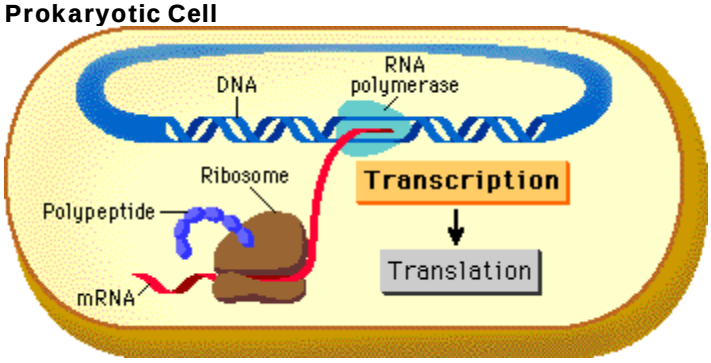
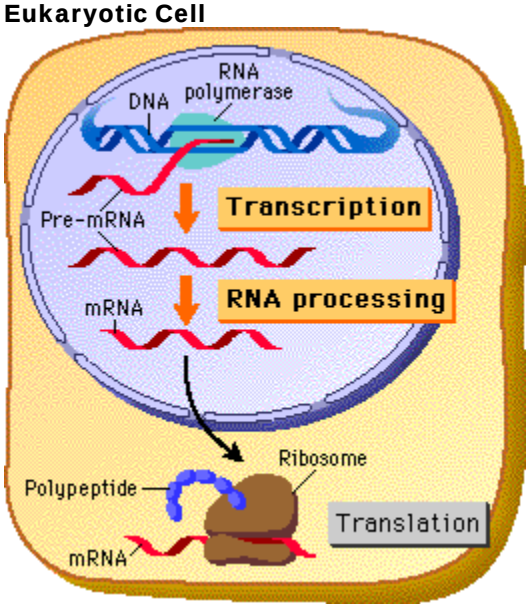
In a prokaryotic cell, transcription and translation are coupled; that is, translation begins while the mRNA is still being synthesized. In a eukaryotic cell, transcription occurs in the nucleus, and translation occurs in the cytoplasm.

Prokaryotic Cell



Because there is no nucleus to separate the processes of transcription and translation, when bacterial genes are transcribed, their transcripts can immediately be translated.

Eukaryotic Cell

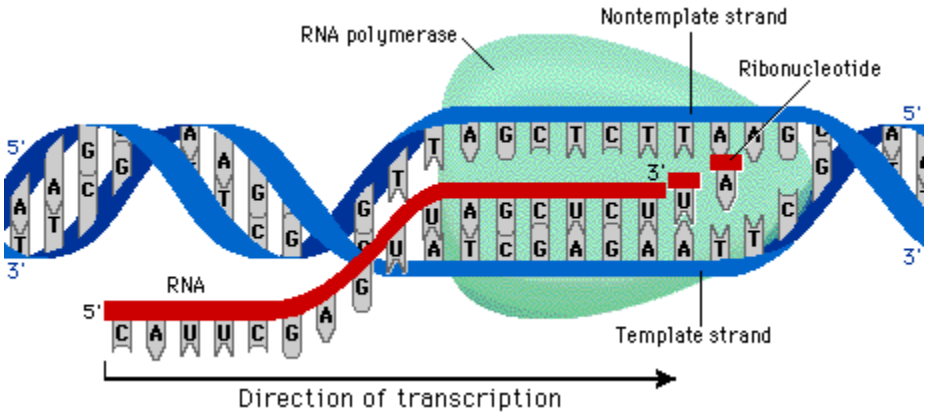


Transcription and translation are spatially and temporally separated in eukaryotic cells; that is, transcription occurs in the nucleus to produce a pre-mRNA molecule.

The pre-mRNA is typically processed to produce the mature mRNA, which exits the nucleus and is translated in the cytoplasm

RNA synthesis involves separation of the DNA strands and synthesis of an RNA molecule in the 5' to 3' direction by RNA polymerase, using one of the DNA strands as a template.

In complementary base pairing, A, T, G, and C on the template DNA strand specify U, A, C, and G, respectively, on the RNA strand being synthesized.

