CHAPTER 17
FROM GENE TO PROTEIN

OUTLINE

I. The Connection between Genes and Proteins
   A. The study of metabolic defects provided evidence that genes specify proteins: *science as a process*
   B. Transcription and translation are the two main processes linking gene to protein: *an overview*
   C. In the genetic code, nucleotide triplets specify amino acids
   D. The genetic code must have evolved very early in the history of life

II. The Synthesis and Processing of RNA
   A. Transcription is the DNA-directed synthesis of RNA: *a closer look*
   B. Eukaryotic cells modify RNA after transcription

III. The Synthesis of Protein
   A. Translation is the RNA-directed synthesis of a polypeptide: *a closer look*
   B. Signal peptides target some eukaryotic polypeptides to specific locations in the cell
   C. RNA plays multiple roles in the cell: a review
   D. Comparing protein synthesis in prokaryotes and eukaryotes: *a review*
   E. Point mutations can affect protein structure and function
   F. What is a gene? *revisiting the question*

OBJECTIVES

After reading this chapter and attending lecture, the student should be able to:

1. Give early experimental evidence that implicated proteins as the links between genotype and phenotype.
2. Describe Beadle and Tatum's experiments with *Neurospora*, and explain the contribution they made to our understanding of how genes control metabolism.
3. Distinguish between "one gene—one enzyme" hypothesis and "one gene—one polypeptide," and explain why the original hypothesis was changed.
4. Explain how RNA differs from DNA.
5. In their own words, briefly explain how information flows from gene to protein.
6. Distinguish between transcription and translation.
7. Describe where transcription and translation occur in prokaryotes and in eukaryotes; explain why it is significant that in eukaryotes, transcription and translation are separated in space and time.
8. Define codon, and explain what relationship exists between the linear sequence of codons on mRNA and the linear sequence of amino acids in a polypeptide.
9. List the three stop codons and the one start codon.
10. Explain in what way the genetic code is redundant and unambiguous.

11. Explain the evolutionary significance of a nearly universal genetic code.
12. Explain the process of transcription including the three major steps of initiation, elongation, and termination.
13. Describe the general role of RNA polymerase in transcription.
14. Explain how RNA polymerase recognizes where transcription should begin.
15. Specifically, describe the primary functions of RNA polymerase II.
16. Distinguish among mRNA, tRNA, and rRNA.
17. Describe the structure of tRNA and explain how the structure is related to function.
18. Given a sequence of bases in DNA, predict the corresponding codons transcribed on mRNA and the corresponding anticodons of tRNA.
19. Describe the wobble effect.
20. Explain how an aminoacyl-tRNA synthetase matches a specific amino acid to its appropriate tRNA; describe the energy source that drives this endergonic process.
21. Describe the structure of a ribosome, and explain how this structure relates to function.
22. Describe the process of translation including initiation, elongation, and termination and explain what enzymes, protein factors, and energy sources are needed for each stage.
23. Explain what determines the primary structure of a protein and describe how a polypeptide must be modified before it becomes fully functional.
24. Describe what determines whether a ribosome will be free in the cytosol or attached to rough ER.
25. Explain how proteins can be targeted for specific sites within the cell.
26. Describe the difference between prokaryotic and eukaryotic mRNA.
27. Explain how eukaryotic mRNA is processed before it leaves the nucleus.
28. Describe some biological functions of introns and gene splicing.
29. Explain why base-pair insertions or deletions usually have a greater effect than base-pair substitutions.
30. Describe how mutagenesis can occur.

KEY TERMS

auxotroph
one gene–one polypeptide
transcription
messenger RNA (mRNA)
translation
RNA processing
primary transcript
triplet code
template strand
codon
reading frame
RNA polymerase
transcription unit
transcription factors
transcription initiation complex
TATA box
terminator
5' cap
poly (A) tail
RNA splicing
intron
exon
spliceosome
domain
transfer RNA (tRNA)
anticodon
wobble
aminoacyl-tRNA synthetases
ribosomal RNA (rRNA)
P site
A site
E site
polyribosome
signal peptide
signal-recognition particle (SRP)
mutation
point mutation
point mutation
base-pair substitution
missense mutation
nonsense mutation
insertion
deletion
frameshift mutation
mutagens
Ames test

LECTURE NOTES

Inherited instructions in DNA direct protein synthesis. Thus, proteins are the links between genotype and phenotype, since proteins are directly involved in the expression of specific phenotypic traits.

I. The Connection between Genes and Proteins

A. The study of metabolic defects provided evidence that genes specify proteins: *science as a process*

Archibald Garrod was the first to propose the relationship between genes and proteins (1909).

- He suggested that genes dictate phenotypes through enzymes that catalyze reactions.
- As a physician, Garrod was familiar with inherited diseases which he called "inborn errors in metabolism." He hypothesized that such diseases reflect the patient's inability to make particular enzymes.
- One example he studied was *alkaptonuria*, which causes the afflicted person's urine to turn black.
  - People with alkaptonuria accumulate alkapton in their urine, causing it to darken on contact with air.
  - Garrod reasoned that alkaptonurics, unlike normal individuals, lack the enzyme that breaks down alkapton.

1. How genes control metabolism

Garrod's hypothesis was confirmed several decades later by research which determined that specific genes direct production of specific enzymes.

- Biochemists found that cells synthesize and degrade organic compounds via metabolic pathways, with each sequential step catalyzed by a specific enzyme.
- Geneticists George Beadle and Boris Ephrussi (1930s) studied eye color in *Drosophila*. They speculated that mutations affecting eye color block pigment synthesis by preventing enzyme production at certain steps in the pigment synthesis pathway.

George Beadle and Edward Tatum were later able to demonstrate the relationship between genes and enzymes by studying mutants of a bread mold, *Neurospora crassa* (see Campbell, Figure 17.1).

- Wild-type *Neurospora* in laboratory colonies can survive on *minimal medium*. All other molecules needed by the mold are produced by its own metabolic pathways from this minimal nutrient source.
- Beadle and Tatum searched for mutants or *auxotrophs* that could not survive on minimal medium because they lacked the ability to synthesize essential molecules.
- Mutants were identified by transferring fragments of growing fungi (in complete medium) to vials containing minimal medium. Fragments that didn't grow were identified as auxotrophic mutants.

