From Gene to Protein

I. Transcription and translation are the two main processes linking gene to protein.
   A. RNA is chemically similar to DNA, except that it contains ribose as its sugar and substitutes the nitrogenous base uracil for thymine. An RNA molecule almost always consists of a single strand.
   B. The specific sequence of hundreds or thousands of nucleotides in each gene carries the information for the primary structure of proteins, the linear order of the 20 possible amino acids.
   C. During **transcription,** a DNA strand provides a template for the synthesis of a complementary RNA strand or **messenger RNA (mRNA)** molecule.
   D. During **translation,** there is a change of language from nucleic acid language to amino acid language.
   E. Although transcription might seem unnecessary, it provides benefits.
      1. The use of an RNA intermediate provides protection for DNA and its genetic information.
      2. Using an RNA intermediate allows more copies of a protein to be made simultaneously, since many RNA transcripts can be made from one gene. Also, each gene transcript can be translated simultaneously.
   F. Because bacteria lack nuclei, their DNA is not segregated from ribosomes and other protein-synthesizing equipment. Ribosomes attach to the leading end of an mRNA molecule while transcription is still in progress.
   G. In a eukaryotic cell, transcription occurs in the nucleus, and translation occurs at ribosomes in the cytoplasm.
      1. The transcription of a protein-coding eukaryotic gene results in *pre-mRNA.*
      2. The initial RNA transcript of any gene is called a primary transcript.
      3. **RNA processing** yields the finished mRNA.

II. In the genetic code, nucleotide triplets specify amino acids.
   A. Triplets of nucleotide bases are the smallest units of uniform length that can code for all the amino acids. With a **triplet code,** three consecutive bases specify an amino acid, creating 4^3 (64) possible code words.
   B. The genetic instructions for a polypeptide chain are written in DNA as a series of nonoverlapping three-nucleotide words.
   C. During transcription, one DNA strand, the **template strand,** provides a template for building the RNA transcript.
      1. The complementary RNA molecule is synthesized according to base-pairing rules, except that uracil is the complementary base to adenine.
      2. Like a new strand of DNA, the RNA molecule is synthesized in an antiparallel direction to the template strand of DNA.
      3. The mRNA base triplets are called **codons,** and they are written in the 5’ to 3’ direction.
   D. A given DNA strand can be the template strand for some genes along a DNA molecule, while for other genes in other regions, the complementary strand may function as the template.
   E. During translation, the sequence of codons along an mRNA molecule is translated into a sequence of amino acids making up the polypeptide chain.
      1. Each codon specifies which one of the 20 amino acids will be incorporated at the corresponding position along a polypeptide.
      2. Sixty-one of 64 triplets code for amino acids.
      3. The codon AUG not only codes for the amino acid methionine, but also indicates

the “start” of translation.

4. Three codons do not indicate amino acids but are “stop” signals marking the termination of translation.

III. The synthesis and processing of RNA

A. **RNA polymerase** separates the DNA strands at the appropriate point and bonds the RNA nucleotides as they base-pair along the DNA template.

B. Like DNA polymerases, RNA polymerases can only assemble a polynucleotide in its 5’ to 3’ direction.

C. Unlike DNA polymerases, RNA polymerases are able to start a chain *de novo*; they don’t need a primer.

D. Specific sequences of nucleotides along the DNA mark where gene transcription begins and ends.

1. RNA polymerase attaches and initiates transcription at the **promoter.** The presence of a promoter sequence determines which strand of the DNA helix is the template.
2. In prokaryotes, the sequence that signals the end of transcription is called the **terminator.**
3. Molecular biologists refer to the direction of transcription as “downstream” and the other direction as “upstream.”

E. The stretch of DNA that is transcribed into an RNA molecule is called a **transcription unit.**

F. Within the promoter is the starting point for the transcription of a gene. The promoter also includes a binding site for RNA polymerase several dozen nucleotides “upstream” of the start point.

