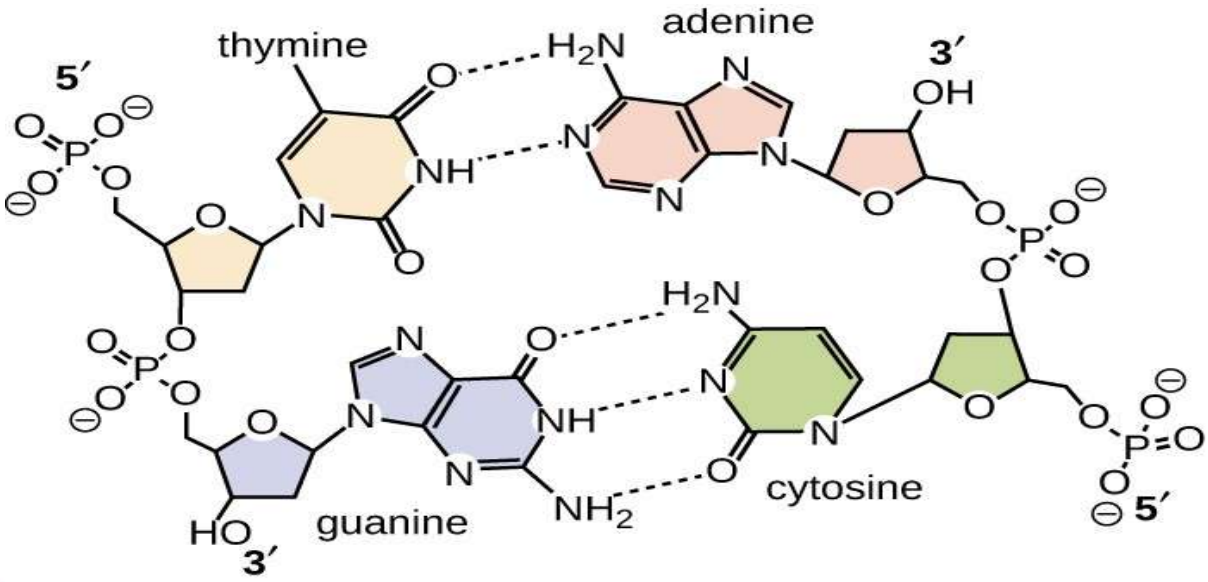


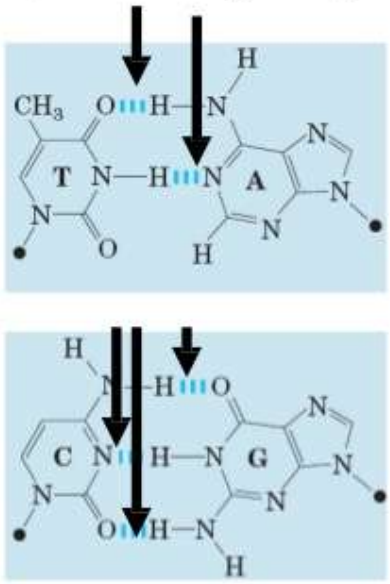
Fig-Correct mode of base pairing in DNA

Replication proceeds with an unprecedented degree of fidelity. In E. coli, an error is formed just one occasion for each 10^9 to 10^{10} nucleotides added. For the E. coli chromosome of ~4.6 106 bp, this suggests that a mistake occurs only once per 1,000 to 10,000 replications. During polymerization, discrimination between correct and incorrect nucleotides relies not just on the hydrogen bonds that specify the right pairing between complementary bases but also on the common geometry of the standard nucleotide base pairing. A binds with T with 2 hydrogen bond and G binds with C with 3 hydrogen bonds (Fig.). The DNA polymerase accommodates only base pairs with this geometry. An incorrect nucleotide could also be ready to hydrogen-bond with a base within the template, but it generally will not fit into the active site of DNA polymerase. **Incorrect bases can be rejected before the phosphodiester bond is made**. The accuracy of the polymerization reaction itself, however, is insufficient to account for the high

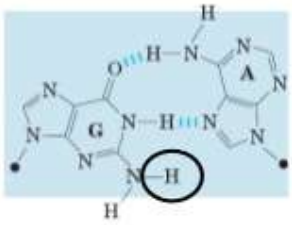
degree of fidelity in replication. Careful measurements in vitro have shown that DNA polymerases insert one incorrect nucleotide for each 10^4 to 10^5 correct ones.

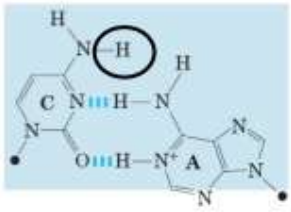
Contribution of base-pair geometry to the fidelity of DNA replication

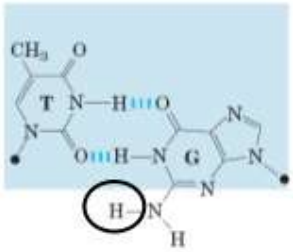
The standard A=T and G≡C base pairs have very similar geometries, and an active site sized to fit one (blue shading) will generally accommodate the other



The geometry of incorrectly paired bases can exclude them from the active site, as occur on DNA polymerase

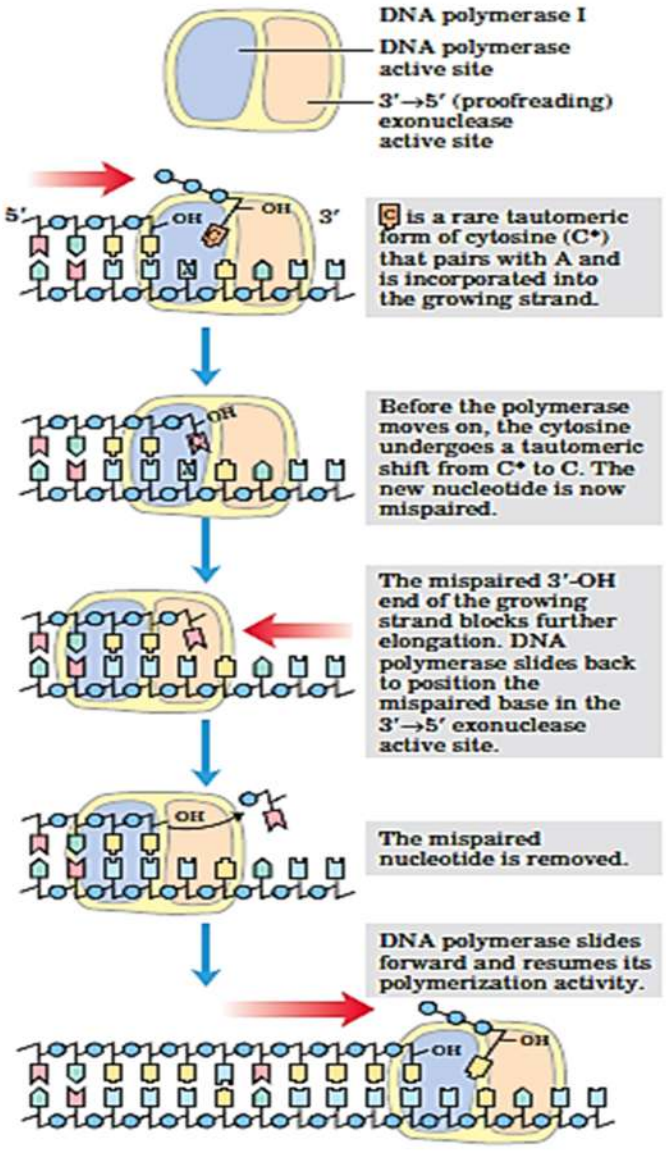






One mechanism intrinsic to virtually all DNA polymerases is a separate **3′-5′ exonuclease** activity that double-checks each nucleotide after it′s added. This nuclease activity permits the enzyme to get rid of a newly added nucleotide and is very specific for mismatched base pairs. If the polymerase has added the wrong nucleotide, translocation of the enzyme to the position where subsequent nucleotide is to be added is inhibited. This kinetic pause provides the chance for a correction. The **3′-5′ exonuclease activity** removes the miss-paired nucleotide, and therefore the polymerase begins again. This activity, referred to as proofreading, is not simply the reverse of the polymerization reaction because

pyrophosphate isn't involved. The polymerizing and proofreading activities of a DNA polymerase can be measured separately. Proofreading improves the inherent accuracy of the polymerization reaction 10^2 to 10^3 fold. Within the monomeric DNA polymerase I, the polymerizing and proofreading activities have **separate active sites** within an equivalent polypeptide. When base selection and proofreading are combined, DNA polymerase leaves behind one net error for every 10^6 to 10^8 bases added. Yet the measured accuracy of replication in E. coli is higher still. the extra accuracy is provided by a separate enzyme system that repairs the mismatched base pairs remaining after replication. As like prokaryotic E.coli DNA polymerases, **δ and ε DNA polymerase** in eukaryotic cell used for DNA replication also shows proof reading activity



An example of error correction by the 3′-5′ exonuclease

Activity of DNA polymerase I. Structural analysis has located the exonuclease activity ahead of the polymerase activity as the enzyme is oriented in its movement along the DNA. A mismatched base (here, a C–A mismatch) impedes translocation of DNA polymerase I to the next site. Sliding backward, the enzyme corrects the mistake with its 3′-5′ exonuclease activity, then resumes its polymerase activity in the 5′-3′ direction.

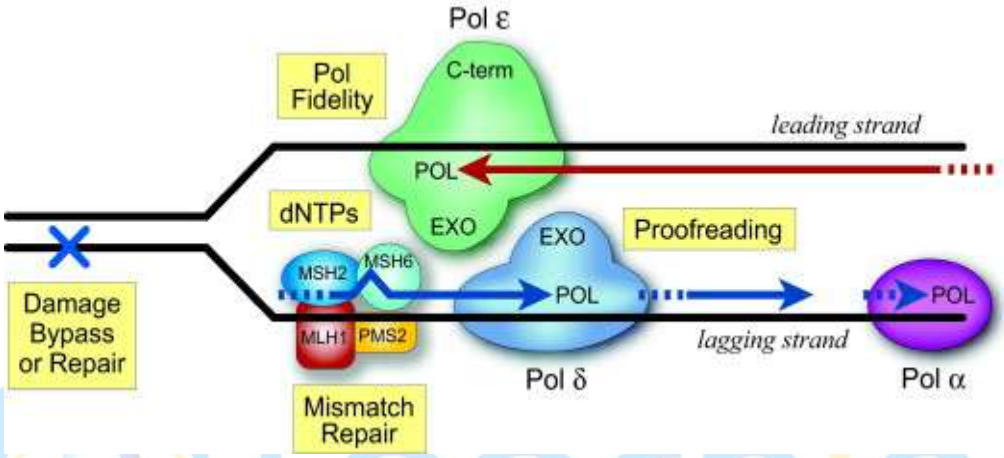
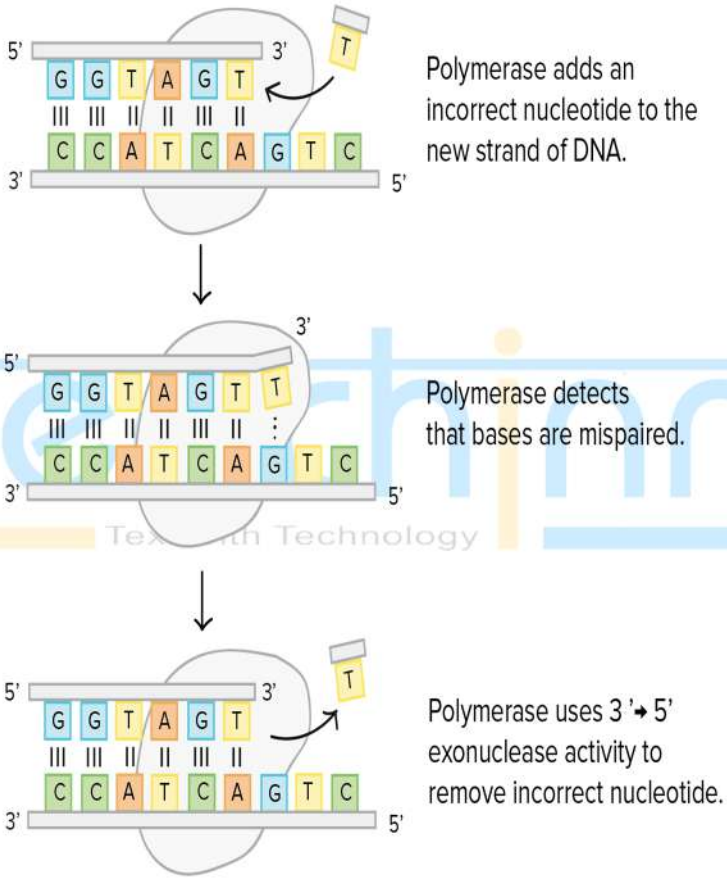


Fig: Proofreading in eukaryotic cell

3.2.6 Important Key Concept:

1. **Nuclease :** An enzyme that can degrade nucleotides
2. **Endonuclease:** A nuclease enzyme that can degrade nucleotides in the middle area of strand (DNA strand)
3. **Exonuclease:** nuclease enzyme that can bind at the end point of the DNA or RNA strand and degrade the nucleotides. For example a 5′-3′ exonuclease enzyme binds at 5′- end of the strand and degrade nucleotides in 5′-3′ direction.

4. **DNA Proofreading: DNA polymerases** are the enzymes that build DNA in cells. During DNA replication (copying), most DNA polymerases can "check their work" with each base that they add. This process is called **proofreading**. If the polymerase detects that a wrong (incorrectly paired) nucleotide has been added, it will remove and replace the nucleotide right away, before continuing with DNA synthesis.



3.2.7 DNA DAMAGE, MUTATION AND DNA REPAIR:

3.2.7.1 MUTATION:

Mutation is the process by which the sequence of base pairs in a DNA molecule is altered. A mutation may result in a change to either a DNA base pair or a chromosome. A cell with a mutation may be a mutant cell. If a mutation happens to occur during a vegetative cell or simply occurs in a somatic cell (in multicellular organisms), it's a **somatic mutation**—the mutant characteristic affects **only the individual during which the mutation occurs** and isn't passed on to the succeeding generation. In contrast, a mutation within the germ line of sexually reproducing organisms—**a germ-line mutation**—may be transmitted by the gametes to **subsequent generation**, producing a private with the mutation in both its somatic and its germ-line cells.

3.2.7.2 TYPE OF MUTATIONS:

Point mutations fall under two general categories: base-pair substitutions and base-pair insertions or deletions. A base-pair substitution mutation is a change from one nucleotide to a different in DNA, and there are two general types. A transition mutation **(Figure.a)** is a mutation from one purine–pyrimidine nucleotide to the other purine–pyrimidine nucleotide, like A–T to G–C. Specifically, this suggests that the purine on one strand of the DNA (A within the example) is modified to the opposite purine, while the pyrimidine on the complementary strand (T, the base paired to the A) is modified to the opposite pyrimidine. A transversion mutation (**Figure.b**) may be a mutation from a purine–pyrimidine nucleotide to a pyrimidine–purine base pair, like G–C to C–G, or A–T to C–G. Specifically, this means that the purine on one strand of the

DNA (A within the second example) is modified to a pyrimidine (C during this example), while the pyrimidine on the complementary strand (T, the bottom paired to the A) is modified to the purine that base pairs with the altered pyrimidine (G during this example). Base-pair substitutions in protein-coding genes also are defined consistent with their effects on amino acid sequences in proteins. counting on how a base-pair substitution is translated via the ordering , the mutations can result in no change to the protein, an insignificant change, or a clear change.

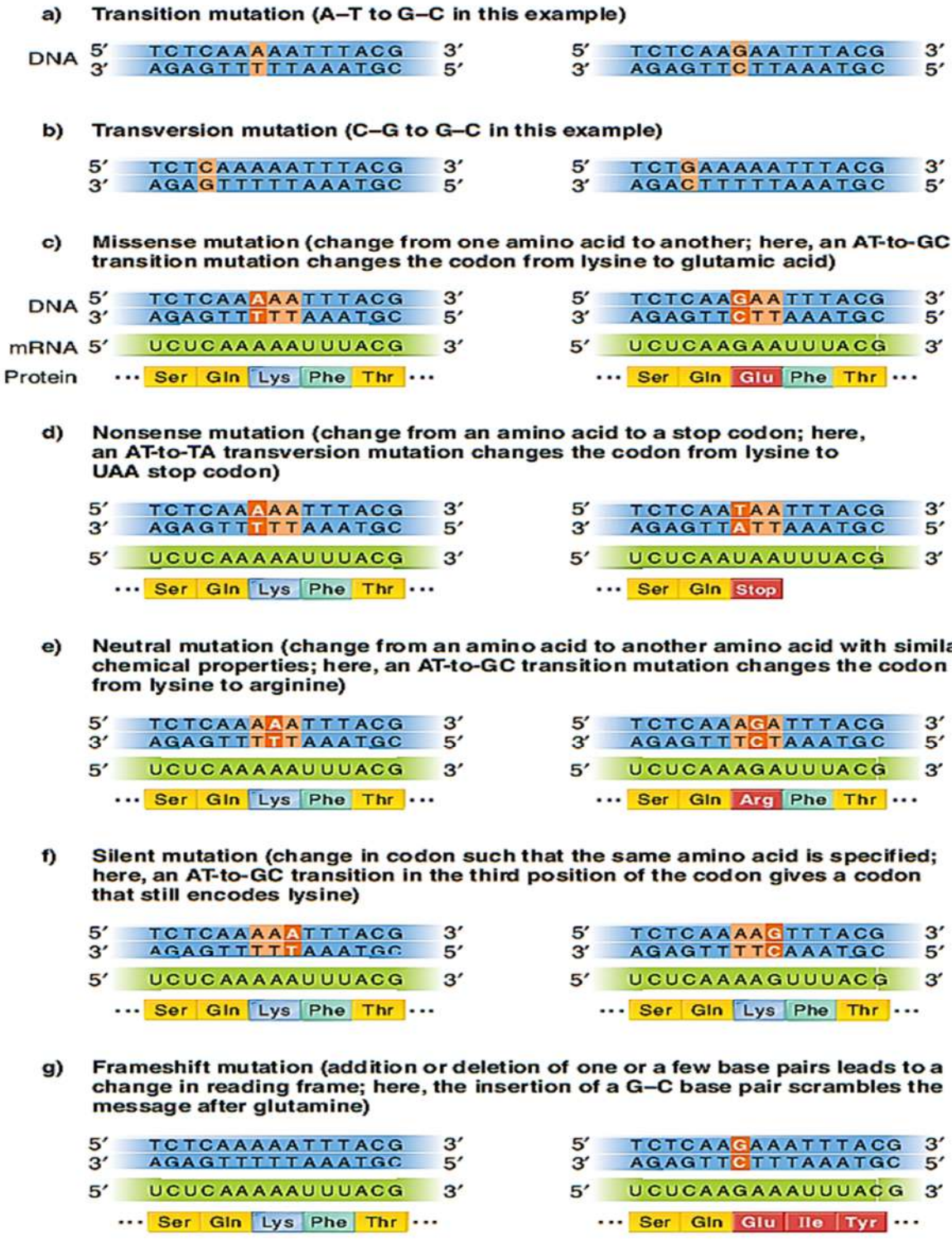
A **missense mutation** may be a point mutation during which a base-pair change causes a change in an mRNA codon in order that a special amino acid is inserted into the polypeptide. A phenotypic change may or might not result, counting on the amino acid change involved

A **nonsense mutation** may be a point mutation during which a base-pair change alters an mRNA codon for an amino acid to a stop (**nonsense**) codon (**UAG, UAA, or UGA**).

A **neutral mutation** may be a base-pair change during a gene that changes a codon within the mRNA such the resulting amino acid substitution produces no detectable changes within the function of the protein translated from that message. A neutral mutation may be a subset of missense mutations during which the new codon codes for a special amino acid that's chemically like the first or the amino acid isn't functionally important and thus doesn't affect the protein's function. Consequently, the phenotype doesn't change. An AT-to-GC transition mutation changes the codon from 5′-AAA-3′ **(lysine)** to 5′-AGA-3′ (**arginine**). Because arginine and lysine have similar properties —both are basic amino acids—the protein's function might not alter significantly.

