COURSE: MSc Part -II

PAPER – XI

TOPIC- Molecular Biology (3)

PREAPARED BY: Ajai Kishore Sharan

Topic- 7

Transcription and Translation in Prokaryotes

A cell is a chemical factory in which protein are the workhorses. These proteins control every chemical reaction which regulates the function of life. Proteins are long chains of amino acids, and the sequence of the amino acids determines the final structure and function of the protein. Instructions for the sequence of the amino acid are encoded in genes. In order to make a particular protein, a messenger ribonucleic acid (mRNA) copy is copied from the DNA (the process is transcription). Two ribosomal sub units (rRNA) link into mRNA to begin protein synthesis called as translation. As soon as mRNA links with rRNA initiation begins. This is followed by elongation in which successive amino acids are added to the growing chain brought in by transfer RNAs (t RNAs). mRNA copies the message from DNA, a three lettered language called as codons and this forms a bond with three nucleotides of tRNA, called the anticodon. After the chain elongation is completed, the ribosome (rRNA) unbinds from the mRNA, and the process is called as termination. The protein thus formed is called as single helix protein, which folds variously to make the secondary structure the tertiary structure to make the final, functional protein. Prokaryotes do not have membrane-enclosed nuclei. Therefore, the processes of transcription, translation, and mRNA degradation can all occur simultaneously. The intracellular level of a bacterial protein can quickly be amplified by multiple transcription and translation events occurring concurrently on the same DNA template. Prokaryotic transcription often covers more than one gene and produces polycistronic mRNAs that specify more than one protein.

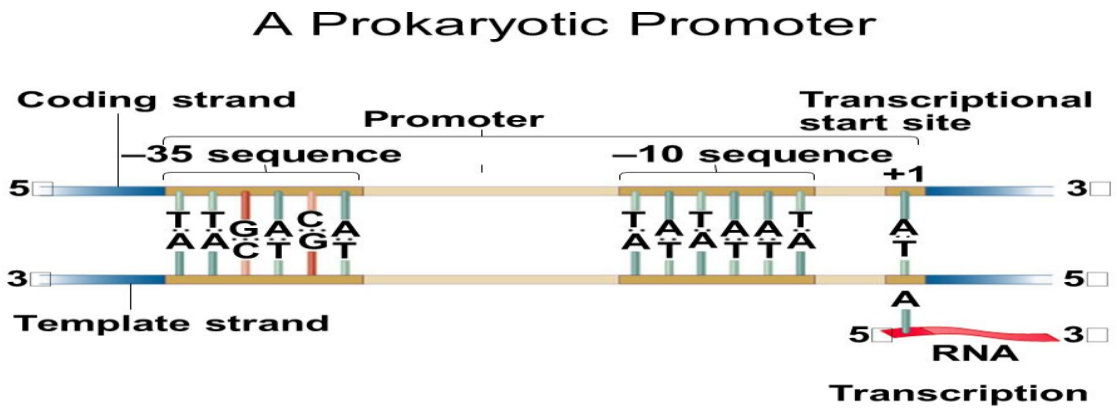
Transcription is the process by which the information in a strand of DNA is copied into a new molecule of messenger RNA (mRNA). In prokaryotic organism's transcription occurs in three phases known as initiation, elongation and termination. RNA is synthesized by a single RNA polymerase enzyme which contains multiple polypeptide subunits. In *E. coli*, the RNA polymerase has subunits: two α, one β, one β' and one ω and σ subunit (α2ββ'ωσ). This complete enzyme is called as the holoenzyme. The σ subunit may dissociate from the other subunits to leave a form known as the core enzyme.

The main components that take part in protein synthesis at cellular level are: 20 different amino acids, different types of RNAs, enzymes, amino acid activating enzymes, poly peptide , polymerase and energy liberating molecules, such as ATP and GTP.

Prokaryotes use the same RNA polymerase to transcribe all of their genes. In E. coli, the polymerase is composed of five polypeptide subunits, two of which are identical. Four of these subunits, denoted α, α, β, and β', comprise the polymerase core enzyme. These subunits assemble each time a gene is transcribed; they disassemble once transcription is complete. Each subunit has a unique role: the two α-subunits are necessary to assemble the polymerase on the DNA; the β-subunit binds to the ribonucleoside triphosphate that will become part of the nascent "recently-born" mRNA molecule; and the β' binds the DNA template strand. The fifth subunit, σ, is involved only in transcription initiation. It confers transcriptional specificity such that the polymerase begins to synthesize mRNA from an appropriate initiation site. Without σ, the core enzyme would transcribe from random sites and would produce mRNA molecules that specified protein gibberish. The polymerase comprised of all five subunits is called the holoenzyme.

