BIO-321 ABIN HW Assignment

Plagiarism: All submitted codes are expected to be the result of your individual effort. You should never misrepresent someone else's work as your own. In case of any plagiarism institute policy will be followed.

Instructions:
1) Submit a Python notebook file as ABIN_HA_<Name>_<rollno>.ipnyb. Inside python notebook, write problem number and the solution with the output.

Question 1: String Reversal Problem [marks 5]

Given a DNA sequence A. Return the string A after reversing the string base by base.
For example, the reverse complement of Pattern = "GTCA" is Pattern = "ACTG".
Find the reverse complement of a DNA string.

Given: A DNA string Pattern.
Return: Pattern, the reverse of Pattern.

Sample Dataset:
ATGCCGTAGGATCATTGACTTAC

Sample Output:
GCATTCAGTTACTAGGATGCCGTA

Question 2: Reverse Complement Problem [marks 5]

In DNA strings, symbols 'A' and 'T' are complements of each other, as are 'C' and 'G'. Given a nucleotide p, we denote its complementary nucleotide as p. The reverse complement of a DNA string Pattern = p1…pn is the string Pattern = pn … p1 formed by taking the complement of each nucleotide in Pattern, then reversing the resulting string.

For example, the reverse complement of Pattern = "GTCA" is Pattern = "TGAC".
Find the reverse complement of a DNA string.

Given: A DNA string Pattern.
Return: Pattern, the reverse complement of Pattern.

Sample Dataset
AAAACCCGGT

Sample Output
ACCGGGTTTT

Question 3: Finding K-mers [marks 10]

The k-mers in a string are all the substrings of length k. So, given the string "ELEPHANT", and k-mer length k=4, the k-mers are:

ELEP
LEPH
EPHA
PHAN
HANT

Find all possible k-mer occurrences in a given read.

Given: read : A single DNA sequence,
k : The value of k for which to count kmers.

Return: List of k-mers

Sample Dataset
count_kmers("GATGAT", 3)

Sample Output
ATG
GAT
TGA

Question 4: Counting DNA Nucleotides **[marks 5]**

A string is simply an ordered collection of symbols selected from some alphabet and formed into a word; the length of a string is the number of symbols that it contains.
An example of a length 21 DNA string (whose alphabet contains the symbols 'A', 'C', 'G', and 'T') is "ATGCTTCAGAAAGGTCTTACG."

Given: A DNA string s.
Return: Four integers (separated by spaces) counting the respective number of times that the symbols 'A', 'C', 'G', and 'T' occur in s.

Sample Dataset
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTG
TCTGATAGCAGC

Sample Output
A:20 C:12 G:17 T:21

Question 5: Finding position of a Motif in DNA [marks 5]

Given two strings s and t, t is a substring of s if t is contained as a contiguous collection of symbols in s. The position of a symbol in a string is the total number of symbols found to its left, including itself (e.g., the positions of all occurrences of 'U' in "AUGCUUCAGAAAGGUCUUACG" are 2, 5, 6, 15, 17, and 18). The symbol at position i of s is denoted by s[i].

Given: Two DNA strings s and t.

Return: All locations of t as a substring of s.

Sample Dataset

GATATATGCATATACTT
ATAT

Sample Output

2
4
10

Question 6: Find K-mer of length 5 and implement De-Bruijn Graph Assembly for the following sequence with minimum overlap of 2:
TGTAGAAAGTACCCAGTGCTCAGTATAG [marks 10]