A **silent mutation** also referred to as a synonymous mutation—is a mutation that changes a nucleotide in a gene, but the altered codon within the mRNA specifies an equivalent/same amino acid within the protein. During this case, the protein obviously features a wild-type function. For instance, in Figure, a silent mutation results from an AT-to-GC transition mutation that changes the codon from 5′-AAA-3′ to 5′-AAG-3′, both of which code for lysine. Silent mutations most frequently occur by changes like this at the third—wobble—position of a codon. This is sensible from the degeneracy patterns of the ordering

If one or more base pairs are added to or deleted from a protein-coding gene, the reading frame of an mRNA can change downstream of the mutation. An addition or deletion of 1 nucleotide, for instance, shifts the mRNA′s downstream reading frame by one base so that incorrect amino acids are added to the polypeptide chain after the mutation site. This sort of mutation, called a frameshift mutation usually leads to a non-functional protein. Frameshift mutations may generate new stop codons, leading to a shortened polypeptide; they'll end in longer-than-normal proteins because the traditional stop codon is now during a different reading frame; or they'll end in a big alteration of the amino acid sequence of a polypeptide. In Figure, an insertion of a G–C nucleotide scrambles the message after the codon specifying glutamine. Since each codon consists of three bases, a frameshift mutation is produced by the insertion or deletion of any number of base pairs within the DNA that's not divisible by three. Frame shift mutations were instrumental in scientists′ determining that the genetic code is a triplet code.



Different types of mutations

3.2.7.3 MUTAGENS:

Mutagens are those factors that induce the onset of mutation in genome and the process is called mutagenesis.

3.2.8 REPAIR OF DNA DAMAGE:

Mutagenesis involves damage to DNA. Especially with high doses of mutagens, the mutational damage can be considerable. What we see as mutations are DNA alterations that are not corrected by various DNA damage repair systems; that is, "mutation= DNA damage-DNA repair." Both prokaryotic and eukaryotic cells have a number of enzyme based systems that repair DNA damage. If the repair systems cannot correct all the lesions, the result is a mutant cell (or organism) or, if too many mutations remain, death of the cell (or organism).

DNA repair is possible largely because the DNA molecule consists of two complementary strands. DNA damage in one strand can be removed and accurately replaced by using the undamaged complementary strand as a template. We consider here the principal types of repair systems, beginning with those that repair the rare nucleotide mismatches that are left behind by replication.

3.2.8.1 Type of DNA DAMAGE:

DNA damage are often categorized into two main classes supported its origin: endogenous and exogenous. The bulk of the endogenous DNA damage arises from the chemically active DNA engaging in hydrolytic and oxidative reactions with water and reactive oxygen species (ROS), respectively, that are naturally present within cells. Such inherently predisposed reactions of DNA with molecules from its immediate surroundings fuel the event of hereditary diseases and sporadic cancers. Exogenous DNA damage, on the opposite hand, occurs when environmental, physical and chemical agents damage the DNA. Examples include UV and radiation, alkylating agents, and crosslinking agents. we provide here a quick summary of the most endogenous and environmental agents that produce the various classes of DNA damage that then become substrates for the precise DNA repair pathways discussed within the subsequent section.

3.2.8.2 Endogenous DNA damage

Every time a person′s cell replicates, approximately 3 X 109 bases are copied over by hi-fi replicative polymerases (δ and ε). However, A battery of other DNA polymerases (α, β, σ, γ, λ, REV1, ζ, η, ι, κ, θ, ν, μ, Tdt and PrimPol) can perform lower fidelity DNA synthesis during DNA replication or repair. hi-fi DNA synthesis may be a consequence of structural and biochemical attributes of replicative DNA polymerases, which make sure the insertion of an accurate complementary deoxynucleotide opposite the template base. this is often accomplished, as an example , by: 1) the thermodynamic stability and base-pair energetics of the incoming dNTP and template base, 2) the geometric selection of a correctly shaped and sized dNTP within the polymerase′s site , and 3) removing an incorrectly inserted deoxynucleotide by a 3′-5′ deoxynucleotide

exonuclease. additionally , the mismatch repair (MMR) pathway contributes to replication fidelity by quite 100-fold by correcting the rare errors that have escaped proofreading by replicative polymerases

Another source of endogenous DNA damage results from the action of topoisomerase enzymes (for example: TOP I, TOP II, TOP III; 7 TOP genes are found within the human genome), which primarily remove superhelical tension on DNA during replication and transcription. TOP1, for instance , transiently nicks the supercoiled DNA and facilitates rotation of the broken strand round the TOP1-bound DNA strand to relax the DNA. Thereafter, TOP1 religates the breaks by aligning the 5′-OH group of the DNA with the tyrosine-DNA phosphodiester bond to resolve the complex. Misalignment of the 5′-OH DNA end stabilizes the cleavage complex to make a DNA lesion. Interestingly, anticancer drugs like camptothecin and lots of naturopathic compounds are known to stabilize the TOP1-DNA cleavage complexes. Additionally, DNA adducts (from UV and benzene derivatives) and aberrant DNA structures (nicks, mismatches, abasic sites) also can irreversibly trap the TOP1-DNA cleavage complex into DNA lesions called suicidal complexes. TOP1-associated DNA damage is typically repaired by reversal of those complexes or is excised by TDP1 (tyrosyl DNA phosphodiesterase) and endonucleases.

3.2.8.3 Exogenous DNA damage include DNA damage by Ionizing radiation, composed of alpha, beta, gamma, neutrons and X-rays. Ionizing radiations produces a spectrum of base lesions that is similar to that generated by ROS species .UV rays also cause severe DNA damage. Environmental sources of stress such as extreme cold or extreme heat, hypoxia, and oxidative stress have been shown to cause DNA damage in human cells.

Other everyday use biological products are also found to associated with DNA damage. For example, butyl paraben (BP) and bisphenol A (BPA), found in cosmetics, pharmaceuticals, food-products and beverage processing, are associated to DNA damage in sperm cells. Food preservatives [(sodium benzoate (SB), potassium benzoate (PB) and potassium sorbate (PS)] and food additives [(citric acid (CA), phosphoric acid (PA), brilliant blue (BB) and sunset yellow (SY)] are also known to cause DNA damage

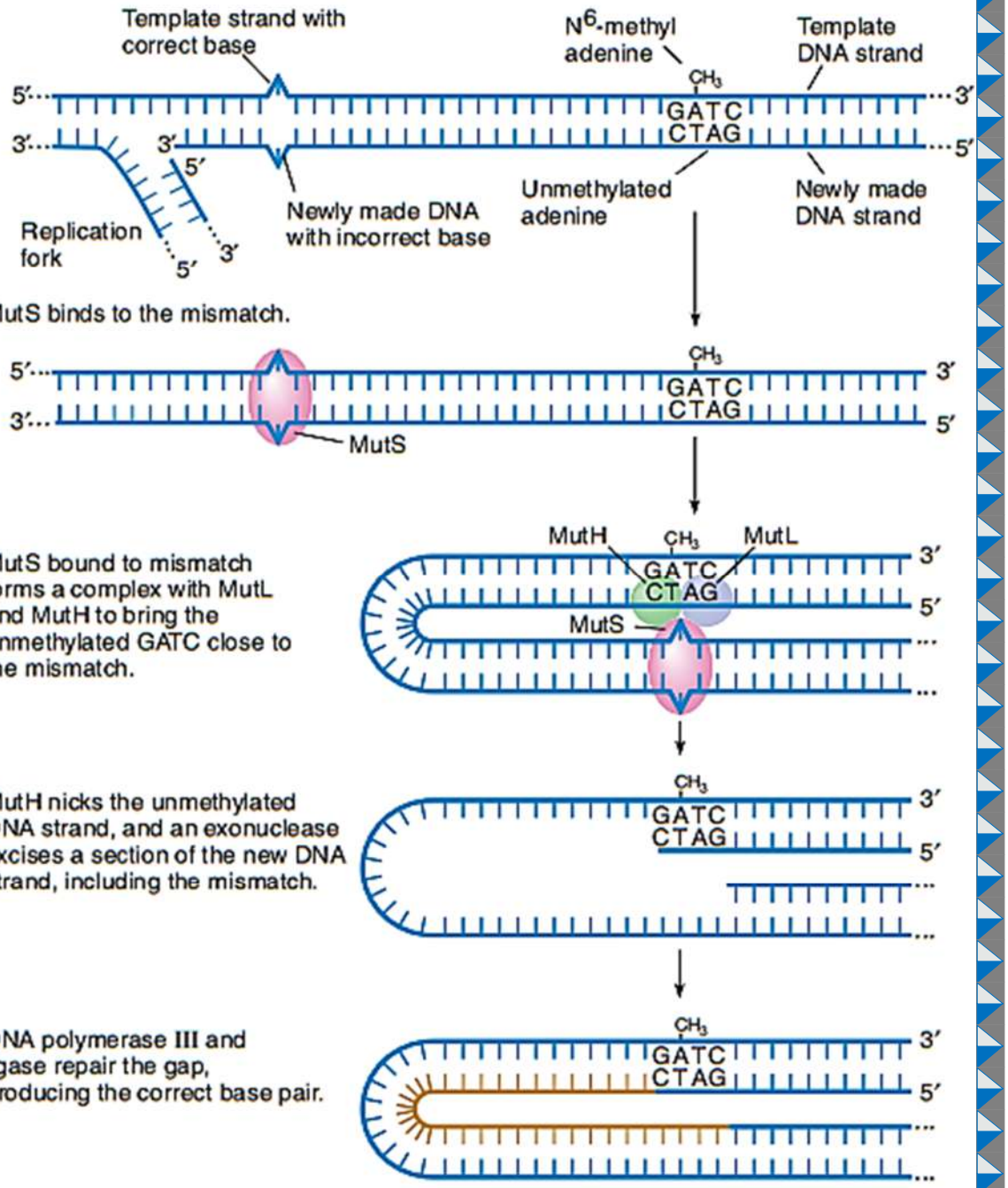
3.2.8.4 IMPORTANT DNA REPAIR MECHANISM:

1. **Mismatch Repair by DNA Polymerase Proofreading:** The frequency of base-pair substitution mutations in bacterial genes ranges from 10^{-7} to 10^{-11} errors per generation. However, DNA polymerase inserts incorrect nucleotides at a frequency of 10^{-5}. Most of the difference between the two values is accounted for by the 3′-to-5′ exonuclease proofreading activity of the DNA polymerase in both bacteria and eukaryotes. When an incorrect nucleotide is inserted, the polymerase often detects the mismatched base pair and corrects the area by “backspacing” to remove the wrong nucleotide and then resuming synthesis in the forward direction. The mutator mutations in E. coli illustrate the importance of the 3′-to-5′ exonuclease activity of DNA polymerase for maintaining a low mutation rate. Mutator mutants have a much higher than normal mutation frequency for all genes.

These mutants have mutations in genes for proteins whose normal functions are required for accurate DNA replication. For example, the mutD mutator gene of E. coli encodes the (epsilon) subunit of DNA polymerase III, the primary replication enzyme of E. coli. The mutD mutants are defective in 3′-to-5′ proofreading activity, so that many incorrectly inserted nucleotides are left unrepaired.

2. **Methyl-Directed Mismatch Repair**: Despite proofreading by DNA polymerase, a number of mismatched base pairs remain uncorrected after replication has been completed. In the next round of replication, these errors will become fixed as mutations if they are not repaired. Many mismatched base pairs left after DNA replication can be corrected by methyl-directed mismatch repair. This system recognizes mismatched base pairs, excises the incorrect bases, and then carries out repair synthesis. In E. coli, the products of three genes—mutS, mutL, and mutH—are involved in the initial stages of mismatch repair. First, the mutS-encoded protein, MutS, binds to the mismatch (Figure, step 1). Then the repair system determines which base is the correct one (the base on the parental DNA strand) and which is the erroneous one (the base on the new DNA strand). In E. coli, the two strands are distinguished by methylation of the A nucleotide in the sequence GATC. Both A nucleotides in the sequence usually are methylated. However, after replication, the parental DNA strand has a methylated A in the GATC sequence, whereas the A in the GATC of the newly replicated DNA strand is not methylated until a short time after its synthesis. Therefore, the MutS protein bound to the mismatch forms a complex with the mutL- and mutH-encoded proteins, MutL and MutH, to bring the unmethylated GATC sequence close to the mismatch

(Figure, step 2). The MutH protein then nicks the unmethylated DNA strand at the GATC site, the mismatch is removed by an exonuclease (Figure, step 3), and the gap is repaired by DNA polymerase III and ligase (Figure, step 4). Mismatch repair also takes place in eukaryotes. However, it is unclear how the new DNA strand is distinguished from the parental DNA strand (no methylation is involved). In humans, four genes, respectively named hMSH2, hMLH1, hPMS1, and hPMS2, have been identified; hMSH2 is homologous to E. coli mutS, and the other three genes have homologies to E. coli mutL. The genes are known as mutator genes, because loss of function of such a gene results in an increased accumulation of mutations in the genome. Mutations in any one of the four human mismatch repair genes confer a phenotype of hereditary predisposition to a form of colon cancer called hereditary nonpolyposis colon cancer.


Template strand with correct base
N6-methyl adenine
Template DNA strand
CH3
GATC
CTAG
5′
3′
Replication fork
Newly made DNA with incorrect base
Unmethylated adenine
Newly made DNA strand
1 MutS binds to the mismatch.
MutS
2 MutS bound to mismatch forms a complex with MutL and MutH to bring the unmethylated GATC close to the mismatch.
MutH
MutL
3 MutH nicks the unmethylated DNA strand, and an exonuclease excises a section of the new DNA strand, including the mismatch.
4 DNA polymerase III and ligase repair the gap, producing the correct base pair.

3.3 TRANSCRIPTION

Final expression of gene is the formation of protein. According to central dogma theory genes in DNA segment first form RNA molecules by means of Transcription. mRNA molecules translated into specific protein depending upon the genetic sequences. In the process of transcription other types of RNA like rRNA, tRNA etc also synthesized, which helps in decoding the genetic sequences in mRNA to synthesis protein molecules. Few types of RNA molecules discussed in the below.

1. mRNA (messenger RNA) encodes the amino acid sequence of a polypeptide. mRNAs are the transcripts of protein-coding genes. Translation of mRNA produces a polypeptide.

2. rRNA (ribosomal RNA), with ribosomal proteins, makes up the ribosomes the structures on which mRNA is translated.

3. tRNA (transfer RNA) brings amino acids to ribosomes during translation.

4. snRNA (small nuclear RNA), with proteins, forms complexes that are used in eukaryotic RNA processing to produce functional mRNAs.

During replication the entire chromosome is usually copied, but transcription is more selective. Only particular genes or groups of genes can be transcribed at any certain time, and some parts of the DNA genome are never transcribed. The cell restricts the expression of genetic information to the formation of gene products needed at any particular moment. Specific regulatory sequences mark the beginning and end of the DNA segments to be transcribed and designate which strand in duplex DNA is to be used as the template.

3.3.1 SIGNALS FOR TRANSCRIPTION:

There are no such biological mechanisms in which can starts by itself. There always be a specific signal that induce or influence the onset of a certain biological mechanism. As like this transcription is depends upon certain signal. According to central dogma theory transcription leads to generate specific mRNA which then translated to form polypeptide chain, the ultimate form of phenotype of a certain gene. Whenever there is any requirement of certain phenotypic character body secrete signalling molecules like hormone, enzyme etc. these signalling molecules act through specific signalling pathway like GPCR, JAK-STAT etc. Ultimately these signalling pathways ended up in the process called gene expression. These gene expression means synthesis of specific type of protein depending upon the signalling molecule. Protein is synthesized by translation of mRNA molecule and active gene in DNA transcribed to mRNA molecule and other RNA molecule which helps in translation process. So requirement of protein influence the translation that ultimately induce the onset of transcription process.

3.3.2 TRANSCRIPTION PROCESS:

Transcription is the process by which ribonucleic acid polymerized to form RNA molecules. As like DNA replication polymerase enzyme involve in this mechanism, the RNA polymerase enzyme. As like replication process transcription also need a template to transcript RNA. RNA polymerase can use DNA as well as RNA strand as template strand. RNA synthesis using DNA molecule as template is called DNA dependent RNA synthesis and the polymerase enzyme used is known as DNA dependent RNA polymerase enzyme. On the other hand RNA synthesis using RNA strand as a template strand is RNA dependent RNA synthesis and the polymerase enzyme act in this process is called

RNA dependent RNA polymerase. We will first focus on the DNA dependent Transcription in prokaryotic and eukaryotic cell.

3.3.3 DNA Dependent RNA synthesis:

Transcription proceeds through three main stages **initiation, elongation and termination.** The direction of RNA synthesis is similar to that of replication process that is the direction of transcription is in 5'-3' direction. But unlike DNA polymerase enzyme RNA polymerase enzyme does not require any primer to synthesize RNA strand. (DNA polymerase require RNA primer to start replication). In case of replication the whole part of DNA strand is replicated but the specific part of DNA (gene) is transcribed not the full length of a DNA molecule.

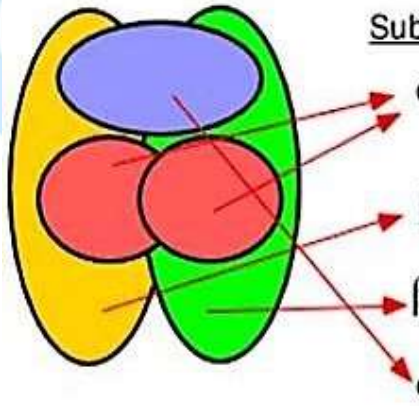
In transcription process the 3'-5' DNA strand that is read to generate the RNA strand called the **template strand**. The DNA strand complementary to the template strand, and having the same polarity as the resulting RNA strand, is called the **non-template strand**. The RNA precursors for transcription are the ribonuclroside triphosphate ATP, GTP, CTP, UTP. These are collectively known as the NTPs or nucleoside triphosphate. Transcription process is also based upon the complementary base pairing mechanism. If the template strand is 3'-ATACTGGAC-5' then the RNA chain synthesized will be 5'-UAUGACCUG-3'

3.3.3.1 RNA POLYMERASE:

The DNA dependent RNA polymerase enzyme in E.coli (prokaryotic) is a large complex structure enzyme containing five core subunits ($\alpha_2\beta\beta'\omega$) and a sixth subunit the σ subunit. The σ subunit binds to the core part and directs the enzyme at the specific binding site of the template DNA strand. These six subunits collectively known as the holoenzyme of the RNA polymerase. Different types of RNA polymerase enzyme present depending upon the σ subunit. The most common type of σ subunit is the σ^{70}.

Beside that DNA dependent RNA polymerase enzyme does not have proof reading exonuclease activity (present in DNA polymerase)

Prokaryotic RNA Polymerase: Holoenzyme Enzyme



Subunit	Size	#/Molecule	Function
α	36.5 kD	2	chain initiation and interaction with regulatory proteins
β	151 kD	1	chain initiation and elongation
β'	155 kD	1	DNA binding
σ	70 kD	1	promoter recognition

3.3.3.2 INITIATION OF TRANSCRIPTION:

The first step of transcription is the initiation process. In this process the main function is to recognition of the site of transcription in the template DNA strand as it was informed earlier that specific portion of DNA can transcribed not the whole DNA. The region of DNA from where RNA sequence starts generating is called +1 base pair. The previous base pair is called upstream base pair. Mutational studies show a crucial role of specific upstream region. These upstream sequences are known as promoter region. RNA polymerase first interact with these promoter region. These sequences are generally found surrounding the -35 and -10 from +1 bp point of template strand.

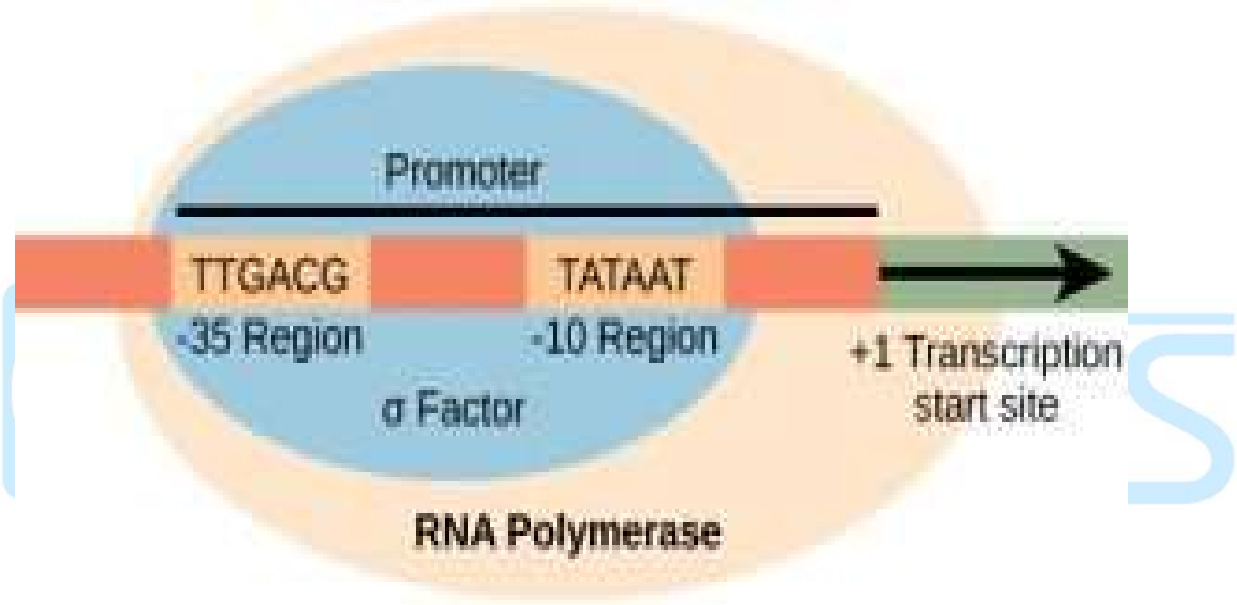
The **consensus sequence (**Certain nucleotides that are particularly common at each position form a **consensus sequence**) for the -35 region or the **-35** box is **5'-TTGACA-3'** and the consensus sequence for the -10 region or the **-10** box or **Pribnow box** is **5'TATAAT-3'**.