Prokaryotic Promoters

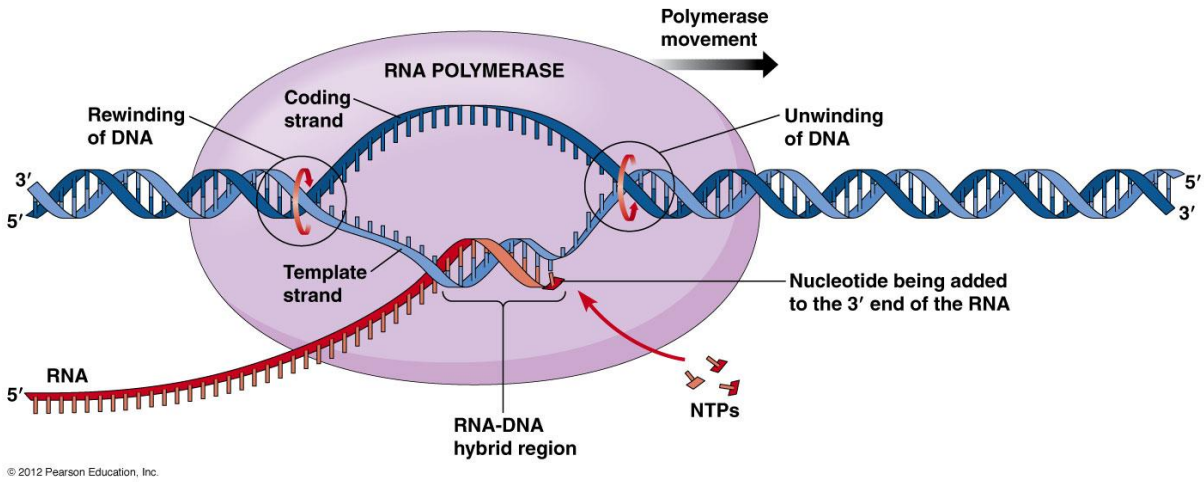
A promoter is a DNA sequence onto which the transcription machinery binds and initiates transcription. Although promoters vary among prokaryotic genomes, a few elements are conserved. At the -10 and -35 regions upstream (before)of the initiation site, there are two promoter consensus sequences, or regions that are similar across all promoters and across various bacterial species. The -10-consensus sequence, called the -10 region, is TATAAT. The -35 sequence, TTGACA, is recognized and bound by σ. Once this interaction is made, the subunits of the core enzyme bind to the site. The A–T-rich -10 region facilitates unwinding of the DNA template, and several phosphodiester bonds are made. The transcription initiation phase ends with the production of abortive transcripts, which are polymers of approximately 10 nucleotides that are made and released.



The σ subunit of prokaryotic RNA polymerase recognizes consensus sequences found in the promoter region upstream of the transcription start sight. The σ subunit dissociates from the polymerase after transcription has been initiated..

Elongation and Termination in Prokaryotes

The transcription elongation phase begins with the release of the σ subunit from the polymerase. The dissociation of σ allows the core enzyme to proceed along the DNA template, synthesizing mRNA in the 5′ to 3′ direction at a rate of approximately 40 nucleotides per second. As elongation proceeds, the DNA is continuously unwound ahead of the core enzyme and rewound behind it. The base pairing between DNA and RNA is not stable enough to maintain the stability of the mRNA synthesis components. Instead, the RNA polymerase acts as a stable linker between the DNA template and the nascent RNA strands to ensure that elongation is not interrupted prematurely.



During elongation, the prokaryotic RNA polymerase tracks along the DNA template, synthesizes mRNA in the 5′ to 3′ direction, and unwinds and rewinds the DNA as it is read.