Auxotroph = (Auxo = to augment; troph = nourishment); nutritional mutants that can only be grown on *minimal medium*augmented with nutrients not required by the wild type

Minimal medium= Support medium that is mixed only with molecules required for the growth of wild-type organisms

- Minimal medium for *Neurospora* contains inorganic salts, sucrose, and the vitamin biotin.
- Nutritional mutants cannot survive only on minimal medium.

Complete growth medium= Minimal medium supplemented with all 20 amino acids and some other nutrients

- Nutritional mutants can grow on complete growth medium, since all essential nutrients are provided.

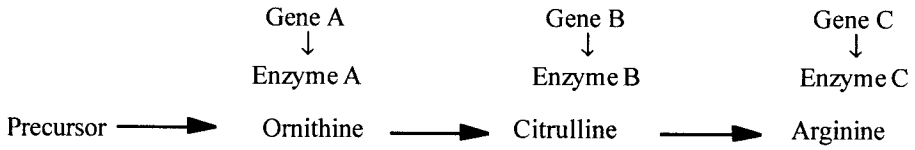
Beadle and Tatum then identified specific metabolic defects (from mutations) by transferring fragments of auxotrophic mutants growing on complete growth medium to vials containing minimal medium each supplemented with only *one* additional nutrient.

- Vials where growth occurred indicated the metabolic defect, since the single supplement provided the necessary component.
- For example, if a mutant grew on minimal medium supplemented with only arginine, it could be concluded that the mutant was defective in the arginine synthesis pathway.

Experiment:

Beadle and Tatum experimented further to more specifically describe the defect in the multistep pathway that synthesizes the amino acid arginine.

- Arginine synthesis requires three steps each catalyzed by a specific enzyme:



- They distinguished between three classes of arginine auxotrophs by adding either arginine, citrulline, or ornithine to the medium and seeing if growth occurred.

Results:

Some mutants required arginine, some either arginine or citrulline, and others could grow when any of the three were added.

	Minimal Medium (MM)	MM plus Ornithine	Mm plus Citrulline	MM plus Arginine
Wild Type	+	+	+	+
Class I Mutants	–	+	+	+
Class II Mutants	–	–	+	+
Class III Mutants	–	–	–	+

+ = growth, – = no growth

Conclusions:

Beadle and Tatum deduced from their data that the three classes of mutants each lacked a different enzyme and were thus blocked at different steps in the arginine synthesis pathway.

- Class I mutants lacked enzyme A; Class II mutants lacked enzyme B; and Class III mutants lacked enzyme C.

- Assuming that each mutant was defective in a single gene, they formulated the *one gene-one enzyme* hypothesis, which states that the function of a gene is to dictate the production of a specific enzyme.

2. One gene—one polypeptide

Beadle and Tatum's one gene-one enzyme hypothesis has been slightly modified:

- While most enzymes are proteins, many proteins are not enzymes. Proteins that are not enzymes are still, nevertheless, gene products.
- Also, many proteins are comprised of two or more polypeptide chains, each chain specified by a different gene (e.g., globulin chains of hemoglobin).

As a result of this new information, Beadle and Tatum's hypothesis has been restated as *one gene-one polypeptide*.

> As we will see later, even this notion is no longer tenable given that a) differential processing of a single RNA transcript can lead to the synthesis of numerous different proteins, and b) not all RNA is translated into protein.

B. Transcription and translation are the two main processes linking gene to protein: *an overview*

Ribonucleic acid (RNA) links DNA's genetic instructions for making proteins to the process of protein synthesis. It copies or transcribes the message from DNA and then translates that message into a protein.

- RNA, like DNA, is a nucleic acid or polymer of nucleotides.
- RNA structure differs from DNA in the following ways:
  - The five-carbon sugar in RNA nucleotides is *ribose* rather than deoxyribose.
  - The nitrogenous base *uracil* is found in place of thymine.

The linear sequence of nucleotides in DNA ultimately determines the linear sequence of amino acids in a protein.

- Nucleic acids are made of four types of nucleotides which differ in their nitrogenous bases. Hundreds or thousands of nucleotides long, each gene has a specific linear sequence of the four possible bases.
- Proteins are made of 20 types of amino acids linked in a particular linear sequence (the protein's primary structure).
- Information flows from gene to protein through two major processes, *transcription* and *translation* (see Campbell, Figure 17.2).

Transcription = The synthesis of RNA using DNA as a template

- A gene's unique nucleotide sequence is transcribed from DNA to a complementary nucleotide sequence in messenger RNA (mRNA).
- The resulting mRNA carries this transcript of protein-building instructions to the cell's protein-synthesizing machinery.

Translation = Synthesis of a polypeptide, which occurs under the direction of messenger RNA (mRNA)

- During this process, the linear sequence of bases in mRNA is translated into the linear sequence of amino acids in a polypeptide.
- Translation occurs on *ribosomes*, complex particles composed of ribosomal RNA (rRNA) and protein that facilitate the orderly linking of amino acids into polypeptide chains.

Prokaryotes and eukaryotes differ in how protein synthesis is organized within their cells.

- Prokaryotes lack nuclei, so DNA is not segregated from ribosomes or the protein-synthesizing machinery. Thus, transcription and translation occur in rapid succession.

- Eukaryotes have nuclear envelopes that segregate transcription in the nucleus from translation in the cytoplasm; mRNA, the intermediary, is modified before it moves from the nucleus to the cytoplasm where translation occurs. This *RNA processing* occurs only in eukaryotes.

C. In the genetic code, nucleotide triplets specify amino acids

There is not a one-to-one correspondence between the nitrogenous bases and the amino acids they specify, since there are only 4 nucleotides and 20 amino acids.

- A two-to-one correspondence of bases to amino acids would only specify 16 (4^2) of the 20 amino acids.
- A three-to-one correspondence of bases to amino acids would specify 64 (4^3) amino acids.

Researchers have verified that the flow of information from a gene to a protein is based on a triplet code (see Campbell, Figure 17.3).

- Triplets of nucleotides are the smallest units of uniform length to allow translation into all 20 amino acids with plenty to spare.
- These three-nucleotide "words" are called *codons*.