The σ factor is responsible for RNA polymerase to recognition and attachment with specific site of DNA double helix strand or at promoter region of the DNA. However the σ factor is not require in elongation process of transcription and termination steps. Changes in the consensus sequences or mutation in the promoter region highly effect the transcription mechanism.

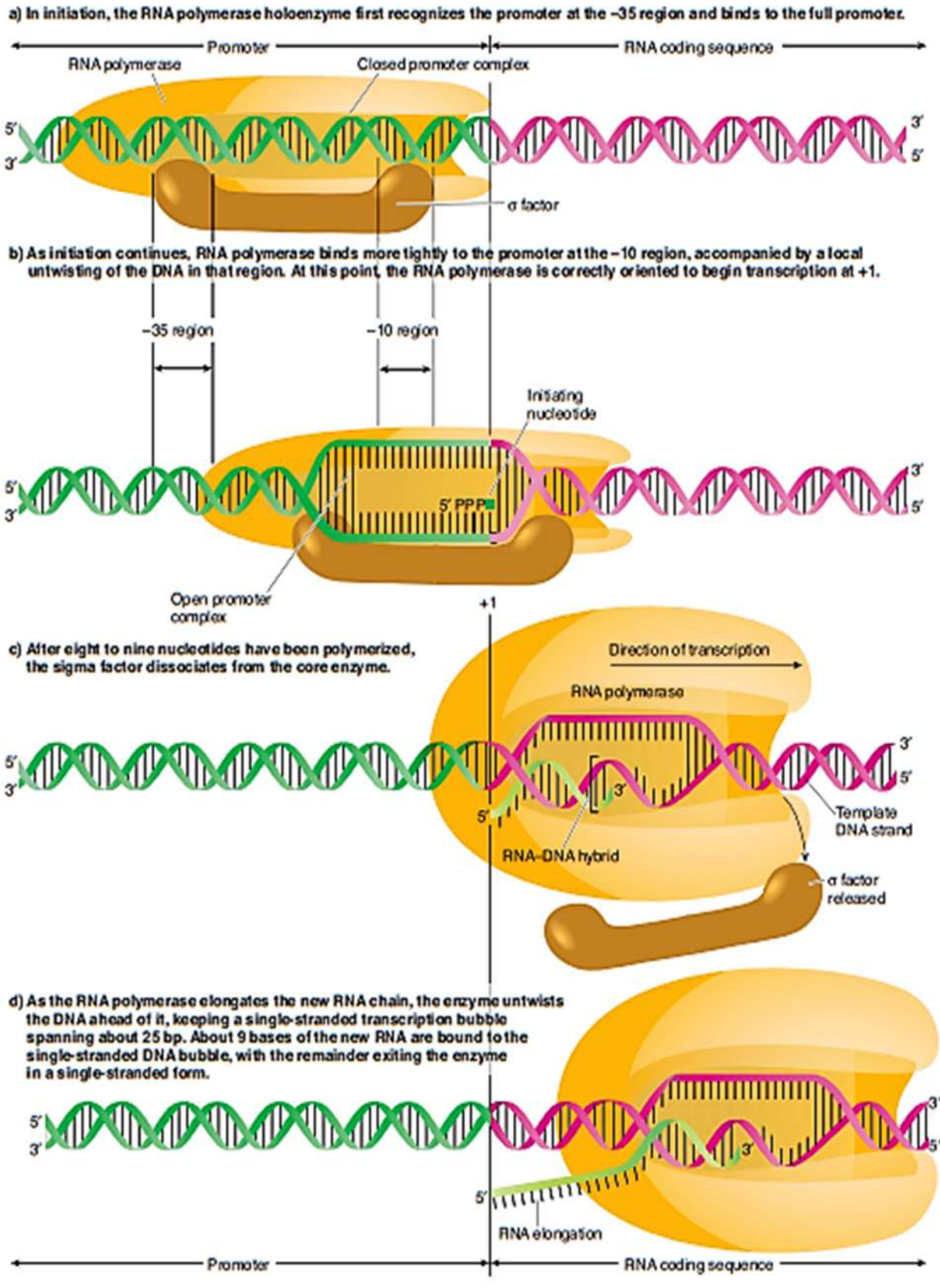
After recognition of RNA polymerase binding site by σ factor the RNA polymerase enzyme bind at the specific region on the DNA strand with the help of alpha subunit of the RNA polymerase enzyme.

As already mentioned, the promoters of most genes in E. coli have the-35 and-10 recognition sequences. Those promoters are recognized by a sigma factor with a molecular weight of 70,000 Da, called 70. There are other sigma factors in E. coli with important roles in regulating gene expression. Each type of sigma factor

binds to the **core RNA polymerase** and permits the holoenzyme to recognize different promoters. For example, under conditions of high heat (heat shock) and other forms of stress, 32 (molecular weight 32,000 Da) increases in amount, directing some RNA polymerase molecules to bind to the promoters of genes that encode proteins needed to cope with the stress. Such promoters have consensus recognition sequences specific to the 32 factor at-39 and-15. There are few other types of sigma factors in E.coli.



Sigma factor	Promoters recognized	-35 region	-10 region
σ^{70}	Housekeeping genes, most genes in exponentially replicating cells	**TTGACA**	**TATAAT**
σ^{S}	Stationary phase genes and general stress response	**TTGACA**	**TATAAT**
σ^{32}	Induced by unfolded proteins in the cytoplasm. Genes encoding chaperones that refold unfolded proteins and protease systems leading to the degradation of unfolded proteins in the cytoplasm	**TCTCNCCCTTGAA**	**CCCCATNTA**


a) In initiation, the RNA polymerase holoenzyme first recognizes the promoter at the –35 region and binds to the full promoter.
Promoter
RNA coding sequence
RNA polymerase
Closed promoter complex
5′
3′
3′
5′
σ factor
b) As initiation continues, RNA polymerase binds more tightly to the promoter at the –10 region, accompanied by a local untwisting of the DNA in that region. At this point, the RNA polymerase is correctly oriented to begin transcription at +1.
–35 region
–10 region
Initiating nucleotide
5′ PPP
Open promoter complex
+1
c) After eight to nine nucleotides have been polymerized, the sigma factor dissociates from the core enzyme.
Direction of transcription
RNA polymerase
Template DNA strand
RNA–DNA hybrid
σ factor released
d) As the RNA polymerase elongates the new RNA chain, the enzyme untwists the DNA ahead of it, keeping a single-stranded transcription bubble spanning about 25 bp. About 9 bases of the new RNA are bound to the single-stranded DNA bubble, with the remainder exiting the enzyme in a single-stranded form.
RNA elongation
Promoter
RNA coding sequence

3.3.4 ELONGATION PHASE:

RNA polymerase enzyme makes a complex when bound to promoter region. At -10 region RNA polymerase binds most tightly. This complex of promoter and RNA polymerase is known as closed complex. The promoter region bound tightly but not became unwind. Around 12-15 bp region of DNA from -10 region to +2 or +3 position became unwound to form a complex known as transcription bubble. Now elongations of RNA strand synthesis using the unwounded DNA strand as template for the synthesis of RNA. Unwounded DNA strand enter into the active site of the RNA polymerase that catalyses the formation of phosphodiester bond between ribonucleoside triphosphates that are complementary to the promoter template strand at the start site of transcription. After 8-10 nucleotides have been polymerised the σ (sigma) factor released from the RNA polymerase. Subsequently the RNA polymerase enzyme leaves the promoter region and moves along the direction of transcription. During this movement RNA polymerase add one nucleotide to the 3'- end of the growing (nascent) RNA chain and simultaneously template strand become unwind.

3.3.5 TERMINATION OF TRANSCRIPTION:

Transcription process termination is controlled by a special protein called Rho factor (ρ). Although termination of all genes are not depend upon Rho-factor. Those gene transcription termination controlled by Rho-protein is known as Rho-dependent termination and those genes which are independent of Rho-protein is known as Rho-independent termination. In Rho-independent termination RNA polymerase enzyme itself terminate the transcription process.

Rho-independent terminators consist of an inverted repeat sequence that is about **16 to 20** base pairs upstream of the transcription termination point, followed by a string of about **4 to 8 A–T** base pairs. The RNA polymerase transcribes the terminator sequence, which is part of the initial RNA-coding sequence of the gene. Because of the inverted repeat arrangement, the RNA folds into **a hairpin loop** structure (Figure). The hairpin loop structure causes the RNA polymerase to slow the rate of transcription and then pause in its catalysis of RNA synthesis. The string of U nucleotides downstream of the hairpin destabilizes the complementary base pairing between the new RNA chain and the DNA template strand, and RNA polymerase dissociates from the template; transcription has terminated. Mutations that disrupt the hairpin cause partial or complete prevention of termination.

Rho-dependent terminators are **C-rich, G-poor** sequences that have no hairpin structures like those of rho-independent terminators. Termination at these terminators is as follows: **Rho binds to the C-rich** terminator sequence in the transcript upstream of the transcription termination site. Rho then moves along the transcript until it reaches the RNA polymerase, where the most recently synthesized RNA is base paired with the template DNA. Rho protein is a **helicase** enzyme, meaning that it can **unwind double-stranded nucleic acids**. When Rho reaches the RNA polymerase, helicase **unwinds the helix formed between the RNA and the DNA template strand**, using ATP hydrolysis to provide the needed energy. The new RNA strand is then released, the DNA double helix reforms, and the RNA polymerase and Rho dissociate from the DNA; transcription has terminated.

After the transcription process completed, the sigma factor rebind with the RNA polymerase enzyme.

3.3.6 TRANSCRIPTION IN EUKARYOTIC CELLS:

Transcription is more complicated in eukaryotes than in bacteria. This is because eukaryotes possess three different classes of RNA polymerases and because of the way in which transcripts are processed to their functional forms.

3.3.6.1 EUKARYOTIC RNA POLYMERASE:

There are three types of RNA polymerase enzyme in eukaryotic cells, RNA polymerase-I, RNA polymerase-II, RNA polymerase-III. All three RNA polymerase enzymes catalyse transcription of genes encoding different types of RNA. RNA polymerase-I is responsible for synthesis of rRNA, RNA polymerase-II synthesize mRNA

RNA POLYMERASE-I: RNA polymerase enzyme is responsible for the synthesis of pre-rRNA or pre-ribosomal RNA which contains the precursor for the 28S rRNA, 18S rRNA, 5.8S rRNA. Promoters for RNA polymerase-I vary greatly depending upon species.

RNA POLYMERASE-II: RNA polymerase enzyme is responsible for the synthesis of mRNA or messenger RNA. Some other types of RNAs are also synthesized with the help of RNA polymerase-II enzyme, like, snRNA, miRNA. In general RNA polymerase-II transcribes protein coding genes. Many RNA polymerase-II promoters have few sequence in common, including the TATA box, the eukaryotic consensus sequence TATAAA near -30 bp region and Inr sequence the initiator sequence near the transcription start site +1 position of template strand.

RNA POLYMERASE-III: RNA polymerase-III is responsible for the synthesis tRNA. It also transcribes 5S rRNA, 7S rRNA, snRNA U6 and other stable specialized short RNAs. Some of the promoter sequences for RNA polymerase-III are located within the gene itself, whereas others are in more conventional locations upstream of the RNA start site.

3.3.7 TRANSCRIPTION OF PROTEIN CODING GENE BY RNA POLYMERASE-II:

As discussed that RNA polymerase is responsible for the transcription of protein coding gene that is RNA polymerase-II is responsible for the synthesis of the mRNA.

Two large subunits **RPB1 and RPB2** of all three eukaryotic RNA polymerase are related to each other and are similar to the bacterial (E.coli) **β′ and β** subunits respectively. Each eukaryotic polymerase also contains an ω like and two non-identical α like subunits.

The carboxyl end of the **largest subunit of RNA polymerase-II (RPB1)** contains a stretch of 7 amino acids that is nearly precisely repeated multiple times. Neither RNA polymerase-I nor RNA polymerase-III contains these repeating units. These **heptapeptide repeat** with a consensus sequence of **Tyr-Ser-Pro-Thr-Ser-Pro-Ser,** is known as the **Carboxyl terminal domain (CTD)**. Yeast RNA polymerase-II contains 26 or more repeats, vertebrate RNA polymerase-II enzymes have 52 repeats. The CTD is very crucial for viability. For example at least 10 copies of the repeat must be present for Yeast to survive.

3.3.7.1 PROMOTER:

Two general regions of the promoter are: (1) the core promoter; and (2) promoter-proximal elements. The **core promoter** is the set of cis-acting sequence elements required for the transcription machinery to initiate RNA synthesis at the correct site. ('Cis' means "on the same side." A cis-acting sequence element affects the activity only of a gene on the same molecule of DNA.) These elements are typically within not more than 50 bp upstream of that site. The best-characterized core promoter elements are: (1) a short sequence element known as Inr (initiator), which spans the transcription initiation start site **(defined as +1)**; and (2) the TATA box, or TATA element (also called the Goldberg-Hogness box, after its discoverers), located at about -30 region. The TATA box has the **7-nucleotide** consensus sequence 5'-TATAAAA-3'. The Inr and TATA elements indicate where the transcription machinery assembles and determine where transcription will start. However, in the absence of other elements, transcription will occur only at a very low rate.

Promoter-proximal elements are upstream from the TATA box, in the region from-50 to-200 nucleotides from the transcription start site. Examples of these elements are the **CAAT** (pronounced as "cat") box, named as its consensus sequence and located at about-75 region; and the GC box, with consensus sequence **5'-GGGCGG-3'**, located at about -90 bp region. Both these CAAT box and the GC box work in either orientation (meaning with the sequence element oriented either toward or away from the direction of transcription). Mutations in either of these elements (or other promoter-proximal elements not mentioned) significantly decrease transcription initiation from the promoter, indicating that they play a crucial role in determining the efficiency of the promoter.

Promoters contain various combinations of core promoter elements and promoter-proximal elements that together determine promoter function. The promoter-proximal elements are important for determination of how and when a gene get expressed. Key to this regulation are transcription regulatory proteins known as **activators,** which determine the efficiency of transcription initiation process. For example, genes that are expressed in all cell types for basic cellular functions that is—"housekeeping genes"—have promoter-proximal elements that are recognized by activators found in all cell types. Examples of housekeeping genes are the actin gene and the gene for the enzyme glucose 6-phosphate dehydrogenase. By contrast, genes that are expressed only in particular cell types or at particular times have promoter-proximal elements recognized by activators in those cell types or at those particular times.

3.3.7.2 ENHANCER:

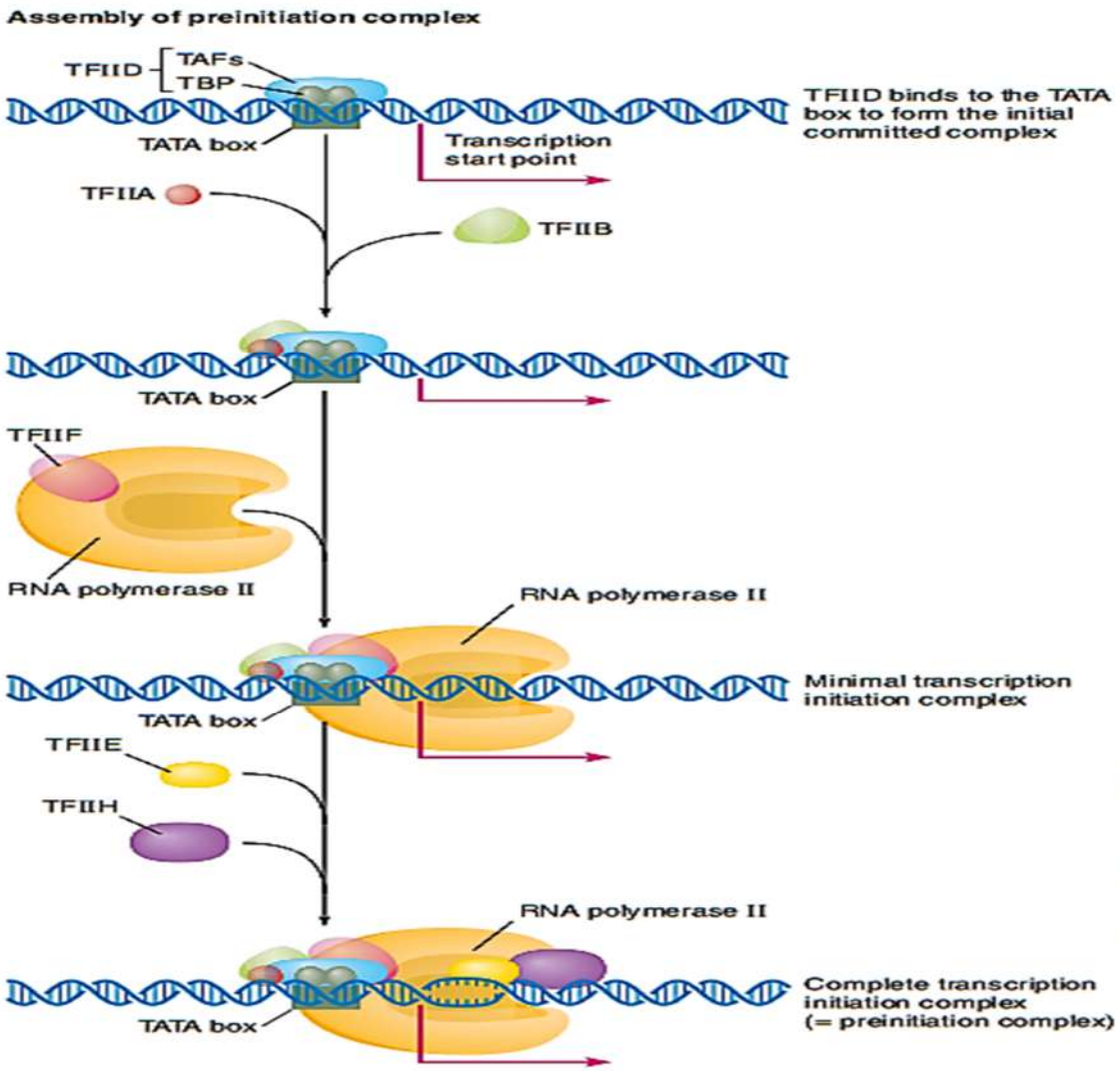
Other sequences—the **enhancers**—are required for the maximal transcription of a gene. Enhancers are another type of **cis-acting** element. By definition, enhancers function either upstream or downstream from the transcription initiation site—although, commonly, they located upstream of the genes they control, sometimes it may found thousands of base pairs away. In other words, enhancers regulate transcription from a distance. Enhancers contain various types of short sequence elements, some of them are same as those found in the promoter. Activators also bind to these elements as well as with other protein complexes. The DNA containing the enhancer is brought closer to the promoter DNA to which the transcription machinery is bound, stimulating transcription to the maximal level for the particular gene.

3.3.8 TRANSCRIPTION INITIATION:

Accurate initiation of transcription of a protein-coding gene involves the assembly of RNA polymerase II and a number of other important proteins called general transcription factors or **GTFs** on the core promoter. In contrast to bacterial RNA polymerase enzymes, none of these three eukaryotic RNA polymerases can bind directly to DNA. Instead, particular GTFs bind first and recruit the RNA polymerase to form a complex. Other GTFs then bind, and transcription can be start. The GTFs are numbered for the RNA polymerase with which they work and are lettered to reflect their order of discovery. For example, TFIID is the fourth general transcription factor (D) discovered that works with RNA polymerase II.

