Chapter 17- From Gene to Protein

(Key Concepts are Underlined)

The Connection Between Genes and Proteins

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

One gene-one enzyme hypothesis- in the 1930's, Beadle and Tatum studied nutritional mutants of bread molds and concluded the various mutations were abnormal variations of different genes- each gene dictating the production of one enzyme; each mutant had a different defective enzyme in the metabolic pathway of the amino acid arginine

One gene-one polypeptide hypothesis- takes into account not all proteins are enzymes, with many proteins being made up of more than one polypeptide coded by specific genes

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

- genes are typically hundreds or thousands of nucleotides long with a specific sequence of bases

Transcription- synthesis of RNA under the direction of DNA (which provides a template for assembling a sequence of RNA nucleotides)

Messenger RNA (mRNA)- a *transcript* of the gene's protein-building instructions which carries the genetic message to the protein-synthesizing machinery of the cell

Translation- synthesis of a polypeptide at ribosomes (the facilitator) under the direction of mRNA

- transcription and translation are coupled in prokaryotes, as DNA and ribosmes are not separated by the nucleus

- transcription of eukaryotic genes results in *pre-mRNA* (or **primary transcript**), where **RNA processing** yields the finished mRNA

<u>In the genetic code, nucleotide triplets specify amino acids</u>

Triplet code- the genetic instructions for a polypeptide are written in DNA as a series of three-nucleotide words- one base triplet codes for one amino acid

Template strand- strand of DNA transcribed- other is the *complimentary strand* (varies between different regions of the DNA molecule

Codons- mRNA base triplets (compliment to DNA triplet); each codon specifies one of the 20 amino acids; mRNA is a sequence of codons

- the number of nucleotides making up a genetic message is three times the number of amino acids making up the protein product
- in 1961, the first codon was deciphered (UUU = phenylalanine); all 64 codons were deciphered by the mid-1960's
- genetic messages begin with the mRNA "start" codon AUG which also codes for methionine, thus polypeptide chains begin with methionine (may be subsequently removed)
- translation will continue until a termination codon is reached
- there is redundancy in the genetic code, but no ambiguity (i.e. more than one codon may specify the same amino acid, but they will not code for any other amino acid); those codons that specify the same amino acid usually only differs by the third base

- in summary, “genetic information is encoded as a sequence of nonoverlapping base triplets, or codons, each of which is translated into a specific amino acid during protein synthesis”

<u>The genetic code must have evolved very early in the history of life</u>
- the genetic code is universal, shared by all organisms- “it is a reminder of kinship that bonds all life on Earth”

The Synthesis and Processing of RNA

<u>Transcription is the DNA-directed synthesis of RNA: *a closer look*</u>

RNA polymerase- pries the two strands of DNA and combines RNA nucleotides as they base-pair along the DNA template; three different types (RNA polymerase II synthesizes mRNA); can only add nucleotides to the 3´ end of the growing RNA strand (elongates in its 5´to 3´ direction)

Transcription unit- stretch of DNA transcribed

Three Stages of Transcription:

Initiation

Promoter- region of DNA where RNA polymerase attaches and initiates transcription; determines which DNA strand is used as the template

Transcription factors- in eukaryotes, they are a collection of proteins mediating the binding of RNA polymerase to the promoter

Transcription initiation complex- the assembly of transcription factors and RNA polymerase bound to the promoter; once assembled, the complex may begin transcribing the DNA template strand

TATA box- DNA "initiating" sequence which aids in the formation of the initiation complex

Elongation

- RNA polymerase unwinds the double helix, exposing about 10 – 20 bases at a time; RNA nucleotides are added to the 3′ end of the growing RNA transcript, while the DNA helix reforms behind; the rate of transcription is about 60 nucleotides per second in eukaryotes
- a single gene can be transcribed simultaneously by several RNA polymerases (effectively increases production of the protein/polypeptide)

Termination

Terminator- DNA sequence responsible for stopping transcription (actually, the transcribed RNA sequence gives the "stop" signal)
- transcription in prokaryotes stops at the end of the termination signal, whereas eukaryotes continues past the termination signal for about 10 – 35 nucleotides (mRNA ends in a AAUAAA sequence)