Initiation

The initiation of protein synthesis begins with the formation of an initiation complex. In E. coli, this complex involves the small 30S ribosome, the mRNA template, three initiation factors that help the ribosome assemble correctly, guanosine triphosphate (GTP) that acts as an energy source, and a special initiator tRNA carrying N-formyl-methionine (fMet-tRNAfMet) . The initiator tRNA interacts with the start codon AUG of the mRNA and carries a formylated methionine (fMet). Because of its involvement in initiation, fMet is inserted at the beginning (N terminus) of every polypeptide chain synthesized by *E. coli.* In *E. coli* mRNA, a leader sequence upstream of the first AUG codon, called the Shine-Dalgarno sequence (also known as the ribosomal binding site AGGAGG), interacts through complementary base pairing with the rRNA molecules that compose the ribosome. This interaction anchors the 30S ribosomal subunit at the correct location on the mRNA template. At this point, the 50S ribosomal subunit then binds to the initiation complex, forming an intact ribosome.

Prokaryotic Termination Signals

Once a gene is transcribed, the prokaryotic polymerase needs to be instructed to dissociate from the DNA template and liberate the newly made mRNA. Depending on the gene being transcribed, there are two kinds of termination signals. One is protein-based and the other is RNA-based. Rho-dependent termination is controlled by the rho protein, which tracks along behind the polymerase on the growing mRNA chain. Near the end of the gene, the polymerase encounters a run of G nucleotides on the DNA template and it stalls. As a result, the rho protein collides with the polymerase. The interaction with rho releases the mRNA from the transcription bubble.

Rho-independent termination is controlled by specific sequences in the DNA template strand. As the polymerase nears the end of the gene being transcribed, it encounters a region rich in C–G nucleotides. The mRNA folds back on itself, and the complementary C–G nucleotides bind together. The result is a stable hairpin that causes the polymerase to stall as soon as it

begins to transcribe a region rich in A–T nucleotides. The complementary U–A region of the mRNA transcript forms only a weak interaction with the template DNA. This, coupled with the stalled polymerase, induces enough instability for the core enzyme to break away and liberate the new mRNA transcript.

Upon termination, the process of transcription is complete. By the time termination occurs, the prokaryotic transcript would already have been used to begin synthesis of numerous copies of the encoded protein because these processes can occur concurrently. The unification of transcription, translation, and even mRNA degradation is possible because all of these processes occur in the same 5′ to 3′ direction, and because there is no membranous compartmentalization in the prokaryotic cell (Figure 3). In contrast, the presence of a nucleus in eukaryotic cells precludes simultaneous transcription and translation.

Prokaryotic Translation

During initiation, the mRNA–ribosome complex is formed and the first codon (always AUG) binds the first aminoacyl tRNA (called initiator tRNA). During the elongation phase, the other codons are read sequentially and the polypeptide grows by addition of amino acids to its C-terminal end. This process continues until a termination codon (Stop codon), which does not have a corresponding aminoacyl-tRNA with which to base pair, is reached. At this point, protein synthesis ceases (termination phase) and the finished polypeptide is released from the ribosome.

Synthesis of aminoacyl-tRNA

Synthesis of aminoacyl-tRNAs is crucially important for two reasons:

Each amino acid must be covalently linked to a tRNA molecule in order to take part in protein synthesis, which depends upon the ‘adaptor’ function of tRNA to ensure that the correct amino acids are incorporated.The covalent bond that is formed between the amino acid and the tRNA is a high energy bond that enables the amino acid to react with the end of the growing polypeptide chain to form a new peptide bond.For this reason, the synthesis of aminoacyl-tRNA is also referred to as amino acid activation.Each tRNA molecule has a cloverleaf secondary structure with the anticodon accessible at the end of the anticodon stem loop.During synthesis of the aminoacyl-tRNA, the amino acid is covalently bound to the A residue of the CCA sequence at the 3’ end.Each tRNA molecule carries only a single amino acid.The attachment of an amino acid to a tRNA is catalyzed by an enzyme called aminoacyl-tRNA synthetase.A separate aminoacyl-tRNA synthetase exists for every amino acid, making 20 synthetases in total.

The synthesis reaction occurs in two steps.

a) The first step is the reaction of an amino acid and ATP to form an aminoacyl-adenylate (also known as aminoacyl-AMP).
b) In the second step, without leaving the enzyme, the aminoacyl group of aminoacyl-AMP is transferred to the 3’ end of the tRNA molecule to form aminoacyl-tRNA

The overall reaction is:

Amino acid + ATP + tRNA → aminoacyl-tRNA + AMP + PPi

Initiation of Protein Synthesis

The first codon translated in all mRNAs is the start codon or initiation codon, AUG which codes for methionine.Two different tRNAs are used for the two types of AUG codon; tRNAfMet is used for the initiation codon and is called the initiator tRNA whereas tRNAm Met is used for internal AUG codons.In prokaryotes the first amino acid of a new protein is N-formylmethionine (abbreviated fMet). Hence the aminoacyl-tRNA used in initiation is fMet-tRNAfMet.A short sequence rich in purines (5'-AGGAGGU-3'), called the Shine–Dalgarno sequence, lies 5' to the AUG initiation codon and is complementary to part of the 16S rRNA in the small ribosomal subunit.Therefore, this is the binding site for the 30S ribosomal subunit which then migrates in a 3' direction along the mRNA until it encounters the AUG initiation codon.