For protein-coding genes, the GTFs and RNA polymerase II bind to promoter elements in a particular order in vitro to produce the complete transcription initiation complex, also called the **pre-initiation complex (PIC)** as it is ready to begin transcription. As mentioned earlier, the binding of activators to promoter-proximal elements and to enhancer elements determines the overall efficiency of transcription initiation at a particular promoter.

While in vitro experiments indicate a sequential order of actions of GTFs and RNA polymerase II onto the promoter, the situation is less clear in vivo. Some experiments indicate that the initiation complex comes to the promoter in a single complex. Whether or not that is the case, transcription initiation in vivo is clearly more complicated because of the nucleosome organization of chromosomes



About figure: Assembly of the transcription initiation machinery. First, TFIID binds to the TATA box to form the initial committed complex. The multi-subunit TFIID has one subunit called the TATA binding protein (TBP), which recognizes the TATA box sequence and several other proteins called TBP-associated factors (TAFs). In vitro, the TFIID–TATA box complex acts as a binding site for the sequential addition of other transcription factors. Initially, TFIIA and then TFIIB bind followed by RNA polymerase II and TFIIF, to produce the minimal transcription initiation complex. (RNA polymerase II, like all eukaryotic RNA polymerases, cannot directly recognize and bind to promoter elements.) Next, TFIIE and TFIIH bind to produce the complete transcription initiation complex, also called the pre-initiation complex (PIC). TFIIH's helicase-like activity now unwinds the promoter DNA, and transcription is ready to begin.

Table: Proteins Required for Initiation of Transcription at the RNA polymerase II (Pol II) Promoters of Eukaryotes

Transcription protein	*Number of subunits*	*Subunit(s) M_r*	*Function(s)*
Initiation			
Pol II	12	10,000–220,000	Catalyzes RNA synthesis
TBP (TATA-binding protein)	1	38,000	Specifically recognizes the TATA box
TFIIA	3	12,000, 19,000, 35,000	Stabilizes binding of TFIIB and TBP to the promoter
TFIIB	1	35,000	Binds to TBP; recruits Pol II–TFIIF complex
TFIIE	2	34,000, 57,000	Recruits TFIIH; has ATPase and helicase activities
TFIIF	2	30,000, 74,000	Binds tightly to Pol II; binds to TFIIB and prevents binding of Pol II to nonspecific DNA sequences
TFIIH	12	35,000–89,000	Unwinds DNA at promoter (helicase activity); phosphorylates Pol II (within the CTD); recruits nucleotide-excision repair proteins
Elongation*			
ELL†	1	80,000	
p-TEFb	2	43,000, 124,000	Phosphorylates Pol II (within the CTD)
SII (TFIIS)	1	38,000	
Elongin (SIII)	3	15,000, 18,000, 110,000	

***The function of all elongation factors is to suppress the pausing or arrest of transcription by the Pol II–TFIIF complex.**

†Name derived from *e*leven-nineteen lysine-rich leukaemia. The gene for ELL is the site of chromosomal recombination events frequently associated with acute myeloid leukaemia.

3.3.9 ELONGATION AND TERMINATION:

After the initiation process completed RNA polymerase enzyme move in the direction of 5'-3' direction. During this process it **unwind** the double stranded DNA to make template strand available for the RNA synthesis. RNA synthesize on basis of complementary base pairing principle. The A of template strand code for U on the RNA strand, T on template strand code for A on RNA strand. G and C on template strand code for C and G respectively on the RNA strand. As the transcription bubble move, the nascent RNA strand starts leaving the template strand. The upstream part of transcription bubble starts rewinding again as DNA double strand as the transcription bubble moves.

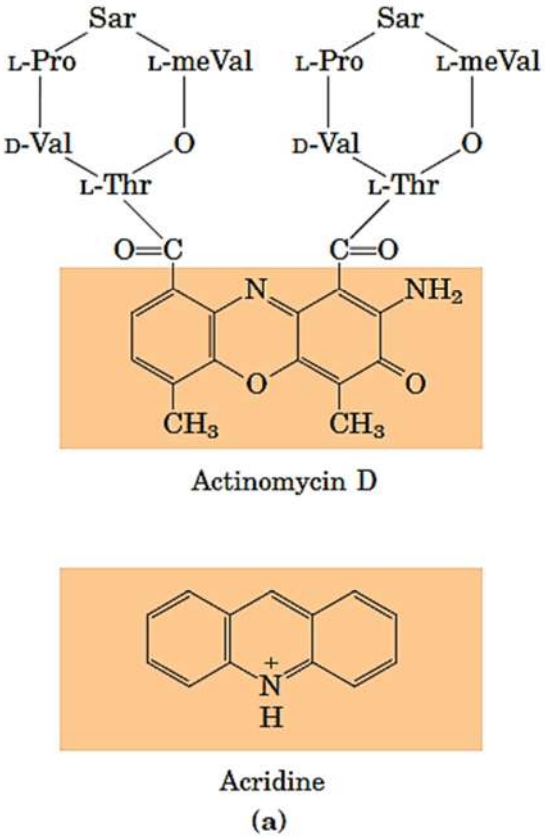
Termination of the eukaryotic transcription is variable depending upon the RNA polymerase enzyme involves in the transcription process.

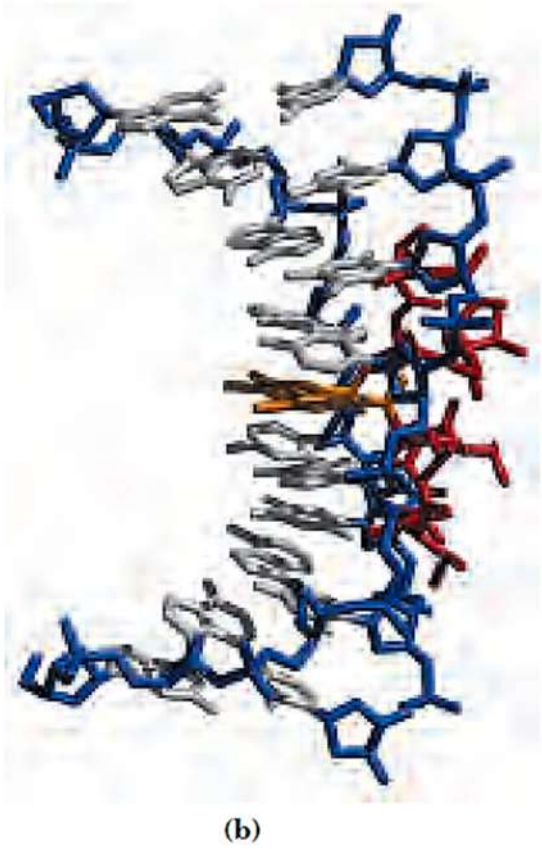
In case of mammalian protein coding gene transcription that is transcription by RNA polymerase-II enzyme, once the transcription of about 50 bases completed further elongation of transcription process become highly processive and does not terminated until any specific sequence transcribed that directs the cleavage and polyadenylation of the RNA at the 3' of the encoded mRNA. RNA polymerase-II then can terminate the transcription process at multiple site located over a distance of 0.5 to 2 kbp beyond this site of polyadenylation. After the termination process completed the RNA polymerase-II enzyme **dephosphorylated** at CTD region at it recycled back for initiation of another transcription process.

3.3.10 INHIBITION OF DNA DEPENDENT RNA POLYMERASE:

The elongation of RNA strands by RNA polymerase in both bacteria and eukaryotes is inhibited by the antibiotic **actinomycin D**. The planar portion of this molecule inserts (intercalates) into the doublehelical DNA between successive GqC base pairs, deforming the DNA. This prevents movement of the polymerase along the template. Because actinomycin D inhibits RNA elongation in intact cells as well as in cell extracts, it is used to identify cell processes that depend on RNA synthesis. **Acridine** inhibits RNA synthesis in a similar fashion. **Rifampicin** inhibits bacterial RNA synthesis by binding to the _ subunit of bacterial RNA polymerases, preventing the promoter clearance step of transcription. It is sometimes used as an antibiotic.

The mushroom *Amanita phalloides* has evolved a very effective defense mechanism against predators. It produces **-amanitin,** which disrupts mRNA formation in animal cells by blocking Pol II and, at higher concentrations, Pol III. Neither Pol I nor bacterial RNA polymerase is sensitive to _-amanitin—nor is the RNA polymerase II of *A. phalloides* itself.





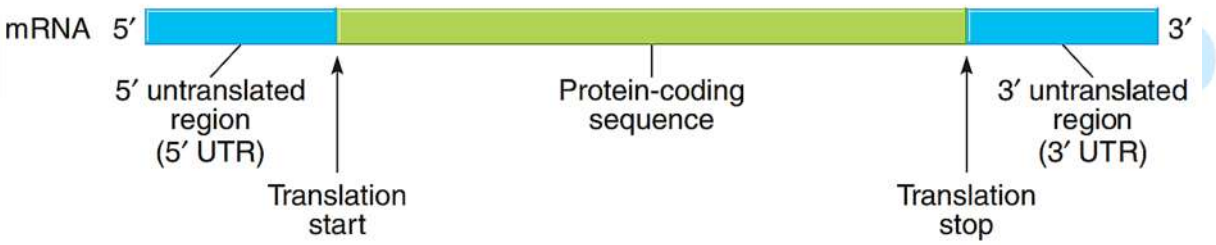
About figure: Actinomycin D and acridine, inhibitors of DNA transcription. (a) The shaded portion of actinomycin D is planar and intercalates between two successive GqC base pairs in duplex DNA. The two cyclic peptide structures of actinomycin D bind to the minor groove of the double helix. Sarcosine (Sar) is *N*-methylglycine; meVal is methylvaline. Acridine also acts by intercalation in DNA. **(b)** A complex of actinomycin D with DNA (PDB ID 1DSC). The DNA backbone is shown in blue, the bases are white, the intercalated part of actinomycin (shaded in **(a)**) is orange, and the remainder of the actinomycin is red. The DNA is bent as a result of the actinomycin binding.

3.3.11 POST TRANSCRIPTIONAL MODIFICATION- PROCESSING OF PRE-mRNA to FUBCTIONAL mRNA:

At the end of transcription process the mRNA generated is not the functional mRNA that can be translated to generate protein to shoe the genetic expression. DNA transcribed to form pre-mRNA molecule in eukaryotic cells. These pre-mRNA molecules then pass through post transcriptional modification to form the functional mRNA molecule.

Three changes occur in pre-mRNA molecules

1. 5' capping to generate 5' untranslated region or 5'-UTR or may also called leader sequence.
2. RNA splicing to generate protein coding region.
3. 3'-polyadenylatio to form 3'-UTR or poly-A tail

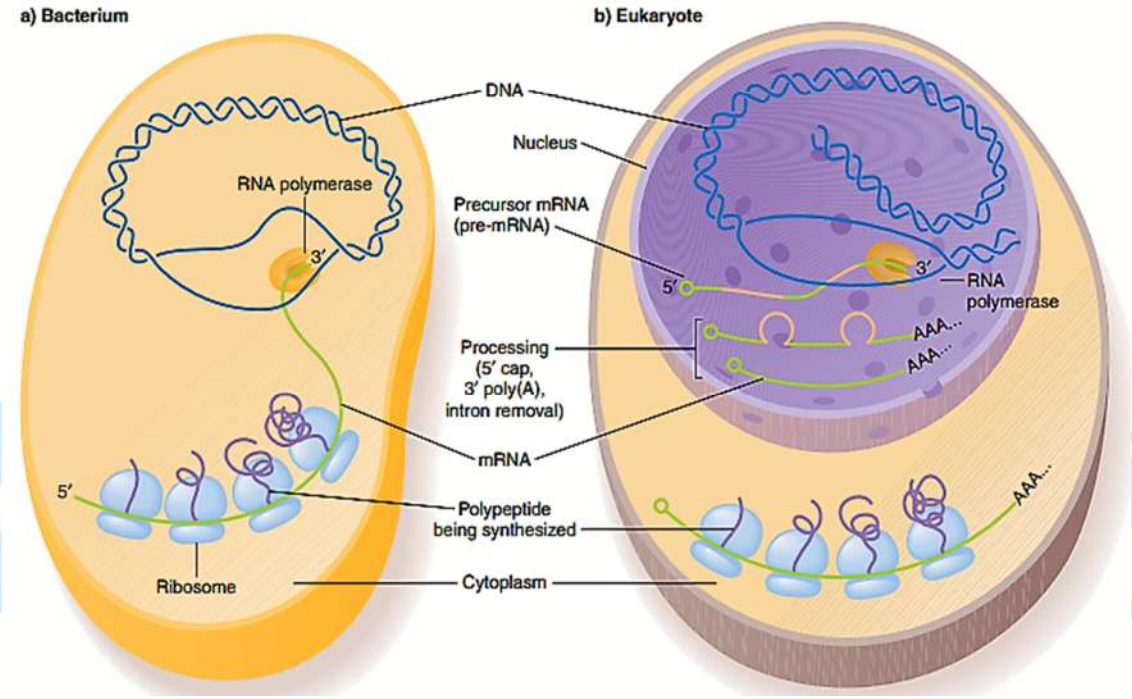


General structure of mRNA found in both bacterial and eukaryotic cell

In prokaryotic cell nucleus is absent so translation of mRNA into protein can begin from the 5'- end of the mRNA while the 3' end is still being synthesized by RNA polymerase. In other words, the transcription and translation process in prokaryotic cells occur simultaneously. However in eukaryotic cells site of transcription and site of translation is not same. Transcription takes place in nucleus and the translation in cytoplasm. In eukaryotic cells the primary transcript (pre-mRNA) is the precursor for functional mRNA molecule. Pre-mRNA

molecules must go through several modifications that collectively called as RNA processing.

Processes for the synthesis of functional mRNA in bacteria and eukaryotes. (a) In bacteria, the mRNA synthesized by RNA polymerase does not have to be processed before it can be translated by ribosomes. Also, because there is no nuclear membrane, mRNA translation can begin while transcription continues, resulting in a coupling of transcription and translation. (b) In eukaryotes, the primary RNA transcript is a precursor-mRNA (pre-mRNA) molecule, which is processed in the nucleus by the addition of a 5′ cap and a 3′ poly(A) tail and the removal of introns. Only when that mRNA is transported to the cytoplasm can translation occur.



5'-Capping: As soon as the nascent RNA transcript emerges out from the RNA channel of **RNA polymerase-II** and reaches a length of **20-25** nucleotides, a protective cap composed of 7-methylguanosine and methylated ribose is added to the 5' end of eukaryotic pre-mRNAs. This 5' cap marks the RNA molecule as the precursor for functional mRNA processing and serves to protect them from RNA digesting enzyme **(5'-exoribonuclease)** in the nucleus and cytoplasm. This initial step of RNA processing is catalysed by a dimeric capping enzyme, which is associated with the phosphorylated **carboxyl terminal** domain **(CTD)** of RNA polymerase-II. Binding to the phosphorylated CTD stimulates the activity of capping enzyme so that it can focus only on the RNAs with 5'- triphosphate that

emerges out RNA polymerase-II and not RNA polymerase-I and RNA polymerase-III that do not have any CTD.

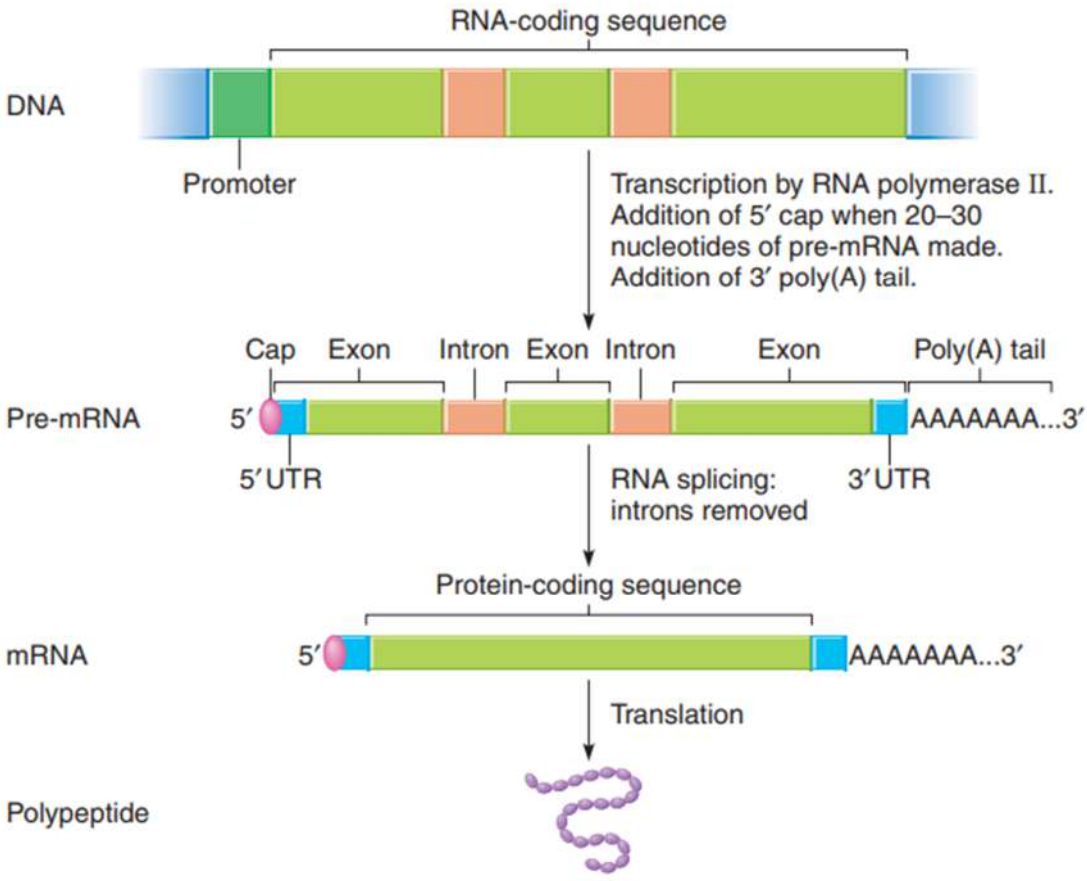
RNA SPLICING:

Before going to discuss more about RNA splicing mechanism we should first look after what is intron and what exon is.

Intron: The **non-coding** (no gene expression or no protein synthesis) region of a pre-mRNA is known as intron.

Exon: The **coding** sequence or the gene (responsible for protein synthesis) on a pre-mRNA is known as the exon.

During formation of mature or functional mRNA from a transcript the **introns** are removed and the **exons** are spliced together.



Introns typically begin with **5'-GU** and end with **AG-3'**, although more than just those nucleotides are needed to specify a junction between an intron and an exon. Introns in pre-mRNAs are removed and exons joined in the nucleus by mRNA splicing. The splicing events occur in a spliceosome, a complex of the pre-mRNA bound to small nuclear ribonucleoprotein particles (snRNPs; pronounced snurps). **snRNPs are small nuclear RNAs** (snRNAs) associated with proteins. The five principal snRNAs **are U1, U2, U4, U5, and U6**; each is associated with a number of proteins to form the snRNPs. U4 and U6 snRNAs are found within the **same snRNP** (U4/U6 snRNP), and the others are found within their own special snRNPs. Each snRNP type is abundant in the nucleus, with at least 105 copies per cell.

1. U1 snRNP binds to the **5' splice** junction of the intron. This binding is primarily the result of base pairing of U1 snRNA in the snRNP to the 5' splice junction.
2. U2 snRNP binds to a sequence called the branch point sequence, which is **located upstream of the 3' splice junction**. This binding occurs as a result of the base pairing of U2 snRNA in the snRNP to the branch-point sequence.
3. A U4/U6 snRNP and a U5 snRNP interact, and the combination binds to the U1 and U2 snRNPs, causing the intron to loop and thereby bringing its two junctions close together.
4. U4 snRNP dissociates, resulting in the formation of the active spliceosome.
5. The snRNPs in the spliceosome cleave the intron from exon 1 at the 5' splice junction, and the now free 5' end of the intron bonds to a particular A nucleotide in the branch-point sequence. Because of its resemblance to the rope cowboys use, the looped back structure is called an RNA lariat structure. The branch point in the RNA that produces the lariat structure involves an unusual 2'–5' phosphodiester bond formed between the 2'-OH of the adenine nucleotide in the

branch-point sequence and the 5’ phosphate of the guanine nucleotide at the end of the intron. The A itself remains in normal 3’–5’ linkage with its adjacent nucleotides of the intron.

6. Next, the spliceosome excises the intron (still in lariat shape) by cleaving it at the 3’ splice junction and then ligates exons 1 and 2 together. The snRNPs are released at this time. The process is repeated for each intron.

