**Exercise – structural bioinformatics**

**Part A**

Write a Python class "Seq" for a protein sequence (string of amino acids). This class should have three methods:

1. "pI" – returns the pI of the protein (see explanation below).

2. "mass" - returns the molecular mass (in Daltons, by the attached table).

3. "numRes" - returns the number of residues.

Using "seq" class, calculate these three values for each sequence in the sequences file protSeq.txt.

**pI calculation:**

The charge of ionizable residue in a given pH can be calculated using eq. 1 for basic residues and eq. 2 for acidic residues. Don’t forget to relate to the C and N terminals.

http://isoelectric.ovh.org/files/pI1.pnghttp://isoelectric.ovh.org/files/pI2.png

Eq. 1 Eq. 2

The pI is the pH at which the protein with no (0) net charge.

For example the pI value of the protein "HKRSDGE" is 6.8 (check also [here](http://web.expasy.org/compute_pi/))

**Note:**

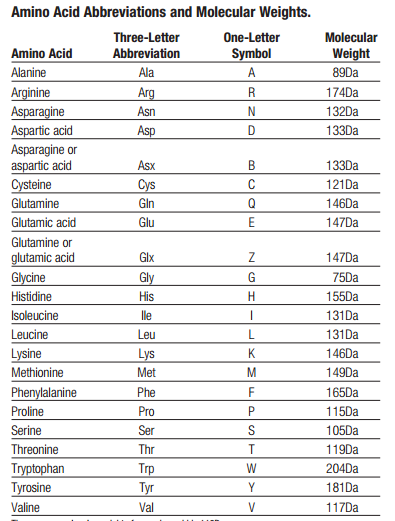
You can use Biopython package (SeqIO) for parsing the fasta file and for calculating the methods 2,3 (ProteinAnalysis class at the sub package ProtParam). Anyway, the method "pI" should be implemented in this exercise (i.e. you can inherit ProteinAnalysis class and just add the method pI).

**Table of pKa:**


**Table of charges:**

charges = {'D':-1, 'E':-1, 'H':1, 'C':-1, 'Y':-1, 'K':1, 'R':1, 'N-ter':1, 'C-ter':-1}



**Table of amino acid weights:**

**Part B**

In this part we will use the class you created in the previous part in order to calculate the pI for the entire proteome of human. Do the following steps in your home linux directory:

1. Download all Swiss-Prot sequences - "Reviewed (Swiss-Prot)" fasta format from [here](https://www.uniprot.org/downloads).
2. Unzip and take only the sequences that contain os=Homo sapiens
3. Calculate pI of these sequences with Seq class.
4. Draw histogram of the results.
5. What is the highest pI protein ?
6. What is the lowest pI protein?