Lab 7 Name _____

Transcription, translation and mutations

OBJECTIVES

--To review organic compounds and macromolecules
--To gain a better understanding of transcription and translation.
--To use the genetic code to convert DNA and RNA sequence into polypeptides.
--To explore the possible consequences of DNA mutations on polypeptide sequence and protein structure.

Cells contain 4 classes of organic macromolecules. One class of organic macromolecules is the nucleic acids. Both DNA and RNA are nucleic acids, which consist of long chains of nucleotides. One DNA or RNA nucleotide consists of a nitrogenous base, a sugar, and a phosphate group. Neighboring nucleotides are covalently linked, the sugar of one is connected to the phosphate of the next, forming a sugar-phosphate backbone, with the bases protruding from the sugars.

There are notable differences between DNA and RNA molecules. DNA molecules are double stranded, whereas RNA is typically single stranded. DNA nucleotides contain a deoxyribose sugar; RNA nucleotides contain a ribose sugar. The nitrogenous bases of DNA nucleotides are adenine (A), thymine (T), cytosine (C), and guanine (G). RNA has the nitrogenous base uracil (U) instead of thymine. The nitrogenous bases of a nucleotide participate in hydrogen bonding with one specific partner. Adenine can best form hydrogen bonds with thymine, and guanine with cytosine. In 'shorthand' we might say "A pairs with T", and "G pairs with C". Where RNA is involved, uracil will substitute for thymine, thus "A pairs with U". These hydrogen-bonded base pairings, or attractions between certain bases, is referred to as **complementary base pairing**.

Complementary base pairing dictates the bases that exist side-by-side at the center of the DNA double helix. They do not, however, dictate the sequence of the nucleotides *along* the length of a DNA strand. In fact, the sequence of bases can vary almost infinitely. These differing base sequences represent the **genetic information.** In this activity, we will explore how different base sequences 'code' for different RNA molecules and thus different polypeptides (short chains of amino acid monomers). Remember that polypeptides, or proteins, serve critically important roles in the body. The human body has between 50,000-100,000 proteins, each with a unique **structure and function**. Some of the proteins we have discussed so far include: enzymes, membrane receptors (like dopamine or serotonin receptors), hemoglobin (the protein in RBCs that carries O2 in the blood), insulin, and keratin (structural protein in hair). Recall that the primary sequence of a protein (the amino acid sequence along the chain) dictates the molecules secondary, tertiary and quarternary structure. In simpler terms, the sequence of amino acids in a proteins determines its 3D structure, which is imperative for its ability to function.

In this activity you will be given a DNA sequence (represented by a long strand of nucleotide bases). From this sequence you will **transcribe** the DNA sequence into the complementary RNA sequence. With the RNA sequence in hand you will then need to hunt for the start codon within the sequence. A **codon** is a set of three adjacent nucleotides (a triplet), and is the shortest sequence that can code for an amino acid or a start or stop codon. The start codon will dictate the 'reading frame' of the sequence. Use the genetic code (see back page) to then **translate** the RNA sequence into amino acid sequence.

A **mutation** is any change in the nucleotide sequence of DNA. Mutations typically arise when DNA is exposed to radiation or chemical agents. Mutations can involve **base substitutions**, where one (or more) base is changed to a different base. **Deletion and insertion mutations** occur when one or more bases are removed or added to the original DNA sequence. Deletion and insertion mutations often result in a change in the reading frame (or triplet grouping) of the genetic message. In this activity, you will explore the consequences of DNA mutation on protein sequences and structure.

Review Questions:

1. Cells contain organic compounds. The term **organic** means that the compound

contains _____.

2. Some of the organic compounds found in cells exist as **polymers** (repeating linked units). Identify **three** types of polymers found in cells and the composition of the "linked units", or **monomers**.

3. Cells contain two different kinds of nucleic acids, _____ and _____.
Nucleic acids are composed of **nucleotides** which are made up of what **three** parts? Sketch a nucleotide.