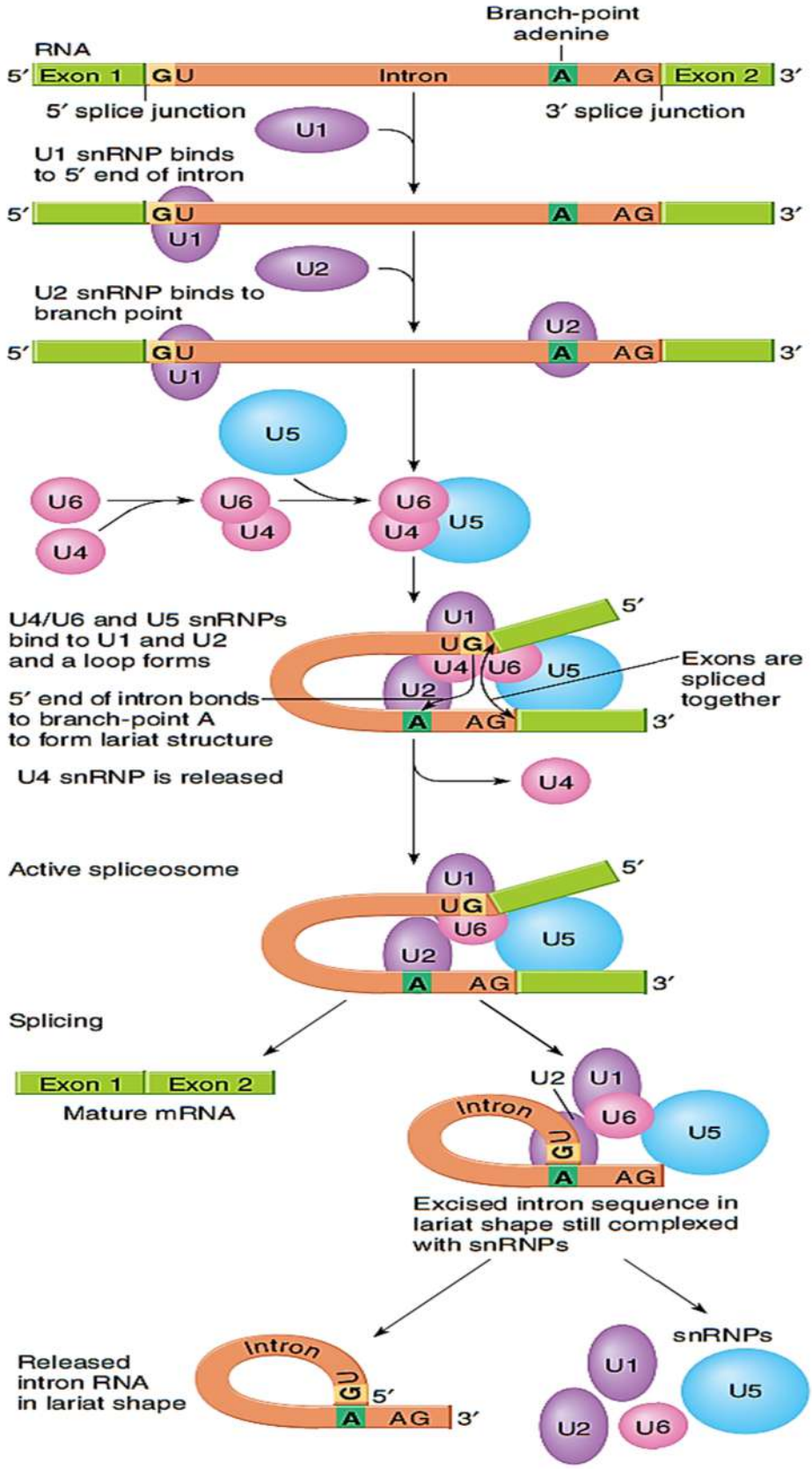

Branch-point adenine
RNA
5′ Exon 1 GU Intron A AG Exon 2 3′
5′ splice junction
3′ splice junction
U1
U1 snRNP binds to 5′ end of intron
U2
U2 snRNP binds to branch point
U5
U6
U4
U4/U6 and U5 snRNPs bind to U1 and U2 and a loop forms
5′ end of intron bonds to branch-point A to form lariat structure
Exons are spliced together
U4 snRNP is released
Active spliceosome
Splicing
Exon 1 Exon 2
Mature mRNA
Intron
Excised intron sequence in lariat shape still complexed with snRNPs
Released intron RNA in lariat shape
snRNPs

Figure- Model for intron removal by the spliceosome. At the 5' end of an intron is the sequence GU and at the 3' end is the sequence AG. Near the 3' end of the intron is an A nucleotide located within the branch-point sequence, which in mammals is YNCURAY, where Y = pyrimidine, N = any base, R = purine, and A = adenine, and in yeast is UACUAAC (the italic A is where the 5' end of the intron bonds). With the aid of snRNPs, intron removal begins with a cleavage at the first exon–intron junction. The G at the released 5' of the intron folds back and forms an unusual 2'–5' bond with the A of the branch-point sequence. This reaction produces a lariat shaped intermediate. Cleavage at the 3' intron–exon junction and ligation of the two exons completes the removal of the intron.

Polyadenylation at 3'-End:

Most eukaryotic mRNAs have a string of 50-250 adenine (A) residues, making up the **poly A tail.** This tail serves as the binding site for one or more specific protein and also it protect mRNA from enzymatic destruction.

The Poly A tail is added in a multistep process. The transcript is extended beyond the site of the polyadenylation, then this extended site is cleaved at the point of poly A addition site by an endonuclease component of a large enzyme complex, again it associated with the CTD of RNA polymerase-II. The mRNA site where the cleavage occurs is marked by two sequence elements one is the highly conserved sequence 5'-AAUAAA-3', a 10 to 30 nucleotides on the 5' side (upstream) of the cleavage site. The other is a less well-defined sequence which is rich in G and U residues, 20 to 40 nucleotides downstream of the cleavage site. As a result of cleavage the free 3' hydroxyl group is generated which defines the end of the mRNA to which A residues are immediately added by **polyadenylate polymerase** which catalyses the reaction

$$RNA + nATP \longrightarrow RNA\text{-}(AMP)_n + nPP_i$$

Where the n = 50 to 250. This enzyme does not require a template but does require the cleaved mRNA as a primer.

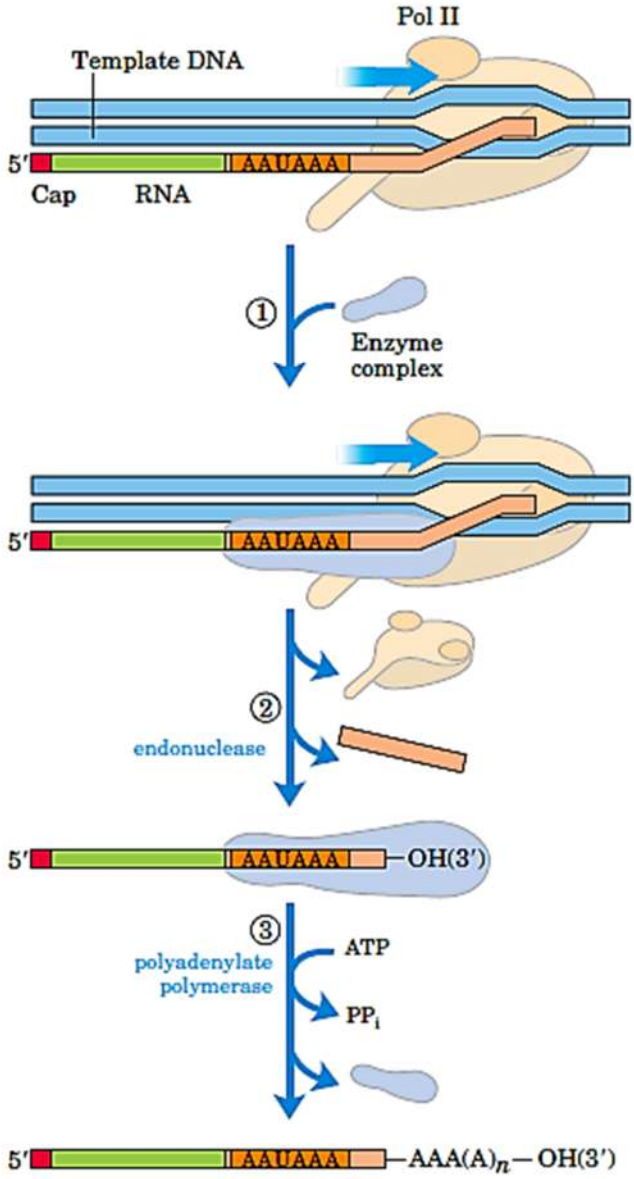


Figure: Addition of the poly(A) tail to the primary RNA transcriptof eukaryotes.
Pol II synthesizes RNA beyond the segment of the transcript containing the cleavage signal sequences, including the highly conserved upstream sequence 5'AAUAAA.

1. The cleavage signal sequence is bound by an enzyme complex that includes an endonuclease, a polyadenylate polymerase, and several other multisubunit proteins involved in sequence recognition, stimulation of cleavage, and regulation of the length of the poly(A) tail.
2. 2 The RNA is cleaved by the endonuclease at a point 10 to 30 nucleotides 3' to (downstream of) the sequence AAUAAA.
3. 3 The polyadenylate polymerase synthesizes a poly(A) tail 80 to 250 nucleotides long, beginning at the cleavage site.

3.4 TRANSLATION

Synthesis of **protein** from information encoded on **mRNA** molecules is known as the translation.

We all know that DNA store the information for protein synthesis and the **mRNA transcribed from DNA** convey the information encoded on DNA molecule. The properties and functions of protein is strictly depend upon its **three dimensional structure**. The linear arrangements of amino acid control the three dimensional structure, so the accurate arrangement of amino acids of a polypeptide chain is very crucial for ultimate **gene expression**. Translation process is the process by which the nucleotide sequences of an mRNA is used as a **template** to join the **amino acids** in a polypeptide chain in **correct order**. Translation process takes place in **cytoplasm**.

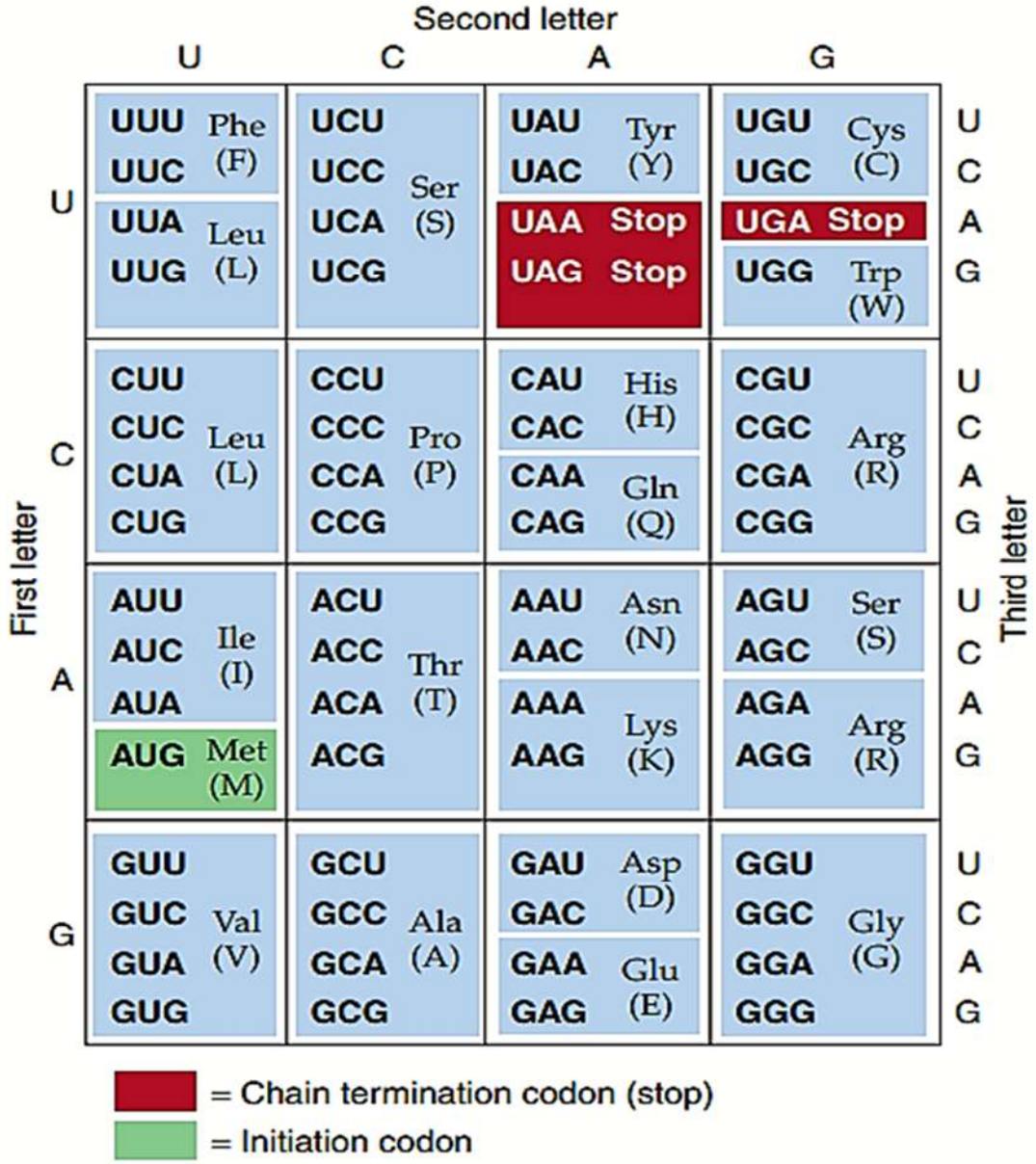
Before going to discuss about translation process we should first have a clear concept about rRNA, Ribosome, tRNA and Codon as each of these plays a very crucial role in translation process.

3.4.1 CODON:

There are about 20 standard amino acids present in nature. As mentioned above the nucleotide sequences present in **mRNA** molecule encode the **amino acid** sequences of a polypeptide chain. The genetic code used by a cell to synthesis amino acid is a **triplet code** or simply **codon** made up of **three** nucleotide sequences. There are 4 types of nucleotide in mRNA molecule **(A,U,G,C)** and three nucleotide combine to generate a codon. So the maximum number of possible codon will be $\mathbf{4^3=64}$. Among these 64 codons **61** code for amino acids. **Three** codons does not code for any amino acid rather they indicate the **termination** of translation process and so these three codons are known as stop codon. **UAA, UAG, UGA** are stop codon. 61 codons code for 20 amino acids, so

there are several codons which code for a single amino acids like **GUU, GUC,GUA, GUG code for a single amino acid Valine** whereas UGG is the only codon which code for tryptophan, AUG is also the only codon which code for methionine.

In prokaryotic and eukaryotic cells synthesis of a polypeptide chain always begins with **methionine** amino acids. The codon which codes for methionine is known as **initiator codon**. In prokaryotic cells **AUG** act as the initiator codon but **CUG** occasionally is used as initiator codon for **methionine in eukaryotic** cell. In few bacterial cells **GUG** is used as initiator codon. Codons are read always from 5'-3' direction. The sequence of codons that runs from a specific start codon to a stop codon is called as a reading frame.


Second letter
U
C
A
G
First letter
Third letter
UUU Phe (F)
UUC
UUA Leu (L)
UUG
UCU Ser (S)
UCC
UCA
UCG
UAU Tyr (Y)
UAC
UAA Stop
UAG Stop
UGU Cys (C)
UGC
UGA Stop
UGG Trp (W)
CUU
CUC Leu (L)
CUA
CUG
CCU Pro (P)
CCC
CCA
CCG
CAU His (H)
CAC
CAA Gln (Q)
CAG
CGU
CGC Arg (R)
CGA
CGG
AUU
AUC Ile (I)
AUA
AUG Met (M)
ACU Thr (T)
ACC
ACA
ACG
AAU Asn (N)
AAC
AAA Lys (K)
AAG
AGU Ser (S)
AGC
AGA Arg (R)
AGG
GUU
GUC Val (V)
GUA
GUG
GCU Ala (A)
GCC
GCA
GCG
GAU Asp (D)
GAC
GAA Glu (E)
GAG
GGU
GGC Gly (G)
GGA
GGG
= Chain termination codon (stop)
= Initiation codon

3.4.2 t-RNA:

tRNA or transfer RNA is the key participant to deciphering the codon in mRNA during translation process or simply protein synthesis process. Each type of amino acids has its own specific tRNA molecule. Depending upon the codon on mRNA, specific tRNA molecule activated and add specific amino acid bound to tRNA molecule to the growing polypeptide chain. To understand the process of understanding the translating language of nucleotides we must know the structure of tRNA molecule in detail.

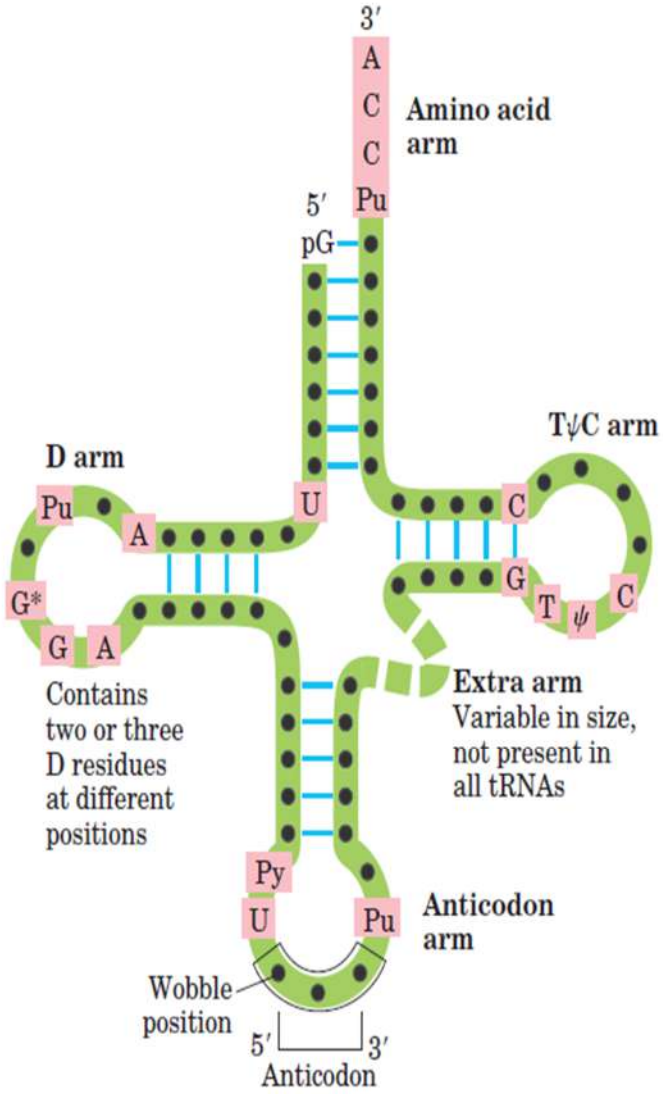
3.4.2.1 STRUCTURE OF tRNA:

There about 73-93 nucleotide residue in a tRNA structure of prokaryotic and eukaryotic cells. tRNA molecule present in the cytosol. Minimum one tRNA molecule is available for one amino acid. Some tRNA can recognize more than one codon. There are at least 32 tRNA to adapt 61 codon. Some cells may contain more than 32 codon, on an average it is said that there are about 40 types tRNA molecules for total 61 codon. Yeast alanine tRNA was the first tRNA to sequence its nucleotide contains. It contain 76 nucleotide chain. The two dimensional structure of a tRNA molecule is like a cloverleaf structure which contains four arms.

Two of the four arms of a tRNA are critical for its adaptor function. One is the anticodon arm which binds with specific codon of mRNA and the other arms is specific for binding of amino acid. The **anticodon arm** contain three nucleotide molecule which pairs with specific **codon** of mRNA during translation process. The **codon-anticodn** binding is a very crucial steps during translation.

The **amino acid** arm contain 5'-CCA-3' sequence at the 3' end of the tRNA molecule. The amino acid arm contain a specific amino esterified by its carboxyl group to the 2' or 3' hydroxyl group of the amino acid residue at the 3' end of the tRNA. These two arm mainly indicate or direct the addition of amino acid with a growing polypeptide chain.