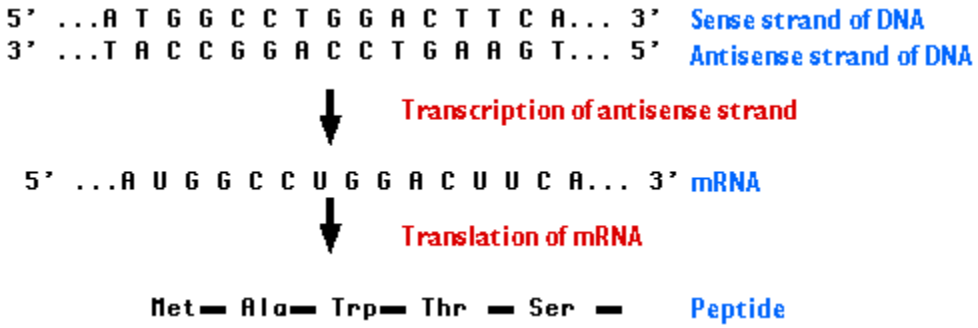


The majority of genes are expressed as the proteins they encode. The process occurs in two steps:

- **Transcription = DNA ⟶ RNA**
- **Translation = RNA ⟶ protein**

Taken together, they make up the "central dogma" of biology: DNA ⟶ RNA ⟶ protein.

Here is an overview.



This page examines the first step:

Gene Transcription: DNA ⟶ RNA

DNA serves as the template for the synthesis of RNA much as it does for its own replication.

The Steps

- Some 50 different protein **transcription factors** bind to **promoter** sites, usually on the 5′ side of the gene to be transcribed.

- An enzyme, an **RNA polymerase**, binds to the complex of transcription factors.
- Working together, they open the DNA double helix.
- The RNA polymerase proceeds down one strand moving in the 3′ → 5′ direction.
- In eukaryotes, this requires — at least for protein-encoding genes — that the nucleosomes in front of the advancing RNA polymerase (RNAP II) be removed. A complex of proteins is responsible for this. The same complex replaces the nucleosomes after the DNA has been transcribed and RNAP II has moved on.
- As the RNA polymerase travels along the DNA strand, it assembles **ribo**nucleotides (supplied as triphosphates, e.g., ATP) into a strand of RNA.
- Each ribonucleotide is inserted into the growing RNA strand following the rules of base pairing. Thus for each C encountered on the DNA strand, a G is inserted in the RNA; for each G, a C; and for each T, an A. However, each A on the DNA guides the insertion of the pyrimidine uracil (**U**, from uridine triphosphate, UTP). There is no T in RNA.
- Synthesis of the RNA proceeds in the 5′ → 3′ direction.
- As each nucleoside triphosphate is brought in to add to the 3′ end of the growing strand, the two terminal phosphates are removed.
- When transcription is complete, the transcript is released from the polymerase and, shortly thereafter, the polymerase is released from the DNA.

Note that at any place in a DNA molecule, either strand may be serving as the template; that is, some genes "run" one way, some the other (and in a few remarkable cases, the same segment of double helix contains genetic information on both strands!). In all cases, however, RNA polymerase proceeds along a strand in its 3′ → 5′ direction.

A report in the 4 January 2001 issue of Nature shows that RNA polymerase actually tracks around the double helix of DNA. In vitro, at least, when RNA polymerase is immobilized, it spins the DNA molecule around and around as it moves along the molecule. Whether it is the polymerase or the DNA that does the spinning in vivo remains to be determined.

Types of RNA

Several types of RNA are synthesized:

- **messenger RNA** (**mRNA**). This will later be **translated** into a polypeptide.
- **ribosomal RNA** (**rRNA**). This will be used in the building of ribosomes: machinery for synthesizing proteins by translating mRNA.
- **transfer RNA** (**tRNA**). RNA molecules that carry amino acids to the growing polypeptide.
- **small nuclear RNA** (**snRNA**). DNA transcription of the genes for mRNA, rRNA, and tRNA produces large precursor molecules (**"primary transcripts"**) that must be processed within the nucleus to produce the functional molecules for export to the cytosol. Some of these processing steps are mediated by snRNAs.