Initiation of protein synthesis requires proteins called initiation factors (IFs).

In prokaryotes, three initiation factors (IF-1, IF-2 and IF-3) are essential.

Because of the complexity of the process, the exact order of binding of IF-1, IF-2, IF-3, fMet-tRNAf is controversial.

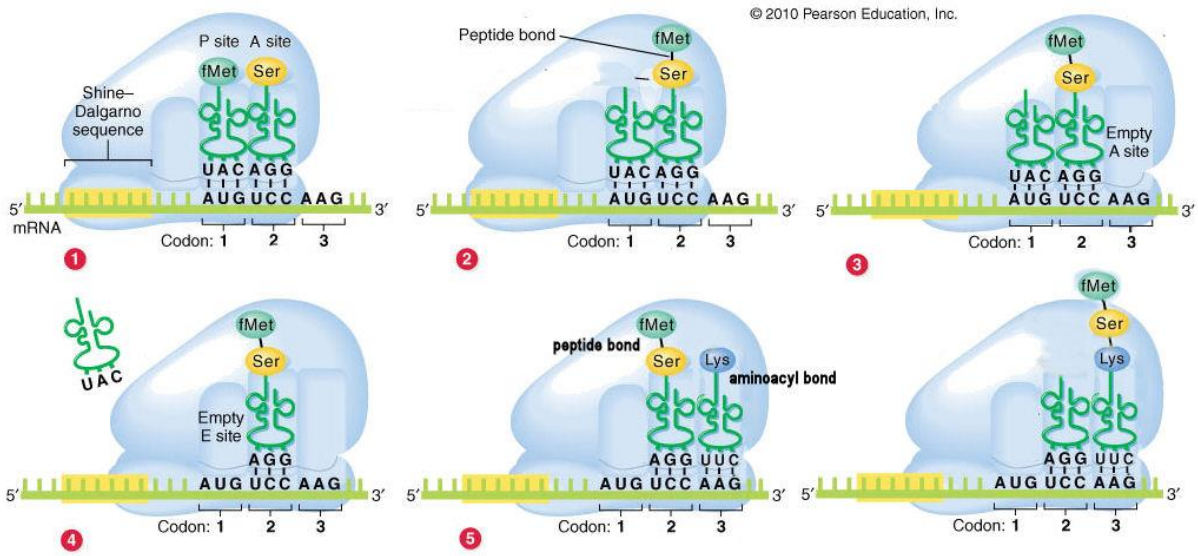
Steps Involved

Initiation begins with the binding of IF-1 and IF-3 to the small (30S) ribosomal subunit.

Their role is to stop the 30S subunit binding to the 50S subunit in the absence of mRNA and fMet-tRNAf Met which would result in a nonfunctional ribosome.The small subunit then binds to the mRNA via the Shine–Dalgarno sequence and moves 3' along the mRNA until it locates the AUG initiation codon.The initiator tRNA charged with N-formylmethionine and in a complex with IF-2 and GTP (fMet-tRNAfMet/IF-2/GTP) now binds.IF-3 is released.The complex of mRNA, fMet-tRNAf Met, IF-1, IF-2 and the 30S ribosomal subunit is called the 30S initiation complex.The large (50S) ribosomal subunit now binds, with the release of IF-1 and IF-2 and hydrolysis of GTP, to form a 70S initiation complex.

Elongation of Protein Synthesis

At the start of the first round of elongation, the initiation codon (AUG) is positioned in the P site with fMet-tRNAfMet bound to it via codon–anticodon base pairing.The next codon in the mRNA is positioned in the A site.Elongation of the polypeptide chain occurs in three steps called the elongation cycle, namely aminoacyl-tRNA binding, peptide bond formation and translocation:



Aminoacyl-tRNA binding. The corresponding aminoacyl-tRNA for the second codon binds to the A site via codon–anticodon interaction.Binding of the aminoacyl-tRNA requires elongation factor EF-Tu and GTP which bind as an aminoacyl-tRNA/EF-Tu/GTP complex. Following binding, the GTP is hydrolyzed and the EF-Tu is released, now bound to GDP. Before the EF-Tu molecule can catalyze the binding of another charged tRNA to the ribosome, it must be regenerated by a process involving another elongation factor, EF-Ts.This regeneration is called the EF-Tu–EF-Ts exchange cycle.

