Molecular Biology Through Discovery
Problem Set 8: The Genetic Code

1. 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.

2. 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

2a. What should convince you that there is an error in your sequence?

2b. Presuming that there is only one error, a deletion or insertion of a single base, which do you suppose it is and where?

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**T**AGTCATGGACTCTAAGCAGCGATCCACAAAG...
Individual **b**	5'-GGAGTTGAGTCAT**T**GACTCTAAGCAGCGATCCACAAAG...
Individual **c**	5'-GGAGTTGAGTCATGGACTCT**T**AGCAGCGATCCACAAAG...
Individual **d**	5'-GGAGTTGAGTCATGGACTCTAAGCAGC**T**ATCCACAAAG...
Individual **e**	5'-GGAGTTGAGTCATGGACTCTAAGCAGCGATCCAC**T**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. Suppose that two mutants (X1 and X2) of T4 strains, both defective in gene X and both obtained by proflavin, are crossed and phenotypically wild type T4 progeny are selected. Protein X from many of these progeny are purified and sequenced.[1] It is found that the protein from about 50% of the progeny have sequence **A** and protein from the remainder have sequence **B**, both shown below (the amino acids before and after those shown are identical for **A** and **B**).

---

[1] A ridiculously laborious process! In real life, the genes would surely be amplified by PCR and sequenced.

A: `...Tyr-Ser-Trp-Ser-Ser-His-Pro-Arg-Gln...`
B: `...Tyr-Gln-Leu-Val-Ile-Thr-Pro-Lys-Gln...`

4a. As far as you can, deduce the nucleotide sequences for gene X from the two recombinant strains.

4b. What can you say regarding the original mutants?

5. Belozersky and Spirin published a listing of nucleic acid base compositions from a large number of bacteria [(1958) Nature 182:111-112]. Part of the list is shown below.

Species	DNA Base Composition					RNA Base Composition				
	Bases (moles per cent)				G+C	Bases (moles per cent)				G+C
	G	A	C	T	A+T	G	A	C	U	A+U
Proteus vulgaris	19.8	30.1	20.7	29.4	0.68	31.0	26.3	24.0	18.7	1.22
Escherichia coli	26.0	23.9	26.2	23.9	1.09	30.7	26.0	24.1	19.2	1.21
Erwinia carotovora	27.1	23.3	26.9	22.7	1.17	29.5	26.5	23.7	20.3	1.14
Mycobacterium vadosum	29.2	20.7	28.5	21.6	1.37	31.7	23.8	23.5	21.0	1.23
Pseudomonas aeruginosa	33.0	16.8	34.0	16.2	2.03	31.6	25.1	23.8	19.5	1.24

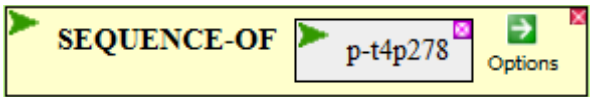
5a. What striking feature is evident in the comparison of DNA and RNA compositions?

5b. Given our present knowledge of RNA, how do you account for their findings?

6. Determine as much of the genetic code as you can of the aliens you mashed up in the investigation *Alien Genetic Code*.

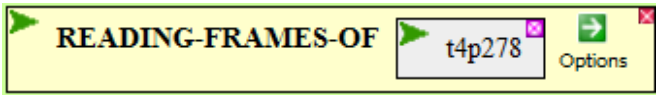
7. Why were no suppressors of FC0 found in segments B1b2 and B2 (see Fig. 2 of Crick et al, 1961)? Let's find out. In doing so, I'll make use of two pieces of special knowledge. BioBIKE/PhAnToMe calls the T4 rIIB gene T4p278 and calls its protein p-T4p278.

7a. Go to BioBIKE/Phantome through the BioBIKE portal, and from the **All** menu, bring down the SEQUENCE-OF function. Display the amino acid sequence of the rIIB ***protein*** by executing the function below (double-click SEQUENCE-OF):



Does the protein sequence begin with a methionine? How big is the protein? From the map of rIIB found in Crick (1962)[2] estimate the amino acid coordinates of the FC0 mutation and of the beginning of the B1b2 segment.

7b. From the **All** menu, bring down the READING-FRAMES-OF function. Display the reading frames of the *rIIB* **gene** by double-clicking READING-FRAMES-OF in:



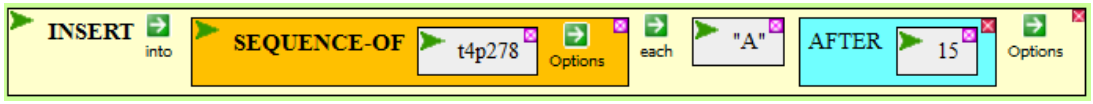
What is the length of the DNA sequence shown? How does that compare to the length of the amino acid sequence of the rIIB protein you found in **7a**?

[2] Crick FHC (1962). The genetic code. Sci Amer 207:66-77 (October, 1962)

7c. How many reading (translation) frames are displayed? Why that many? Which one gives an amino acid sequence that matches the amino acid sequence of the rII protein?

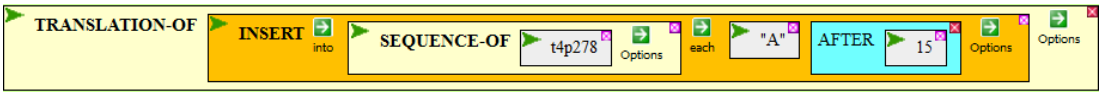
7d. What is the significance of the second reading frame? Hint: Note that the M of the first reading frame lies directly under the A of the first triplet. Where does the C of the second reading frame lie?

7e. Make an insertion mutant of rIIB. Here's how. From the **All** menu, bring down the INSERT function. Drag the SEQUENCE-OF function into the *target* box and edit the entity within SEQUENCE-OF so that it's the gene t4p278 (not the protein p-t4p278). Into the *insert-stuff* box, type "A" and press the Enter key. Finally mouse over the Options icon, select AFTER, and then type 15. Close the box, producing:



7e1. Execute the function. How does the resulting DNA sequence (in the result pane) compare with the sequence of t4p278 you obtained from READING-FRAMES-OF?

Now bring down from the **All** menu the TRANSLATION-OF function. Drag the INSERT function (and all that's in it) into the *entity* box of TRANSLATION-OF.



7e2. Execute the function. How does the resulting amino acid sequence (in the result pane) compare with the amino acid sequences in the different reading frames you found in **7b-7d**?

7e3. What is the significance of reading frame #2 relative to reading frame #1? Of reading frame #3 relative to reading frame #1?

7f. What is the significance of the asterisks (*) that occasionally appear in the amino acid sequences of some reading frames? (check to make sure your guess is correct) Do any appear in the first reading frame? Why not?

7g. The positions of the FC0 mutation and the FC9 mutation are very close together, according to Fig. 2 of Crick et al (1961), but the mutations that can suppress FC0 extend into B1b1 but no further, while the mutations that can suppress FC9 extend into B2. Considering your answers to the above questions (particularly **7a**, **7e2**, and **7f**), why were no suppressors of FC0 found in segments B1b2 and B2? (and is FC0 an insertion or a deletion)?

8. Use the simulation of Crick et al (1961) to create a triple mutant similar to the one that was described in the article. List the steps you took, including the number of plaques you got at each step, the map positions of each mutant, and the phenotypes of each.