- **small nucleolar RNA** (**snoRNA**). These RNAs within the nucleolus have several functions (described below).
- **microRNA** (**miRNA**). These are tiny (~22 nts) RNA molecules that appear to regulate the expression of messenger RNA (mRNA) molecules
- **XIST** RNA. This inactivates one of the two X chromosomes in female vertebrates.

Ribosomal RNA (rRNA)

There are 4 kinds. In eukaryotes, these are

- **18S rRNA**. One of these molecules, along with some 30 different protein molecules, is used to make the **small subunit** of the ribosome.
- **28S, 5.8S, and 5S rRNA**. One each of these molecules, along with some 45 different proteins, are used to make the **large subunit** of the ribosome.

The S number given each type of rRNA reflects the rate at which the molecules sediment in the ultracentrifuge. The larger the number, the larger the molecule (but not proportionally).

The 28S, 18S, and 5.8S molecules are produced by the processing of a single primary transcript from a cluster of identical copies of a single gene. The 5S molecules are produced from a different cluster of identical genes.

Transfer RNA (tRNA)

There are some 32 different kinds of tRNA in a typical eukaryotic cell.

- Each is the product of a separate gene.
- They are small (~4S), containing 73-93 nucleotides.
- Many of the bases in the chain pair with each other forming sections of double helix.
- The unpaired regions form 3 loops.
- Each kind of tRNA carries (at its 3′ end) one of the 20 **amino acids** (thus most amino acids have more than one tRNA responsible for them).
- At one loop, 3 unpaired bases form an **anticodon**.
- Base pairing between the anticodon and the complementary codon on a mRNA molecule brings the correct amino acid into the growing polypeptide chain. Further details of this process are described in the discussion of translation.

Messenger RNA (mRNA)

Messenger RNA comes in a wide range of sizes reflecting the size of the polypeptide it encodes. Most cells produce small amounts of thousands of different mRNA molecules, each to be translated into a peptide needed by the cell.

Many mRNAs are common to most cells, encoding "housekeeping" proteins needed by all cells (e.g. the enzymes of glycolysis). Other mRNAs are specific for only certain types of cells. These encode proteins needed for the function of that particular cell (e.g., the mRNA for hemoglobin in the precursors of red blood cells).

Small Nuclear RNA (snRNA)

Approximately a dozen different genes for snRNAs, each present in multiple copies, have been identified. The snRNAs have various roles in the processing of the other classes of RNA. For example, several snRNAs are part of the **spliceosome** that participates in converting pre-mRNA into mRNA by excising the introns and splicing the exons. [Link down to the discussion of RNA processing.]

Small Nucleolar RNA (snoRNA)

As the name suggests, these RNAs (there are probably over 100 of them) are found in the nucleolus where they are responsible for several functions:

- Some participate in making ribosomes by helping to cut up the large RNA precursor of the 28S, 18S, and 5.8S molecules.
- Others chemically modify many of the nucleotides in these molecules, e.g., by adding methyl groups to ribose.
- Still others serve as the template for the synthesize of telomeres.

In vertebrates, the snoRNAs are made from **introns** removed during RNA processing.