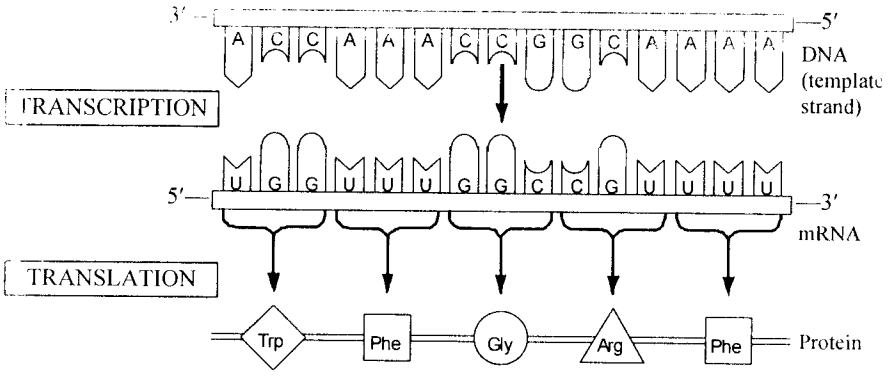
Codon = A three-nucleotide sequence in mRNA that specifies which amino acid will be added to a growing polypeptide or that signals termination; the basic unit of the genetic code

Genes are not directly translated into amino acids, but are first transcribed as codons into mRNA.

- For each gene, only one of the two DNA strands (the *template strand*) is transcribed.
- The complementary nontemplate strand is the parental strand for making a new template when DNA replicates.
- The same DNA strand can be the template strand for some genes and the nontemplate strand for others.

An mRNA is complementary to the DNA template from which it is transcribed.

- For example, if the triplet nucleotide sequence on the template DNA strand is CCG; GGC, the codon for glycine, will be the complementary mRNA transcript.
- Recall that according to the base-pairing rules, uracil (U) in RNA is used in place of thymine (T); uracil thus base-pairs with adenine (A).



During translation, the linear sequence of codons along mRNA is translated into the linear sequence of amino acids in a polypeptide.

- Each mRNA codon specifies which one of 20 amino acids will be incorporated into the corresponding position in a polypeptide.
- Because codons are base triplets, the number of nucleotides making up a genetic message is three times the number of amino acids making up the polypeptide product.

1. Cracking the genetic code

The first codon was deciphered in 1961 by Marshall Nirenberg of the National Institutes of Health.

- He synthesized an mRNA by linking only uracil-bearing RNA nucleotides, resulting in UUU codons.
- Nirenberg added this "poly U" to a test-tube mixture containing the components necessary for protein synthesis. The artificial mRNA (poly U) was translated into a polypeptide containing a string of only one amino acid, phenylalanine.
- Nirenberg concluded that the mRNA codon UUU specifies the amino acid phenylalanine.
- These same techniques were used to determine amino acids specified by the codons AAA, GGG, and CCC.

More elaborate techniques allowed investigators to determine all 64 codons by the mid-1960s (see Campbell, Figure 17.4).

- 61 of the 64 triplets code for amino acids.
- The triplet AUG has a *dual* function—it is the start signal for translation and codes for methionine.
- Three codons do not code for amino acids, but signal termination (UAA, UAG, and UGA).

There is redundancy in the genetic code, but no ambiguity.

- *Redundancy* exists since two or more codons differing only in their third base can code for the same amino acid (UUU and UUC both code for phenylalanine).
- *Ambiguity* is absent, since codons code for only one amino acid.

The correct ordering and grouping of nucleotides is important in the molecular language of cells. This ordering is called the *reading frame.*

Reading frame = The correct grouping of adjacent nucleotide triplets into codons that are in the correct sequence on mRNA.

- For example, the sequence of amino acids Trp–Phe–Gly–Arg–Phe can be assembled in the correct order only if the mRNA codons UGGUUUGGCCGUUUU are read in the correct sequence and groups.
- The cell reads the message in the correct frame as a series of *nonoverlapping* three-letter words: UGG–UUU–GGC–CGU–UUU.

D. The genetic code must have evolved very early in the history of life

The genetic code is shared nearly universally among living organisms.

- For example, the RNA codon CCG is translated into proline in all organisms whose genetic codes have been examined.
- The technology exists to transfer genes from one species to another. For example, the human gene for insulin can be inserted into bacteria where it is successfully expressed. Campbell, Figure 17.5 shows the incorporation of a firefly gene into a tobacco plant.

There are some exceptions to this universality:

- Several ciliates (e.g., *Paramecium* and *Tetrahymena*) depart from standard code; codons UAA and UAG are not stop signals, but code for glutamine.
- Mitochondria and chloroplasts have their own DNA that codes for some proteins.
- Mitochondrial genetic codes vary even among organisms; for example, CUA codes for threonine in yeast mitochondria and leucine in mammalian mitochondria.

The fact that the genetic code is shared nearly universally by all organisms indicates that this code was established very early in life's history.

II. The Synthesis and Processing of RNA

A. Transcription is the DNA-directed synthesis of RNA: *a closer look*

Transcription of messenger RNA (mRNA) from template DNA is catalyzed by *RNA polymerases*, which:

- Separate the two DNA strands and link RNA nucleotides as they base-pair along the DNA template.
- Add nucleotides only to the 3′ end; thus, mRNA molecules grow in the 5′ to 3′ direction.

There are several types of RNA polymerase.

- Prokaryotes have only one type of RNA polymerase that synthesizes all types of RNA—mRNA, rRNA, and tRNA.
- Eukaryotes have three RNA polymerases that transcribe genes. *RNA polymerase II* is the polymerase that catalyzes mRNA synthesis; it transcribes genes that will be translated into proteins.

Specific DNA nucleotide sequences mark where transcription of a gene begins (*initiation*) and ends (*termination*). Initiation and termination sequences plus the nucleotides in between are called a transcription unit.

Transcription unit = Nucleotide sequence on the template strand of DNA that is transcribed into a single RNA molecule by RNA polymerase; it includes the initiation and termination sequences, as well as the nucleotides in between

- In eukaryotes, a transcription unit contains a single gene, so the resulting mRNA codes for synthesis of only one polypeptide.
- In prokaryotes, a transcription unit can contain several genes, so the resulting mRNA may code for different, but functionally related, proteins.

Transcription occurs in three stages: a) polymerase binding and initiation; b) elongation; and c) termination (see Campbell, Figure 17.6).