First, EF-Ts binds to EF-Tu and displaces the GDP. Then GTP binds to the EF-Tu and displaces EF-Ts. The EF-Tu-GTP is now ready to take part in another round of elongation.

Peptide bond formation

The second step, peptide bond formation, is catalyzed by peptidyl transferase.In this reaction the carboxyl end of the amino acid bound to the tRNA in the P site is uncoupled from the tRNA and becomes joined by a peptide bond to the amino group of the amino acid linked to the tRNA in the A site.

Translocation

In the third step, a complex of elongation factor EF-G (also called translocase) and GTP (i.e. EF-G/GTP) binds to the ribosome. Three concerted movements now occur, collectively called translocation:

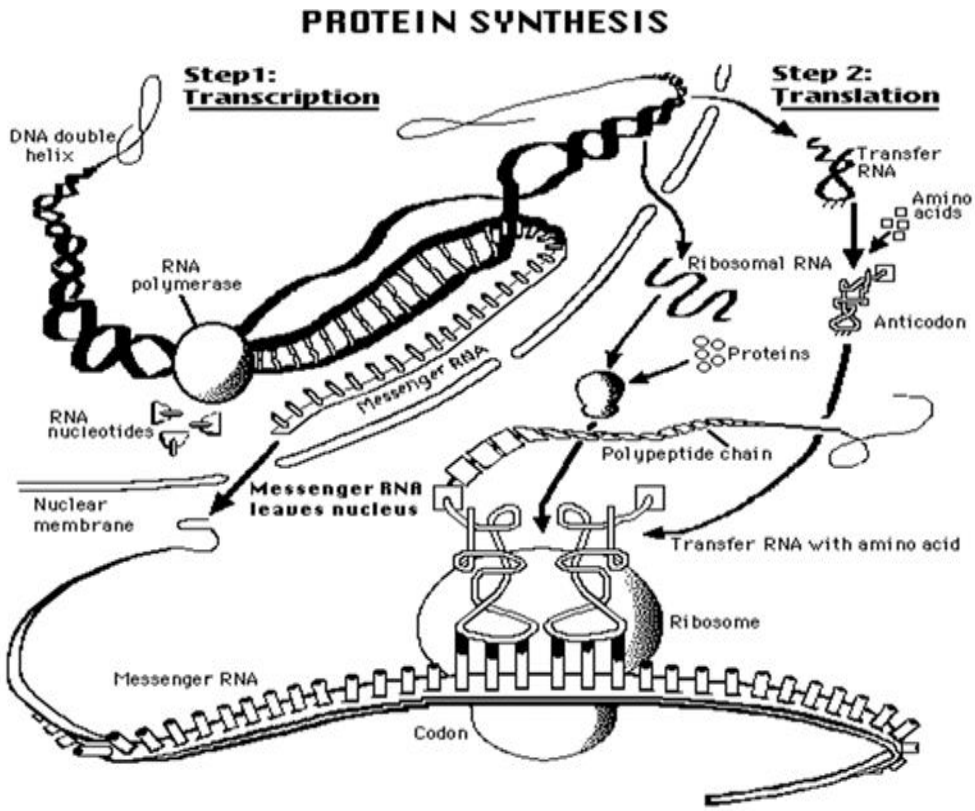
the deacylated tRNA moves from the P site to the E site the dipeptidyl-tRNA in the A site moves to the P site, and the ribosome moves along the mRNA (5’ to 3’) by three nucleotides to place the next codon in the A site.

During the translocation events, GTP is hydrolyzed to GDP and inorganic phosphate, and EF-G is released ready to bind more GTP for another round of elongation. After translocation,

the A site is empty and ready to receive the next aminoacyltRNA. The A site and the E site cannot be occupied simultaneously. Thus the deacylated tRNA is released from the E site before the next aminoacyl-tRNA binds to the A site to start a new round of elongation. Elongation continues, adding one amino acid to the C-terminal end of the growing polypeptide for each codon that is read, with the peptidyl-tRNA moving back and forth from the P site to the A site as it grows.

