The flow of information from DNA → RNA → protein was called the Central Dogma by Francis Crick in 1956.

Synthesis of an RNA molecule using a DNA template is called transcription. Only one of the DNA strands is transcribed. The enzyme used is RNA polymerase.

There are four major types of RNA molecules:
- a. Messenger RNA (mRNA) encodes the amino acid sequence of a polypeptide.
- b. Transfer RNA (tRNA) brings amino acids to ribosomes during translation.
- c. Ribosomal RNA (rRNA) combines with proteins to form a ribosome, the catalyst for translation.
- d. Small nuclear RNA (snRNA) combines with proteins to form complexes used in eukaryotic RNA processing.

Transcription, or gene expression, is regulated by gene regulatory elements associated with each gene. DNA unwinds in the region next to the gene, due to RNA polymerase in bacteria and other proteins in eukaryotes.

RNA is transcribed 5'-to-3'. The template DNA strand is read 3'-to-5'.

RNA polymerization is similar to DNA synthesis except:
- a. The precursors are NTPs (not dNTPs).
- b. No primer is needed to initiate synthesis.
- c. Uracil is inserted instead of thymine.

Transcription is divided into three steps for both bacteria and eukaryotes.
- a. Initiation,
- b. Elongation,
- c. Termination.

The process of elongation is highly conserved between bacteria and eukaryotes, but initiation and termination are somewhat different.

A prokaryotic gene is a DNA sequence in the chromosome. The gene has three regions, each with a function in transcription:
- a. A promoter sequence that attracts RNA polymerase to begin transcription
- b. The transcribed sequence, called the RNA-coding sequence.
- c. A terminator region that specifies where transcription will stop.

The common *E. coli* promoter
- a. For the −35 region the consensus is 5’-TTGACA-3’.
- b. For the −10 region the consensus is 5’-TATAAT-3’.

Transcription initiation requires the RNA polymerase holoenzyme (only one type in bacteria) to bind to the promoter DNA sequence. Holoenzyme consists of:
- a. Core enzyme of RNA polymerase, containing four polypeptides (two α, one β, and one β’).
- b. Sigma factor (σ), which binds the core enzyme and recognizes promoters.

RNA polymerase holoenzyme binds promoter in two steps:
- a. First, it loosely binds to the −35 sequence of dsDNA (closed promoter complex).
- b. Second, it binds tightly to the −10 sequence, un twisting about 17 bp of DNA at the site, and in position to begin transcription (open promoter complex).

The common *E. coli* promoter
- a. For the −35 region the consensus is 5’-TTGACA-3’.
- b. For the −10 region the consensus is 5’-TATAAT-3’.

Once initiation is done, RNA synthesis begins and the sigma factor is released.

Core enzyme untwists DNA helix locally, allowing a small region to denature. Newly synthesized RNA forms an RNA-DNA hybrid.

RNA polymerase has two types of proofreading:
- a. Similar to DNA polymerase editing, newly inserted nucleotide is removed by reversing synthesis reaction.
- b. Enzyme moves back one or more nucleotides, cleaves RNA, and then resumes synthesis in a forward direction.
**Termination of an RNA Chain**
- Terminator sequences are used to end transcription. Two types in *E. coli*:
  - a. Rho–independent (r–independent) or type I terminators have twofold symmetry and hairpin loop forms. The palindrome is followed by 4–8U residues in the transcript.
  - b. Rho–dependent (r–dependent) or type II terminators require the protein Rho binds to the C-rich sequence in the RNA upstream of the termination site. It then acts as a helicase, using ATP hydrolysis for energy to move along the transcript and destabilize the RNA–DNA hybrid.

**Transcription in Eukaryotes:**
- Eukaryotes contain three different RNA polymerases:
  - a. RNA polymerase I, located in the nucleolus, transcribes the three major rRNAs (28S, 18S, and 5.8S).
  - b. RNA polymerase II, located in the nucleoplasm, transcribes mRNAs and some snRNAs.
  - c. RNA polymerase III, located in the nucleoplasm, transcribes tRNAs, 5S rRNA, and the remaining snRNAs.
- Eukaryotic RNA polymerases are harder to study as present at low concentrations.
- Eukaryotic RNA polymerases have multiple subunits. An example is yeast RNA pol II with 12 subunits, 5 of which are also in its RNA pol III.

