
BT 3051 — Data Structures and Algorithms for Biology

Jul–Nov 2017

Assignment 1

17th August 2017

Due date: 24th August, 2017 @ 17:00

Maximum marks: 20

Instructions: Write Python codes to solve the problems mentioned below. If you need any assistance, feel free to write to me or the TAs via Piazza (private note). Evaluation will be based on the codes and the logic.

Academic Integrity: You are allowed to discuss the problems verbally with your friends, but copying or looking at codes (either from your friend or the Web) is not permitted. Transgressions are easy to find, and will be reported to the “Sub-committee for the Discipline and Welfare of Students” and will be dealt with very strictly. Mention any collaboration (discussions only!) in your solutions.

Late submission penalties: 1 second – 24 h: 20%; 24–48 h: 40%; > 48h: 60%

Early submission bonuses: > 24h: 5%, > 48h: 10%, > 72h: 20%

Evaluation: Assignments will be evaluated by the TAs within two weeks of the due date. You can check out your marks and contest them, if needed, for at most one more week post-evaluation, i.e. three weeks from the due date of the assignment.

Problem Statement

1a. (10 marks) Calculating Protein Mass

(The problem is adapted from <http://rosalind.info/problems/prtm/>.)

Given: A protein sequence in FASTA format. See <http://rosalind.info/glossary/fasta-format/> for details on the FASTA format.

Return: The total weight of each of the proteins in the input file. Consult the monoisotopic mass table from the file `PROT_MASS.txt`.

Your code should read something like:

```
# Homework header as usual
#
#
#
import sys
def read_fasta(fname):
    """ (str) -> (list of tuples)
    # function body with documentation

    """
```

```

    return sequences # a list of (sequence_name, sequence) tuples

def compute_protein_mass(protein_str):
    """
    #function body including documentation and test cases
    >>> compute_protein_mass('SKADYEK')
    821.392
    """

if __name__ == '__main__':
    #DO NOT CHANGE THE FOLLOWING STATEMENTS
    for seq_name, seq in read_fasta("hw1a_dataset.faa"):
        print (seq_name, compute_protein_mass(seq))

```

1b. (10 marks) Translating mRNA to protein

The 20 commonly occurring amino acids are abbreviated by using 20 letters from the English alphabet (all letters except for B, J, O, U, X, and Z). Protein strings are constructed from these 20 symbols. The RNA codon table dictates the details regarding the encoding of specific codons into the amino acid alphabet.

Given: An RNA string s corresponding to a strand of mRNA (of length at most 10 kbp).

Return: The protein string encoded by s .

See <http://rosalind.info/problems/prot/> for the complete description of the problem. You can read the translation table from the file `RNA_TABLE.txt`. Your code should read something like:

```

# Homework header as usual
#
#
#
import doctest

def translate_DNA(mrna, translation_table = 'RNA_TABLE.txt'):
    """
    #function body including documentation and test cases
    >>> translate_DNA('AUGUAUGAUGCGACCGCGAGACCCGUGCACCCGCGAAAGCUGA')
    MYDATASTRCTRES
    """

if __name__ == '__main__':
    doctest.testmod(verbose = True)

```

How to Submit your Homework

- Submit your assignment ONLY via the submission link: <http://tinyurl.com/bt3051-submit>.
- You should not be signed into Dropbox while uploading this file (or use an incognito window to open the link), so that you can enter the following details during submission, instead of Dropbox auto-filling it:
 - First Name: Roll Number
 - Last Name: Your Full Name
 - E-mail: Your email id

-
- Save your solution files as `hw1a.py` and `hw1b.py`. Do not use different filenames!
 - Each of your submission files, `hw1a.py` and `hw1b.py` should begin with the **header information** shown below — the number of the assignment, your roll number, your collaborators' roll number(s), and approximately how much time you took to solve the problems in that part of the assignment.
 - **Submissions not adhering to any of the above instructions will not be evaluated.**
 - Also do not send the files by e-mail — obviously, they will not be evaluated.

```
#BT3051 Assignment 1a
#Roll number: BE15B001
#Collaborators: CH15B001, EE15B001
#Time: 1:15
```

Attention: This assignment is fairly simple; the main purpose of this assignment is to ensure that you can write simple Python programs, document them well and write sensible test cases.