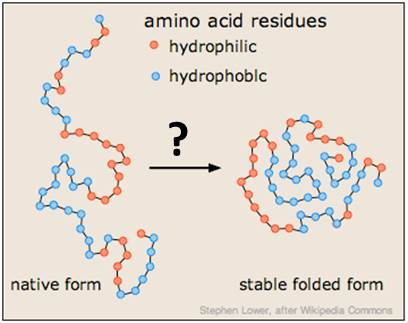
**Part A: Exercise – Radius of Gyration - Biopython**

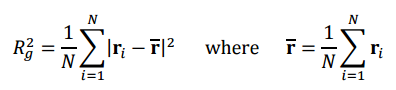
Proteins are biopolymers made of covalently bonded chains of the natural twenty amino acids. Unlike synthetic polymers, most proteins spontaneously organize or “fold” to a unique three dimensional structure in aqueous solution, the dominant

environment in the cell. This process is driven by the protein’s desire to tuck hydrophobic amino acids in the core of the structure away from water (which is ~70% of cells).

In this assignment, we will check if indeed in real protein structures the hydrophobic residues form the core and the hydrophilic residues lay on the surface.



You will test it using the so-called radius of gyration. Radius of gyration describes the overall spread of the molecule (Radius of gyration is indicator of compactness of the protein structure) and is defined as the root mean square distance of the collection of atoms from their common center of gravity.



For example the Rg of protein structure at 1vii pdb code is: 8.817567

Download from the course website the collection of protein structures (.pdb files) that will be used in this example. The files are compressed in an archive; unzip this archive to a new directory on your computer where you will store your code and perform your computations.

1. Write a class "protein":

Variables:

Tuple with x, y, z coordinate of **only the Ca** (alpha carbon) atoms.

Tuple with the protein residue names.

Methods:

Constructor: gets pdb file name and parse it using Biopython

Rg: returns the radius of gyration of the protein.

Rg\_phobic: returns the radius of gyration of only the hydrophobic residues.

Rg\_philic: returns the radius of gyration of only the hydrophilic residues.

1. Compute Rg, Rg\_philic and Rg\_phobic for all the

Proteins and Show on excel graph:

1. General Rg as a function of chain length (number of residues). T
2. Rg\_phobic and Rg\_philicas a function of chain length.
3. The ratio Rg\_philic/ Rg\_phobicas a function of chain length.

* What is your conclusion?
* In an old paper the authors showed this relation of Rg(N). (N is the chain length). Did your results agree with this?

Rg = 0.395\*N3/5 + 7.257

**Part B:**

שני החלבונים 1AFL ו-*2I5S* הם אנזימים מסוג RNAse A. בתרגיל זה נבחן את הקשר בין דמיון ברצף לבין דמיון במבנה השלישוני של החלבון.

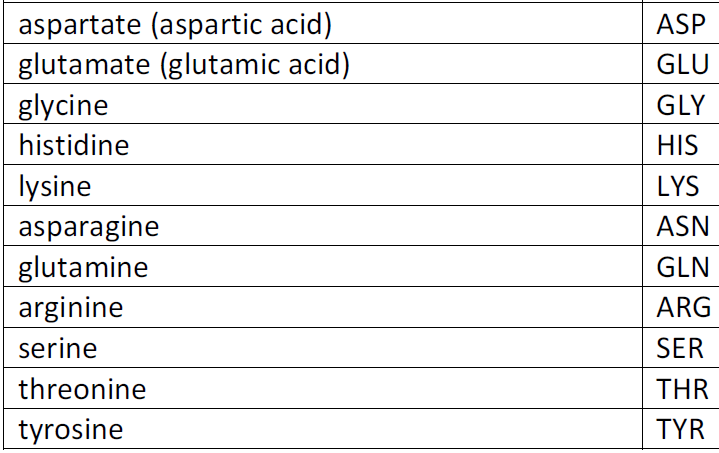
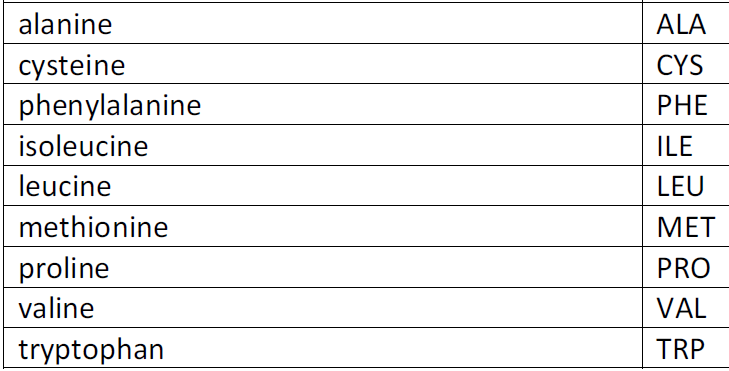
1. בצעי alignment לרצפים של 2 החלבונים. כמה % identity מצאת?
2. פתחי את web app MSDfold (עדיף בדפדפן אקלפלורר) והריצי השוואת מבנה של 1AFL מול כל החלבונים ב pdb. את התוצאות יש לסדר לפי % identity ולחפש את *2I5S* שנמצא בעמוד האחרון. מהו ה RMSD בין שני המבנים? כמוכ צרפי תמונה שלהם לאחר סופרפוזיציה (View superpose). מהי המסקנה?
3. מהו הקיטלוג של scop של שני המבנים הנ"ל? נמקי את הקיטלוג שמצאת לאור סעיף א ו-ב (התיחסי להגדרות fold,superfamily,family)

**Appendix:**

* This exercise can be done in our linux machine "bioinfo": after ssh to linapp do "ssh bioinfo". In bioinfo we have python 2.7+Biopython and numpy installed.

Hydrophobic Residues

Hydrophilic Residues



**Biopython BioPDB**



Add this command to cancel the warnings of biopython:

import warnings

warnings.filterwarnings("ignore")

**Mean and sum calculations:**

import numpy as np

print np.mean([1,2,3])

print np.sum([1,2,3])

output

>>>

2.0

6

>>>

**Iteration of all the files in the directory:**

import glob

for pdbfile in glob.glob('\*.pdb'):