

NUCLEOTIDE-TO-AMINO ACID TRANSLATOR

HOMEWORK 3, DUE THURSDAY, FEBRUARY 28, 2013

Minimum Requirements (maximum grade: B). Please write a Python script that will take a DNA nucleotide sequence as input and use the standard genetic codon table to translate it into an amino acid sequence. Your script must also calculate the number of codons and the GC/AT percentage and output that information. Your program should raise a `ValueError` exception if it detects malformed input data, such as non GCAT nucleotides or an input nucleotide sequence that isn't evenly divisible by 3.

Additional Requirements (functionality needed to get an A). There are two requirements for an A grade. First, your script must somehow allow for the user to input their sequence to your program without having to open up your program for editing and copy/pasting a FASTA into it. Your program might prompt the user for input (good), or it might prompt the user for a file where the input can be found and read the file (better), or it might take the raw nucleotide sequence or the path to the file containing the sequence as a command line argument, a.k.a., start the program and specify input to that program all in one step (best). (Hint: google `sys.argv`) Second, your script must also keep track of the codons used and print out a summary of codon usage as part of the output.

Concepts used. Using a `dict`, checking if a key exists within a `dict`, methodically taking a substring from one larger string, counting instance of an item in a string, updating a `dict` of counters, taking raw user input.

Python Packages to be Used. You shouldn't need to import any additional packages, core nor 3rd party, however the core package `argparse` may be useful to allow the user to input her input as a command line argument.

Additional Optional Features. If you'd like to challenge yourself, here are some additional features of your homework that might make it more fun:

- You could restrict the AT/GC percentage output to one or two decimal places.
- You could give 6 translations for all 6 reading frames.
- You could allow the user to use RNA codons as well as DNA.

- You could allow the user to specify whether he would like to use the standard codon table, or some other codon translating regime such as bacteria, yeast, or mitochondria.