

BNFO 301: Introduction to Bioinformatics
Introduction to Molecular Biology: Translation - Problem Set

1. Complete the following table:

DNA double helix						A		G				A	
				T				G			T		
mRNA transcribed	5'			A						U			
Appropriate tRNA anticodon						U			G				5'
Amino acids incorporated into protein		met											

2. List the changes that can be produced by a single basepair mutation in the AGA codon encoding arginine and label each silent (no effect on protein structure), conservative (mild effect on protein structure), hydrophobic-to-hydrophilic, hydrophilic-to-hydrophobic, or other.
3. Hemophilia A is an X-linked disease associated with the absence of an essential blood clotting factor, factor VIII (if you don't have any idea what an X-linked trait is, don't worry about it). Factor VIII is encoded by the gene called *FACTOR8*. This gene was cloned from several individuals -- some affected, some not -- and sequenced. A portion of each sequence that you're sure contains the beginning of the gene (i.e., the start codon) was compared with the same portion of the wild-type sequence, as shown below. Each sequence contains only one mutation, shown emphasized.

Wild-type	5'-GGAGTTGAGTCATGGACTCTAAGCAGCGATCCACAAAG...
Individual a	5'-GGAGTT I AGTCATGGACTCTAAGCAGCGATCCACAAAG...
Individual b	5'-GGAGTTGAGTCAT I GACTCTAAGCAGCGATCCACAAAG...
Individual c	5'-GGAGTTGAGTCATGGACTCT I AGCAGCGATCCACAAAG...
Individual d	5'-GGAGTTGAGTCATGGACTCTAAGCAGC I ATCCACAAAG...
Individual e	5'-GGAGTTGAGTCATGGACTCTAAGCAGCGATCCAC I AAG...

For each individual, choose from the list below to describe what you predict would be the severity of the phenotype, and give the reason for your choice.

- A. Severe hemophilia
- B. Mild hemophilia
- C. No hemophilia

4. It is unfortunately quite common for humans or computers to make errors in determining the sequence of nucleic acids. It is particularly common for a string of G's or C's (e.g., CCCCC) to be read erroneously as one too few or one too many. Suppose that you have sequenced some DNA and you are certain that the sequence listed below contains the translational start of a large protein (greater than 300 amino acids), beginning with the normal start codon.

GGGGAGGATAGCCATGCCAGCCCCCTAATTAGGGGGAGTTTCTCTGCAAAA

- a. What should convince you that there is an error in your sequence?
- b. Presuming that there is only one error, a deletion or insertion of a single base, which do you suppose it is and where?