Two other arms are **D arm** and **TψC arm. D arm** contain the unusual nucleotide dihydriuridine (D), and the **TψC** arm contains ribothymidine (T) and pseudouridine (ψ). The D and TψC arms are important for for overall folding of a tRNA molecule. TψC arm crucial for interaction with the large subunit of the rRNA

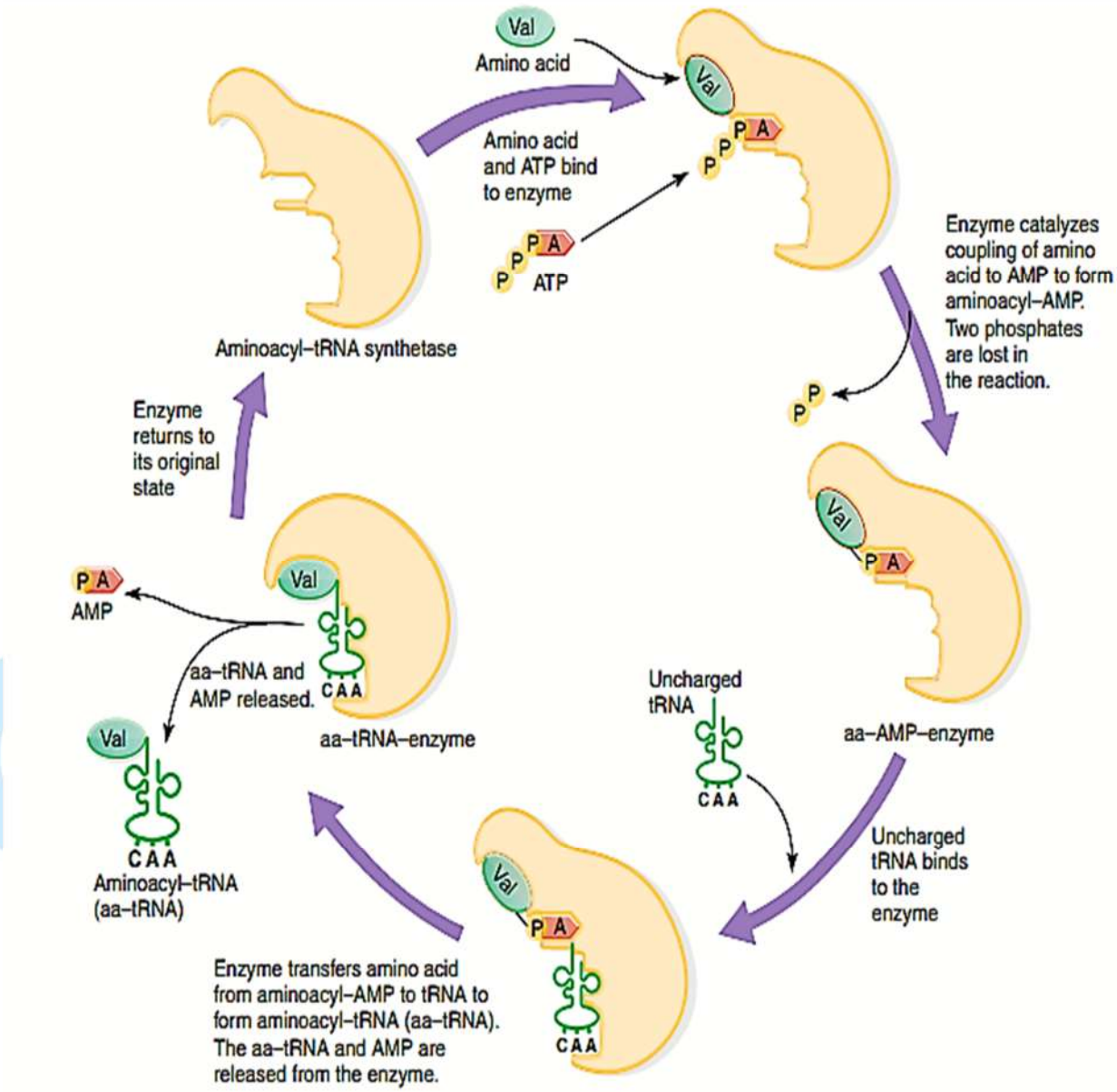


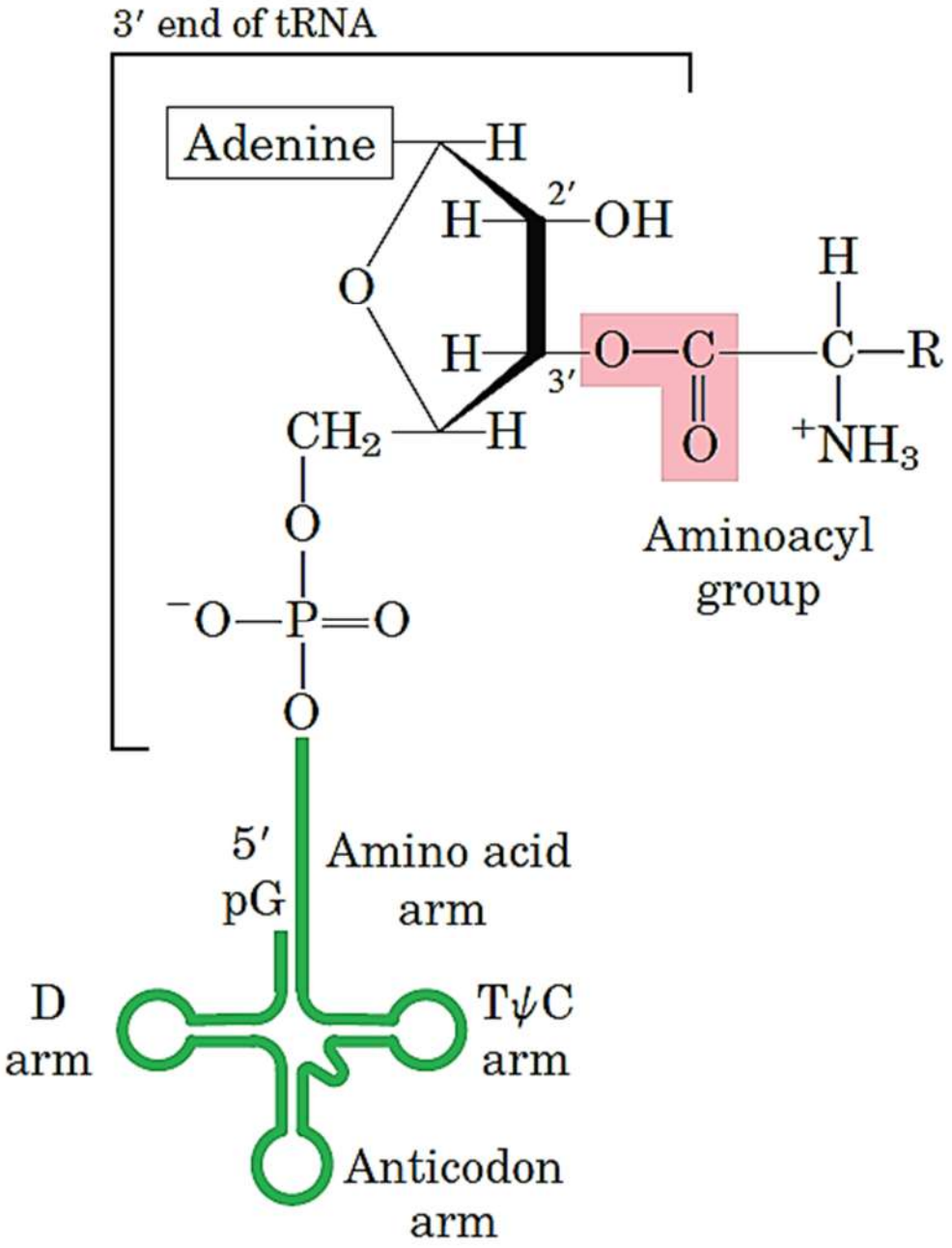
General cloverleaf secondary structure of tRNAs.

The large dots on the backbone represent nucleotide residues; the blue lines represent base pairs. Characteristic and or invariant residues common to all tRNAs are shaded in pink. Transfer RNAs vary in length from 73 to 93 nucleotides. Extra nucleotides occur in the extra arm or in the D arm. At the end of the anticodon arm is the anticodon loop, which always contains seven unpaired nucleotides. The D arm contains two or three D (5,6-dihydrouridine) residues, depending on the tRNA. In some tRNAs, the D arm has only three hydrogen-bonded base pairs. **Pu,** purine nucleotide; **Py**, pyrimidine nucleotide; **G*,** guanylate *or* 2_-*O*-methylguanylate.

3.4.2.2 Mechanism Of Amino Acid Addition At 3'-End:

The correct amino acid is attached to the tRNA by an enzyme called **aminoacyl–tRNA synthetase**. The process is called aminoacylation, or charging, and produces an aminoacyl–tRNA (or charged tRNA). Aminoacylation uses energy from ATP hydrolysis. There are 20 different aminoacyl–tRNA synthetases, one for each of the 20 different amino acids. Each enzyme recognizes particular structural features of the tRNA or tRNAs it aminoacylates. Figure below shows the charging of a tRNA molecule to produce valine–tRNA (Val–tRNA). First, the amino acid and ATP bind to the specific aminoacyl–tRNA synthetase enzyme. The enzyme then catalyzes a reaction in which the **ATP is hydrolyzed to AMP**, which joins to the amino acid as AMP to form **aminoacyl–AMP**. Next, the tRNA molecule binds to the enzyme, which transfers the amino acid from the aminoacyl–AMP to the tRNA and displaces the AMP. The enzyme then releases the aminoacyl–tRNA molecule. Chemically, the amino acid attaches at the end of the tRNA by a covalent linkage between the carboxyl group of the amino acid and the or group of the ribose of the adenine nucleotide found at the end of every tRNA


Val
Amino acid
Amino acid and ATP bind to enzyme
P A
P P
ATP
Aminoacyl–tRNA synthetase
Enzyme catalyzes coupling of amino acid to AMP to form aminoacyl–AMP. Two phosphates are lost in the reaction.
P P
aa–AMP–enzyme
Uncharged tRNA
CAA
Uncharged tRNA binds to the enzyme
Enzyme transfers amino acid from aminoacyl–AMP to tRNA to form aminoacyl–tRNA (aa–tRNA). The aa–tRNA and AMP are released from the enzyme.
aa–tRNA–enzyme
aa–tRNA and AMP released.
AMP
Aminoacyl–tRNA (aa–tRNA)
Enzyme returns to its original state



General structure of aminoacyl-tRNAs. The aminoacyl group is esterified to the 3' position of the terminal A residue. The ester linkage that both activates the amino acid and joins it to the tRNA is shaded pink.

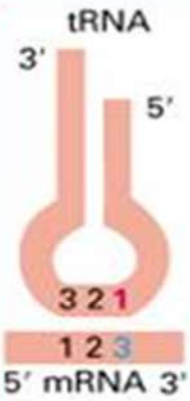
3.4.2.3 Wobble Hypothesis- Mechanism of Codon Binding at Anticodon Arm:

According to the wobble hypothesis proposed by Francis Crick, the complete set of 61 sense codons can be read by fewer than 61 distinct tRNAs, because of pairing properties of the bases in the anticodon.

According to the Wobble hypothesis the third (3') base in the mRNA codon pairs with the corresponding first (5') base in the tRNA anticodon region. These specific position for base pairing in the anticodon arm during translation is called **wobble position**.

There is a peculiar means of codon- anticodon pairing to bind with 61 codons with the help of about 40 tRNA molecules. For example, a given anticodon in tRNA with G in the first wobble position can base pairs with the two corresponding codons that have either pyrimidine (C or U) in the third position of the of the mRNA molecule. The phenylalanine codons 5'-UUU-3' and 5'-UUC-3' are both recognized by the tRNA that has 5'-GAA-3' as the anticodon. Infact any two codons of the type 5'-NNPyr-3' { N= any base; Pyr= pyrimidine} encode a single amino acid and are decoded by a single tRNA with G in the first (wobble) position of the anticodon.

Although adenine rarely is found in the anticodon wobble position. Many tRNA molecules in animals and plant cells contains **Inosine (I),** a **deaminated** product of **adenine**, at this position. Inosine can base pair with A,C and U in a non-standard manner. A tRNA which contain inosine at the wobble position can recognize the corresponding mRNA codons with the A,C or U in the third (wobble) position. For example four of the six codons for leucine (CUA, CUC, CUU and UUA) are all recognized by the same tRNA with the anticodon 3'-GAI-5'.



If these bases are in first, or wobble, position of anticodon

C	A	G	U	I	
G	U	C U	A G	C A U	then the tRNA may recognize codons in mRNA having these bases in third position

If these bases are in third, or wobble, position of codon of an mRNA

C	A	G	U	
G I	U I	C U	A G I	then the codon may be recognized by a tRNA having these bases in first position of anticodon

3.4.3 rRNA and Ribosome:

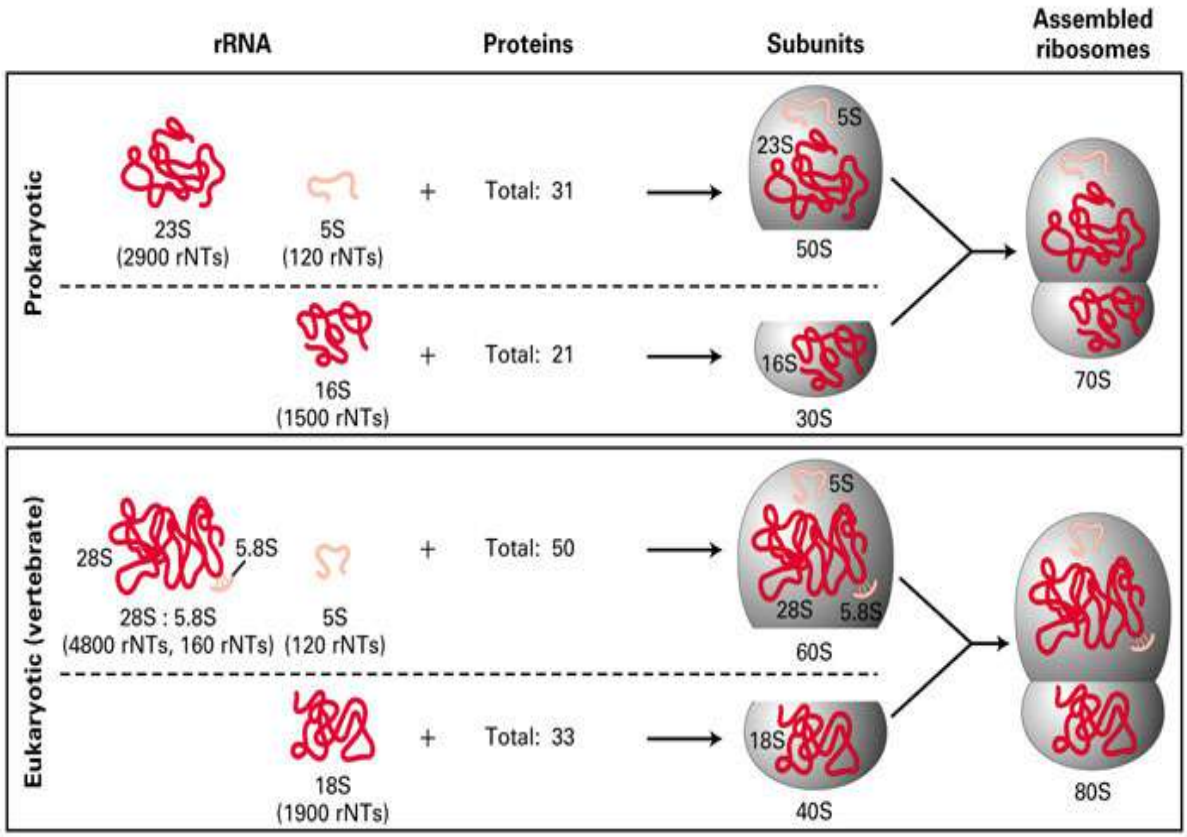
Ribosomes of all species shows some similar structure. A ribosome is composed of three (prokaryote) or four (eukaryote) different rRNA molecules and it also contains as many as **83 proteins** (eukaryotic) organized into a large and a small subunit structure. The ribosomal subunits and the rRNAs are generally designated in **Svedberg units (S)** depending upon the sedimentation rate. Each subunit contains one or more rRNA molecules and a large number of ribosomal proteins. The bacterial or the prokaryotic ribosome has a size of 70S and consists of two subunits of sizes 50S (large subunit) and 30S (small subunit). Eukaryotic ribosomes are larger and more complex than prokaryotic ribosome, and they vary in size and composition among eukaryotic organisms. Mammalian ribosomes, for example, have a size of 80S and the large subunit is of 60S and the small subunit is of 40S. Each ribosomal subunit contains one or more specific rRNA molecules and a number of ribosomal proteins. Bacterial ribosomes contain three rRNA molecules—the 23S rRNA and 5S rRNA in the large subunit, and the 16S rRNA in the small subunit. Eukaryotic ribosomes contain four rRNA molecules—the 28S rRNA, 5.8S rRNA, and 5S rRNA in the large subunit, and the 18S rRNA in the small subunit. The 5.8S rRNA base paired with the 28S rRNA in the eukaryotic large ribosomal subunits.

During translation, the mRNA passes through the small subunit of the ribosome. Specific sites of the ribosome bind tRNAs at different stages of polypeptide synthesis: the A (aminoacyl) site is where an incoming aminoacyl–tRNA binds, the P (peptidyl) site is where the tRNA carrying the growing polypeptide chain is located, and the E (exit) site is where a tRNA binds on its path from the P site to leaving the ribosome. The P and A sites consist of regions of both the large and small subunits, whereas the E site is a region of the large subunit. We will learn

more about these sites in the discussion of the steps of translation in the next three sections

The overall efficiency of the translation process is greatly increased by the interaction of mRNA and specific aminoacyl-tRNAs to aribosome molecule. Ribosome is the most abundant protein complex structure which directs the elongation process during translation process at a rate of three to five amino acids added with the help of ribosome to the growing polypeptide chain at a minute. So it takes about 1minute or less time to translate a polypeptide chain containing 100-200 amino acids residue. On the other hand it takes about 2-3 hours to synthesize the largest protein **titin**, found in the muscle cell.

During translation process ribosome molecule moves along the length of mRNA molecule and interact with various protein factors and tRNAs.



3.4.4 TRANSLATION MECHANISM:

Translation process is quite similar in prokaryotic and eukaryotic cells, though the interacting translation factors may little bit different. In both eukaryotic and prokaryotic cells translation process completed in three steps. **Initiation, Elongation and Termination.**

3.4.4.1 MECHANISM OF PROTEIN SYNTHESIS:

The first step of the protein synthesis in both eukaryotic and prokaryotic cells is the activation tRNA that bind with the initiator codon AUG on the mRNA molecule. Protein synthesis starts from the amino terminal of the growing polypeptide chain. The initiator codon AUG which code for methionine amino acid, provide the amino terminal of the growing polypeptide chain.

1.**Activation of tRNA:**

Generally AUG is the only codon which code for methionine amino acid. But the tRNA for addition of methionine to the initiator site and the tRNA which add the methionine in any site other than the initiator site of a polypeptide chain are different. The tRNA which add the methionine in the initiator region is the $tRNA^{fMet}$ or tRNA.fMet and the tRNA which add methionine in the interior part of the polypeptide chain is denoted as the $tRNA^{Met}$ or tRNA.Met.

In bacteria, the initiator tRNA is tRNA.fMet, which has the anticodon 5¿-CAU-3¿ to bind to the AUG start codon. This tRNA carries a modified form of methionine, **formylmethionine (fMet).** The tRNA containing this fMet reaches to the ribosome in the form of fMet-$tRNA^{fMet}$ or fMet.tRNA.fMet (**N-formylmethionyl-$tRNA^{fMet}$).** Formation of fMet.tRNA.fMet is a two steps reaction. First methionine added to the tRNA.fMet y the help of Met-tRNA synthatase (a catalase enzyme)

Methionine + tRNA.fMet + ATP ⟶ Met.tRNA.fMet + AMP + PP_i.

Next a transformylase enzyme transfers a formyl group from N^{10}-Formyltetrahydroflorate to the amino group of the Met residue.

N^{10}-Formyltetrahydroflorate + Met.tRNA.fMet ⟶ fMet.tRNA.fMet + tetrahydroflorate.

