Transcription

Prepared By: Dr. Bhaskar Sarma

Transcription is a process of formation of transcript (RNA). It takes place by the usual process of complementary base pairing, catalyzed and scrutinized by the enzyme RNA polymerase. It occurs unidirectionally in which RNA chain (transcript) is synthesized from the 5′ to 3′ direction.

ASPECTS OF PROKARYOTIC TRANSCRIPTION.

ENZYMATIC SYNTHESIS OF RNA

The essential chemical characteristics of the synthesis of RNA are the following:

1. The precursors in the synthesis of RNA are the four ribonucleoside 5′- triphosphates (rNTP) --- ATP, GTP, CTP, and UTP.
2. In the polymerization reaction the 3'-OH terminus of the RNA being synthesized reacts with the 5′- triphosphate of a precursor rNTP; a pyrophosphate is released, and a phosphodiester bond results.
3. The sequence of bases in an RNA molecule is determined by the base sequence of the DNA.
4. The DNA molecule is double stranded, yet at the same time any particular region is transcribed only one strand serves as a template.
5. The RNA chain grows in the 5′ → 3′ direction.
6. RNA polymerase (the enzyme that catalyzes RNA synthesis) does not need a primer; it can initiate transcription de novo.
7. Only ribonucleoside 5′-triphosphate participate in RNA synthesis, and the first base to be laid down in the initiation event retains its triphosphate. Its 3′-OH group is the point of attachment for the subsequent nucleotide. Thus, the 5′ end of a growing RNA molecule terminates with a triphosphate.
8. Bacteria have only a single RNA polymerase, while in eukaryotic cells there are three: RNA Pol I, II, and III.
9. RNA polymerase II is the enzyme responsible for all mRNA synthesis. Pol I and Pol III are each involved in transcribing specialized, RNA encoding genes. Specifically, Pol I transcribes the large ribosomal RNA precursor gene, whereas Pol III transcribes tRNA genes, some small nuclear RNA genes, and the 5S rRNA gene.
10. E. coli RNA polymerase consists of five subunits--- two identical α subunits and one each of types β, β′, and σ. The σ subunit dissociates from the enzyme easily and in fact does so shortly after polymerization is initiated. The term holoenzyme is used to describe the complete enzyme and core enzyme for the enzyme which lacks the σ subunit.

TRANSCRIPTION BY RNA POLYMERASE PROCEEDS IN A SERIES OF STEPS

To transcribe a gene, RNA polymerase proceeds through a series of well-defined steps which are grouped into three phases: initiation, elongation, and termination.

A. **Initiation**—A promoter is the DNA sequence that initially binds the RNA polymerase (together with initiation factors in many cases). Once formed, the promoter-polymerase complex undergoes structural changes required for initiation to proceed. The DNA around the point where transcription will start unwinds, and the base pairs are disrupted, producing a "bubble" of single-stranded DNA. Transcription always occur in a 5′ to 3′ direction and only one of the DNA strands acts as a template on which the RNA strand is built. The choice of promoter determines which of DNA is transcribed and is the main step at which regulation is imposed.
B. **Elongation**—Once the RNA polymerase has synthesized a short stretch of RNA (approximately 10 bases), it shifts into the elongation phase. This transition requires further conformational changes in polymerase that lead it to grip the template more firmly. During elongation, the enzyme performs an impressive range of tasks in addition to the catalysis of RNA synthesis. It unwinds the DNA in front and re-anneals it behind, it dissociates the growing RNA chain from the template as it moves along, and it performs proofreading functions.
C. **Termination**—Once the polymerase has transcribed the length of the gene (or genes), it must stop and release the RNA product. This step is called termination. In some cells there are specific, well- characterized, sequences that trigger termination; in others it is less clear what instructs the enzyme to cease transcribing and dissociate from the template.

TRANSCRIPTION INITIATION INVOLVES THREE DEFINED STEPS

The first phase in the transcription cycle—initiation – can itself be broken down into a series of defined steps.