1. RNA polymerase binding and initiation of transcription

RNA polymerases bind to DNA at regions called *promoters*.

Promoter = Region of DNA that includes the site where RNA polymerase binds and where transcription begins (*initiation site*). In eukaryotes, the promoter is about 100 nucleotides long and consists of:

a. The initiation site, where transcription begins (including which DNA strand serves as template).

b. A few nucleotide sequences recognized by specific DNA-binding proteins (transcription factors) that help initiate transcription.

In eukaryotes, RNA polymerases cannot recognize the promoter without the help of *transcription factors* (see Campbell, Figure 17.7).

Transcription factors = DNA-binding proteins which bind to specific DNA nucleotide sequences at the promoter and help RNA polymerase recognize and bind to the promoter region, so transcription can begin.

- RNA polymerase II, the enzyme that synthesizes mRNA in eukaryotes, usually cannot recognize a promoter unless a specific transcription factor binds to a region on the promoter called a TATA box.

TATA box = A short nucleotide sequence at the promoter which is rich in thymine (T) and adenine (A) and located about 25 nucleotides upstream from the initiation site.

- RNA polymerase II recognizes the complex between the bound TATA transcription factor and the DNA binding site.
- Once RNA polymerase recognizes and attaches to the promoter region, it probably associates with other transcription factors before RNA synthesis begins.

When active RNA polymerase binds to a promoter, the enzyme separates the two DNA strands at the initiation site, and transcription begins.

2. Elongation of the RNA strand

Once transcription begins, RNA polymerase II moves along DNA and performs two primary functions:

a. It untwists and opens a short segment of DNA exposing about ten nucleotide bases; one of the exposed DNA strands is the template for base-pairing with RNA nucleotides.

b. It links incoming RNA nucleotides to the 3′ end of the elongating strand; thus, RNA grows one nucleotide at a time in the 5′ to 3′ direction.

During transcription, mRNA grows about 30 to 60 nucleotides per second. As the mRNA strand elongates:

- It peels away from its DNA template.
- The nontemplate strand of DNA re-forms a DNA-DNA double helix by pairing with the template strand.

Following in series, several molecules of RNA polymerase II can simultaneously transcribe the same gene.

- Cells can thus produce particular proteins in large amounts.
- The growing RNA strands hang free from each polymerase. The length of each strand varies and reflects how far the enzyme has traveled from the initiation site on template DNA.

3. Termination of transcription

Transcription proceeds until RNA polymerase transcribes a DNA sequence called a terminator. The transcribed terminator functions as the actual termination signal.

- Additional proteins may cooperate with RNA polymerase in termination.
- In eukaryotes, the most common terminator sequence is AAUAAA.

Prokaryotic mRNA is ready for translation as soon as it leaves the DNA template. Eukaryotic mRNA, however, must be processed before it leaves the nucleus and becomes functional.

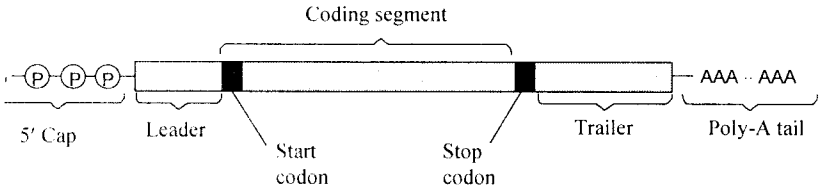
B. Eukaryotic cells modify RNA after transcription

RNA transcripts in eukaryotes are modified, or processed, before leaving the nucleus to yield functional mRNA. Eukaryotic RNA transcripts can be processed in two ways: a) covalent alteration of both the 3' and 5' ends and b) removal of intervening sequences.

Primary transcript = General term for initial RNA transcribed from DNA

Pre-mRNA = Primary transcript that will be processed to functional mRNA

1. Alteration of pre-mRNA ends



During pre-mRNA processing, both the 5′ and 3′ ends are covalently modified.

5′ cap = Modified guanine nucleotide (guanosine triphosphate) that is added to the 5′ end of mRNA shortly after transcription begins; has two important functions:

- Protects the growing mRNA from degradation by hydrolytic enzymes.
- Helps small ribosomal subunits recognize the attachment site on mRNA's 5′ end. A *leader* segment of mRNA may also be part of the ribosome recognition signal.

Leader sequence = Noncoding (untranslated) sequence of mRNA from the 5′ end to the start codon.

The 3′ end, which is transcribed last, is modified by enzymatic addition of a *poly-A tail*, before the mRNA exits the nucleus.

Poly (A) tail = Sequence of about 30 to 200 adenine nucleotides added to the 3′ end of mRNA before it exits the nucleus.

- May inhibit degradation of mRNA in the cytoplasm.
- May facilitate attachment to small ribosomal subunit
- May regulate protein synthesis by facilitating mRNA's export from the nucleus to the cytoplasm.
- Is not attached directly to the stop codon, but to an untranslated *trailer* segment of mRNA.

Trailer sequence = Noncoding (untranslated) sequence of mRNA from the stop codon to the poly (A) tail.

2. Split genes and RNA splicing

Genes that code for proteins in eukaryotes may not be continuous sequences.

- Coding sequences of a gene are interrupted by noncoding segments of DNA called intervening sequences, or introns.

Introns = Noncoding sequences in DNA that intervene between coding sequences (exons); are initially transcribed, but not translated, because they are excised from the transcript before mature RNA leaves the nucleus. Not all genes possess introns.

- Coding sequences of a gene are called *exons*, because they are eventually expressed (translated into protein).

Exons = Coding sequences of a gene that are transcribed and expressed

- In 1977, Richard Roberts and Philip Sharp independently found evidence for "split genes"; they received a Nobel Prize in 1993 for their discovery.

Introns and exons are both transcribed to form pre-mRNA, but the introns are subsequently removed and the remaining exons linked together during the process of RNA splicing.

RNA splicing = RNA processing that removes introns and joins exons from eukaryotic pre-mRNA; produces mature mRNA that will move into the cytoplasm from the nucleus (see Campbell, Figure 17.9).

- Enzymes excise introns and splice together exons to form an mRNA with a continuous coding sequence.

- RNA splicing also occurs during post-transcriptional processing of tRNA and rRNA.

