Molecular Biology Through Discovery

Problem Set 6: 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. 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 and sequenced from several individuals -- some affected, some not. 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...

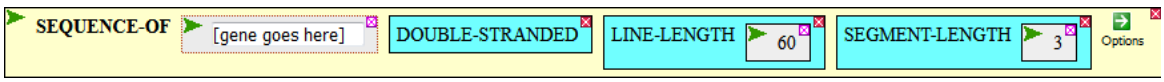
For each individual, choose from the list below to describe what you predict would be the severity of the phenotype, and give the reason for your choice.

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

3. In one of the experiments reported by Brenner et al (1965),[1] *amber* mutations within the *rIIB* gene of bacteriophage T4 were produced through mutagenesis by hydroxylamine. Let's re-examine that experiment given the actual sequence of the gene (unknown to Brenner et al).

3a. Go to ViroBIKE (not CyanoBIKE), and find the rIIB of phage T4 (nicknamed ente-T4 for no particularly good reason).

3b. Display the sequence of the gene, double-stranded, in units of three (to make codons more apparent):

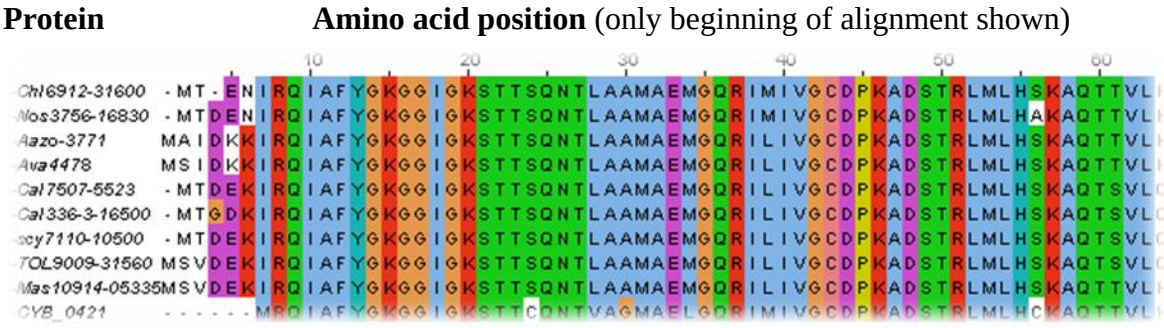


3c. Identify at least one site that when subjected to hydroxylamine could produce a class K *amber* mutant. Find a different site that could produce a class B *amber* mutant. For the purposes of this question, presume that any mutation leading to an amino acid change produces a mutant rII phenotype and any mutation that does not lead to an amino acid change produces a wild-type rII phenotype.

4. Determine as much of the genetic code as you can of the aliens you mashed up in the investigation *Alien Genetic Code*, using the methods of Jones and Nirenberg (1962).[2]

5. You want to determine whether nitrogen fixation is taking place in a lake deep below the Antarctic permafrost. You could try to measure nitrogen fixation directly, but it is impossible to get to the lake with the necessary equipment, and a laboratory measurement of fixation in lake water sampled from the lake would be fraught with uncertainty. You therefore decide an indirect approach: isolating DNA from a lake water sample and determining whether there are any genes present that encode nitrogenase, the enzyme responsible for nitrogen fixation.

   To do this, you need to amplify the genes using PCR… but how do you design PCR primers to amplify a gene whose DNA sequence is not known? There is hope. Nitrogenase is a highly conserved protein. Below you'll find an alignment of the amino acid sequences of nitrogenase subunits from 10 bacteria. You can see that there is unanimity at most positions. It's reasonable to expect that even weird Antarctic bacteria will have nitrogenase proteins that



   are similar to all others.

   From an alignment of the sequences from more nitrogenase subunit proteins, a consensus sequence was determined, part of which is shown below. Letters are given when all sequences have the same amino acid at that position. A hyphen indicates that the position shows variability amongst the sequences.

```
         10        20        30        40        50
RQIAFYGKGGIGKSTT-QNT-A—A-----RI-IVGCDPKADSTRL-L—K
AQ---L—AAE-G-VED-EL—V---G-----CVESGGPEPGVGCAGRGI
IT-INFLEE-GAY-D—FV-YDVLGDVVCGGFAMPIRE-KAQEIYIV-SG
EMMAMYAANNIARG-LKYA—GGVRLGGLICNSR
```

   Your goal is to design a forward primer and a backward primers to amplify a part of the nitrogenase genes of any nitrogen-fixing cyanobacteria that happen to be in the lake water. The primers should:

   - be based on the consensus amino acid sequence shown above
   - each be at least 14 nucleotides in length
   - amplify a DNA fragment at least 100 nucleotides in length

- guarantee amplification of any target DNA that encodes a nitrogenase that matches the consensus amino acid sequence

You will find that no one DNA sequence will fit the last criterion, so you're permitted to specify ***degenerate*** primers with ambiguous positions. For example, AG[CG] is a degenerate sequence, because the third position can be either C or G. This leads to the fourth condition

- The degenerate primer sequence should minimize ambiguity as much as possible. For example AG[CG] (two possibilities) is less ambiguous than [GT]C[AT] (four possibilities)

Provide the two degenerate primer sequences that fulfill these conditions and calculate how many possible sequences each degenerate primer matches.

REFERENCES

1. Brenner S, Stretton AOW, Kaplan S (1965). Genetic Code: The 'nonsense' triplets for chain termination and their suppression. Nature 206:994-998
2. Jones OW, Nirenberg MW (1962). Qualitative Survey of RNA Codewords. Proc Natl Acad Sci USA 48:2115-2123