1. In eukaryotes, proteins called **transcription factors** mediate the binding of RNA polymerase and the initiation of transcription.
2. Only after certain transcription factors are attached to the promoter does RNA polymerase II bind to it.
3. The completed assembly of transcription factors and RNA polymerase II bound to a promoter is called a **transcription initiation complex.**
4. A crucial promoter DNA sequence is called a **TATA box.**

G. RNA polymerase then starts transcription. As RNA polymerase moves along the DNA, it untwists the double helix, 10 to 20 bases at time. The enzyme adds nucleotides to the 3' end of the growing strand. Transcription progresses at a rate of 60 nucleotides per second in eukaryotes.

H. Behind the point of RNA synthesis, the double helix re-forms and the RNA molecule peels away.

I. Transcription proceeds until after the RNA polymerase transcribes a terminator sequence in the DNA.

1. In eukaryotes, the pre-mRNA is cleaved from the growing RNA chain while RNA polymerase II continues to transcribe the DNA.
    a. Specifically, the polymerase transcribes a DNA sequence called the polyadenylation signal sequence that codes for a polyadenylation sequence (AAUAAA) in the pre-mRNA.
    b. At a point about 10 to 35 nucleotides past this sequence, the pre-mRNA is cut from the enzyme.
    c. The polymerase continues transcribing for hundreds of nucleotides.
    d. Transcription is terminated when the polymerase eventually falls off the DNA.

IV. Enzymes in the eukaryotic nucleus modify pre-mRNA after transcription before it is sent to the cytoplasm.

A. During RNA processing, both ends of the primary transcript are usually altered.
B. Certain interior parts of the molecule are cut out and the remaining parts spliced together.
    1. At the 5’ end of the pre-mRNA molecule, a modified form of guanine is added, the **5’ cap.**
    2. At the 3’ end, an enzyme adds 50 to 250 adenine nucleotides, the **poly-A tail.**
    3. These modifications share several important functions.
        a. They seem to facilitate the export of mRNA from the nucleus.
        b. They help protect mRNA from hydrolytic enzymes.
        c. They help the ribosomes attach to the 5’ end of the mRNA.
C. The most remarkable stage of RNA processing occurs during the removal of a large portion of the RNA molecule in a cut-and-paste job of **RNA splicing.**
    1. Most eukaryotic genes and their RNA transcripts have long noncoding stretches of nucleotides.
    2. Noncoding segments of nucleotides called intervening regions, or **introns,** lie between coding regions.
    3. The final mRNA transcript includes coding regions, **exons,** which are translated into amino acid sequences, plus the leader and trailer sequences.
    4. RNA splicing removes introns and joins exons to create an mRNA molecule with a continuous coding sequence.
    5. This splicing is accomplished by a **spliceosome.**
        a. Spliceosomes consist of a variety of proteins and several *small nuclear ribonucleoproteins (snRNPs)* that recognize the splice sites.
        b. snRNPs are located in the cell nucleus and are composed of RNA and protein molecules.
        c. Each snRNP has several protein molecules and a *small nuclear RNA molecule (snRNA)* about 150 nucleotides long.
    6. The spliceosome interacts with certain sites along an intron, releasing the introns and joining together the two exons that flanked the introns.
    7. Introns and RNA splicing appear to have several functions.
        a. Some introns play a regulatory role in the cell. These introns contain sequences that control gene activity in some way.
        b. Splicing itself may regulate the passage of mRNA from the nucleus to the cytoplasm.
        c. One clear benefit of split genes is to enable one gene to encode for more than one polypeptide.
            (1) Alternative RNA splicing gives rise to two or more different polypeptides, depending on which segments are treated as exons.
            (2) This phenomenon may be common in humans, and may explain why we have a relatively small number of genes.
        d. The presence of introns in a gene may facilitate the evolution of new and potentially useful proteins as a result of a process known as *exon shuffling.* In many cases, different exons code for different domains of a protein.
        e. The presence of introns increases the probability of potentially beneficial crossing over between genes. Introns also increase the opportunity for recombination between two alleles of a gene. This raises the probability that a crossover will switch one version of an exon for another version found on the homologous chromosome.

V. Translation is the RNA-directed synthesis of a polypeptide.
    A. The interpreter in this process is **transfer RNA (tRNA),** which transfers amino acids from the cytoplasmic pool to a ribosome. The tRNA molecule is a translator, because it can read

a nucleic acid word (the mRNA codon) and translate it to a protein word (the amino acid).