4. What is the composition of the **nucleotides** found in DNA? in RNA?

5. Who is considered to have discovered the structure of DNA?

The DNA molecule is a polymer comprised of repeating units called nucleotides. For convenience, DNA is said to be "read" (i.e. transcribed) from the **3'** (pronounced "three prime") end to the **5'** (five prime) end (the terminology is related to organic chemistry). Don't get confused, DNA is 'read' by RNA polymerase from **3' to 5',** but the new strand of RNA is formed by adding nucleotides in the **5' to 3'** direction.

A **gene** is the portion of the DNA molecule that codes for an actual **protein**, which are the workhorse molecules within a cell. The word "**base**" is sometimes used in place of the word "nucleotide". Thus, the DNA double helix is said to comprise **base-pairs**. You will need the genetic code (back page).

Suppose you have the following strand of DNA (oriented with 3' on the left and 5' on the right):

TTTTTTTACCCCCCTGGGGGAATTTTACTTAAAAAAAAATTTTATTTT

1. Write the **complementary** strand of DNA, which attaches to the segment above.

2. Write the strand of **RNA** generated from the **original** strand of DNA.

3. a. Find the **start codon** (a three RNA nucleotide sequence) and mark it. Once you have found the start codon, 'break up' the RNA strand into groups of three bases and then translate the groups.

b. Mark on the RNA sequence you have generated, the portion which is the **gene**. Hint: you will need to find the **stop** or **end** codon.

4. What is the **amino acid sequence** (protein) for which this RNA codes?

5. What **percentage** of the RNA (and therefore the DNA) *actually* codes for a protein? (Hint: don't count the start and stop signals)

A **mutation** is any change in the base (nucleotide) sequence of DNA, brought about by radiation or chemical interactions.

6. a. Suppose an X-ray strikes the original strand of DNA and alters **one** base (nucleotide). This type of mutation is called a **substitution**. What is the **probability** that the change will occur in the gene? Express the probability as a fraction or a decimal and don't count the start and stop signals as part of the gene. Hint: count the number of codons in the whole sequence, then figure out the fraction of codons that will actually be translated into amino acids.

b. Suppose the X-ray does strike somewhere in the gene and alters **one** base (nucleotide). What is the probability that the change will affect the resulting amino acid sequence? (Hint: the answer is not 100% — look at the redundancies in the genetic code).

c. What is the probability that a cosmic ray will cause an **actual** change in this protein? (Hint: this is not as hard as it sounds; simply multiply the fractions or decimals from the previous two questions)

7. Looking at the **DNA**, suppose the change (mutation) occurs in the second letter (position) of the first coding codon (i.e. CCC becomes CAC). What will the resulting amino acid sequence be?

8. Recall that glycine is a small molecule and that the other amino acids are all larger (back page). Glycine is used in proteins that need to be long and flexible, like keratin in hair and fingernails. Other amino acids tend to be used in proteins that are bulky, like the protein hemoglobin in blood. What will the **shape change** in the protein be, due to the mutation?

Perhaps more frightening are the **reading frame shift** mutations brought on by either the **insertion** or **deletion** of a base or bases.

9. a. Suppose, **in the original DNA sequence**, a **mutagen** (a chemical which induces mutations) removes the first A base from the left (just that base). In other words, the original strand of DNA begins with "TTTTTTTCCC...". Will the start sequence still be recognized?

b. In fact, write the new complementary RNA sequence (it should not be a lot different from what you wrote in question 6), and mark the new start codon.

10. What is the **amino acid sequence** for which this RNA codes?

11. Is this sequence significantly different from the sequence in question 4? More importantly, is the **shape** of the protein going to be significantly different?

By removing or adding one base to the DNA sequence, then, results in a reading frame shift, so whole genes can become inactive, and hitherto inactive genes can become activated. In fact, there are plenty of inactive genes in the human **genome** which await only a frame shift to become active again.

In sexual reproduction, a double strand of DNA from the mother and a double strand of DNA from the father unite to form the chromosomes that will generate a unique individual. However, the mitochondria of the first cell of the new individual all come from the mother. Furthermore, the mitochondria, having been an independent organism at one time, retains its own DNA and replicates using this DNA.

12. In principle, then, if there were no mutations, should your **mitochondrial DNA** (mDNA) differ at all from your great-great-great grandmother's (on your mother's mother's mother's side)?

