

I. Biology (35 points total)

The following questions cover some of the core concepts in biology that will be indispensable for understanding the lectures. Please be concise, but be sure to include all of the keywords in your answers.

- a) a) Compare the structure of DNA and RNA. Keywords: nucleotides, 5 prime, 3 prime, deoxy-ribonucleotides, ribonucleotides, anti-parallel double-helix (5 points)
- b) b) What's the basic subunit that's used to build a protein? Describe the 3 levels of protein structure? Keywords: amino acid, N terminus, C terminus, primary-, secondary, tertiary-structure (5 points)
- c) c) Explain the Central Dogma in biology. Keywords: DNA, RNA, protein, transcription, translation (5 points)
- d) d) Explain what is genetic code. Keywords: translation, tri-nucleotide codon, tRNA (5 points)
- e) e) Compare the basic structure of eukaryotic and prokaryotic cells. Keywords: nucleus, organelle, mitochondria, ribosome (5 points)
- f) f) Describe the structure of a eukaryotic gene. Keywords: promoter, exon, intron (5 points)
- g) g) Describe a biological problem that you think is worth looking into using some computational methods. This could be speculation of your possible final project topic. (5 points)

Bonus: (5 points total)

- a) Explain several ways in which the expression of genes can be regulated. (2 points)
- b) Which techniques could be utilized to look into one (or more) ways of gene regulation globally? In which of the above mentioned regulation steps could systematic quantitation and/or computational modeling be applied, and how? (3 points)

II. Perl Program (35 points total)

A Navy Seals unit retrieved a mysterious sample from a suspected Al Qaeda hideout, and to support the president's war on terror, you are asked to prove that it's related to biological weapon research. So far, a crack team of lab techs at the CDC extracted RNA from the sample, which turned out to be low quality. They made what cDNA they could out of it, and then amplified it using non-specific PCR. The DNA was run out on a gel, and a large band was found that turned out to be a 188 base pair fragment with the following sequence:

```
CATTACGATGCATTG ATTTTCAAAGGAAT GTACTATCGAAATCA
CAAGTCGTGGACTAC
GGTTTGCAGTGGAGG AATCGCAGTCTTTGC AGGCTCACGCCTTTC
TTGATAAGTCGTTGT
TTCAAACGTTTAATT TTCAGGGTGATTCAG ATGGGGATACATATA
TGTTCCAGACGATGA
TTTCACCT
```

The initial results didn't support the right conclusion, so now its up to you. The only problem is that you have no idea about the correct reading frame or direction for this sequence. Write a perl program that can translate this sequence in all six reading frames, and use the results to infer the most likely amino acid sequence that this fragment encodes. You can use the skeleton code (see ps1_skeleton.pl file) as a guideline. Your code should print the following to output:

1. Transcribed sequence (i.e., converted to RNA) (5 points)
2. Translated sequence in the first three reading frames (10 points)
3. The reverse-complement RNA sequence (10 points)
4. Translation of the reverse-complement RNA sequence (5 points)

Submit this output, as well as your source code, with your problem set.

What does the `translate_codon` subroutine do and how? Annotate each line and submit the annotated version with your problem set (5 points)

If you don't finish the whole thing on time, a sound plan partially implemented with correct, working code will earn a generous amount of partial credit.

Bonus: (5 points)

Create a display of the amino acid frequency counts in all the reading frames. Feel free to use and improve the example code. Highlight your improvements and your own re-usable code in your solution, and post them in the course discussion forum, so we can build up a nice library for the projects and for future assignments. If you are comfortable with advanced perl, check out bioperl.org.

III Mathematica (30 points)

The Mathematica problem is in the Mathematica "notebook" file Warmup.nb (see associated files for this set). It is meant to be a self-contained introduction to discrete data sets and functions. First go over the tutorial(see associated files for this set) until you are comfortable with typing and evaluating basic expressions.