- The first step is the initial binding of polymerase to a promoter to form what is called a closed complex. In this form the DNA remains double stranded, and the enzyme is bound to one face of the helix.
- In the second step of initiation, the closed complex undergoes a transition to the open complex in which the DNA strands separate over a distance of some 14 bp around the start site to form the transcription bubble.
- The opening of the DNA frees the template strand. The first two ribonucleotides are brought into the active site, aligned on the template strand, and joined together. The enzyme then begins to move along the template strand, opening the DNA helix ahead of the site of polymerization and allowing it to reseal behind. In this way, subsequent ribonucleotides are incorporated into the growing RNA chain. Incorporation of the first ten or so ribonucleotides is a rather inefficient process, and at that stage the enzyme often releases short transcripts (each of less than ten or so nucleotides) and then begins synthesis again. Once an enzyme gets further than the 10 bp, it is said to have escaped the promoter. At this point it has formed a stable ternary complex, containing enzyme, DNA, and RNA. This is the transition to the elongation phase.

THE TRANSCRIPTION CYCLE IN BACTERIA

- The first step in transcription is binding of RNA polymerase to a DNA molecule.
- Binding occurs at particular sites called promoters, which are specific sequences of about 40 base-pairs at which several interactions occur.
- The most crucial interactions for positioning of E-coli RNA polymerase at a promoter occur at two short- sequence patches on the DNA. These patches are located about 10 and 35 base-pairs before the first base, which is copied into RNA. The first base copied is called the start point of transcription, and is assigned as position +1. The two patches therefore are at positions -10 and-35, with respect to the start point.
- E-coli RNA polymerase is a large protein that covers as much as 70 to 75 base-pairs of DNA from position -55 to position +20 when it is bound to a promoter. It can therefore easily make the two separate contacts at the same time.
- The sequences around -10 are considered to be variants of a basic sequence, TATAAT, called the -10 consensus sequence, while the sequence around -35 is a variant of TTGACA, called the -35 consensus sequence.
- With some promoters, the sequence at one of the two contact regions differs greatly from the consensus. In this case, RNA polymerase recognizes the sequence very poorly, and the promoter does not function very well. For such genes, however, there is often an accessory protein that helps RNA polymerase bind to the promoter. This kind of accessory protein is called a gene activator protein and is basically a positive effector. Gene activator proteins usually bind to specific sequences very near or even within the promoter sequence and they appear to have surfaces to which RNA polymerase can attach, when correctly positioned on the promoter. Thus a gene-activator protein helps the RNA polymerase bind to a promoter with a poor recognition signal. An important effector protein, is the CAP protein; is needed for the activation of many promoters for genes required for sugar metabolism.
- After RNA polymerase binds tightly at the promoter region, it manages to unwind a small part of the DNA, from base-pairs +3 to -10 to form the open-promoter complex. This unwinding is necessary for pairing of the incoming ribonucleoside triphosphate, the building blocks for synthesis of the RNA. The base composition between the start site through the -10 region is generally rich in A+T, which renders the DNA especially susceptible to unpairing.
- Once an open-promoter complex has formed, RNA polymerase is ready to initiate synthesis. RNA polymerase contains two nucleotide binding sites, called the initiation site and the elongation site, respectively.
- The initiation site prefers to bind purine nucleoside triphosphate, namely ATP and GTP, and one of these is usually the first nucleotide in the chain. The first base, therefore, on the DNA template that is transcribed is usually a thymine or a cytosine.
- The initiating nucleoside triphosphate binds to the enzyme in the open-promoter complex and forms a hydrogen bond with the complementary DNA base. The elongation site is then filled with a nucleoside triphosphate that is selected by its ability to form a hydrogen-bond with the next base in the DNA strand. The two nucleotides are then joined together.

- During the joining process, RNA polymerase moves along the DNA to the next base-pair. As it moves, it separates the base-pair at position +4, and allows the base-pair at position -10 to come together again, thus maintaining the same stretch of 13 base-pairs unwound.
- RNA polymerase motion, DNA base unpairing and DNA base reannealing gradually elongates the RNA chain.
- After several nucleotides (between four and eight) are added to the growing RNA chain, three important changes occur. First, the nucleotide at the 5′ end of the RNA chain (the first one to be incorporated) becomes unpaired from the base on the template strand to allow that base to pair again with its complement on the other DNA strand. Second, RNA polymerase changes its structure and loses the σ subunit. Third, a protein aiding elongation (called NusA protein), binds. The rest of the elongation process is carried out by the core enzyme NusA complex. Each new nucleotide that is added at the 3' end of the growing RNA allows another nucleotide to be released from its binding to the template. Hence, as elongation continues a section of 4 to 8 nucleotides at the 3' end of the newly-made RNA is paired to the DNA template strand, while the rest of the RNA emerges from RNA polymerase as a single strand.
- Termination of RNA synthesis occurs at specific base-sequences in the DNA molecule, called terminators. Some termination sequences allow RNA polymerase to terminate elongation spontaneously. These sequences are called intrinsic terminators. Other terminators require the action of a protein called Rho; they are called rho-dependent terminators.