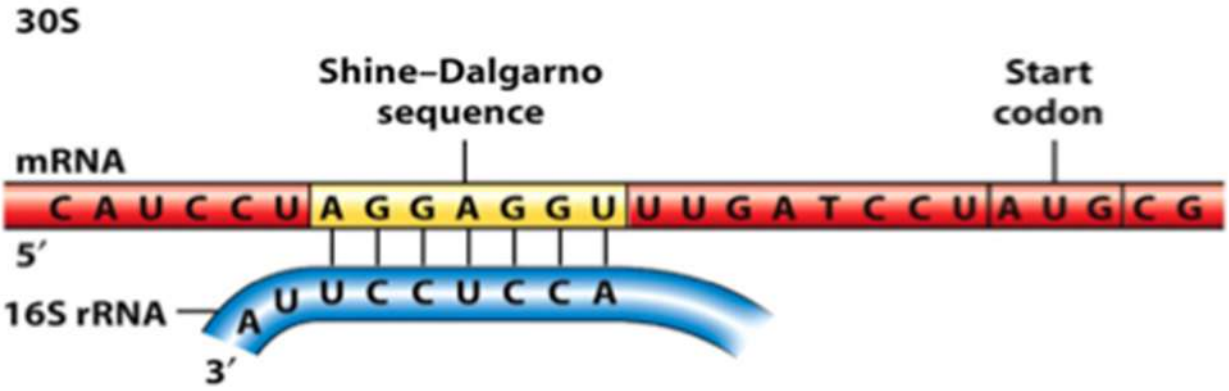
The enzyme transformylase is very selective for only addition of methionine in the initiation site where as the Met.tRNA synthatase enzyme selective for the addition of methionine in any interior site of the growing polypeptide chain.

3.4.4.2 INITIATION: the translational initiation process in bacterial cells requires

i. The 30S ribosomal subunit
ii. The mRNA chain
iii. Initiating fMet-tRNA
iv. Three important regulatory protein IF-1,IF-2,IF-3
v. GTP
vi. 50S ribosomal subunits
vii. Magnesium ion (Mg^{2+})

The AUG initiation codon binds with the 30S subunit of the ribosome molecule in bacteria. AUG codon alone is not sufficient to indicate where the 30S subunit should bind to the mRNA; a sequence upstream (to the 5' side in the leader of the mRNA) of the AUG called the **ribosome-binding site (RBS)** is also needed. In the 1970s, John Shine and Lynn Dalgarno hypothesized that the purine-rich RBS sequence **(5'-AGGAG-3'** or some similar sequence) and sometimes other nucleotides in this region could pair with a complementary pyrimidine-rich region (always containing the sequence **5'-UCCUCC-3'**) at the 3' end of 16S

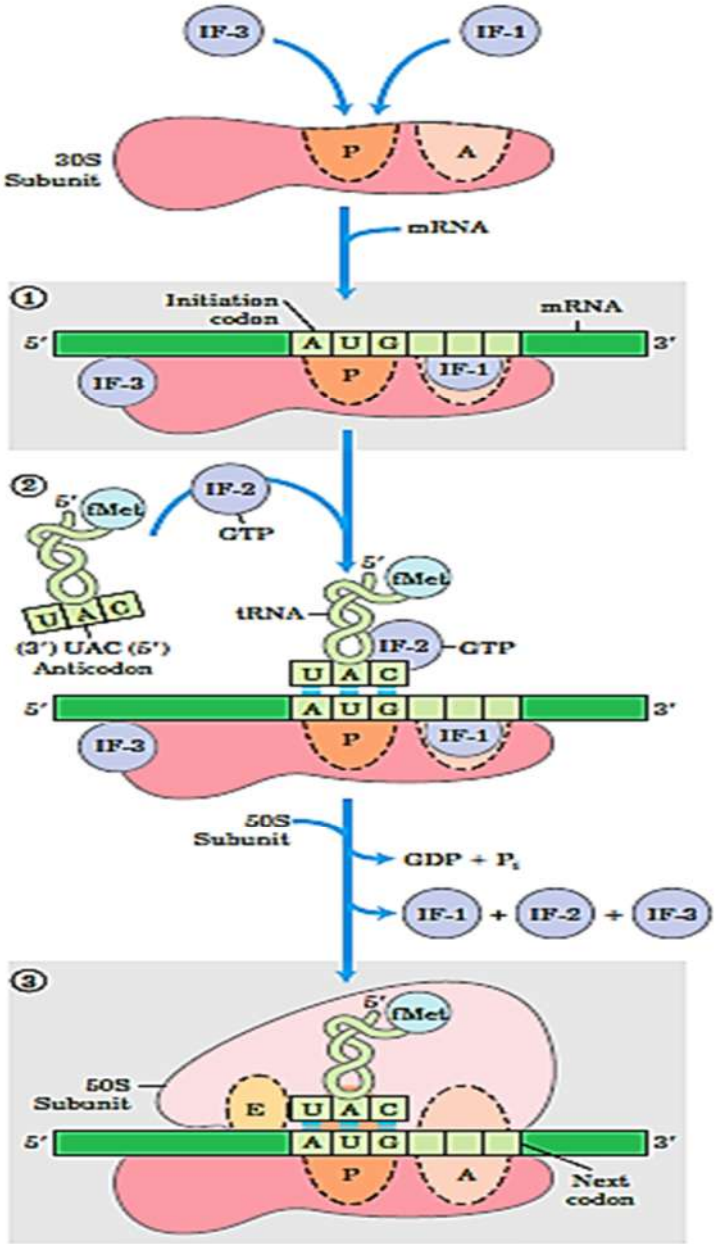
rRNA. Joan Steitz was the first to demonstrate this pairing experimentally. The mRNA RBS region is now commonly known as the **Shine–Dalgarno** sequence.



3.4.5 Initiation In Bacteria:

1. 30S ribosomal subunit first binds with the two initiation factor IF-1 and IF-3.
2. Then the mRNA binds to the 30S subunit. As it was earlier mention that the ribosomes have three sites that **bind aminoacyl-tRNAs, the aminoacyl (A) site, the peptidyl (P) site, and the exit (E) site.** Both the 30S and 50S subunit contain A and P site. But E site is only present in 50S subunit of the ribosomes. The initiation codon AUG binds at the P site of the 30S subunit
3. Then the fMet.tRNA.fMet bind with the AUG codon of mRNA with the help of anticodon 5'-CAU-3' at the P site. Along with this tRNA mRNA is also joined by GTP bound IF-2.
4. During elongation phase all other tRNAs along with Met.tRNA.Met bind first to the A site.
5. Initiation factor IF-1remain binds at the A site to prevent binding of tRNAs at A site.

6. Completion of these steps forms a functional 70S ribosome called the **Initiation complex,** containing the mRNA and the initiation fMet.tRNA.fMet



3.4.6 Initiation In Eukaryotic Cell:

The initiation of translation is similar in eukaryotes, although the process is more complex and involves many more initiation factors, called eukaryotic initiation factors (eIF), than is the case in bacteria. The main differences are that: (1) the initiator amino acid methionine is unmodified, although a special initiator tRNA still brings it to the ribosome; and (2) Shine–Dalgarno sequences are not found in eukaryotic mRNAs. Instead, the eukaryotic ribosome uses another way to find the AUG initiation codon. First, a eukaryotic initiator factor eIF-4F— a polymer of several proteins, including eIF-4E, a special cap binding protein (CBP)—binds to the cap at the 5' end of the mRNA molecule. Then, a complex of the 40S ribosomal subunit with the initiator Met–tRNA, several eIF proteins, and GTP binds, together with other eIFs, and moves along the mRNA, scanning for the initiator AUG codon. The **AUG** codon is embedded in a short sequence—called the **Kozak sequence**, after Marilyn Kozak—which indicates that it is the initiator codon. This process is called the scanning model for initiation. The AUG codon is almost always the first AUG codon from the 5' end of the mRNA; but, to be an initiator codon, it must be in an appropriate sequence context. Once the 40S subunit finds this AUG, it binds to it, and then the **60S** ribosomal subunit binds, displacing the eIFs (**except for eIF-4F**, which is needed for the subsequent initiation of translation), producing the **80S** initiation complex with the initiator **Met–tRNA** bound to the mRNA in the **P** site of the ribosome. The poly(A) tail of the eukaryotic mRNA also plays a role in translation. Poly(A) binding protein II (**PABPII**) bound to the poly(A) tail also binds to **eIF-4G**, one of the proteins of eIF-4F at the cap, thereby looping the 3' end of the mRNA close to the 5' end. In this way, the poly(A) tail stimulates the initiation of translation.

Factor	Function
Bacterial	
IF-1	Prevents premature binding of tRNAs to A site
IF-2	Facilitates binding of fMet-tRNA^{fMet} to 30S ribosomal subunit
IF-3	Binds to 30S subunit; prevents premature association of 50S subunit; enhances specificity of P site for fMet-tRNA^{fMet}
Eukaryotic*	
eIF2	Facilitates binding of initiating Met-tRNA^{Met} to 40S ribosomal subunit
eIF2B, eIF3	First factors to bind 40S subunit; facilitate subsequent steps
eIF4A	RNA helicase activity removes secondary structure in the mRNA to permit binding to 40S subunit; part of the eIF4F complex
eIF4B	Binds to mRNA; facilitates scanning of mRNA to locate the first AUG
eIF4E	Binds to the 5′ cap of mRNA; part of the eIF4F complex
eIF4G	Binds to eIF4E and to poly(A) binding protein (PAB); part of the eIF4F complex
eIF5	Promotes dissociation of several other initiation factors from 40S subunit as a prelude to association of 60S subunit to form 80S initiation complex
eIF6	Facilitates dissociation of inactive 80S ribosome into 40S and 60S subunits

3.4.7 ELONGATION PHASAE:

After the completion of formation of initiation complex the elongation of polypeptide chain starts. This phase has three steps.

1. Activated tRNA binds to the ribosome in the A site.
2. A peptide bond forms
3. The ribosome moves (translocate) along the mRNA codon.

These three steps are repeated as many times as there are amino acid residues to be added.

Several protein factors are also essential in this phase. These are called **EF** or **elongation factors**. **GTP** is also plays an important role in elongation process. There are three elongation factors in bacterial cells the EF-Tu, EF-Ts, EF-G.

Binding of Aminoacyl-tRNA: As said earlier that during elongation phase aminoacyl tRNAs binds with the codon in the A site of ribosome not in the P site.

For example we are considering Ser-tRNA.Ser (addition of Serine amino acid in the growing polypeptide chain) this aminoacyle tRNA binds with the complex of GTP bound EF-Tu. Now this resulting aminoacyle-tRNA-EF-Tu-GTP complexbinds to the A site of the 70S initiator complex. When the aminoacyl-tRNA binds to the codon in the A site, GTP hydrolysis releases EF-Tu–GDP. Now the EF-Tu is recycled. First, a second elongation factor, EF-Ts, binds to EF-Tu and displaces the GDP. Next, GTP binds to the EF-Tu–EF-Ts complex to produce an EF-Tu–GTP complex simultaneously with the release of EF-Ts. An aminoacyl-tRNA binds to the EF-Tu–GTP, and that complex can bind to the A site in a ribosome when the complementary codon is exposed. The process is highly similar in eukaryotes, with eEF-1A playing the role of EF-Tu, and eEF-1B playing the role of EF-Ts.

3.4.8 Peptide Bond Formation: The ribosome maintains the two aminoacyl–tRNAs in the P and A sites in the correct positions, so that a peptide bond can be formed between the two amino acids. Two steps are involved in the formation of this peptide bond. First, the bond between the amino acid and the tRNA in the P site is degraded. In this case, the breakage is between the fMet and its tRNA. Second, the peptide bond is formed between the fMet and the Ser attached to the aminoacyle-tRNA in the A site. this reaction is catalyzed by peptidyl transferase enzyme. Once the peptide bond has formed, a tRNA without an attached amino acid (an uncharged tRNA) is left in the P site. The tRNA in the A site, now called peptidyl–tRNA, has the first two amino acids of the polypeptide chain attached to it—in this case, fMet–Ser.

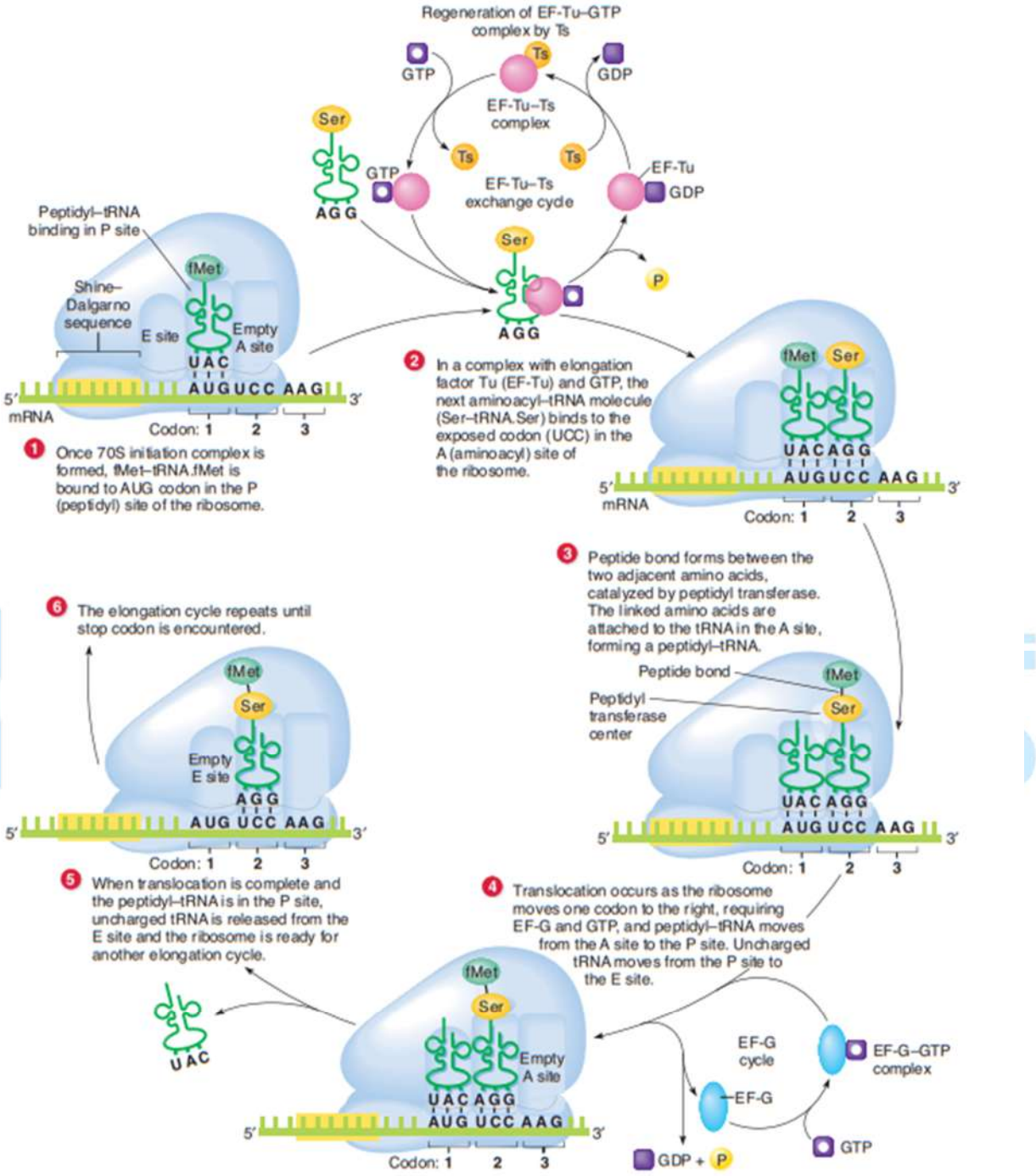
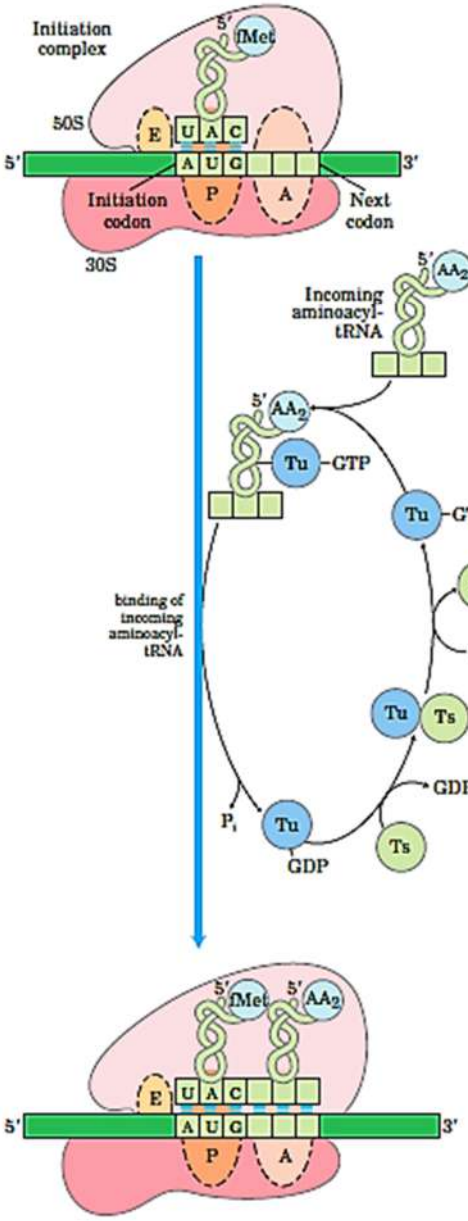
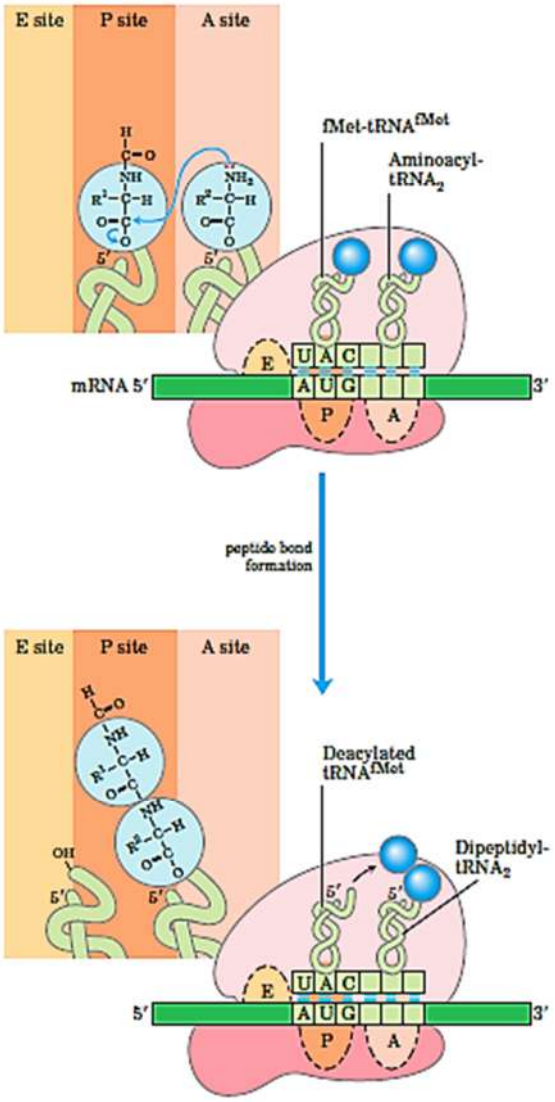


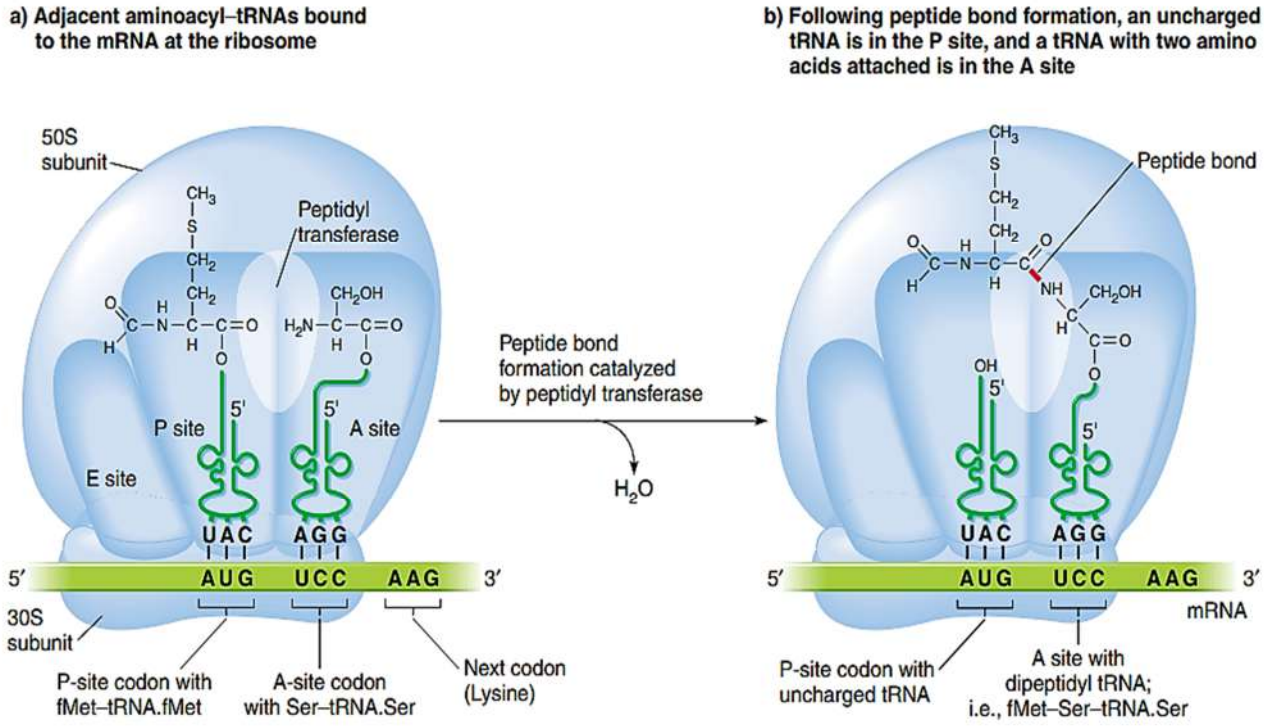
Figure: Elongation stage of translation in bacteria. For the EF-Tu and EF-Ts proteins, the "u" stands for unstable, while the "s" stands for stable.