Though there is much left to be discovered, some details of RNA splicing are now known.

- Each end of an intron has short boundary sequences that accurately signal the RNA splicing sites.
- Small nuclear ribonucleoproteins (snRNPs), play a key role in RNA splicing.

Small nuclear ribonucleoproteins (snRNPs) = Complexes of proteins and small nuclear RNAs that are found only in the nucleus; some participate in RNA splicing; snRNPs is pronounced "snurps".

- These small nuclear particles are composed of:
  1. *Small nuclear RNA (snRNA).* This small RNA molecule has less than 300 nucleotides—much shorter than mRNA.
  2. *Protein.* Each snRNP possesses several different proteins.
- involved in RNA splicing are part of a larger, more complex assembly called There are various types of snRNPs with different functions; those a spliceosome.

Spliceosome = A large molecular complex that catalyzes RNA splicing reactions; composed of small nuclear ribonucleoproteins (snRNPs) and other proteins (see Campbell, Figure 17.10)

- As the spliceosome is assembled, one type of snRNP base pairs with a complementary sequence at the 5′ end of the intron.
- The spliceosome precisely cuts the RNA transcript at specific splice sites at either end of the intron, which is excised as a lariat-shaped loop.
- The intron is released and the adjacent exons are immediately spliced together by the spliceosome.

3. Ribozymes

Other kinds of RNA primary transcripts, such as those giving rise to tRNA and rRNA, are spliced by mechanisms that do not involve spliceosomes; however, as with mRNA splicing, RNA is often involved in catalyzing the reactions.

Ribozymes = RNA molecules that can catalyze reactions by breaking and forming covalent bonds; called ribozymes to emphasize their catalytic activity.

- Ribozymes were first discovered in *Tetrahymena*, a ciliated protozoan that has self-splicing rRNA. That is, intron rRNA itself catalyzes splicing, which occurs completely without proteins or extra RNA molecules.
- Since RNA is acting as a catalyst, it can no longer be said that "All biological catalysts are proteins."
- It has since been discovered that rRNA also functions as a catalyst during translation.

4. The functional and evolutionary importance of introns

Introns may play a regulatory role in the cell.

- Intron DNA sequences may control gene activity.
- The splicing process itself may help regulate the export of mRNA to the cytoplasm.

Introns may allow a single gene to direct the synthesis of different proteins.

- This can occur if the same RNA transcript is processed differently among various cell types in the same organism.

- For example, all introns may be removed from a particular transcript in one case; but in another, one or more of the introns may be left in place to be translated. Thus, the resulting proteins in each case would be different.

Introns play an important role in the evolution of protein diversity; they increase the probability that recombination of exons will occur between alleles.

- In split genes, coding sequences can be separated by long distances, so they have higher recombination frequencies than continuously coded genes without introns.
- Exons of a "split gene" may code for different *domains* of a protein that have specific functions, such as, an enzyme's active site or a protein's binding site.

Protein domains = Continuous polypeptide sequences that are structural and functional units in proteins with a modular architecture

- Genetic recombination can occur in just one exon resulting in the synthesis of a novel protein with only one altered domain.

Introns also may increase the likelihood of genetic exchange between and among non-allelic genes.

III. The Synthesis of Protein

A. Translation is the RNA-directed synthesis of a polypeptide: *a closer look*

During translation, proteins are synthesized according to a genetic message of sequential codons along mRNA (see Campbell, Figure 17.11).

- *Transfer RNA* (*tRNA*) is the interpreter between the two forms of information—base sequence in mRNA and amino acid sequence in polypeptides.
- tRNA aligns the appropriate amino acids to form a new polypeptide. To perform this function, tRNA must:
  - Transfer amino acids from the cytoplasm's amino acid pool to a ribosome.
  - Recognize the correct codons in mRNA.

Molecules of tRNA are specific for only one particular amino acid. Each type of tRNA associates a distinct mRNA codon with one of the 20 amino acids used to make proteins.

- One end of a tRNA molecule attaches to a specific amino acid.
- The other end attaches to an mRNA codon by base-pairing with its anticodon.

Anticodon = A nucleotide triplet in tRNA that base pairs with a complementary nucleotide triplet (codon) in mRNA.

tRNAs decode the genetic message, codon by codon. For example:

- The mRNA codon UUU is translated as the amino acid phenylalanine (see Campbell, Figure 17.4)
- The tRNA that transfers phenylalanine to the ribosome has an anticodon of AAA.
- When the codon UUU is presented for translation, phenylalanine will be added to the growing polypeptide.
- As tRNAs deposit amino acids in the correct order, ribosomal enzymes link them into a chain.

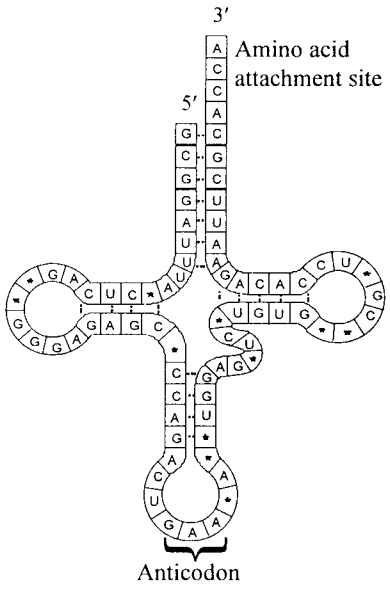
1. The structure and function of transfer RNA

All types of RNA, including tRNA, are transcribed from template DNA.

- In eukaryotes, tRNA, like mRNA, must travel from the nucleus to the cytoplasm, where translation occurs.
- In prokaryotes and eukaryotes, each tRNA molecule can be used repeatedly.

The ability of tRNA to carry specific amino acids and to recognize the correct codons depends upon its structure; its form fits function (see also Campbell, Figure 17.12).

- tRNA is a single-stranded RNA only about 80 nucleotides long.
- The strand is folded, forming several double-stranded regions where short base sequences of hydrogen bond with complementary base sequences.
- A single-plane view reveals a clover leaf shape.