Termination of Protein Synthesis

Eventually, one of three termination codons (also called Stop codons) becomes positioned in the A site. These are UAG, UAA and UGA. Unlike other codons, prokaryotic cells do not contain aminoacyl-tRNAs complementary to Stop codons. Instead, one of two release factors (RF-1 and RF-2) binds instead. RF-1 recognizes UAA and UAG whereas RF-2 recognizes UAA and UGA. A third release factor, RF-3, is also needed to assist RF-1 or RF-2 interaction with the ribosome. Thus either RF-1 + RF-3 or RF-2 + RF-3 bind depending on the exact termination codon in the A site.RF-1 (or RF-2) binds at or near the A site whereas RF-3/GTP binds elsewhere on the ribosome. The release factors cause the peptidyl transferase activity to transfer the polypeptide to a water molecule instead of to aminoacyl-tRNA, effectively cleaving the bond between the polypeptide and tRNA in the P site.



Topic-8

Notes on

(a) Chargaff rule

Chargaff’s Rules

It was known that DNA is composed of nucleotides, each of which contains a nitrogen-containing base, a five-carbon sugar (deoxyribose), and a phosphate group. In these nucleotides, there is one of the four possible bases: adenine (A), guanine (G), cytosine (C), or thymine (T) . Adenine and guanine are purine bases, and cytosine and thymine are pyrimidine bases.

Erwin Chargaff (1905-2002), an Austrian-American biochemist from Columbia University, analyzed the base composition of the DNA of various species. This led him to propose two main rules that have been appropriately named Chargaff's rules.

Rule 1

Chargaff determined that in DNA, the amount of one base, a purine, always approximately equals the amount of a particular second base, a pyrimidine. Specifically, that in any double-stranded DNA the number of guanine units equals approximately the the number of cytosine units and the number of adenine units equals approximately the number of thymine units.

Human DNA is 30.9% A and 29.4% T, 19.9% G and 19.8% C. The rule constitutes the basis of base pairs in the DNA double helix: A always pairs with T, and G always pairs with C. He also demonstrated that the number of purines (A+G) always approximates the number of pyrimidines (T+C), an obvious consequence of the base-pairing nature of the DNA double helix.Rule 2

In 1947 Chargaff showed that the composition of DNA, in terms of the relative amounts of the A, C, G and T bases, varied from one species to another. This molecular diversity added evidence that DNA could be the genetic material.

Base Composition of DNA-Chargaff Rule

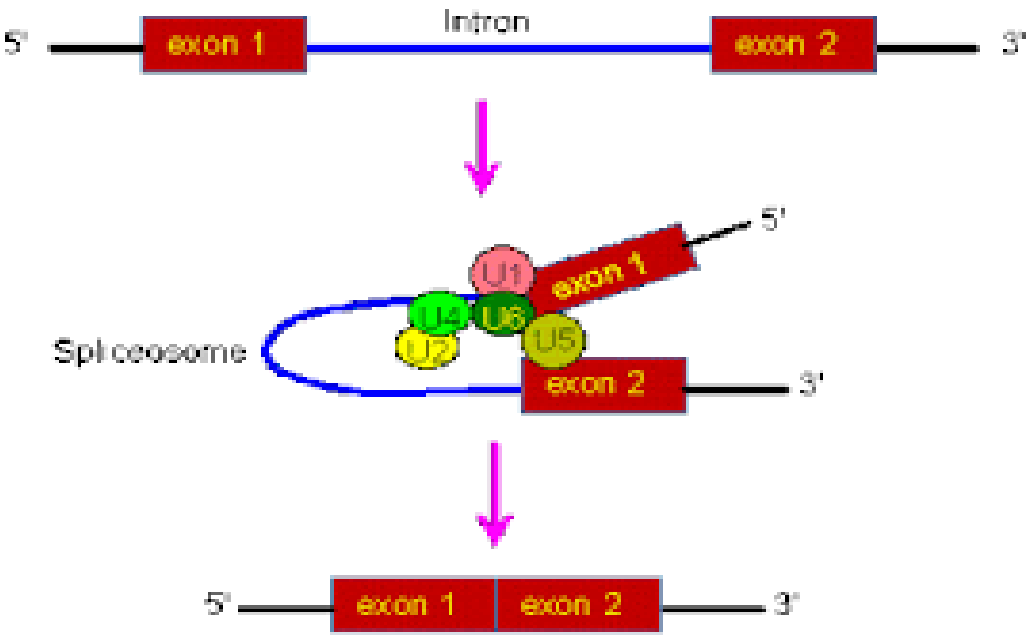
- ✓ The sum of purines (Pu) is equal to the sum of pyrimidines (Py), i.e., Pu/Py = 1. In other words, A + G = T + C
- ✓ The ratio of adenine to thymine is also one, i.e., A/T = 1.
- ✓ The ratio of guanine to cytosine is also one, i.e., G/C= 1.
- ✓ Bases with 6-amino groups are equal to bases with 6-keto (hydroxyl) groups, i.e., A + C = G + T.
- ✓ The ratio of A + T/G + C, known as dissymmetry ratio, varies greatly from one species of DNA to the other and is characteristic of that species.
- ✓ Chargaff's data suggest that A is always paired with T and G is always paired with C.

