

Lab Exercise 1

1)

Calculate the GC content of the following DNA sequence:

```
ATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAA  
AAATATAAATTAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGA  
AGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGC  
AACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAAGACACCAAGGAAGCT
```

2)

Translate the above DNA sequence to protein sequence using the codon table:

```
codon_table = {  
    'UUU': 'F', 'UUC': 'F', 'UUA': 'L', 'UUG': 'L',  
    'UCU': 'S', 'UCC': 'S', 'UCA': 'S', 'UCG': 'S',  
    'UAU': 'Y', 'UAC': 'Y', 'UAA': '*', 'UAG': '*',  
    'UGU': 'C', 'UGC': 'C', 'UGA': '*', 'UGG': 'W',  
    'CUU': 'L', 'CUC': 'L', 'CUA': 'L', 'CUG': 'L',  
    'CCU': 'P', 'CCC': 'P', 'CCA': 'P', 'CCG': 'P',  
    'CAU': 'H', 'CAC': 'H', 'CAA': 'Q', 'CAG': 'Q',  
    'CGU': 'R', 'CGC': 'R', 'CGA': 'R', 'CGG': 'R',  
    'AUU': 'I', 'AUC': 'I', 'AUA': 'I', 'AUG': 'M',  
    'ACU': 'T', 'ACC': 'T', 'ACA': 'T', 'ACG': 'T',  
    'AAU': 'N', 'AAC': 'N', 'AAA': 'K', 'AAG': 'K',  
    'AGU': 'S', 'AGC': 'S', 'AGA': 'R', 'AGG': 'R',  
    'GUU': 'V', 'GUC': 'V', 'GUA': 'V', 'GUG': 'V',  
    'GCU': 'A', 'GCC': 'A', 'GCA': 'A', 'GCG': 'A',  
    'GAU': 'D', 'GAC': 'D', 'GAA': 'E', 'GAG': 'E',  
    'GGU': 'G', 'GGC': 'G', 'GGA': 'G', 'GGG': 'G',  
}
```

Calculate the frequency of each amino-acid, and find the most frequent one.

3)

Are there other possible start sites in the sequence? How many?

For each start site, what would be the sequence of the translated protein? What would be the lengths of the DNA and protein sequences? What are the reading frames (0, 1 or 2)?

4)

Run:

import this