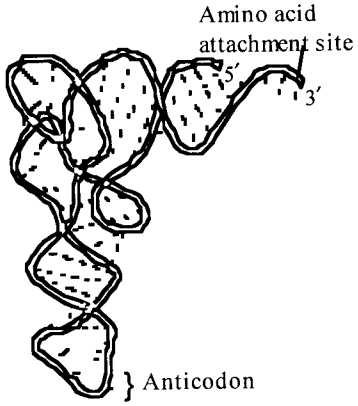


The three-dimensional structure is roughly L-shaped.

- A loop protrudes at one end of the L and has a specialized sequence of three bases called the *anticodon*.
- At the other end of the L protrudes the 3′ end of the tRNA molecule —the attachment site for an amino acid.

There are only about 45 distinct types of tRNA. However, this is enough to translate the 64 codons, since some tRNAs recognize two or three mRNA codons specifying the same amino acid.

- This is possible because the base-pairing rules are relaxed between the third base of an mRNA codon and the corresponding base of a tRNA anticodon.
- This exception to the base-pairing rule is called *wobble*.



Wobble = The ability of one tRNA to recognize two or three different mRNA codons; occurs when the third base (5′ end) of the tRNA anticodon has some play or wobble, so that it can hydrogen bond with more than one kind of base in the third position (3′ end) of the codon.

- For example, the base U in the wobble position of a tRNA anticodon can pair with either A or G in the third position of an mRNA codon.
- Some tRNAs contain a modified base called inosine (I), which is in the anticodon's wobble position and can base pair with U, C, or A in the third position of an mRNA codon.
- Thus, a single tRNA with the anticodon CCI will recognize three mRNA codons: GGU, GGC, or GGA, all of which code for glycine.

2. Aminoacyl-tRNA synthetases

The correct linkage between tRNA and its designated amino acid must occur before the anticodon pairs with its complementary mRNA codon. This process of correctly pairing a tRNA with its appropriate amino acid is catalyzed by an aminoacyl-tRNA synthetase.

Aminoacyl-tRNA synthetase = A type of enzyme that catalyzes the attachment of an amino acid to its tRNA

- Each of the 20 amino acids has a specific aminoacyl-tRNA synthetase.
- In an endergonic reaction driven by the hydrolysis of ATP, the proper synthetase attaches an amino acid to its tRNA in two steps (see Campbell, Figure 17.13):
  1. *Activation of the amino acid with AMP.* The synthetase's active site binds the amino acid and ATP; the ATP loses two phosphate groups and attaches to the amino acid as AMP (adenosine monophosphate).
  2. *Attachment of the amino acid to tRNA.* The appropriate tRNA covalently bonds to the amino acid, displacing AMP from the enzyme's active site.
- The aminoacyl-tRNA complex releases from the enzyme and transfers its amino acid to a growing polypeptide on the ribosome.

3. Ribosomes

Ribosomes coordinate the pairing of tRNA anticodons to mRNA codons.

- Ribosomes have two subunits (small and large) which are separated when not involved in protein synthesis (see Campbell, Figure 17.14a).
- Ribosomes are composed of about 60% *ribosomal RNA* (*rRNA*) and 40% protein.

The large and small subunits of eukaryotic ribosomes are:

- Constructed in the nucleolus
- Dispatched through nuclear pores to the cytoplasm
- Once in the cytoplasm, are assembled into functional ribosomes only when attached to an mRNA

Compared to eukaryotic ribosomes, prokaryotic ribosomes are smaller and have a different molecular composition.

- Selection of effective drug therapies against bacterial pathogens capitalizes on this difference.

- For example, the antibiotics tetracycline and streptomycin can be used to combat bacterial infections, because they inhibit bacterial protein synthesis without affecting the ribosomes of the eukaryotic host.

In addition to an mRNA binding site, each ribosome has three tRNA binding sites (P, A, and E) (see Campbell, Figure 17.14b).

- The *P site* holds the tRNA carrying the growing polypeptide chain.
- The *A site* holds the tRNA carrying the next amino acid to be added.
- Discharged tRNAs exit the ribosome from the *E site*.

As the ribosome holds the tRNA and mRNA molecules together, enzymes transfer the new amino acid from its tRNA to the carboxyl end of the growing polypeptide (see Campbell, Figure 17.14c).

4. Building a polypeptide

The building of a polypeptide, or translation, occurs in three stages: 1) initiation, 2) elongation, and 3) termination.

- All three stages require enzymes and other protein factors.
- Initiation and elongation also require energy provided by GTP (a molecule closely related to ATP).

a. Initiation

Initiation brings together mRNA, a tRNA attached to the first amino acid (initiator tRNA; the first amino acid is always methionine), and the two ribosomal subunits.

The first step of initiation involves the binding of the small ribosomal subunit to mRNA and initiator tRNA (see Campbell, Figure 17.15a).

- In prokaryotes, rRNA in the small subunit base-pairs with specific nucleotides in the leader sequence of the mRNA.
- In eukaryotes, the 5' cap of the mRNA aids in binding of the leader sequence to the small ribosomal subunit.
- With help from the small ribosomal subunit, methionine-bound initiator tRNA finds and base-pairs with the initiation or start codon on mRNA. This start codon, AUG, marks the place where translation will begin and is located just downstream from the leader sequence.
- Assembly of the initiation complex—small ribosomal subunit, initiator tRNA and mRNA—requires:
  - Protein *initiation factors* that are bound to the small ribosomal subunit
  - One GTP molecule that probably stabilizes the binding of initiation factors, and upon hydrolysis, drives the attachment of the large ribosomal subunit.

In the second step, a large ribosomal subunit binds to the small one to form a functional translation complex (see Campbell, Figure 17.15b).

- Initiation factors attached to the small ribosomal subunit are released, allowing the large subunit to bind with the small subunit.
- The initiator tRNA fits into the P site on the ribosome.
- The vacant A site is ready for the next aminoacyl-tRNA.

b. Elongation

Several proteins called *elongation factors* take part in this three-step cycle which adds amino acids one by one to the initial amino acid (see Campbell, Figure 17.16).

