Gene Expression

Gene expression is the process by which the instructions in our DNA are converted into a functional product, such as a protein.

- When the information stored in our DNA[?] is converted into instructions for making proteins[?] or other molecules, it is called gene expression[?].
- Gene expression is a tightly regulated process that allows a cell to respond to its changing environment.
- It acts as both an on/off switch to control when proteins are made and also a volume control that increases or decreases the amount of proteins made.
- There are two key steps involved in making a protein, transcription and translation.

Genetic code

Genetic code, the sequence of nucleotides in deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) that determines the amino acid sequence of proteins. Though the linear sequence of nucleotides in DNA contains the information for protein sequences, proteins are not made directly from DNA. Instead, a messenger RNA (mRNA) molecule is synthesized from the DNA and directs the formation of the protein. RNA is composed of four nucleotides: adenine (A), guanine (G), cytosine (C), and uracil (U). Three adjacent nucleotides constitute a unit known as the codon, which codes for an amino acid. For example, the sequence AUG is a codon that specifies the amino acid methionine. There are 64 possible codons, three of which do not code for amino acids but indicate the end of a protein. The remaining 61 codons specify the 20 amino acids

that make up proteins. The AUG codon, in addition to coding for methionine, is found at the beginning of every mRNA and indicates the start of a protein. Because most of the 20 amino acids are coded for by more than one codon, the code is called degenerate.

The genetic code, once thought to be identical in all forms of life, has been found to diverge slightly in certain organisms and in the mitochondria of some eukaryotes. Nevertheless, these differences are rare, and the genetic code is identical in almost all species, with the same codons specifying the same amino acids. The deciphering of the genetic code was accomplished by the American biochemists Marshall W. Nirenberg, Robert W. Holley, and Har Gobind Khorana in the early 1960s.

Transcription :

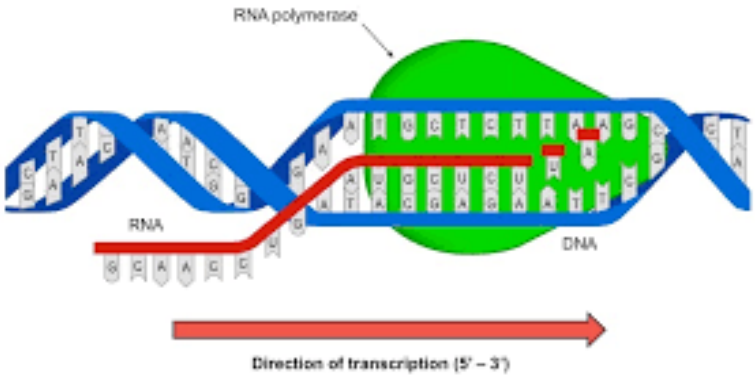
Transcription, the synthesis of RNA from DNA. Genetic information flows from DNA into protein, the substance that gives an organism its form. This flow of information occurs through the sequential processes of transcription (DNA to RNA) and translation (RNA to protein).

Transcription proceeds in the following general steps:

1. RNA polymerase, together with one or more general transcription factors, binds to promoter DNA.
2. RNA polymerase creates a transcription bubble, which separates the two strands of the DNA helix. This is done by breaking the hydrogen bonds between complementary DNA nucleotides.
3. RNA polymerase adds RNA nucleotides (which are complementary to the nucleotides of one DNA strand).

4. RNA sugar-phosphate backbone forms with assistance from RNA polymerase to form an RNA strand.
5. Hydrogen bonds of the RNA–DNA helix break, freeing the newly synthesized RNA strand.
6. If the cell has a nucleus, the RNA may be further processed. This may include polyadenylation, capping, and splicing.
7. The RNA may remain in the nucleus or exit to the cytoplasm through the nuclear pore complex.

Transcription is divided into initiation, promoter escape, elongation, and termination.



Initiation

Transcription begins with the binding of RNA polymerase, together with one or more general transcription factors, to a specific DNA sequence referred to as a "promoter" to form an RNA polymerase-promoter "closed complex". In the "closed complex" the promoter DNA is still fully double-stranded.

RNA polymerase, assisted by one or more general transcription factors, then unwinds approximately 14 base pairs of DNA to form an RNA

polymerase-promoter "open complex". In the "open complex" the promoter

DNA is partly unwound and single-stranded. The exposed, single-stranded DNA is referred to as the "transcription bubble."

RNA polymerase, assisted by one or more general transcription factors, then selects a transcription start site in the transcription bubble, binds to an initiating NTP and an extending NTP (or a short RNA primer and an extending NTP) complementary to the transcription start site sequence, and catalyzes bond formation to yield an initial RNA product.

