1. Please compare the structures and functions of DNA and RNA. Explain how the structures of these molecules facilitate their required functions within the cell. Also indicate how they are used in the DNA-protein paradigm.

Chemically DNA and RNA are similar with some differences. Like RNA, DNA is made up of four nucleotides; these nucleotides are linked together covalently by phosphodiester bonds. For DNA the nucleotide contains the sugar deoxyribose (hence deoxyribonucleotides) and for RNA the sugar is ribose (ribonucleotides). RNA similar to DNA contains 4 bases: adenine (A), guanine (G) and cytosine (C); and the base uracil (U) instead of thymine (T) in DNA. Unlike DNA, RNA is usually a single-stranded molecule. In DNA, a base in one strand is complementary paired to another base of the other strand with a hydrogen bond. Since RNA molecules are copied during transcription from a limited DNA region, they are shorter than DNA molecules. And single-stranded RNA can fold up in a variety of shape (cloverleaf shape of tRNA) compared to the two strands of the DNA which has an arrangement to maximize the efficiency of base-pair packing (DNA has a minor and a major grove). DNA is found in the nucleus and also in mitochondria, RNA is formed in the nucleus then moves to cytoplasm. Each strand of DNA acts a template for producing a complementary DNA strand during replication to pass the genetic information at each cell division but also is the template for the synthesis of an RNA molecule. Unlike DNA polymerases, RNA polymerases can start an RNA chain without a primer and the transcription does not need to be as accurate as DNA replication. RNA molecules serve as a template for protein production for a limited time before they are degraded. There are several types of RNAs (mRNAs, rRNAs, tRNAs, snRNAs, snoRNAs, and others) with a variety of functions. During protein synthesis, the mRNA nucleotide sequence, is translated into an amino acid sequence by matching the codons of the mRNA molecule to the anticodons of the tRNAs, using complementary base pairing, adding each amino acid carried by the tRNA molecule to a growing polypeptide chain. When the stop codon is decoded, the finished protein is released.

2. You are interested in locating homologs of a human protein "W" in other species. You run a protein sequence similarity search and obtain the following results:

Human protein W vs. :

	E Statistic Value (Low indicating high similarity)
Human Protein W	$< 10^{-84}$
Chimpanzee	$< 10^{-23}$
Dog	$< 10^{-21}$
C.Elegans	$< 10^{-19}$

All of these are statistically significant.
Assuming this is a housekeeping protein vital for cellular function and other proteins have been identified as demonstrating true homology, indicate the most likely species each match is from by filling in the table above with C. Elegans**, Chimpanzee or Dog.**

If you wanted to try to identify more distally related homologs perhaps in E. Coli, would a DNA similarity search be helpful? Briefly describe why or why not.

The human genome is thought to contain about 25 times more genes than E. coli. Housekeeping genes are involved in the basic functions and maintenance of a cell or a set of cells. It will be faster and more. Efficient to run a computer analysis starting with the E. coli gene proteins and find homologs using human gene databases to find sequence similarities.

3. You ask a BME student is working in your lab to design an oligonucleotide to incorporate into a vector for the construction of a custom peptide. The student designs the following sequence. The student does NOT tell you if this is the coding or non-coding strand.

5' TGTTAACTTAGTTATCTCCTCTGCATGGCATGCCTTCA 3'

Reminder – the coding strand corresponds to the codon sequence. A codon table is provided below and available in the slides from lecture 2.1.

a. What is the RNA sequence of the longest reading frame that can be read. This is the desired sequence to translate on the ribosome (begin with the start codon and stop with the stop if any). Be sure to clearly indicate the 5' and 3' ends.

Assume we were given the coding strand[1]:

- If we choose the first nucleotide as the first base in the codon, we obtain:

Coding strand	5'	TGT TAA CTT AGT TAT CTC CTC TGC ATG GCA TGC CTT	3'

- If we choose the second nucleotide as the first base in the codon:

Coding strand	5'	GTT AAC TTA GTT ATC TCC TCT GCA TGG CAT GCC TTC	3'

- The third nucleotide as the first base in the codon:

Coding strand	5'	TTA ACT TAG TTA TCT CCT CTG CAT GGC ATG CCT TCA	3'

Assume we were given the non-coding strand[2]:

Non coding strand	5'	TGTTAACTTAGTTATCTCCTCTGCATGGCATGCCTTCA	3'
Reversed strand	3'	ACTTCCGTACGGTACGTCTCCTCTATTGATTCAATTGT	5'
Coding strand	5'	TGAAGGCATGCCATGCAGAGGAGATAACTAAGTTAACA	3'

- First nucleotide as the first base in the codon:

Coding strand	5'	TGA AGG CAT GCC ATG CAG AGG AGA TAA CTA AGT TAA	3'

- Second nucleotide as the first base in the codon:

Coding strand	5'	GAA GGC ATG CCA TGC AGA GGA GAT AAC TAA GTT AAC	3'
Reading frame	5'	ATG CCA TGC AGA CGA GAT AAC TAA	3'

- Third nucleotide as the first base in the codon:

Coding strand	5'	AAG GCA TGC CAT GCA GAG GAG ATA ACT AAG TTA ACA	3'

The longest reading frame is the frame:

Reading frame	5'	ATG CCA TGC AGA CGA GAT AAC TAA	3'
RNA transcript	5'	AUG CGA UGC AGA CGA GAU AAC UAA	3'

b. Which sequence did the student give you? How do you know?

To design a successful oligonucleotide corresponding to the longest reading frame, the student gave us the sequence of the non-coding strand as detailed in question a.

---

[1] In yellow: start codon; and green: stop codon
[2] In yellow: start codon; and green: stop codon

c. What is the DNA sequence read by the RNA polymerase when assembling the portion of the transcription for part a. What direction does the RNA polymerase read in? Be sure to clearly indicate the 5' and 3' ends.

The RNA polymerase moves along the template strand or non-coding strand in the 3' => 5' direction, and the RNA molecule grows in the 5' => 3' direction.

Direction in which the RNA molecule grows (5' to 3')

Coding strand	5'	TGAAGGCATGCCATGCAGAGGAGATAACTAAGTTAACA	3'
Template strand	3'	ACTTCCGTACGGTACGTCTCCTCTATTGATTCAATTGT	
RNA	5'	UGAAGGCAUGCCAUGCAGAGGAGAUAACUAAGUUAACA	3'

d. The student had some foresight so this sequence can be unidirectionally placed in a vector. This means upon proper construction of the vector and infection in an insect cell, transcription will only occur in the one direction, which produces the longest transcript you have identified by using a promotor upstream and terminator downstream. There is still a possibility that the ribosome could occasionally start translation at a downstream start codon in the same direction, yielding an additional and undesired protein product. How could you eliminate this possibility without impacting the amino acid sequence?

To eliminate the possibility of starting translation downstream missing the start codon we are targeting, we may add 2 or 3 stop codons after this codon for the synthesis of the desired protein. Optionally we may also add a repressor to the start codon located downstream.