  Rho-independent terminators, also called intrinsic terminators, consists of two sequence elements: a short inverted repeat (of about 20 nucleotides) followed by a stretch of about eight A:T base pairs. These elements do not affect the polymerase until after they have been transcribed—that is, they function in the RNA rather than in the DNA. Thus, when polymerase transcribes an inverted repeat sequence, the resulting RNA can form a stem-loop structure (often called a “hairpin”) by base-pairing with itself. The hairpin is believed to cause termination by disrupting the elongation complex.

  The hairpin only works as an efficient terminator when it is followed by a stretch of A:U base pairs. This is because, under those circumstances, at the time the hairpin forms, the growing RNA chain will be held on the template at the active site by only A:U base pairs. As A:U base pairs are the weakest of all base pairs (weaker even than A:T base pairs), they are more easily disrupted by the effects of the stem loop on the transcribing polymerase, and so the RNA will more readily dissociate.

  Rho acts by binding to a special sequence on the nascent RNA and forcibly pulls the RNA away from its contact with the DNA in the transcription-elongation complex with RNA polymerase. Rho derives its force from the hydrolysis of ATP molecules.

The final step in the termination process is dissociation of RNA polymerase from the DNA. This comes about by the rebinding of a σ subunit to the core RNA polymerase after the RNA is released. The released holoenzyme is free to search for a new promoter where it can start synthesis of a new RNA.

EUKARYOTIC TRANSCRIPTION

Initiation of transcription is a crucial regulation point for gene expression in all organisms. Although eukaryotes and bacteria use some of the same regulatory mechanisms, the regulation of transcription in the two systems is fundamentally different.

The components and steps involved in eukaryotic transcription can be summarized as follows:

1. There are three different gene class specific RNA polymerases each of which consists of at least 10 subunits.
2. A number of additional proteins, the so called general or basal transcription factors, together with the RNA polymerase, form a minimal complex that recognizes a promoter and can initiate transcription in vitro.
3. The relative level of transcriptional initiation is regulated by various specific transcription factors. These bind to upstream regions of promoters or to other regulatory sequences called **enhancers** that are often located very far away from the coding region of a given gene. Furthermore, additional factors, so called **coactivators**, which themselves do not bind DNA, may interact with DNA-bound transcription factors.
4. The specific transcription factors bound to distal promoter or enhancer sequences most likely stimulate transcription by directly interacting with the basal transcription apparatus while looping out the intervening DNA. In addition, at least some of these factors seem to also help reorganize chromatin into a more accessible, (loose) form. Conversely, repressing factors binding so called silencer DNA may instead disturb interactions within the transcription apparatus and/or reorganize the chromatin into a tight, non accessible form.

DNA Binding Activators and Coactivators Facilitate Assembly of the General Transcription Factors

Many (but not all) Pol II promoters include the **TATA box** and **Inr (initiator) sequences**, with their standard spacing, they vary greatly in both the number and the location of additional sequences required for the regulation of transcription. These additional regulatory sequences are usually called enhancers in higher eukaryotes and **upstream activator sequences (UASs) in yeast**. A typical enhancer may be found hundreds or even thousands of base pairs upstream from the transcription start site, or may even be downstream, within the gene itself. When bound by the appropriate regulatory proteins, an enhancer increases transcription at nearby promoters regardless of its orientation in the DNA.

Successful binding of active RNA polymerase II holoenzyme at one of its promoters usually requires the action of other proteins of four types:

1. Transcription activators, which bind to enhancers or UASs and facilitate transcription;
2. Chromatin modification and remodeling proteins
3. Coactivators and
4. Basal transcription factors, required at every pol II promoter.

The coactivators act indirectly-not by binding to the DNA—and are required for essential communication between and the complex composed of Pol II and the basal (or general) transcription factors. Furthermore, a variety of repressor proteins can interfere with communication between the RNA polymerase and the activators, resulting in repression of transcription.