<u>Eukaryotic cells modify RNA after transcription</u>

5′ cap- modified guanine nucleotide attached to the nontranslated *leader* segment of the 5′ end of the pre-mRNA molecule (the end made first during transcription); protects the mRNA from degradation of hydrolytic enzymes and serves as a point of attachment of ribosomes in the cytoplasm

Poly(A) tail- consists of 30 to 200 adenine nucleotides added to the nontranslated *trailer* segment of the 3′ end of the pre-mRNA molecule; like the 5′ cap, it inhibits degradation of the RNA and aids in ribosome attachment;

in addition, it seems to facilitate the export of mRNA from the nucleus

RNA splicing- the removal of large noncoding regions of pre-mRNA called **introns** (or intervening sequences), leaving behind the coding regions called **exons**

Small nuclear ribonucleoproteins (or *snRNPs*)- recognize "splice sites" (i.e. short nucleotide sequences signaling RNA splicing) at the ends of introns; different snRNPs contain different *small nuclear RNAs* (*snRNAs*) of about 150 nucleotides in length and protein molecules

Spliceosome- a large assembly (nearly the size of a ribosome) of several snRNPs and additional proteins interact with the splice sites in the excision of introns and joining of exons

Ribozymes- RNA molecules that function as enzymes; intron RNA may even catalyze its splicing

- introns and RNA splicing may play a regulatory role in gene activity (i.e. export of mRNA into the cytoplasm); a number of genes are known to give rise to different proteins, depending on which segments are treated as introns and exons
- introns increases the opportunity of crossing over (increases the distance between exons- lengthens DNA), thus giving rise to new proteins

The Synthesis of Protein

<u>Translation is the RNA-directed synthesis of a polypeptide: *a closer look*</u>

Transfer RNA (tRNA)- transfers amino acids from the cytoplasm to a ribosome by linking a particular amino acid to a particular mRNA codon; at one end of the tRNA molecule there is a specific amino acid and the other end a nucleotide triplet called an **anticodon**, which is complimentary to mRNA codon

- the tRNA anticodon hydrogen bonds with the mRNA molecule at the ribosome where its amino acid will be added to the polypeptide chain; this process of "translation" will repeat itself, codon by codon
- tRNA is used repeatedly
- single RNA strand of about 80 nucleotides and like mRNA, they are transcribed from DNA templates
- certain regions of nucleotides hydrogen bond with complementary bases of other regions to yield a three dimensional L-shape after it folds back upon itself; one end of the "L" is the anticodon and the other is the 3′ end for the attachment of an amino acid
- some tRNAs have anticodons able to recognize two or more different codons due to the relaxation of the base-pairing rules in the THIRD base of (codon-anitcodon) referred to as **wobble** (e.g. U of a tRNA anticodon can base-pair with A or G in mRNA codon; I, or inosine, of tRNA in the wobble position can base pair with U, C, or A); explains why there are synonymous codons for a given amino acid and only about 45 tRNAs (as opposed to 61)

Aminoacyl-tRNA synthetase- enzymes specific to each amino acid, in joining an amino acid to the correct tRNA (about 20 of these enzymes in cells); the result is amino-acyl tRNA ("activated amino acid") fueled by the hydrolysis of ATP (i.e. phosphorylated)

- ribosomes facilitate the coupling of tRNA anticodons with mRNA codons during protein synthesis by holding/positioning the tRNA and mRNA together for the addition of a new amino acid to the carboxyl end of a

growing polypeptide; they are made up of large and small subunits constructed of protein and RNA molecules called **ribosomal RNA** (**rRNA**); rRNA is also DNA transcribed; ribosomal subunits are assembled in the nucleolus (about 60% RNA by weight, which makes rRNA the most abundant in type of RNA) and exported to the cytoplasm via nuclear pores; large and small subunits join to form a functional ribosome only when they are attached to mRNA
- ribosomes have three binding-sites for tRNA, the **P site** (peptidyl-tRNA site- holds the growing polypeptide), the **A site** (aminoacyl-tRNA site- holds the next amino acid), and **E site** (exit site)

Three Stages of Translation:

Initiation

- brings together mRNA, tRNA (aminoacyl-tRNA), and the two ribosomal subunits
- first, the small ribosomal subunit binds to both mRNA (the leader segment of the 5′end) and a special initiator tRNA (carrying the amino acid methionine)
- the start of translation is signaled by the initiation codon AUG located "downstream" on mRNA
- the union of the mRNA, initiator tRNA, and the small ribosomal subunit is followed by the attachment of the large subunit (a.k.a translation initiation complex) via *initiation factors* and GTP (guanosine triphosphate); the initiator tRNA sits in the P site awaiting for another aminocyl tRNA to occupy the A site (which begins the synthesis of the polypeptide)

Elongation (driven by elongation factors and GTP)

1) Codon recognition- an incoming aminoacyl tRNA hydrogen bonds to the codon in the A site (requires GTP)

2) Peptide bond formation- the rRNA of the large ribosomal subunit catalyzes the formation of the peptide bond between the carboxyl end of the polypeptide in the P site with the amino acid in the A site; polypeptide separates from the tRNA in the P site and remains attached to the tRNA in the A site
3) Translocation- the tRNA in the A site remains hydrogen-bonded to the codon and is translocated to the P site while mRNA moves through the ribosome (5′ end first); tRNA in the P site moves to the E site and is released from the ribosome

Termination

- elongation continues until a *release factor* protein attaches to the stop codon UAA, AUG, or UGA in the A site of the ribosome; the release factor adds water to the polypeptide chain instead of an amino acid, which hydrolyzes and frees the completed polypeptide from the tRNA in the P site

Polyribosomes- several ribosomes simultaneously translating a single mRNA molecule

- a polypeptide chain begins to coil and fold into its functional conformation during and after its synthesis; *posttranslational modifications* my be required before the protein may be fully functional (i.e. attachment of prosthetic groups, enzymatically cleaved, etc.)

<u>Signal peptides target some eukaryotic polypeptides to specific destinations</u>

- free and bound ribosomes are structurally identical (can switch their status from free to bound); bound ribosomes are attached to the cytosolic side the endoplasmic reticulum and are responsible for making proteins of the endomembrane system, as well as secretory proteins

Signal peptide- a sequence of about 20 amino acids at or near the leading end of the growing polypeptide is recognized by a protein-RNA complex called a **signal-recognition particle (SRP)**, which brings the ribosome to a receptor protein built into the ER; the growing polypeptide crosses the ER membrane into the cisternal space (if the protein is to be a secretory protein, it will dissolve into the cisternal solution; if it is to be a membrane protein, it remains partially imbedded in the ER membrane)

RNA plays multiple roles in the cell: *a review*

- RNA is versatile!

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

- RNA polymerases are different; eukaryotes require transcription factors
- transcription is terminated differently
- eukaryotic cell's nuclear envelope segregates transcription from translation and has extensive RNA processing
- eukaryotic cells have complicated protein targeting mechanisms

Point mutations can affect protein structure and function

Mutations- changes in genetic material of a cell

Point mutations- changes in one or few base pairs in a single gene; may cause a genetic disorder if located in gametes (e.g. sickle-cell disease)

Base-pair substitution- replacement of one nucleotide and its complement with another; usually disastrous if it occurs in the active site of the protein

> **Silent mutation-** has no effect on the encoded protein due to the redundancy of the genetic code (i.e. either same amino acid is inserted or the different amino acid does not effect the function of the protein)
>
> **Missense mutations-** still codes for an amino acid, whereas **nonsense mutations** change an amino acid codon to a stop codon (i.e. does not make sense)

Insertions and **deletions-** addition or loss of one or more pairs of nucleotides in a gene; usually have a disastrous effect on the resulting protein due to the altered reading frame

> **Frameshift mutation-** insertion or deletion of nucleotides NOT in a multiple of 3; creates much *missense* and *nonsense*

- spontaneous mutations may result during DNA replication, repair, or recombination

Mutagens- physical (e.g. X-ray and UV radiation) or chemical (e.g. base analogues causing base-pairing errors during DNA replication) agents which interact with DNA to cause mutations

What is a gene?: *revisiting the question*

- "A gene is a region of DNA whose final product is either a polypeptide or an RNA molecule," or serves a regulatory role.