b) Splicing/ Splicesome

The discovery that most protein-coding genes in eukaryotes are interrupted by one or more intervening sequences, or introns (junk gene), was one of the major advances in molecular biology made in the 1970s . The formation of mRNA thus involves a splicing process to remove introns from primary gene transcripts. Splicing takes place in the nucleus before the mRNAs are exported to the cytoplasm for translation and occurs by a two-step mechanism, in which both steps involve transesterification reactions. The first step generates a free 5′ exon and a 'lariat' form of the intron still attached to the 3′ exon as reaction intermediates. The lariat contains a branched RNA structure that arises through the esterification of the 5′ end of the intron to a 2′ ribose hydroxyl of an adenosine residue located usually 20–40 nucleotides upstream of the 3′ end of the intron. The second catalytic step results in the ligation of the exons and release of the lariat intron.

Most eukaryotic genes are expressed as precursor mRNAs (pre-mRNAs) that are converted to mRNA by splicing, an essential step of gene expression in which noncoding sequences (introns) are removed and coding sequences (exons) are ligated together. Whereas some

exons are constitutively spliced—that is, they are present in every mRNA produced from a given pre-mRNA—many are alternatively spliced to generate variable forms of mRNA from a single pre-mRNA species. Alternative splicing is prevalent in higher eukaryotes and it enhances their complexity by increasing the number of unique proteins expressed from a single gene . Unraveling splicing at the molecular level is not only important for understanding gene expression, but it is also of medical relevance, as aberrant pre-mRNA splicing is the basis of many human diseases or contributes to their severity.



Nuclear pre-mRNA splicing is catalyzed by the spliceosome, a multi-megadalton ribonucleoprotein (RNP) complex. Both the conformation and composition of the spliceosome are highly dynamic, affording the splicing machinery its accuracy and at the same time flexibility. Two unique spliceosomes coexist in most eukaryotes: the U2-dependent spliceosome, which catalyzes the removal of U2-type introns, and the less abundant U12-dependent spliceosome, which is present in only a subset of eukaryotes and splices the rare U12-type class of introns.

The biochemical mechanism by which splicing occurs has been studied in a number of systems and is now fairly well characterized. Introns are removed from primary transcripts by cleavage at conserved sequences called splice sites. These sites are found at the 5′ and 3′ ends of introns. Most commonly, the RNA sequence that is removed begins with the dinucleotide GU at its 5′ end, and ends with AG at its 3′ end. These consensus sequences are known to be critical, because changing one of the conserved nucleotides results in inhibition of splicing.

Another important sequence occurs at what is called the branch point, located anywhere from 18 to 40 nucleotides upstream from the 3′ end of an intron. The branch point always contains an adenine, but it is otherwise loosely conserved. A typical sequence is YNYYRAY, where Y indicates a pyrimidine, N denotes any nucleotide, R denotes any purine, and A denotes adenine. Rarely, alternate splice site sequences are found that begin with the dinucleotide AU and end with AC; these are spliced through a similar mechanism.

Splicing occurs in several steps and is catalyzed by small nuclear ribonucleoproteins (snRNPs, commonly pronounced "snurps"). First, the pre-mRNA is cleaved at the 5′ end of the intron following the attachment of a snRNP called U1 to its complementary sequence within the intron. The cut end then attaches to the conserved branch point region downstream through pairing of guanine and adenine nucleotides from the 5′ end and the branch point, respectively, to form a looped structure known as a lariat (Figure 1). The bonding of the guanine and adenine bases takes place via a chemical reaction known as transesterification, in which a hydroxyl (OH) group on a carbon atom of the adenine "attacks" the bond of the guanine nucleotide at the splice site. The guanine residue is thus cleaved from the RNA strand and forms a new bond with the adenine.