First elongation step in bacteria: binding of the second aminoacyle-tRNA.



formation of thefirst peptide bond.



The formation of a peptide bond between the first two amino acids (fMet and Ser) of a polypeptide chain is catalyzed on the ribosome by peptidyl transferase

3.4.9 Translocation: In the last step in the elongation cycle, translocation, the ribosome molecule moves one codon along the mRNA toward the 3' end. In bacteria, translocation requires the activity of another protein elongation factor, **EF-G**. The **EF-G–GTP** complex binds to the ribosome, GTP is then **hydrolysed**, and translocation of the ribosome occurs along with displacement of the uncharged tRNA away from the P site. It is possible that GTP hydrolysis changes the three dimensional structure of EF-G, which facilitates the translocation process. Translocation is similar in eukaryotes; the elongation factor in eukaryotic cells is eEF-2, which acts like bacterial EF-G. The uncharged tRNA then moves from the P site and then binds transiently at the E site in the 50S ribosomal subunit, blocking the next aminoacyle–tRNA from binding to the A site until the

translocation process get completed and the peptidyl–tRNA become bound correctly in the P site. Once that has occurred, the uncharged tRNA is then get released from the ribosome. After translocation, EF-G is released and then reused, as shown in . During the translocation step, the peptidyl–tRNA remains attached to its codon on the mRNA; and because the ribosome has moved, the peptidyl–tRNA is now located in the P site (hence the name peptidyl site). After the completion of translocation, the A site is now vacant. An aminoacyl–tRNA with the correct anticodon binds to the newly **exposed codon of mRNA chain** in the A site, reiterating the process already described. The whole process is then repeated until translation terminates at a stop codon. In both bacteria and eukaryotes, once the ribosome moves away from the initiation site on the mRNA, another initiation event occurs. The process is repeated until, typically, several ribosomes are translating each mRNA simultaneously. The complex between an mRNA molecule and all the ribosomes that are translating it simultaneously is called a **polyribosome**, or **polysome**. Each ribosome in a polysome translates the entire mRNA and produces a single, complete polypeptide. Polyribosomes enable a large number of polypeptides to be produced quickly and efficiently from a single mRNA.

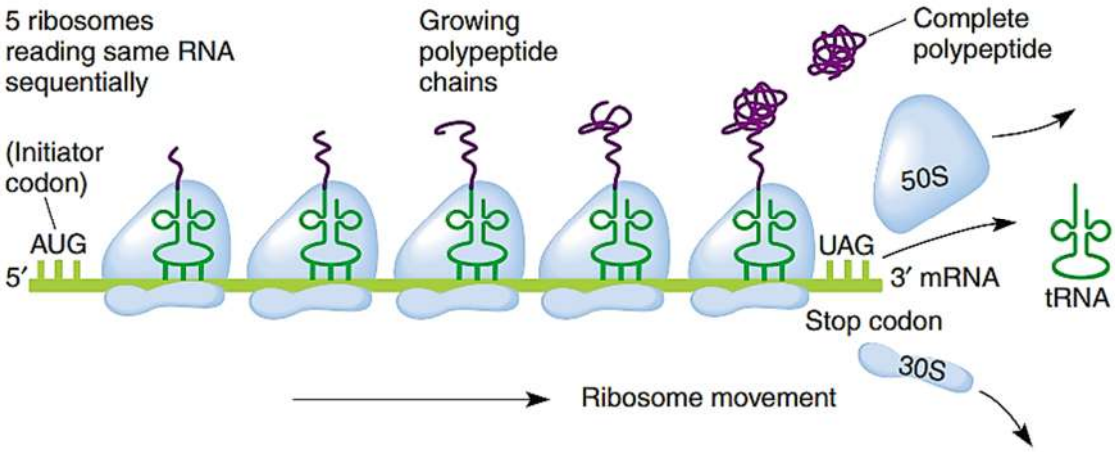


Diagram of a polysome—a number of ribosomes, each translating the same mRNA sequentially

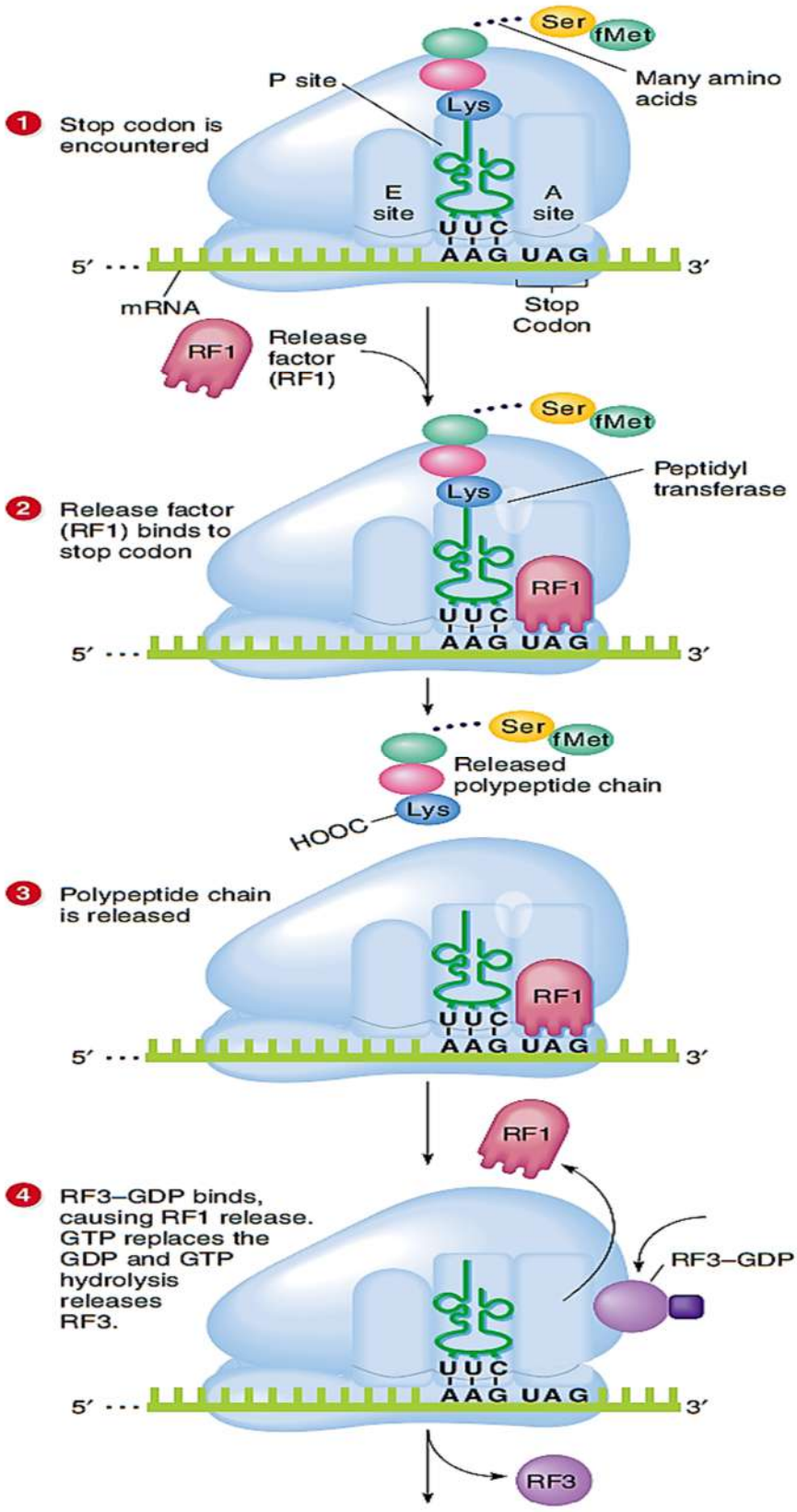
3.4.10 Termination of Translation: The termination of translation is facilitate by one of the three stop codons (UAG, UAA, and UGA), which are the same in prokaryotes and eukaryotes **(step 1).** The stop codons code for no amino acids, so no tRNAs in the cell have anticodons for them. The ribosome recognizes a stop codon with the help of proteins called release factors (RF), which have shapes **mimicking that of a tRNA** including regions that read the codons **(step 2)** and then initiate a series of specialised termination events. In E. coli, there are **three RFs**, two of which read the stop codons: **RF1 recognizes UAA** and **UAG**, and **RF2 recognizes UAA and UGA**— here in the figure RF1 is shown binding to UAG. The binding of RF1 or RF2 to a stop codon activate **peptidyl transferase** to cleave the polypeptide from the tRNA **in the P** site **(step 3).** The polypeptide then released from the ribosome. Next, **RF3–GDP** binds with the ribosome, stimulating the release of the **RF from the stop codon** and the

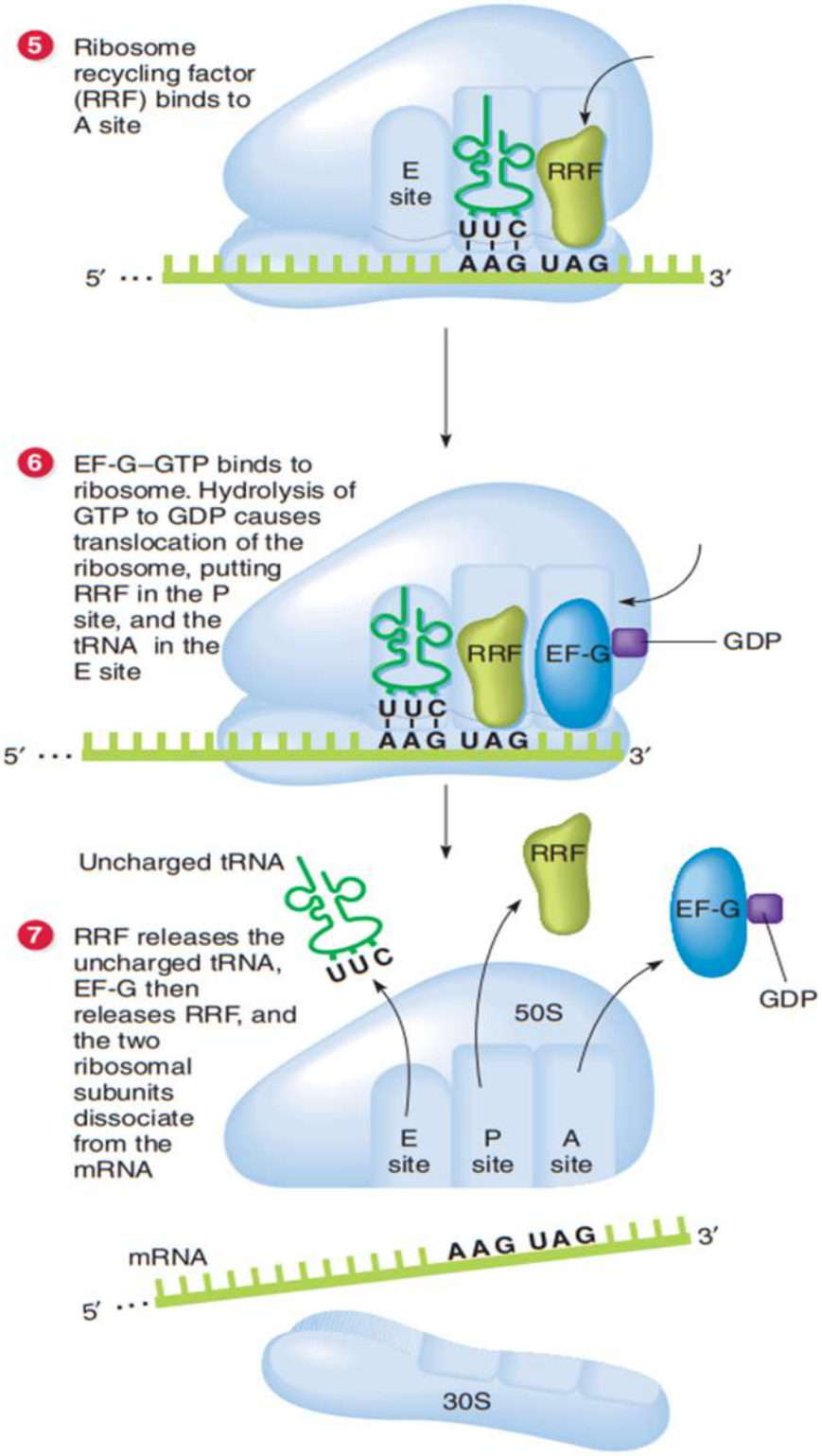
ribosome (step 4). Now the GDP is replaced by GTP on RF3, and RF3 hydrolyses the GTP, which allows RF3 to be released from the ribosome.

An additional important step is the deconstruction of the remaining complex of ribosomal subunits, mRNA, and uncharged tRNA so that the ribosome and tRNA may be recycled. In E. coli, ribosome recycling factor (RRF)—the shape of which is similar to that of a tRNA— binds to the A site **(step 5).** Then EF-G binds, causing translocation of the ribosome and thereby moving RRF to the P site and the uncharged tRNA to the E site **(step 6).** The RRF then releases the uncharged tRNA, and EF-G releases RRF, causing the two ribosomal subunits to dissociate from the mRNA **(step 7).**

In eukaryotes, the termination process is similar to that in bacteria. In this case, a single release factor— **eukaryotic release factor 1 (eRF1)**—recognizes all **three stop codons**, and **eRF3** stimulates the termination events. Ribosome recycling occurs in eukaryotes, but there is no equivalent of RRF.

The nascent polypeptide chain folded during the translation process to form a functional protein molecule.


Ser
fMet
Many amino acids
P site
Lys
1 Stop codon is encountered
E site
A site
U U C
A A G U A G
5′
3′
mRNA
Stop Codon
RF1
Release factor (RF1)
Ser
fMet
Peptidyl transferase
Lys
2 Release factor (RF1) binds to stop codon
RF1
U U C
A A G U A G
5′
3′
Ser
fMet
Released polypeptide chain
Lys
HOOC
3 Polypeptide chain is released
RF1
U U C
A A G U A G
5′
3′
RF1
4 RF3–GDP binds, causing RF1 release. GTP replaces the GDP and GTP hydrolysis releases RF3.
RF3–GDP
U U C
A A G U A G
5′
3′
RF3


5 Ribosome recycling factor (RRF) binds to A site
E site
RRF
UUC
AAG UAG
5′
3′
6 EF-G–GTP binds to ribosome. Hydrolysis of GTP to GDP causes translocation of the ribosome, putting RRF in the P site, and the tRNA in the E site
RRF
EF-G
GDP
UUC
AAG UAG
5′
3′
Uncharged tRNA
7 RRF releases the uncharged tRNA, EF-G then releases RRF, and the two ribosomal subunits dissociate from the mRNA
RRF
EF-G
GDP
UUC
50S
E site
P site
A site
mRNA
AAG UAG
5′
3′
30S

3.4.11 INHIBITORY EFFECTS OF TOXINS AND ANTIBIOTICS ON TRANSLATION PROCESS:

There are several antibiotics which are extremely toxic to harmful bacteria. As every steps of protein synthesis can be precisely inhibited by specific antibiotic or a group of antibiotics, antibiotics have become valuable tools in the study of protein biosynthesis.

Puromycin: made from the mold ***Streptomyces alboniger***. Its structure is very much similar to the **3'** end of an **aminoacyle-tRNA**, which enable it to bind with the **ribosomal A site** and hamper the **peptide bond formation** by forming a peptidyl-puromycin. This permanently terminates polypeptide synthesis. And thus inhibit protein synthesis in harmful bacteria.

Tetracyclines: inhibit protein synthesis in bacterial cells by blocking the **A site** on the ribosome which prevent the binding of **aminoacyl-tRNAs**.

Chloramphenicol: It inhibits protein synthesis by blocking peptidyl transfer on ribosome in bacteria. **It doesn't affect cytosolic protein synthesis** in eukaryotes but it inhibits translation in mitochondria and chloroplast.

Cycloheximide: It blocks the peptidyl transferase on **80S ribosome** (eukaryotic) but **not** that in 70S bacterial ribosome.

Streptomycin: a basic **trisaccharide**, causes misreading of the genetic code in bacteria at relatively **low** concentrations and inhibits **initiation** at higher concentrations.

Diphtheria Toxin: It catalyses the ADP-ribosylation of a dipthamide (a modified histidine) residue of eukaryotic elongation factor **eEF2** and thereby inactivating it in human.

Ricin: an extremely toxic protein of the **castor bean** inactivates the **60S subunit** of eukaryotic ribosomes by **depurinating** a specific adenosine in **23S** rRNA.

KEY CONCEPT ABOUT TRANSLATION:

Translation is the synthesis of polypeptide chain from the mRNA transcribed from DNA by means of transcription process. Codon is a three nucleotide containing sequence in the mRNA which code for an amino acid or stop translation process. There are 64 codon of which 61 code for amino acid and three codon are stop codon which code for the termination of protein synthesis. In the translation process mRNA, tRNA, ribosome-rRNA complex take part.

tRNA structure have a anticodon arm. As like codon anticodon arm contain a special three nucleotide sequence which can bind with codon of the mRNA according to the wobble hypothesis. There are about 32-40 types of tRNA in all organisms. Each tRNA is unique in anticodon sequence. tRNAs are present in the cytoplasm.

Ribosome is present in cytoplasm. Ribosome has two subunit large and small subunits. All subunits contain specific rRNAs and some proteins. Ribosomes structurally have three site A, P and E site. These three sites function in a specific manner during protein synthesis.

Several factors like initiation factors (IF), elongation factors (EF) are essential for the translation process.

Addition of specific amino acid depending upon codon to which anticodon of a tRNA bind make a tRNA molecule charged or active. Activated tRNA or charged tRNAs take part in the protein synthesis. The codon AUG which code for methionine is the initiation codon for translation process. The process of activation of a tRNA molecule is known as the aminoacylation of the tRNA.

Depending upon the codon sequences in the mRNA molecule specific activated tRNAs for specific codon take part to add coded amino acids in the growing

polypeptide chain after the methionine addition completed. This is the elongation process of translation.

Subsequently the completion of amino acid addition in a polypeptide chain there is the process of peptide bond formation starts with the help of specific translation factors.

The completion of elongation process or the termination process of translation is signalled by any of the three stop codon. The direction of translation process 5'-3' direction. During elongation process the ribosome move towards 3'- direction along the length of mRNA chain.

MOST IMPORTANT POINT FOR CSIR-NET EXAM:

1. Concept of chromosome, DNA, Gene.
2. Structure of tRNA, Ribosome
3. Type of ribosomes.
4. Difference between RNA polymerase and DNA polymerase.
5. Role of different enzymes involved in Replication process.
6. Okazaki fragments.
7. Addition of Okazaki fragments.
8. Concept of direction of Replication, Transcription, Translation.
9. Codons and their respective amino acids.
10. Coding strand and noncoding strand.
11. Sigma factor
12. Initiation of translation
13. Aminoacylation of tRNA molecule.
14. Site of addition of mRNA, tRNA on ribosome during translation.
15. Intron, exon, cistron
16. Inhibitors of protein synthesis.
17. Central dogma theory.
18. Wobble hypothesis.
19. Role of different types of RNA
20. Pribnow box, TATA box, consensus sequences.
21. Rho dependent and Rho independent termination process.
22. Movement of Ori-C.
23. Concept of nuclease, exonuclease, endonuclease.
24. Concept of DNA damage mechanism by methylation.
25. Different type of point mutations.