B. Each tRNA arriving at the ribosome carries a specific amino acid at one end and has a specific nucleotide triplet, an **anticodon,** at the other.

C. The anticodon base-pairs with a complementary codon on mRNA.

D. Codon by codon, tRNAs deposit amino acids in the prescribed order, and the ribosome joins them into a polypeptide chain.

E. Like other types of RNA, tRNA molecules are transcribed from DNA templates in the nucleus.

F. Once it reaches the cytoplasm, each tRNA is used repeatedly, picking up its designated amino acid in the cytosol, depositing the amino acid at the ribosome, and returning to the cytosol to pick up another copy of that amino acid.

G. The anticodons of some tRNAs recognize more than one codon.
   1. This is possible because the rules for base pairing between the third base of the codon and anticodon are relaxed (called **wobble**).
   2. At the wobble position, U on the anticodon can bind with A or G in the third position of a codon.
   3. Wobble explains why the synonymous codons for a given amino acid can differ in their third base, but not usually in their other bases.

H. Each amino acid is joined to the correct tRNA by **aminoacyl-tRNA synthetase.** The 20 different synthetases match the 20 different amino acids. Each has active sites for only a specific tRNA-and-amino-acid combination. The synthetase catalyzes a covalent bond between them in a process driven by ATP hydrolysis.

I. Each ribosome is made up of a large and a small subunit. The subunits are composed of proteins and **ribosomal RNA (rRNA),** the most abundant RNA in the cell.
   1. After rRNA genes are transcribed to rRNA in the nucleus, the rRNA and proteins are assembled to form the subunits with proteins from the cytoplasm.
   2. The subunits exit the nucleus via nuclear pores.
   3. The large and small subunits join to form a functional ribosome only when they attach to an mRNA molecule.

J. Each ribosome has a binding site for mRNA and three binding sites for tRNA molecules.
   1. The **P site** holds the tRNA carrying the growing polypeptide chain.
   2. The **A site** carries the tRNA with the next amino acid to be added to the chain.
   3. Discharged tRNAs leave the ribosome at the **E** (exit) **site.**

K. The ribosome holds the tRNA and mRNA in close proximity and positions the new amino acid for addition to the carboxyl end of the growing polypeptide. It then catalyzes the formation of the peptide bond.

VI. Translation can be divided into three stages: initiation, elongation, and termination.

A. **Initiation** brings together mRNA, a tRNA with the first amino acid, and the two ribosomal subunits.
   1. First, a small ribosomal subunit binds with mRNA and a special initiator tRNA, which carries methionine.
   2. The small subunit then moves downstream along the mRNA until it reaches the start codon, AUG, which signals the start of translation. This establishes the reading frame for the mRNA.
   3. The initiator tRNA, already associated with the complex, then hydrogen-bonds with the start codon.
   4. Proteins called *initiation factors* bring in the large subunit so that the initiator tRNA occupies the P site.

B. **Elongation** involves the participation of several protein elongation factors, and consists of a series of three-step cycles as each amino acid is added to the proceeding one.

1. During **codon recognition,** an *elongation factor* assists hydrogen bonding between the mRNA codon under the A site with the corresponding anticodon of tRNA carrying the appropriate amino acid.
2. During **peptide bond formation,** an rRNA molecule catalyzes the formation of a peptide bond between the polypeptide in the P site with the new amino acid in the A site.
3. This step separates the tRNA at the P site from the growing polypeptide chain and transfers the chain, now one amino acid longer, to the tRNA at the A site.
4. During **translocation,** the ribosome moves the tRNA with the attached polypeptide from the A site to the P site.
5. Because the anticodon remains bonded to the mRNA codon, the mRNA moves along with it. The next codon is now available at the A site.
6. The tRNA that had been in the P site is moved to the E site and then leaves the ribosome.
7. Effectively, translocation ensures that the mRNA is "read" 5' to 3' codon by codon.
8. The three steps of elongation continue to add amino acids codon by codon until the polypeptide chain is completed.