**Promoters and Enhancers**
- When protein-coding genes are first transcribed by RNA pol II, the product is a precursor mRNA (pre-mRNA). Pre-mRNA is modified to produce a mature mRNA.
- Promoters for protein-coding genes are analyzed in two ways:
  - Directed mutation.
  - Comparison of sequences from known genes.
- Results of promoter analysis reveal two types of elements:
  - a. Core promoters are located near the transcription start site and specify where transcription begins. Examples include:
    - i. The CAAT box, located near the transcription start site.
    - ii. The GC box, consensus sequence GGGCGG, located at about -90.
- Enhancers are another cis-acting element. They are required for maximal transcription of a gene.
  - i. Glucose-6-phosphate dehydrogenase
  - ii. The CAT box (also known as a TATA element) at -30.
- Enhancers are usually upstream of the transcription initiation site but may also be internal.
- Enhancers are regions in the DNA used to control expression of a gene.
  - a. Enhancers are usually upstream of the transcription initiation site but may also be internal.
  - b. Enhancers may modulate transcription by activating or repressing transcription.
  - c. Enhancers are regions in the DNA used to control expression of a gene.

**Core and Proximal Elements**
- Various combinations of core and proximal elements are found near different genes. Promoter-proximal elements are key to gene expression:
  - a. Actin genes are prominent in transcription regulation, recognized by activator proteins.
  - b. Housekeeping genes have common core elements and are recognized by activator proteins.
- Enhancers are another cis-acting element. They are required for maximal transcription of a gene.
  - a. Enhancers are usually upstream of the transcription initiation site but may also be internal.
  - b. Enhancers may modulate transcription by activating or repressing transcription.
  - c. Enhancers are regions in the DNA used to control expression of a gene.
- Enhancers are regions in the DNA used to control expression of a gene.

**Transcription Initiation**
- Transcription initiation requires assembly of RNA polymerase II and binding of general transcription factors (GTFs) on the core promoter.
  - a. GTFs are needed for initiation by all three RNA polymerases.
  - b. GTFs are numbered to match their corresponding RNA polymerase and lettered in the order of discovery (e.g., TFIIH was the fourth GTF discovered that works with RNA polymerase II).
- Sequence of binding is not completely understood:
  - a. Binding of GTFs and RNA pol II occurs in a set order in vitro experiments to produce the complete transcription initiation complex or preinitiation complex (PIC).
  - b. The situation is less clear in vivo. Some data indicate that initiation complex forms before binding promoters.
- Transcription for eukaryotes is also complicated by the nucleosome organization of chromosomes.

**The Structure and Production of Eukaryotic mRNAs**
- Eukaryotic mRNAs have three main parts:
  - a. The 5' leader sequence, or 5' untranslated region (5' UTR), varies in length.
  - b. The coding sequence, which specifies the amino acid sequence of the protein that will be produced during translation. It varies in length according to the size of the protein that it encodes.
  - c. The trailer sequence, or 3' untranslated region (3' UTR), also varies in length and contains information influencing the stability of the mRNA.
- Eukaryotes and bacteria produce mRNAs somewhat differently.
  - a. Bacteria use the RNA transcript as mRNA without modification.
  - b. Eukaryotes modify pre-RNA into mRNA by RNA processing. The processed mRNA migrates from nucleus to cytoplasm before translation. Messages are always monocistronic. This allows for additional levels of control of gene expression.

**Eukaryotic RNA Pol II with 12 subunits, 5 of which are also in its RNA pol III.**
5’ Modification

- The newly made 5’ end of the mRNA is modified by 5’ capping. A capping enzyme adds a guanine, usually 7-methyl guanosine (m7G), to the 5’ end using a 5’-to-5’ linkage. Sugars of the two adjacent nt are also methylated. The cap is used for ribosome binding to the mRNA during translation initiation.

