Programming assignment 2:

Computer scientists often tackle problems that originate in other fields. One field that has been greatly augmented by computer science is biology, resulting in bioinformatics. For this assignment you will be exploring the potential application of your computer science skills for analyzing protein synthesis.

DNA is composed of four nucleotides: **A**denine, **C**ytosine, **G**uanine, and **T**hymine. The complementary base pairs are A-T and C-G. That is, in the two strands of DNA, every A on one strand is touching a T on the other strand and every C is touching a G.

Cells convert DNA into working proteins by a process called protein synthesis. First, DNA is transcribed to a messenger RNA molecule. During transcription, the DNA molecule functions as a template for complementary base-pairing. RNA however does not have Thymine. Instead Adenine is paired with Uracil. Thus the DNA sequence ATCGAATTCCGG would transcribe to UAGCUUAAGGCC.

You will be given a text file with a DNA sequence. Your program will read in the file, internally convert the DNA sequence to its complementary RNA sequence, and then write a new file with the amino acid sequences that were coded by the DNA. Amino acids are represented by one letter. Every protein will be on its own line (a protein ends when a STOP codon is read).

The file CTable represents an RNA codon table. The format of this file is:

Amino Acid: codon, codon, codon, …

Each line has one amino acid followed by all the codons that code for it.

Please submit your source code, any input files you used, and the resulting output file with the amino acid sequences in a zipped folder or a tarball. Email this folder with the subject line “Project 2, ‘YOUR NAME HERE’” (replacing the placeholder with your name). For example, if I submitted the project the subject line would read “Project 2, Gregory Zhenin”.