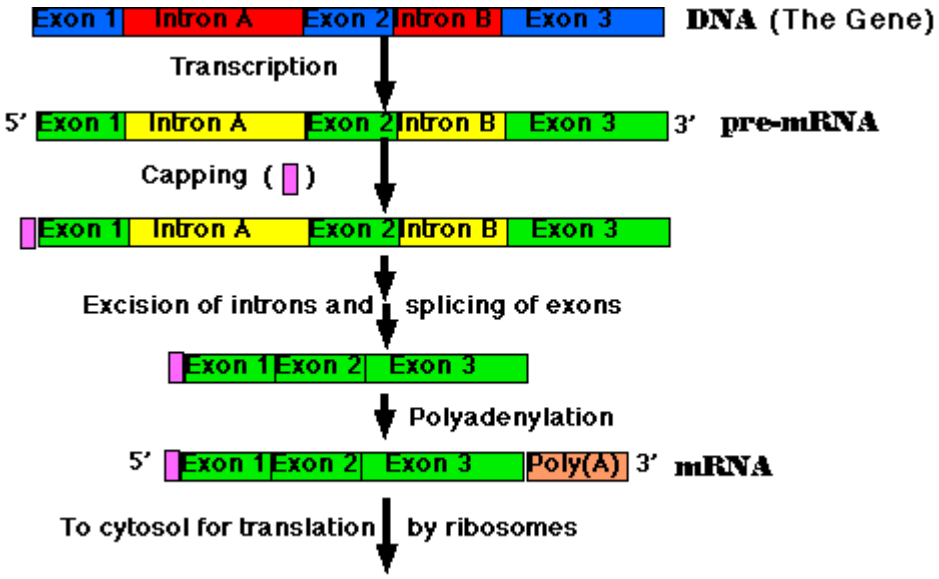
The RNA polymerases

The RNA polymerases are huge multi-subunit protein complexes. Three kinds are found in eukaryotes.

- RNA polymerase I (**Pol I**). It transcribes the **rRNA** genes for the precursor of the 28S, 18S, and 5.8S molecules (and is the busiest of the RNA polymerases).
- RNA polymerase II (**Pol II**; also known as **RNAP II**). It transcribes protein-encoding genes into **mRNA** (and also the **snRNA** genes).
- RNA polymerase III (**Pol III**). It transcribes the 5S rRNA genes and all the **tRNA** genes.

RNA Processing: pre-mRNA ⟶ mRNA

All the primary transcripts produced in the nucleus must undergo processing steps to produce functional RNA molecules for export to the cytosol. We shall confine ourselves to a view of the steps as they occur in the processing of pre-mRNA to mRNA.



The steps:

- Synthesis of the **cap**. This is a modified guanine (G) which is attached to the 5′ end of the pre-mRNA as it emerges from RNA polymerase II (RNAP II). The cap protects the RNA from being degraded by enzymes that degrade RNA from the 5′ end.
- Step-by-step removal of **introns** present in the pre-mRNA and splicing of the remaining **exons**. This step is required because most eukaryotic genes are split. It takes place as the pre-mRNA continues to emerge from RNAP II.
- Synthesis of the **poly(A) tail**. This is a stretch of adenine (A) nucleotides. When a special poly(A) attachment site in the pre-mRNA emerges from RNAP II, the transcript is cut there, and the poly(A) tail is attached to the exposed 3′ end. This completes the mRNA molecule, which is now ready for export to the cytosol. (The remainder of the transcript is degraded, and the RNA polymerase leaves the DNA.)

Split Genes

Most eukaryotic genes are split into segments. In decoding the open reading frame of a gene for a known protein, one usually encounters periodic stretches of DNA calling for

amino acids that do not occur in the actual protein product of that gene. Such stretches of DNA, which get transcribed into RNA but not translated into protein, are called **introns**. Those stretches of DNA that do code for amino acids in the protein are called **exons**. Examples:

- The gene for one type of collagen found in chickens is split into 52 separate exons.
- The gene for **dystrophin**, which is mutated in boys with muscular dystrophy, has 79 exons.
- even the genes for rRNA and tRNA are split.

In general, introns tend to be much longer than exons. An average eukaryotic exon is only 140 nts long, but one human intron stretches for 480,000 nucleotides!

The cutting and splicing of mRNA must be done with great precision. If even one nucleotide is left over from an intron or one is removed from an exon, the **reading frame** from that point on will be shifted, producing new codons specifying a totally different sequence of amino acids from that point to the end of the molecule (which often ends prematurely anyway when the shifted reading frame generates a **STOP** codon).

The removal of introns and splicing of exons is done with the **spliceosome**. This is a complex of several **snRNA** molecules and some 145 different proteins.

The introns in most pre-mRNAs begin with a GU and end with an AG. Presumably these short sequences assist in guiding the spliceosome.