1. *Codon recognition.* The mRNA codon in the A site of the ribosome forms hydrogen bonds with the anticodon of an entering tRNA carrying the next amino acid in the chain.
   - An elongation factor directs tRNA into the A site.
   - Hydrolysis of GTP provides energy for this step.
2. *Peptide bond formation.* A peptide bond is formed between the polypeptide in the P site and the new amino acid in the A site by a *peptidyl transferase.*
   - The peptidyl transferase activity appears to be one of the rRNAs in the large ribosomal subunit (ribozyme).
   - The polypeptide separates from its tRNA and is transferred to the new amino acid carried by the tRNA in the A site.
3. *Translocation.* The tRNA in the A site, which is now attached to the growing peptide, is translocated to the P site. Simultaneously, the tRNA that was in the P site is translocated to the E site and from there it exits the ribosome.
   - During this process, the codon and anticodon remain bonded, so the mRNA and the tRNA move as a unit, bringing the next codon to be translated into the A site.
   - The mRNA is moved through the ribosome only in the 5′ to 3′ direction.
   - GTP hydrolysis provides energy for each translocation step.

Some students have trouble visualizing translocation, especially how the tRNA and mRNA move as a unit, exposing a new codon in the A site. Showing the class an animated sequence would no doubt solve the problem. However, if you do not have a monitor or video-projection capability, a paper simulation is just as effective. Paper models can be tacked to a large bulletin board or small cutouts can be moved on an overhead projector. Students may also actively participate by practicing their own simulations.

c. Termination

Each iteration of the elongation cycle takes less than a tenth of a second and is repeated until synthesis is complete and a termination codon reaches the ribosome's A site (see Campbell, Figure 17.7).

Termination codon (stop codon) = Base triplet (codon) on mRNA that signals the end of translation

- Stop codons are UAA, UAG, and UGA.
- Stop codons do not code for amino acids.

Students often confuse terminator sequence (on DNA), which signals the end of transcription, with termination or stop codons (on mRNA), which signal the end of translation.

When a stop codon reaches the ribosome's A site, a protein *release factor* binds to the codon and initiates the following sequence of events:

- Release factor hydrolyzes the bond between the polypeptide and the tRNA in the P site.
- The polypeptide and tRNA are released from the ribosome.
- The remainder of the translation complex dissociates, including separation of the small and a large ribosomal subunits.

5. Polyribosomes

Single ribosomes can make average-sized polypeptides in less than a minute; usually, however, clusters of ribosomes simultaneously translate an mRNA.

Polyribosome = A cluster of ribosomes simultaneously translating an mRNA molecule (see Campbell, Figure 17.8)

- Once a ribosome passes the initiation codon, a second ribosome can attach to the leader sequence of the mRNA.
- Several ribosomes may translate an mRNA at once, making many copies of a polypeptide.
- Polyribosomes are found in both prokaryotes and eukaryotes.

6. From polypeptide to functional protein

The biological activity of proteins depends upon a precise folding of the polypeptide chain into a native three-dimensional conformation.

- Genes determine *primary structure*, the linear sequence of amino acids.
- Primary structure determines how a polypeptide chain will spontaneously coil and fold to form a three-dimensional molecule with *secondary* and *tertiary structure*; chaperone proteins facilitate polypeptide folding

Some proteins must undergo *post-translational modification* before they become fully functional in the cell. Post-translational modification affects function by affecting protein structure.

- Chemical modification
  - Sugars, lipids, phosphate groups, or other additives may be attached to some amino acids.
- Chain-length modification
  - One or more amino acids may be enzymatically cleaved from the leading (amino) end of the polypeptide chain.
  - Single polypeptide chains may be divided into two or more pieces. The translated product of the insulin gene is a large protein precursor (preproinsulin). The precursor is modified by removal of N-terminal fragments and by internal enzymatic cleavage to yield to separate chains held together by disulfide bonds.
  - Two or more polypeptides may join as subunits of a protein that has quaternary structure (e.g., hemoglobin).

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

Eukaryotic ribosomes function either free in the cytosol or bound to endomembranes.

- Bound and free ribosomes are structurally identical and interchangeable.
- Most proteins made by free ribosomes will function in the cytosol.
- Attached to the outside of the endoplasmic reticulum, bound ribosomes generally make proteins that are destined for:
  - Membrane inclusion in membrane component of the endomembrane system (e.g., membrane-bound enzymes of the nuclear envelope, ER, Golgi, lysosomes, vacuoles, and plasma membrane)
  - Partitioning into the lumenal component of the endomembrane system (e.g., lysozymes)
  - Secretion from the cell (e.g., hormones such as insulin)

There is only one type of ribosome, and synthesis of all proteins begins in the cytosol. What determines whether a ribosome will be free in the cytosol or attached to rough ER?

- Messenger RNA for secretory proteins code for an initial *signal sequence* of 16 to 20 hydrophobic amino acids at the amino end of the newly forming polypeptide (see Campbell, Figure 17.19).
- When a ribosome begins to synthesize a protein with a signal sequence, it moves to the ER membrane by a mechanism that involves two other components.
  - *Signal recognition particle (SRP).* SRPs are a complex of protein and RNA (SRP RNA). They serve as an adaptor between the translation complex and the ER. SRPs first attach to the signal sequence of a growing polypeptide and link the translation complex to a receptor protein on the ER membrane (SRP receptor).
  - *SRP receptor.* This receptor protein is built into the ER membrane. The signal recognition particle docks with the receptor, and the ribosome thus becomes bound to the ER membrane.
- The ribosome continues protein synthesis and the leading end of the new polypeptide (N-terminus) threads into the cisternal space.
- The signal sequence is removed by an enzyme.
- Newly formed polypeptide is released from the ribosome and folds into its native conformation.
- If an mRNA does not code for a signal sequence, the ribosome remains free and synthesizes its protein in the cytosol.

Different signal sequences may also dispatch proteins to specific sites other than the ER. For example, newly formed proteins may be targeted for mitochondria or chloroplasts. In these cases, however, translation is completed in the cytoplasm.

