

Programming Assignment 7

Assignment Goals: understanding problem requirements; effective program design; logical thinking; string operations and methods

You will submit the program on Blackboard. Name it *yourname_PA7.py* (replacing *yourname* with your own name).

Problem:

The following is a short section of genomic DNA:

```
ATCGATCGATCGATCGACTGACTAGTCATAGCTATGCATGTAGCTACTCGATCGATCGATCGATCGATCG
ATCGATCGATCGATCATGCTATCATCGATCGATATCGATGCATCGACTACTAT
```

The given DNA sequence comprises of two exons and an intron. The first exon runs from the start of the sequence to base number 63 (starting counting from zero), and the second exon runs from base 91 (also counting from zero) to the end of the sequence. The intron is the sequence of DNA in between the two exons.

Write a program that counts all substrings including overlapping substrings and determines the coding and non-coding regions of the given DNA sequence. You may hard-code the given DNA sequence into your program. You will write this program using a `main()` and several functions. Your program should contain the following functions and perform the following:

- A function called `count_overlapping_substring(dna, substring)` that returns the number of occurrences of a substring of characters (`substring`) in the DNA sequence (`dna`) **including overlapping substrings**. For example, calling the function with the given DNA sequence and `substring` as 'TCGAT' will return 15. *Note:* You counted substrings before in Programming Assignment 4 but used the `count()` method, which does not count overlapping substrings. This time, you want to count overlapping substrings as well so the `count()` method should not be used here.
- A function called `coding_regions(dna)` that returns only the coding regions (i.e., the exons) of the DNA sequence as one string. You will need to concatenate the coding regions together.
- A function called `percentage(dna)` that computes and returns what percentage of the DNA sequence is coding.
- A function called `dna_coded(dna)` that returns the original genomic DNA sequence with coding bases (i.e., the exons) in uppercase and non-coding bases (i.e., the intron) in lowercase as one string.
- A `main()` function that prompts the user for a substring, calls the functions above, and prints out the information resulting from each function above.