1. **Transcription Activators**—the requirements for activators vary greatly from one promoter to another. A few activators are known to facilitate transcription at hundreds of promoters, whereas others are specific for a few promoters. Many activators are sensitive to the binding of signal molecules, providing the capacity to activate or deactivate transcription in response to a changing cellular environment. Some enhancers bound by activators are quite distant from the promoter's TATA box, in such cases the intervening DNA is looped so that the various protein complexes can interact directly. The looping is promoted by certain nonhistone proteins that are abundant in chromatin and bind nonspecifically to DNA. These High Mobility Group (HMG) proteins play an important structural role in chromatin remodeling and transcriptional activation.
2. **Chromatin modification and remodeling proteins**—transcription of a eukaryotic gene is strongly repressed when its DNA is condensed within heterochromatin. Some, but not all of the euchromatin is transcriptionally active. Transcriptionally active chromosomal regions are characterized not only by the presence of nucleosomes with particular compositions and modifications. The transcription associated structural changes in chromatin are generated by a process called chromatin remodeling. Some enzymes covalently modify the histons of the nucleosome. Others use the chemical energy of ATP to reposition nucleosomes on the DNA. Still others alter the histone composition of the nucleosome. Nucleosome modifiers come in two types ; those that add chemical groups to the tails of histons, such as histone acetyl transferases (HATs), which add acetyl groups; and those that remodel the nucleosomes, such as the ATP dependent activity of SWI/SNF. Remodeling and certain modifications, can uncover DNA-binding sites that would otherwise remain inaccessible within the nucleosome. For example, by increasing the mobility of nucleosomes, remodelers free up binding sites for regulators and for the transcription machinery. Similarly by addition of acetyl groups to histone tails alters the interactions between those tails and adjacent nucleosomes. This modification is also believed to "loosen" chromatin structure, freeing up sites.
3. **Coactivator Protein Complexes**—most transcription requires the presence of additional protein complexes. Some major regulatory protein complexes that interact with Pol II have been defined both genetically and biochmically. These coactivator complexes act as intermediaries between the transcription activators and the Pol II complex.

   The principal eukaryotic coactivator consists of 20 to 30 or more polypeptides in a protein complex called mediator. Mediator binds tightly to the carboxyl-terminal domain (CTD) of the largest subunit of Pol II. The mediator complex is required for both basal and regulated transcription at promoters used by Pol II, and it also stimulates phosphorylation of the CTD by TFIIH (a basal transcription factor). Transcription activators interact with one or more components of the mediator complex, with the precise interaction sites differing from one activator to another. Coactivator complexes function at or near the promoter's TATA box.
4. **TATA-Binding Protein**—The eukaryotic core promoter refers to the minimal set of sequence elements required for accurate transcription initiation by the Pol II machinery, as measured in vitro. A core promoter is typically about 40 nucleotides long, extending either upstream or downstream of the transcription start site. The eukaryotic core promoter consists of TFIIB recognition element (BRE), the TATA element, the initiator (Inr) and the downstream promoter element (DPE). The TATA element is recognized by the general transcription factor called TFIID. The first component to bind in the assembly of a preinitiation complex (PIC) at the TATA box of a typical Pol II promoter is the TATA binding protein (TBP), a component of TFIID. The

complete complex includes the basal transcription factors TFIIB, TFIIE, TFIIF, TFIIH, Pol II and perhaps TFIIA. This minimal PIC, however is often insufficient for the initiation of transcription and generally does not form at all if the promoter is obscured within chromatin. Positive regulation, leading to transcription, is imposed by the activators and coactivators.

Choreography of Transcriptional Activation

During initiation of transcription activator first bind to the appropriate enhancer of the gene to be transcribed. Then it triggers subsequent activation of the promoter. Binding of one activator may enable the binding of others, gradually displacing some nucleosome.