Next, the snRNPs U2 and U4/U6 appear to contribute to positioning of the 5′ end and the branch point in proximity. With the participation of U5, the 3′ end of the intron is brought into proximity, cut, and joined to the 5′ end. This step occurs by transesterification; in this case, an OH group at the 3′ end of the exon attacks the phosphodiester bond at the 3′ splice site. The adjoining exons are covalently bound, and the resulting lariat is released with U2, U5, and U6 bound to it.In addition to consensus sequences at their splice sites, eukaryotic genes with long introns also contain exonic splicing enhancers (ESEs). These sequences, which help position the splicing apparatus, are found in the exons of genes and bind proteins that help recruit splicing machinery to the correct site. Most splicing occurs between exons on a single RNA transcript, but occasionally trans-splicing occurs, in which exons on different pre-mRNAs are ligated together.

The splicing process occurs in cellular machines called spliceosomes, in which the snRNPs are found along with additional proteins. The primary variety of spliceosome is one of the most plentiful structures in the cell, and recently, a secondary type of spliceosome has been identified that processes a minor category of introns. These introns are referred to as U12-type introns because they depend upon the action of a snRNP called U12 (the common introns described above are called U2-type introns). The role of U12-type introns is not yet defined, but their persistence throughout evolution and conservation between homologous genes of widely divergent species suggests an important functional basis

(c) Initiation factor in Eukaryote

Unlike prokaryotes where all RNA is synthesized by a single RNA polymerase, the nucleus of a eukaryotic cell has three RNA polymerases responsible for transcribing different types of RNA.

a) RNA polymerase I (RNA Pol I) is located in the nucleolus and transcribes the 28S, 18S, and 5.8S rRNA genes.

b) RNA polymerase II (RNA Pol II) is located in the nucleoplasm and transcribes protein-coding genes, to yield pre-mRNA, and also the genes encoding small nucleolar RNAs (snRNAs) involved in rRNA processing and small nuclear RNAs (snRNAs) involved in mRNA processing, except for U6 snRNA.

c) RNA polymerase III (RNA Pol III) is also located in the nucleoplasm. It transcribes the genes for tRNA, 5S rRNA, U6 snRNA, and the 7S RNA associated with the signal recognition particle (SRP) involved in the translocation of proteins across the endoplasmic reticulum membrane. Each of the three eukaryotic RNA polymerases contains 12 or more subunits and so these are large complex enzymes.

During initiation, RNA polymerase recognizes a specific site on the DNA, upstream from the gene that will be transcribed, called a promoter site and then unwinds the DNA locally.Most promoter sites for RNA polymerase II include a highly conserved sequence located about 25–35 bp upstream (i.e. to the 5 side) of the start site which has the consensus TATA(A/T)A(A/T) and is called the TATA box.Since the start site is denoted as position +1, the TATA box position is said to be located at about position -25.The TATA box sequence resembles the -10 sequence in prokaryotes (TATAAT) except that it is located further upstream.Both elements have essentially the same function, namely recognition by the RNA polymerase in order to position the enzyme at the correct location to initiate transcription.The sequence around the TATA box is also important in that it influences the efficiency of initiation. Transcription is also regulated by upstream control elements that lie 5′ to the TATA box.

Some eukaryotic protein-coding genes lack a TATA box and have an initiator element instead, centered around the transcriptional initiation site. In order to initiate transcription, RNA polymerase II requires the assistance of several other proteins or protein complexes, called general (or basal) transcription factors, which must assemble into a complex on the promoter in order for RNA polymerase to bind and start transcription.

These all have the generic name of TFII (for Transcription Factor for RNA polymerase II).

The first event in initiation is the binding of the transcription factor IID (TFIID) protein complex to the TATA box via one its subunits called TBP (TATA box binding protein).As soon as the TFIID complex has bound, TFIIA binds and stabilizes the TFIID-TATA box interaction. Next, TFIIB binds to TFIID.

However, TFIIB can also bind to RNA polymerase II and so acts as a bridging protein. Thus,RNA polymerase II, which has already complexed with TFIIF, now binds.This is followed by the binding of TFIIE and H. This final protein complex contains at least 40 polypeptides and is called the transcription initiation complex.