In bacteria, RNA polymerase holoenzyme consists of five subunits: 2 α subunits, 1 β subunit, 1 β' subunit, and 1 ω subunit. In bacteria, there is one general RNA transcription factor known as a sigma factor. RNA polymerase core enzyme binds to the bacterial general transcription (sigma) factor to form RNA polymerase holoenzyme and then binds to a promoter. (RNA polymerase is called a holoenzyme when sigma subunit is attached to the core enzyme which is consist of 2 α subunits, 1 β subunit, 1 β' subunit only).

In archaea and eukaryotes, RNA polymerase contains subunits homologous to each of the five RNA polymerase subunits in bacteria and also contains additional subunits. In archaea and eukaryotes, the functions of the bacterial general transcription factor sigma are performed by multiple general transcription factors that work together. In archaea, there are three general transcription factors: TBP, TFB, and TFE. In eukaryotes, in RNA polymerase II-dependent transcription, there are six general transcription factors: TFIIA, TFIIB , TFIID , TFIIE , TFIIF, and TFIIH. The TFIID is the first component to bind to DNA due to binding of TBP, while TFIIH is the last component to be recruited. In

archaea and eukaryotes, the RNA polymerase-promoter closed complex is usually referred to as the "preinitiation complex."

Transcription initiation is regulated by additional proteins, known as activators and repressors, and, in some cases, associated coactivators or corepressors, which modulate formation and function of the transcription initiation complex.

Promoter escape

After the first bond is synthesized, the RNA polymerase must escape the promoter. During this time there is a tendency to release the RNA transcript and produce truncated transcripts. This is called abortive initiation, and is common for both eukaryotes and prokaryotes. Abortive initiation continues to occur until an RNA product of a threshold length of approximately 10 nucleotides is synthesized, at which point promoter escape occurs and a transcription elongation complex is formed.

Mechanistically, promoter escape occurs through DNA scrunching, providing the energy needed to break interactions between RNA polymerase holoenzyme and the promoter.

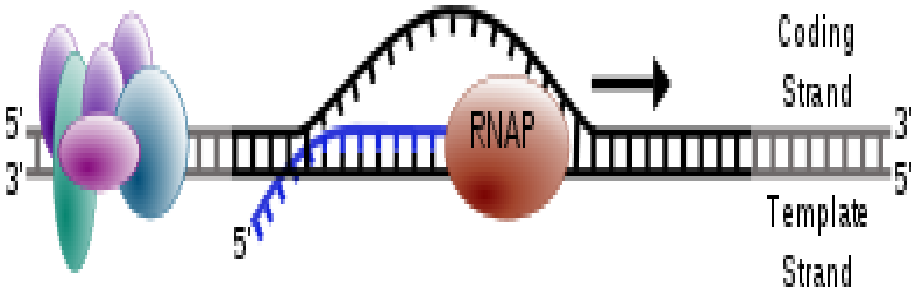
Elongation

One strand of the DNA, the *template strand* (or noncoding strand), is used as a template for RNA synthesis. As transcription proceeds, RNA polymerase traverses the template strand and uses base pairing complementarity with the DNA template to create an RNA copy (which elongates during the traversal). Although RNA polymerase traverses the template strand from 3' → 5', the coding (non-template) strand and newly formed RNA can also be used as reference points, so transcription can be described as occurring 5' → 3'. This produces an RNA molecule from 5' → 3', an exact copy of the coding strand (except that thymines are

replaced with uracils, and the nucleotides are composed of a ribose (5-carbon) sugar where DNA has deoxyribose (one fewer oxygen atom) in its sugar-phosphate backbone).

mRNA transcription can involve multiple RNA polymerases on a single DNA template and multiple rounds of transcription (amplification of particular mRNA), so many mRNA molecules can be rapidly produced from a single copy of a gene. The characteristic elongation rates in prokaryotes and eukaryotes are about 10-100 nts/sec. In eukaryotes, however, nucleosomes act as major barriers to transcribing polymerases during transcription elongation.[13][14] In these organisms, the pausing induced by nucleosomes can be regulated by transcription elongation factors such as TFIIS.

Elongation also involves a proofreading mechanism that can replace incorrectly incorporated bases. In eukaryotes, this may correspond with short pauses during transcription that allow appropriate RNA editing factors to bind. These pauses may be intrinsic to the RNA polymerase or due to chromatin structure.



Termination

Bacteria use two different strategies for transcription termination – Rho-independent termination and Rho-dependent termination. In Rho-independent transcription termination, RNA transcription stops when the newly synthesized RNA molecule forms a G-C-rich hairpin loop followed by a run of Us. When the hairpin forms, the mechanical stress breaks the weak rU-dA bonds, now filling the DNA–RNA hybrid. This pulls the poly-U transcript out of the active site of the RNA polymerase, terminating transcription. In the "Rho-dependent" type of termination, a protein factor called "Rho" destabilizes the interaction between the template and the mRNA, thus releasing the newly synthesized mRNA from the elongation complex.

Transcription termination in eukaryotes is less well understood than in bacteria, but involves cleavage of the new transcript followed by template-independent addition of adenines at its new 3' end, in a process called polyadenylation.