

py_bio_assignment_0

Given the following DNA sequence,

```
'CGAGTAGAGAGTACGGATGCTCCAGTCGTCGTCCTTTTCTTGCGAACTCCTTGAATAGTCCCTGTTTTCGGGATATTGCAACTGCA
CGGTGCATGGCTGATGACCGTTGCACGGTCGCGACGGGCCGACAGGTCACAAAGGGATAAATCAAAACAGAGATCTATAGTTAACCACG
CGGAGTCGACCAACTTGCTAGTTCCCTAAGCCAGCAAATATATAGCTATTACCTGCGTCGCACTTATTTATTCCTGGCCTTTCGCGTATG
TCCTACGTGGCCATACGTTGTCCAAATAAAATCAGTCTTTGCAGGCTGTCACGCTCTACATAATTTCTGTGGCACAACAGGCCTACTCG
TCTCTCTCGACCCCTGCCCGCTTAGGTGAATGAGCTAACCGTTATTACAAGTGACCCGACGAGACCGGGAACTGCTCTGCAGAAGGCCCG
TAGAACGGATAAAAACAAATGACAAACACGTGACCATGCGCCCTGCTGGACAACATGTGTAGTATGCCATCCATCTTACCAATGGGGCTA
GGTGGCAGTTTAAGCGGGATCCTTTGAACCTAATAGCGCGCAATCATGAGCGCGGACTCTACACGTCAGAAATAGATCAGAAAATTATTA
CATCACTAAGTGCTTCTGACCAACCCAGTGAGCCACGCACGCTACTCTTTACCATAAAAACGCATCGGAAGGGCCCGCCTAAACTTCG
CGCCGCGATCTGCATCACTATGTGTACCTATTATCAGAGTCTGTGCAGCACTGTTTTACTGGTAAAGGTTGTGTACTAGCAGTACACATC
TCTTCAGAGTGCCCTGAGGCAATAATAGATCTCGTAATCATAATTCTCGCCACCGCGTAGGCTTCCAGTCCAAGCTGGGTCTATTTATC
GCACAAATGTCGTGTATGTCTGCCCTGAGGGGAGCTGTGCAGACCAAGCGACGACCTTCGGCCGGCGCCTAATAGTACCTGGGATACGGGC
GCTCCGCGGTT'
```

complete the following exercises.

1. Assign the sequence to a variable named **seq**.
2. Determine the length of the sequence, and assign this value to a variable named **length**.
3. Convert the sequence to lowercase, then convert it back to uppercase.
4. Count the number of each nucleotide and store the information in a dictionary named **counts**.

a. *Hint:*

```
counts = {
    'A': 0,
    'T': 0,
    'C': 0,
    'G': 0
}
```

5. Determine the proportion of each nucleotide as a function of the length of the sequence.
6. Determine the GC content of the sequence.

7. Convert all instances of 'T' in the sequence to be 'U', and assign this to a variable named **rna**.
8. Determine the starting indices for all occurrences of the substring 'ATG'.
9. Replace all instances of 'ATG' with M **only if** the starting index is **divisible by 3**.