# Algorithms in Structural Bioinformatics

E.D. Chrysina and I.Z. Emiris

## Assignment 2, version B

Announced: 08.04.2020 / Deadline: Sunday 26.04.2020 23:59

## Distance checking

a) Feruloyl esterases are enzymes of industrial interest that catalyse the hydrolysis of the ester bond between hydroxycinnamic acids such as ferulic acid and sugars present in the plant cell wall. There is strong interest by the industrial sector to exploit them as biocatalysts. The PDB codes for two of these enzymes are 6FAT and 1UWC. These correspond to the same enzyme that is expressed by two different organisms, *Fusarium oxysporum* and *Aspergillus niger*, respectively.

1. Visualize both protein structures using Chimera (https://www.cgl.ucsf.edu/chimera/) showing the secondary structure elements.
2. Prepare two figures:
   - one figure presenting the 2 superimposed overall 3D structures using the relevant tool as implemented in Chimera (following link). Give the c-RMSD between the 2 structures as calculated by Chimera after superposition. This figure will highlight the changes in the overall structures (https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/superposition.html)
   - one figure presenting the two structures superimposed giving emphasis on the calcium binding site which is present only in 6FAT. This figure will reveal changes imposed by the presence of calcium.

5 points

b) Calculate the c-RMSD for the comparison of the structures in (a) using your software from Assignment 1, or any available tool, if you didn't finish your own: Make sure you download the coordinates in .pdb format and NOT .cif format because this might not be readable by selected software. So you shall calculate c-RMSD over:

   i. all atoms,
   ii. over Ca atoms,
   iii. over main chain atoms

10 points

c) Select 3 residues at the calcium binding site namely Val276-Ala277-Asp278 in the 3D structure of Feruloyl esterase with PDB code 6FAT. Consider only their 3 backbone atoms (smallest indices) namely atoms N, Ca, C of each, as points with 3D coordinates. Construct the corresponding Cayley-Menger (border) matrix B of dimension 10x10.

5 points

d) Check that rank(B)=5. Compute the Gram matrix G, and its SVD so as to get 3D point coordinates. Check if it is the same structure, by computing its c-RMSD to (c).

10 points

e) Perturb entries of B by a percentage, then compute G, apply SVD so $G=U\Sigma U^T$. Let S be the diagonal matrix containing zeros and the 3 largest singular values of G. Suppose $G \sim USU^T$, compute the 3D coordinates and report the c-RMSD against the actual structure. What is the

max perturbation that yields a structure whose c-RMSD < R = 1 Angstrom (or another meaningful bound R) with the real structure?

10 points

f) Repeat (e) but instead of perturbing, keep only distances of B which are less than some cutoff T and estimate T such that the computed structure has c-RMSD < R against the input structure. Show the values of T you tried and the resulting c-RMSD.

10 points