13. Mutations, then, are about the only way to change mitochondrial DNA (yeah, there are other mechanisms, but they aren't that significant). This can give us a way to calculate the rate of **genetic evolution**, the change in genes over time. You are lucky enough to find a human hair follicle cell 50,000 years old with its mDNA intact (you get its age by carbon-14 dating). When you compare its mDNA sequence to your own mDNA, you find that the two mDNAs differ by 10 bases. What is the **mutation rate** (or the rate of genetic evolution)? Express this as the number of bases per year.

14. You find another human hair follicle cell, which cannot be dated by carbon-14 because it is too old and there is no detectable amount of carbon-14 left. However, its mDNA is intact and, upon analysis, differs from your mDNA by 100 bases. How many years old do you estimate the sample to be?

The Genetic Code

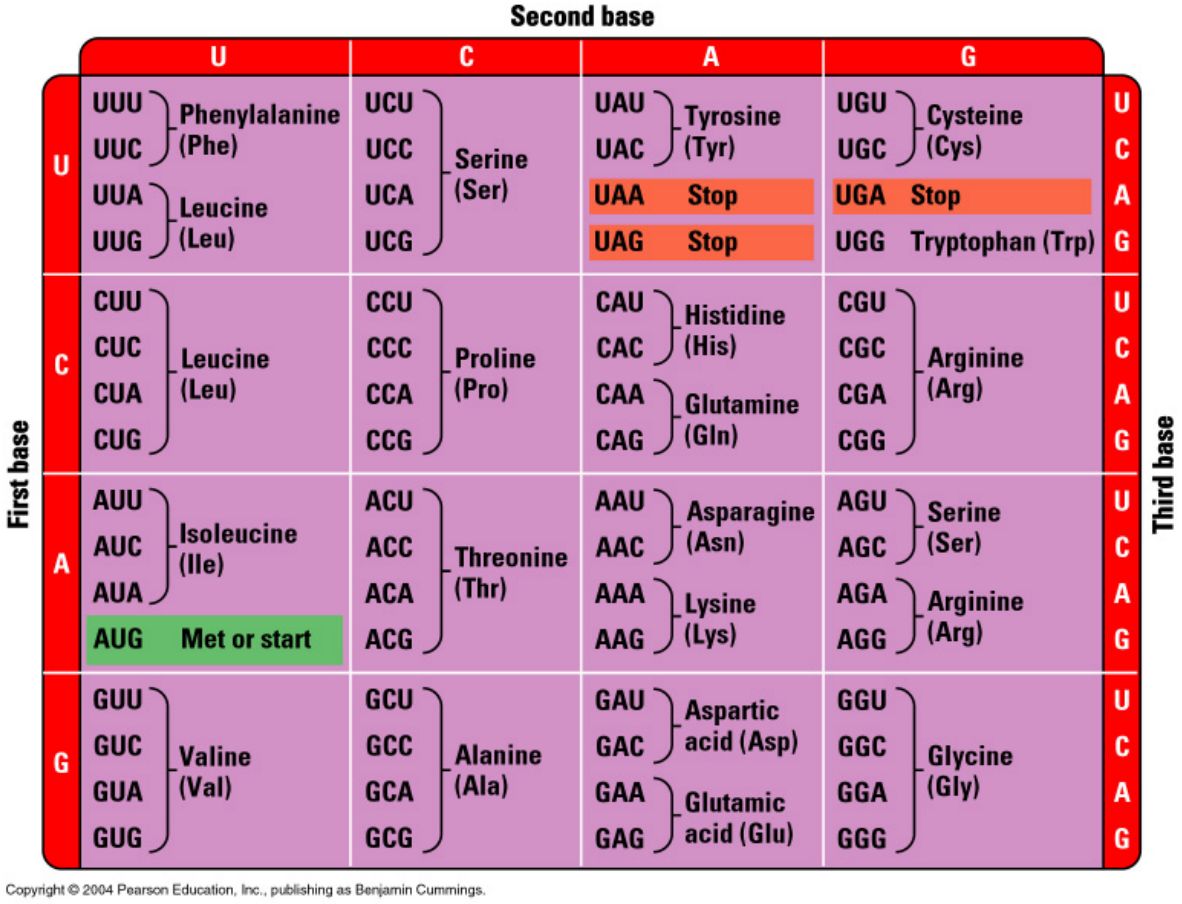
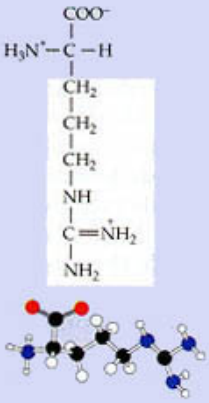


