Molecular Biology Through Discovery

Problem Set 5: The Coding Problem

1. Using a convenient genetic code table, complete the following:

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												

(Table available in DOCX format by clicking here)

2. Consider the RNA sequence below. Suppose that the fourth base, C, were mutated to a U.

GAGCGUGCGAACC

2a. How many amino acids might be affected if the code were nonoverlapping triplet?

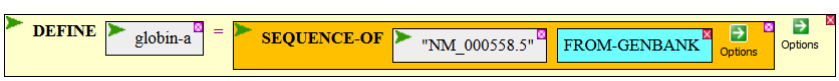
2b. How many if the code were overlapping triplet?

2c. Partially overlapping triplet?

2d. How would your answers be affected if the mutation were a deletion of the C?

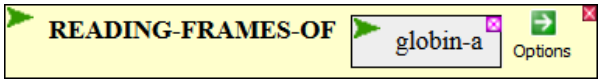
3. Fanconi anemia is an inherited disorder that leads to a variety of developmental defects and has many genetic causes. In one case, symptoms of Fanconi anemia were attributed to a deletion of a nucleotide within the gene encoding alpha globin, leading to a form of hemoglobin known as Hemoglobin Wayne.[1] In this form, the end of the amino acid sequence of alpha globin is markedly different from normal: instead of …Val-Leu-Thr-Ser-Lys-Tyr… (…VLTSKY…) Hemoglobin Wayne has …Val-Leu-Thr-Ser-Asn-Thr… (…VLTSNT…). Your job is to seek a detailed molecular understanding of this mutation.

3a. Download the sequence of alpha globin mRNA from GenBank (accession number NM_000558.5) into BioBIKE as shown below:



Use SEQUENCE-OF globin-a to see what you have (it displays U's as T's – do the mental substitution yourself). How long is the sequence? Can detect where the gene begins and ends?

3b. You probably couldn't find the beginning and end of the gene purely by inspection. The function READING-FRAMES-OF can help. Use it to display the mRNA (with T's instead of U's) in all possible reading frames as shown below:



3c. Explore. How long is the nucleotide sequence shown? How does it relate to the sequence you obtained in **3a**? But there are <u>two</u> nucleotide sequences for each cluster of lines. What is the second one (the one in the middle) about? How does it relate to the one on the top line?

3d. How many reading (translation) frames are displayed? Why that many? What is the significance of the first "T" in Translation Frame 1? How does it relate to a nucleotide sequence? How about the first "L" in Translation Frame 2?

3e. What is the significance of the first "E" in Translation Frame 4? How does relate to a nucleotide sequence?

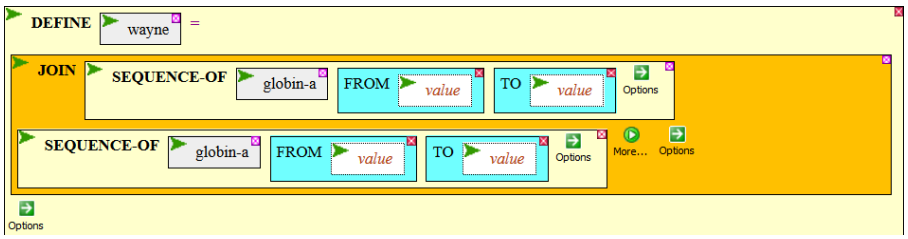
3f. Without looking too hard (yet), which of the translation frames might encode alpha globin? Which could not possibly encode alpha globin? (You might want to recall how big alpha globin is. You've been there before. If you don't remember, you will in a moment…)

3g. Now go and get the amino acid sequence of wild-type human alpha-globin. You learned how to do this a couple of weeks ago. Where in the output of READING-FRAMES-OF globin-a do you find the alpha-globin protein sequence? **Give nucleotide coordinates.**

3h. At what nucleotide coordinate is the first wild-type amino acid in the sequence of alpha-globin that differs from Hemoglobin Wayne? .**Give the coordinate and the amino acid.**

3i. What nucleotide was deleted to produce Hemoglobin Wayne? To what translation frame does the deletion cause translation to shift? **Reproduce the lines relevant to the deletion that show the nucleotide sequence, the original translation frame, and the new translation frame, with the deleted nucleotide highlighted, and the amino acids of Hemoglobin Wayne highlighted.**

3j. (If you're so inclined) Construct the mRNA for Hemoglobin Wayne in the following way:



Fill in the first FROM/TO values with the coordinates 1 and the nucleotide before the deletion. Fill in the second FROM/TO values with the coordinates of the nucleotide after the deletion and the last nucleotide of the sequence. Execute the function. Execute READING-FRAMES-OF `wayne` to see the protein you produced.

4. We live in a world in which genes determine the linear sequence of amino acids that comprise a protein. There are only 20 possible amino acids that may be encoded (putting aside some specialized cases), and there are no restrictions as to what amino acid sequences are possible to encode.

4a. How many possible dipeptides are there? In other words, if you chop up all possible proteins (every conceivable sequence) into two amino acid-segments, how many different kinds of amino acid pairs would you get?

The remaining questions concern an alternate universe in which ***the genetic code consists of overlapping triplets****, each codon overlapping the next by two nucleotides.*

4b. Consider the triplet codon CAG. How many pairs of adjacent codons are possible in which the first codon of the pair is CAG? What is the maximum number of dipeptides that can be encoded by all of those pairs?

4c. How many possible triplet codons are there?

4d. How many possible pairs of adjacent triplet codons are there? What is the maximum number of dipeptides that can be encoded by all of those pairs?

4e. Suppose that the overlapping triplet genetic code we're considering is degenerate, that is more than one triplet may encode the same amino acid. If the dipeptides shown below are found in nature, how many triplets, at minimum, must encode histidine (His)?

His-Lys, His-Ser, His-Leu, His-Thr, His-Phe, His-Pro
Lys-His, Ser-His, Cys-His, Arg-His, Val-His, Phe-His, Glu-His, Gln-His, Ile-His

4f. It is 1957. There are many partial amino acid sequences of proteins known, but DNA sequencing is 20 years in the future. Can you think of a way to use known protein sequences to test the proposition that the genetic code consists of overlapping triplets?

4g. You might enjoy reading Brenner (1957).[2]

5. Suppose that every Virginia resident is to be assigned an ID number, except that it will be in the form of a DNA sequence. How long would the DNA sequence need to be to allow for a unique sequence for every resident? ***Provide details of your calculation plus any assumptions you made.*** Extra credit: Choose the sequence that would be your own ID.

6. Use the simulation of Crick et al (1961) to recreate their experiment

6a. Create mutants that suppress FC0. **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.**

Extra – in case you have the time and inclination

6b. [Not easy!] Use the mutants created in 4a to make double mutants and finally a wild-type triple mutant. Again, 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.

REFERENCES

1. Seid-Akhavan M, Winter WP, Abramson RK, Rucknagel DL (1976). Hemoglobin Wayne: A frameshift mutation detected in human hemoglobin alpha chains. Proc Natl Acad Sci USA 73:882-886.
2. Brenner S (1957). On the impossibility of all overlapping triplet codes in information transfer from nucleic acid to proteins. Proc Natl Acad Sci USA 43:687-694.