Module 02: Quantum Information Processing and Genetic Engineering
Solutions
Total Point Value = 30

Due by midnight on Day 7 of Module 2
This should be submitted to Blackboard as a pdf.

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.

The DNA to protein paradigm begins with DNA, which is transcribed into RNA that is translated into protein. DNA is the most precious commodity in a cell and therefore is the most protected. DNA exists in a double helix where the nucleic acids (A, C, G, and T) are at the center, protected by an outward-facing sugar backbone. When not actively transcribed the long, many-gene containing strands of DNA are condensed into chromatin in the nucleus, again keeping it out of the way of enzymes or other destructive evens within the cytoplasm. In order for transcription to happen special events must decompact the chromatin and unzip the double strand. RNA on the other hand is single stranded – open and ready for translation into protein. Unlike DNA, RNA transcripts contain one or few genes per strand. RNA is comprised of nucleic acids as well (A, C, G, and U). The structure of each molecule aids in its function. For example, DNA's function is to contain all of the genetic code so comes in very long strands that are condensed and protected. RNA's job is to act as a messenger for translation of specific proteins therefore it is short and single stranded.

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	$<10^{-84}$
__*Chimp*_____Protein X	10^{-23}
__*Dog*_ _____Protein Y	10^{-21}
___*C. Elegan* ____Protein Z	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.

No. DNA similarity searches are not as helpful as protein similarity searchers because of the huge amount of "non-sense" DNA.

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.

IA given
T G T T A A C|T T A G|T T A T C T C C T C T G C A T|G G C A T|G C C T T C A 3'

IA compliment
A C A A U U G A A U C A A U A G A G G A G A C G U A C C G U A C G G A A G U 5'
        STOP   STOP   STOP   START   START   STOP

Reading frame 1
Reading frame 2

The stop an start codons are labled above. Remembering that RNA is read from 5' to 3' you can see the START methionine codons (AUG). Now with the STARTS and STOPS labeled you can locate 2 possible reading frames.

Reading frame 2 is the longer frame:
3' A A U|C A A|U A G|A G G|A G A|C G U|A C C|G U A 5'
    STOP   ASN   ASP   GLY   ARG   CYS   PRO   MET

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

The student has given you the non-coding strand. You can tell because the RNA compliment contains the codons for start and stop for the longest reading frame.

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 DNA sequence is read 3' to 5' by RNA polymerase. The DNA sequence for reading frame 2 is:

5' T T A|G T T|A T C|T C C|T C T|G C A|T G G|C A T 3'

We can rewrite this so that e can more easily read it in the 3' to 5' direction

3' T A C|G G T|A C G|T C T|C C T|C T A|T T G|A T T 5'

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?

Using the principle of degeneracy, the MET codon corresponding to the shorter reading frame could be eliminated while still maintaining the amino acid sequence produced by the larger reading frame. The nucleotide that could be changed is the A in the RNA sequence (or T in the DNA sequence). A substitution in this location with a U, C or G in the RNA sequence would still give a proline in that reading frame.

3'	A A U	C A A	U A G	A G G	A G A	C G U	A C C	G U A	5'
	STOP	ASN	ASP	GLY	ARG	CYS	PRO	MET	

desired strand			CYS	PRO	MET
	undesired strand			MET	
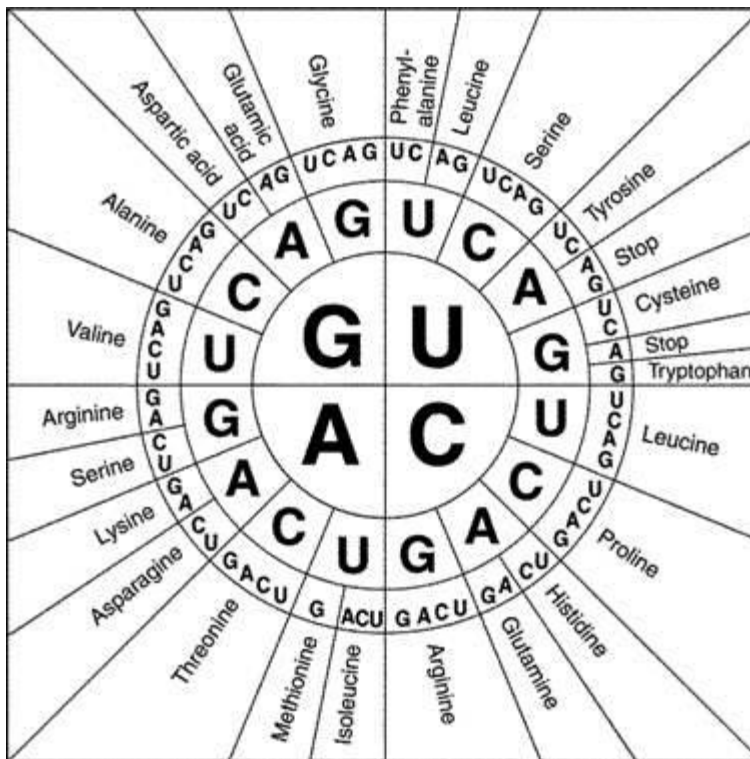
Assignment Rubric

Question	Component	Total Point Value
1	Short answer	6
2	Fill ins	3
	Short answers	3
3	A	5
	B	5
	C	4
	D	4

Total Point Value = 30