C. RNA plays multiple roles in the cell: a review

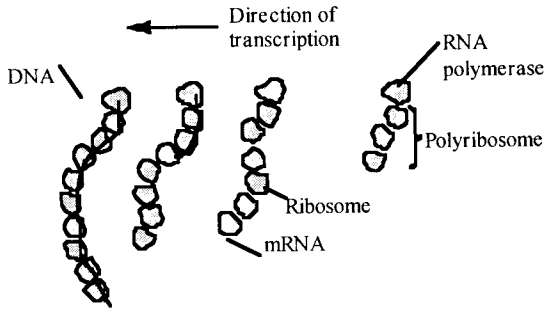
Three-dimensional conformations vary among the types of RNA. These differences in shape give RNA its ability to perform a variety of functions, such as:

1. *Information carrier.* Messenger RNA (mRNA) carries genetic information from DNA to ribosomes; this genetic message specifies a protein's primary structure.
2. *Adaptor molecule.* Transfer RNA (tRNA) acts as an adaptor in protein synthesis by translating information from one form (mRNA nucleotide sequence) into another (protein amino acid sequence). SRP RNA helps direct translation complexes to the ER.
3. *Catalyst and structural molecule.* During translation, ribosomal RNA (rRNA) plays structural and probably enzymatic roles in ribosomes. Small nuclear RNA (snRNA) in snRNP particles also plays structural and enzymatic roles within spliceosomes that catalyze RNA splicing reactions.
4. *Viral genomes.* Some viruses use RNA as their genetic material.

D. Comparing protein synthesis in prokaryotes and eukaryotes: *a review*

While transcription and translation are similar in prokaryotes and eukaryotes, there are some notable differences in the cellular machinery and in some of the details of the processes. The following differ in prokaryotes and eukaryotes:

- RNA polymerases; those of eukaryotes depend on transcription factors
- Termination of transcription
- Ribosomes
- Location (see also Campbell, Figure 17.20)
  - Prokaryotes lack nuclei, so transcription is not segregated from translation; consequently, translation may begin as soon as the 5′ end of mRNA peels away from template DNA, even before transcription is complete.



  - The significance of a eukaryotic cell's compartmental organization is that transcription and translation are segregated by the nuclear envelope. This allows mRNA to be modified before it moves from the nucleus to the cytoplasm. Such *RNA processing* occurs only in eukaryotes.

E. Point mutations can affect protein structure and function

Knowing how genes are translated into proteins, scientists can give a molecular description of heritable changes that occur in organisms.

Mutation = A change in the genetic material of a cell (or virus)

Point mutation = A mutation limited to about one or a few base pairs in a single gene

1. Types of point mutations

There are two categories of point mutations: 1) base-pair substitutions and 2) base-pair insertions or deletions (see Campbell, Figure 17.22).

a. Substitutions

Base-pair substitution = The replacement of one base pair with another; occurs when a nucleotide and its partner in the complementary DNA strand are replaced with another pair of nucleotides according to base-pairing rules.

Depending on how base-pair substitutions are translated, they can result in little or no change in the protein encoded by the mutated gene.

- Redundancy in the genetic code is why some substitution mutations have no effect. A base pair change may simply transform one codon into another that codes for the same amino acid (*silent substitution*).
- Even if the substitution alters an amino acid, the new amino acid may have similar properties to the one it replaces, or it may be in a part of the protein where the exact amino acid sequence is not essential to its activity (*conservative substitution*).

Some base-pair substitutions result in readily detectable changes in proteins.

- Alteration of a single amino acid in a crucial area of a protein will significantly alter protein activity.
- On rare occasions, such a mutation will produce a protein that is improved or has capabilities that enhance success of the mutant organism and its descendants.

- More often, such mutations produce a less active or inactive protein that impairs cell function.

Base-pair substitutions are usually missense mutations or nonsense mutations.

Missense mutation = Base-pair substitution that alters an amino acid codon (sense codon) to a new codon that codes for a different amino acid

- Altered codons make sense (are translated), but not necessarily the right sense.
- Base-pair substitutions are usually missense mutations.

Nonsense mutation = Base-pair substitution that changes an amino acid codon (sense codon) to a chain termination codon, or vice versa

- Nonsense mutations can result in premature termination of translation and the production of a shorter than normal polypeptide.
- Nearly all nonsense mutations lead to nonfunctional proteins.

b. Insertions or deletions

Base-pair insertions or deletions usually have a greater negative effect on proteins than substitutions.

Base-pair insertion = The insertion of one or more nucleotide pairs into a gene

Base-pair deletion = The deletion of one or more nucleotide pairs from a gene

Because mRNA is read as a series of triplets during translation, insertion or deletion of nucleotides may alter the reading frame (triplet grouping) of the genetic message. This type of *frameshift mutation* will occur whenever the number of nucleotides inserted or deleted is not 3 or a multiple of 3.

Frameshift mutation = A base-pair insertion or deletion that causes a shift in the reading frame, so that codons beyond the mutation will be the wrong grouping of triplets and will specify the wrong amino acids

- A frameshift mutation causes the nucleotides following the insertion or deletion to be improperly grouped into codons.
- This results in extensive missense, which will sooner or later end in nonsense (premature termination).
- Frameshift will produce a nonfunctional protein unless the insertion or deletion is very near the end of the gene.

2. Mutagens

Mutagenesis = The creation of mutations

- Mutations can occur as errors in DNA replication, repair, or recombinations that result in base-pair substitutions, insertions, or deletions.
- Mutagenesis may be a naturally occurring event causing *spontaneous mutations* or mutations may be caused by exposure to mutagens.

Mutagen = Physical or chemical agents that interact with genetic material to cause mutations

- Radiation is the most common physical mutagen in nature and has been used in the laboratory to induce mutations.
- Several categories of chemical mutagens are known including *base analogues*, which are chemicals that mimic normal DNA bases, but base-pair incorrectly.
- The Ames test, developed by Bruce Ames, is one of the most widely used tests for measuring the mutagenic strength of various chemicals. Since most mutagens are carcinogenic, this test is also used to screen for chemical carcinogens.

F. What is a gene? ***revisiting the question***

The concept of the gene has emerged as the history of genetics has unfolded.

- The Mendelian concept of a gene was that it served as a discrete unit of inheritance that affected a phenotypic character.
- Morgan and colleagues assigned such units of inheritance (or genes) to specific loci on chromosomes.
- In molecular terms, a gene is a specific sequence of nucleotides at a given location in the genome of an organism. Depending on the gene, the final gene product may be RNA or a specific polypeptide.

REFERENCES

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts and J.D. Watson. *Molecular Biology of the Cell.* 3rd ed. New York: Garland, 1994.

Campbell, N., et al. *Biology.* 5th ed. Menlo Park, California: Benjamin/Cummings, 1998.

Tijan, R. "Molecular Machines that Control Genes." *Scientific American,* February, 1995.