Alternative Splicing

The processing of pre-mRNA for many proteins proceeds along various paths in different cells or under different conditions. For example, early in the differentiation of a B cell (a lymphocyte that synthesizes an antibody) the cell first uses an exon that encodes a transmembrane domain that causes the molecule to be retained at the cell surface. Later, the B cell switches to using a different exon whose domain enables the protein to be secreted from the cell as a circulating antibody molecule.

Alternative splicing provides a mechanism for producing a wide variety of proteins from a small number of genes.

While we humans may turn out to have only 25 to 30 thousand genes, we probably make at least 10 times that number of different proteins. More than 50% of our genes produce pre-mRNAs that are alternatively-spliced.

One of the most dramatic examples of alternative splicing is the **DSCAM** gene in **Drosophila**. This single gene contains some 108 exons of which 17 are retained in the final mRNA. Some exons are always included; others are selected from an array.

Theoretically this system is able to produce 38,016 different proteins. And, in fact, of 50 cDNAs synthesized at random from mRNAs, 49 of them turned out to be unique.

These DSCAM proteins are involved in guiding neurons to their proper destination. Perhaps the incredible diversity of synaptic junctions in the mammalian c.n.s. ($\sim 10^{14}$) is mediated by alternative splicing of a limited number of gene transcripts.

So, whether a particular segment of RNA will be retained as an exon or excised as an intron can vary under different circumstances. Clearly the switching to an alternate splicing pathway must be closely regulated.

Why split genes?

Perhaps during evolution, eukaryotic genes have been assembled from smaller, primitive genes - today's exons. Some proteins, like the **antibodies** mentioned in the previous section, are organized in a set of separate sections or **domains** each with a special function to perform in the complete molecule. Each domain is encoded by a separate exon. Having the different functional parts of the antibody molecule encoded by separate exons makes it possible to use these units in different combinations. Thus a set of exons in the genome may be the genetic equivalent of the various modular pieces in a box of "Lego" for children to assemble in whatever forms they wish.

But the boundaries of other exons do not seem to correspond to domain boundaries of the protein. Furthermore, rRNA and tRNA genes are also split, and these do not encode proteins. So perhaps some exons are simply "junk" DNA that was inserted into the gene at some point in evolution without causing any harm

Script for Translational Initiation

Scene 1

In order for the genetic information in a messenger RNA to be translated accurately into a protein, the ribosome must correctly identify
the precise starting position for decoding. This step is so critical that a complex translational initiation system has evolved just to get the
process started accurately.

Translational initiation occurs as follows in bacteria like E. coli . An inactive 70S ribosome must first be separated into its component
subunits, since the initiation complex forms on the small (30S) subunit alone.

The separation of the subunits is promoted by the translational initiation factor IF-3. (The translation factors are ancillary proteins that
assist in the process of protein synthesis but are not components of the ribosome.)

Initiation factor IF-1 assists IF-3 in isolating the small subunit.

The small subunit, with IF-1 and IF-3 bound, can now associate with mRNA and with the special initiator aminoacyl tRNA, fMet-tRNA.

mRNA and the Ribosome

Before polypeptide synthesis can occur, the correct starting place for decoding (that is, the initiation codon) must be selected. Consider these E. coli mRNA sequences near the start codon. In E. coli a sequence of nucleotides just upstream of the start codon is complementary to a region of the small ribosomal subunit's RNA, the 16S rRNA. This region of the mRNA is called the Shine-Dalgarno sequence. The complement of this sequence is at the 3' end of the 16S rRNA. The base-pairing of these sequences helps the mRNA to bind tightly to the small ribosomal subunit, and it helps to position the start codon in the proper place to find the anticodon of the initiator tRNA.

Aminoacyl tRNA

The initiator tRNA, tRNA-fMet, is specialized to translate only the initiation codon of an mRNA. Unlike the tRNAs that elongate the growing polypeptide, it is charged with an N-blocked amino acid, N-formylmethionine. The formyl group is added after the tRNA has been charged with methionine.

After the polypeptide has been synthesized, the N-terminal formyl group is removed. The initial methionine is frequently removed as well as part of the post-translational processing.

Scene 2

The initiator aminoacyl tRNA is in a complex with another initiation factor, IF-2. The role of IF-2 is to select the initiator tRNA from among the many aminoacyl tRNAs in the cell, and to help it bind to the small ribosomal subunit. IF-2 carries GTP as a ligand.