Crucial remodeling of the chromatin then takes place in stages, facilitated by interactions between activators and HAT’s or enzyme complexes such as SWI/SNP (or both). In this way, a bound activator can draw in other components necessary for further chromatin remodeling to permit transcription of specific genes. The bound activators interact with the large mediator complex. Mediator, in turn, provides an assembly surface for the binding of first TBP (or TFIID), then TFIIB, and then other components of the PIC including RNA polymerase II. Mediator stabilizes the binding of Pol II and its associated transcription factors and generally facilitates formation of PIC. Formation of the pre-initiation complex containing these components is followed by promoter melting. In contrast to bacteria, promoter melting in eukaryotes requires hydrolysis of ATP and is mediated by TFIIH. It is the helicase like activity of that factor which stimulates unwinding of promoter DNA.

In eukaryotes, promoter escape involves phosphorylation of the polymerase. The large subunit of Pol II has a **C-terminal domain (CTD)**,which extends as a “tail”. The CTD contains a series of repeats of the **heptapeptide sequence: Tyr-Ser-Pro-Thr-Ser-Pro-Ser**. There are 27 of these repeats in the yeast Pol II CTD and 52 in the human case. Each repeat contains sites for phosphorylation by specific kinases including one that is a subunit of TFIIH.

The form of Pol II recruited to the promoter initially contains a largely unphosphorylated tail, but the species found in the elongation complex bears multiple phosphoryl groups on its tail. Addition of these phosphates helps polymerase shed most of the general transcription factors used for initiation, and which the enzyme leaves behind as it escapes the promoter. Regulating the phosphorylation state of the CTD of Pol II controls later steps—those involving processing of the RNA—as well. In addition to TFIIH, a number of other kinases have been identified that act on the CTD as well as a phosphatase that removes the phosphates added by those kinases.

The first RNA processing event is capping. This involves the addition of a modified guanine base to the 5’ end of the RNA. Specifically, it is a methylated guanine, and it is joined to the RNA transcript by an unusual 5′-5′ linkage involving three phosphate. The 5′ cap is created in three enzymatic steps. In the first step, a phosphate group is removed from the 5′ of the transcript. Then the GTP is added. And in the final step, that nucleotide is modified by the addition of methyl group. The RNA is capped when it is still only some 20-40 nucleotides long—when the transcription cycle has progressed only to the transition between the initiation and elongation phases. After capping, dephosphorylation of Ser5 within the tail repeats leads to dissociation of the capping machinery, and further phosphorylation(this time of Ser2 within the tail repeats) causes recruitment of the machinery needed for RNA splicing.

The final RNA processing event, polyadenylation of the 3′ end of the mRNA, is intimately linked with the termination of transcription. Just as with capping and splicing, the polymerase CTD tail is involved in recruiting the enzymes necessary for polyadenylation.

Once polymerase has reached the end of a gene, it encounters specific sequences that, after being transcribed into RNA, trigger the transfer of the polyadenylation enzymes to that RNA, leading to three events: cleavage of the message; addition of many adenine residues to its 5′ end; and , subsequently, termination of transcription by polymerase. This process works as follows,

Two protein complexes are carried by the CTD of polymerase as it approaches the end of the gene: CPSF(cleavage and polyadenylation factor) and CstF (cleavage stimulation factor). The sequences which, once transcribed into RNA trigger transfer of these factors to the RNA, are called poly-A signals. Once CPSF and CstF are bound to the RNA, other proteins are recruited as well, leading initially to RNA cleavage and then polyadenylation.Polyadenylation is mediated by an enzyme called poly-A polymerase, which adds about 200 adenines to the RNA's 3′ end produced by the cleavage.This enzyme uses ATP as a precursor and adds the nucleotides using the same chemistry as RNA polymerase, but it does so without a template. The mature mRNA is then transported from the nucleus.

Thus, a mature mRNA is released from polymerase once the gene has been transcribed. But the enzyme does not terminate immediately when the RNA is cleaved and polyadenylated. Rather, it continues to move along the template, generating a second RNA molecule that can become as long as several hundred nucleotides before terminating. The polymerase then dissociates from the template, releasing the new RNA, which is degraded without ever leaving the nucleus.

References

1. J. D. Watson, T. A. Baker, S. P. Bell, A. Gann, M. Levine, R. Losick. Molecular Biology of the Gene, Fifth edition, Pearson Publication.
2. H. Lodish, A. Berk, C. A. Kaiser, M. P. Scott, A. Bretscher, H. Ploegh, P. Matsudaira. Molecular Biology of Cell, Sixth edition, Freeman Publication.
3. P. Kumar, U. Mena. Life Sciences, Part-II, Pathfinder Publication.