Instruction File:

In this exercise, you'll create a Python program to calculate the physical-chemical properties of a protein sequence. Your task is to develop the ProteinParam class included below in template form. The program includes the methods that you will need, along with the completed "main" that does all of the input and output.

When testing, you will type the protein sequence, hit return, the program will respond with the required output. You can then enter a new protein string and a new analysis will be presented.

Your program will read in the protein sequence and print out the:

- Number of amino acids
- Total molecular weight
- Molar extinction coefficient
- Mass extinction coefficient
- Amino acid composition

For example, if I enter the protein sequence: VLSPADKTNVKAAW

then the program should output:

Number of Amino Acids: 14

Molecular Weight: 1499.7

Molar Extinction coefficient: 5500.00

Mass Extinction coefficient: 3.67

Amino acid composition:

A = 21.43%

C = 0.00%

D = 7.14%

E = 0.00%

F = 0.00%

G = 0.00%

H = 0.00%

I = 0.00%

K = 14.29%

L = 7.14%

M = 0.00%

N = 7.14%

P = 7.14%

Q = 0.00%

R = 0.00%

S = 7.14%

T = 7.14%

V = 14.29%

W = 7.14%

Y = 0.00

Design specification:

__init__:

There are a number of ways to design this. Your __init__ method could save an attribute which is just the input string. A more effective solution would compute and save the aaComposition dictionary. All of the protein parameter methods can operate very efficiently using a dictionary. Use of an aaComposition here will save you quite a bit of work.

aaCount():

This method will return a single integer count of valid amino acid characters found. Do not assume that this is the length of the input string, since you might have spaces or invalid characters that are required to be ignored.

aaComposition():

This method is to return a dictionary keyed by single letter Amino acid code, and having associated values that are the counts of those amino acids in the sequence. Make sure to include all 20 amino acids. Proper amino acids that are not represented in the sequence should have a value of zero.

Note: if you have already calculated a composition dictionary in init, then just return that dictionary here.

molecularWeight():

This method calculates the molecular weight (MW) of the protein sequence. If we have the composition of our protein, this is done by summing the weights of the individual Amino acids and excluding the waters that are released with peptide bond formation.

$$MW_{H_2O} + \sum_{}^{aa} (MW_{aa} - MW_{H_2O})$$

Where,

MWH2O – Molecular weight of water

MWaa – Molecular weight of amino acids

molarExtinction():

The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. It is useful to have an estimation of this coefficient for measuring a protein with a spectrophotometer at a wavelength of 280nm. It has been shown by Gill and von Hippel that it is possible to estimate the molar extinction coefficient of a protein from knowledge of its amino acid composition alone. From the molar extinction coefficient of tyrosine, tryptophan and cystine at a given wavelength, the extinction coefficient of the native protein in water can be computed using the following equation.

$$E = N_Y E_Y + N_W E_W + N_C E_C$$

where:

N_Y is the number of tyrosines, N_W is the number of tryptophans, N_C is the number of cysteines

E_Y, E_W, E_C are the molar extinction coefficients for tyrosine, tryptophan, and cysteine respectively.

I have supplied the molar extinction coefficients at 280nm for each of these residues in a dictionary(aa2abs280) in the program template.

N_Y, N_W, N_C - Refer dictionary created for amino acid.

massExtinction():

We can calculate the Mass extinction coefficient from the Molar Extinction coefficient by dividing by the molecularWeight of the corresponding protein.