The mRNA and fMet-tRNA bind to the small ribosomal subunit in either order. The initiation factors help to ensure that the initiation codon of the mRNA pairs with the anticodon bases of fMet-tRNA. This is the 30S Initiation Complex.

The large (50S) ribosomal subunit can now join the complex. IF-3 must leave the initiation complex before this can happen. IF-2 is

stimulated to hydrolyze GTP to GDP, then it and IF-1 depart from the complex too. The result is the 70S Initiation Complex. fMet-tRNA is
bound in the ribosome's P site and the A site is vacant, awaiting the aminoacyl tRNA that will decode the second codon of the mRNA
and incorporate the second amino acid of the polypeptide.

Script for Translational Elongation

Scene 1

Amino acids are added to a growing polypeptide during the elongation stage of translation. The ribosome has a binding site for the tRNA
bearing the nascent peptide. This is called the P site (for "peptidyl"). The ribosome also has a site for the aminoacyl tRNA bearing the
next amino acid to be added to the polypeptide. This is the A site.

Elongation is assisted by several additional proteins known as elongation factors. In E. coli, these are EF-Tu, EF-Ts, and EF-G.

We begin by considering a ribosome with a growing peptide attached to a tRNA in the P site, and a vacant A site.

Scene 2

The next aminoacyl tRNA binds to the ribosome in complex with the elongation factor EF-Tu. EF-Tu also carries a GTP molecule, which
is hydrolyzed in the course of the reaction.

If the anticodon of the tRNA correctly decodes the next, adjacent codon of the mRNA, the ribosome will tightly bind this aminoacyl tRNA.

The anticodon stems of the tRNAs are largely bound to the small ribosomal subunit, while the aminoacyl stems of the tRNAs are largely
bound to the large subunit.

When the aminoacyl tRNA successfully binds to the A site, its associated EF-Tu becomes activated to hydrolyze its bound GTP to GDP.
The EF-Tu then dissociates from the ribosome.

Scene 3

In order for EF-Tu to participate in a further round of elongation, it must replace its bound GDP with GTP. This process is assisted by the
elongation factor EF-Ts. In the presence of EF-Ts, EF-Tu rapidly exchanges these nucleotides. The EF-G·GTP complex then binds
another aminoacyl tRNA preparatory for a new round of elongation.

Scene 4

Once the correct aminoacyl tRNA is bound to the A site, the nascent polypeptide that is linked to the peptidyl tRNA is rapidly transferred to the aminoacyl group of the aminoacyl tRNA. In this so-called transpeptidation reaction, the aminoacyl tRNA's free amino group nucleophilically displaces the tRNA from the peptidyl tRNA ester and thereby forms a new peptide bond. Transpeptidation requires that the aminoacyl stems of the two tRNAs be very close in space and properly oriented.

Scene 5

The final stage of the elongation reaction involves shifting of the two tRNAs and their bound mRNA relative to the ribosome. This moves the mRNA's next codon into the now empty A site, where it can base pair with its corresponding aminoacyl tRNA. This process, which is called translocation, is facilitated by the elongation factor EF-G. EF-G, in complex with GTP, binds to the ribosome after transpeptidation.

The binding of EF-G to the ribosome promotes the motion of tRNAs and mRNA relative to the ribosome in a process that is driven by the hydrolysis of the EF-G's bound GTP to GDP. The newly deacylated tRNA and the EF-G are then successively released from the ribosome.

Scene 6

As with EF-Tu, in order for EF-G to function in an additional round of elongation, it must replace its GDP with GTP. However, unlike EF-Tu, this occurs without the assistance of another factor.

Scene 7

Following translocation, the ribosome is in the same state it was when we started, except that the polypeptide is now one amino acid residue longer. This cycle repeats, just as before, over and over, until the ribosome detects a stop codon in the mRNA, a process that is mediated by release factors. The completed polypeptide is then hydrolyzed off of the final peptidyl tRNA, followed by the release of the final tRNA, the release factors, and the mRNA.