

Programming Assignment 8

Assignment Goals: understanding problem requirements; effective program design; logical thinking; dictionary methods

You will submit the program on Blackboard. Name it *yourname_PA8.py* (replacing *yourname* with your own name).

Problem:

Write a program that prompts the user for a DNA sequence and translates the DNA into protein. Please see below for the codon table to use. Thus, you will have to determine how to split the DNA sequence into codons, look up the amino acid residue for each codon, and append all the amino acids to give a protein.

Some program requirements:

1. When the codon doesn't code for anything (e.g., stop codon), use "*".
2. If the sequence length is not a multiple of 3, ignore the extra bases at the end.
3. If the sequence contains ambiguous codes (i.e., unknown bases), use "X".

Test your program with a few interesting test cases. A sample test case is shown below.

The following is a dictionary which represents the genetic code. The keys are codons and the values are amino acid residues. A list of stop codons is also provided. You may hard-code this dictionary (*gencode*) and list of stop codons (*stop_codons*) into your program.

```
gencode = {
    'TTT': 'F', 'TTC': 'F', 'TTA': 'L', 'TTG': 'L', 'TCT': 'S',
    'TCC': 'S', 'TCA': 'S', 'TCG': 'S', 'TAT': 'Y', 'TAC': 'Y',
    'TGT': 'C', 'TGC': 'C', 'TGG': 'W', 'CTT': 'L', 'CTC': 'L',
    'CTA': 'L', 'CTG': 'L', 'CCT': 'P', 'CCC': 'P', 'CCA': 'P',
    'CCG': 'P', 'CAT': 'H', 'CAC': 'H', 'CAA': 'Q', 'CAG': 'Q',
    'CGT': 'R', 'CGC': 'R', 'CGA': 'R', 'CGG': 'R', 'ATT': 'I',
    'ATC': 'I', 'ATA': 'I', 'ATG': 'M', 'ACT': 'T', 'ACC': 'T',
    'ACA': 'T', 'ACG': 'T', 'AAT': 'N', 'AAC': 'N', 'AAA': 'K',
    'AAG': 'K', 'AGT': 'S', 'AGC': 'S', 'AGA': 'R', 'AGG': 'R',
    'GTT': 'V', 'GTC': 'V', 'GTA': 'V', 'GTG': 'V', 'GCT': 'A',
    'GCC': 'A', 'GCA': 'A', 'GCG': 'A', 'GAT': 'D', 'GAC': 'D',
    'GAA': 'E', 'GAG': 'E', 'GGT': 'G', 'GGC': 'G', 'GGA': 'G',
    'GGG': 'G' }
```

```
stop_codons = ['TAA', 'TAG', 'TGA']
```

Sample test case:

```
Enter DNA sequence: ATGAACCGTATTTAAGCCAUGGATCT
The DNA sequence ATGAACCGTATTTAAGCCAUGGATCT translated into protein is: MNRI*AXD
```