C. **Termination** occurs when one of the three stop codons reaches the A site.
1. A *release factor* binds to the stop codon and hydrolyzes the bond between the polypeptide and its tRNA in the P site.
2. This frees the polypeptide, and the translation complex disassembles.

D. Typically a single mRNA is used to make many copies of a polypeptide simultaneously.
1. Multiple ribosomes, **polyribosomes,** may trail along the same mRNA.
2. Polyribosomes can be found in prokaryotic and eukaryotic cells.
3. A ribosome requires less than a minute to translate an average-sized mRNA into a polypeptide.
4. During and after synthesis, a polypeptide coils and folds to its three-dimensional shape spontaneously.
5. Chaperone proteins may aid correct folding.

E. In addition, proteins may require *posttranslational modifications* before doing their particular job.
1. This may require additions such as sugars, lipids, or phosphate groups to amino acids.
2. Enzymes may remove some amino acids or cleave whole polypeptide chains.
3. Two or more polypeptides may join to form a protein.

VII. Signal peptides target some eukaryotic polypeptides to specific destinations in the cell.

A. Translation in all ribosomes begins in the cytosol, but a polypeptide destined for the endomembrane system or for export has a specific **signal peptide** region at or near the leading end. This consists of a sequence of about 20 amino acids.

B. A **signal recognition particle (SRP)** binds to the signal peptide and attaches it and its ribosome to a receptor protein in the ER membrane. The SRP consists of a protein-RNA complex.

C. After binding, the SRP leaves and protein synthesis resumes with the growing polypeptide snaking across the membrane into the cisternal space via a protein pore. An enzyme usually cleaves the signal polypeptide.

D. Other kinds of signal peptides are used to target polypeptides to mitochondria, chloroplasts, the nucleus, and other organelles that are not part of the endomembrane system.

VIII. **Mutations** are changes in the genetic material of a cell (or virus).

A. These include large-scale mutations in which long segments of DNA are affected (for example, translocations, duplications, and inversions).

B. A chemical change in just one base pair of a gene causes a **point mutation.**

C. If these occur in gametes or cells producing gametes, they may be transmitted to future generations.

D. A point mutation that results in the replacement of a pair of complementary nucleotides with another nucleotide pair is called a **base-pair substitution.** Some base-pair substitutions have little or no impact on protein function.

1. In **silent mutations**, altered nucleotides still code for the same amino acids because of redundancy in the genetic code.
    a. Other changes lead to switches from one amino acid to another with similar properties.
    b. Still other mutations may occur in a region where the exact amino acid sequence is not essential for function.
2. Base-pair substitutions that cause a readily detectable change in a protein are usually detrimental but can occasionally lead to an improved protein or one with novel capabilities.
3. Changes in amino acids at crucial sites, especially active sites, are likely to affect function.
4. **Missense** mutations are those that still code for an amino acid but a different one.
5. **Nonsense** mutations change an amino acid codon into a stop codon, nearly always leading to a nonfunctional protein.

E. **Insertions** and **deletions** are additions or losses of nucleotide pairs in a gene. These have a disastrous effect on the resulting protein more often than substitutions do.

1. Unless insertion or deletion mutations occur in multiples of three, they cause a **frameshift mutation.**
2. All the nucleotides downstream of the deletion or insertion will be improperly grouped into codons.
3. The result will be extensive missense, ending sooner or later in nonsense—premature termination.

F. Mutations can occur in a number of ways.

1. Errors can occur during DNA replication, DNA repair, or DNA recombination.
    a. These can lead to base-pair substitutions, insertions, or deletions, as well as mutations affecting longer stretches of DNA.
    b. These are called *spontaneous mutations.*
    c. Rough estimates suggest that about 1 nucleotide in every 10^{10} is altered and inherited by daughter cells.
2. **Mutagens** are chemical or physical agents that interact with DNA to cause mutations.
    a. Physical agents include high-energy radiation like X-rays and ultraviolet light.
    b. Chemical mutagens fall into several categories.
        (1) Some chemicals are base analogues that may be substituted into DNA, but they pair incorrectly during DNA replication.
        (2) Other mutagens interfere with DNA replication by inserting into DNA and distorting the double helix.
        (3) Still others cause chemical changes in bases that change their pairing properties.