TABLE 3.1
Twenty amino acids found in proteins

A. *Amino acids with electrically charged side chains*

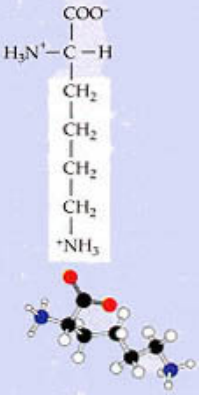
Positive

Arginine (Arg)



Histidine (His)

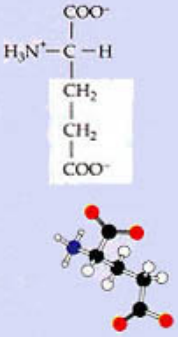
Lysine (Lys)



Negative

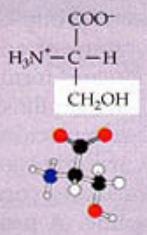
Aspartic acid (Asp)

Glutamic acid (Glu)

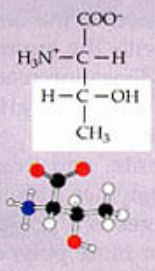


B. *Amino acids with polar but uncharged side chains*

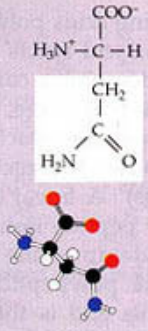
Serine (Ser)



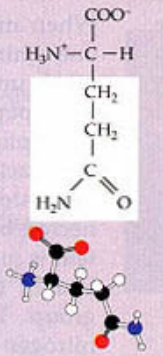
Threonine (Thr)



Asparagine (Asn)

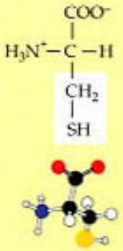


Glutamine (Gln)

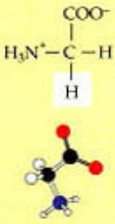


C. *Special cases*

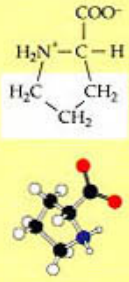
Cysteine (Cys)



Glycine (Gly)

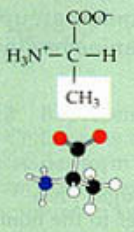


Proline (Pro)

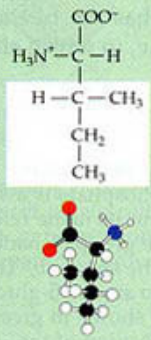


D. *Amino acids with hydrophobic side chains*

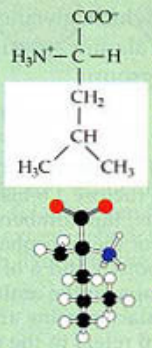
Alanine (Ala)



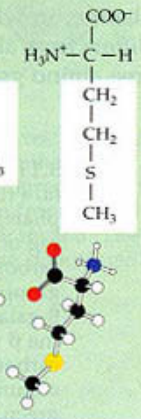
Isoleucine (Ile)



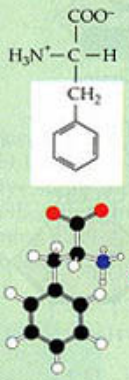
Leucine (Leu)



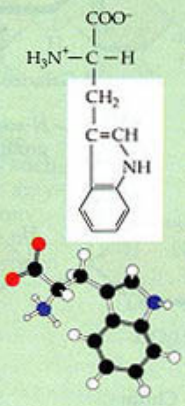
Methionine (Met)



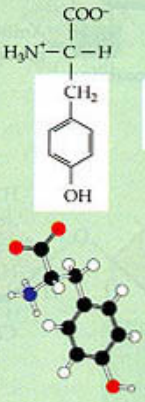
Phenylalanine (Phe)



Tryptophan (Trp)



Tyrosine (Tyr)



Valine (Val)