3’ Modification

- The 3’ end of mRNA is marked by a poly(A) tail.
  - a. Transcription of mRNA continues through the poly(A) consensus sequence (AAUAAA).
  - b. Proteins bind and cleave RNA. These include:
    - i. CPSF (cleavage and polyadenylation specificity factor).
    - ii. Cleavage stimulation factor.
    - iii. Two cleavage factor proteins (CFI and CFII).
  - c. After cleavage, the enzyme poly(A) polymerase (PAP) adds A nucleotides to the 3’ end of the RNA, using ATP as a substrate. PAP is bound to CPSF during this process.
  - d. PABP (poly(A) binding protein II) binds the poly(A) tail as it is produced.

RNA polymerase II continues to synthesize NA. Several models have been proposed for termination. In one model, degradation of the post-poly(A) site reaches the RNA pol II and somehow ends transcription.

Introns

- Eukaryotic pre-RNAs often have introns (intervening sequences) between the exons (expressed sequences) that are removed during RNA processing.
- Removal of introns is necessary for mRNA maturation, as hnRNA (heteronuclear RNA) becomes functional mRNA.
- In Philip Leder’s lab (1978) it was shown that the mouse b-globin pre-mRNA (part of the cell’s hnRNA) is colinear with the gene that encodes it, while the mature b-globin mRNA is shorter than the gene. The missing RNA was an intron that was removed during RNA processing.
- Introns are found in most eukaryotic genes that encode proteins, and they have been found in bacteriophage genes and in some bacterial rRNA and tRNA genes.

Processing of Pre-mRNA to Mature mRNA

- Events in eukaryotic mRNA production include:
  - a. Transcription of the gene by RNA polymerase II.
  - b. Addition of the 5’ cap.
  - c. Addition of the poly(A) tail.
  - d. Splicing to remove introns.

Intron removal

- Introns typically begin with 5’-GU, and end with AG-3’, but mRNA splicing signals involve more than just these two small sequences.
  - a. Spliceosomes are small nuclear ribonucleoprotein particles (snRNPs) associated with pre-mRNAs.
  - b. Spliceosome principal snRNAs are U1, U2, U4, U5, and U6.
    - i. Each snRNA is associated with several proteins.
    - ii. U4 and U6 are part of the same snRNP. Others are in their own snRNPs.
    - iii. Each snRNP type is abundant (~10^5 copies per nucleus).
The steps of splicing:

- a. U1 snRNP binds the 5' splice junction of the intron, as a result of base-pairing of the U1 snRNA to the intron RNA.
- b. U2 snRNP binds by base pairing to the branch-point sequence upstream of the 3' splice junction.
- c. U4/U6 and U5 snRNPs interact and then bind the U1 and U2 snRNPs, creating a loop in the intron.
- d. U4 snRNP dissociates from the complex, forming the active spliceosome.
- e. The spliceosome cleaves the intron at the 5' splice junction, freeing it from exon 1. The free 5' end of the intron bonds to a specific nucleotide (usually A) in the branch-point sequence to form an RNA lariat.
- f. The spliceosome cleaves the intron at the 3' splice junction, liberating the intron lariat. Exons 1 and 2 are ligated, and the snRNPs are released.

Alternative splicing can be used to produce different polypeptides from the same gene.

RNA Editing:

RNA editing adds or deletes nucleotides from a pre-mRNA, or chemically alters the bases, resulting in an mRNA with bases that don't match its DNA coding sequence.

Examples have been found in a number of organisms:

- a. In *Trypanosome brucei* (a protozoan causing sleeping sickness), the cytochrome oxidase subunit III gene (from mitochondrial DNA) does not match its mRNA. Uracil residues have been added and removed, and over 50% of the mature mRNA consists of posttranscriptionally added Us.
- b. In *Physarum polycephalum*, a slime mold, single C nucleotides are added posttranscriptionally at many positions in mRNAs from several mitochondrial genes.
- c. In higher plants, many mitochondrial and chloroplast mRNAs undergo C-to-U editing, including production of an AUG initiation codon from an ACG codon in some chloroplast mRNAs.
- d. In mammals, C-to-U editing occurs in the mRNA for apolipoprotein B, resulting in a tissue-specific stop codon. A-to-G editing occurs in the glutamate receptor mRNA, and pyrimidine editing occurs in some tRNAs.