

Algorithms in Structural Bioinformatics

I.Z. Emiris and E.D. Chrysina

Assignment 2

Announced: 25.04.2021 / Deadline: Sunday 09.05.2021 23:59

Part I.

The crystal structure of the receptor binding domain of SARS-COV-2 Spike glycoprotein in complex with COVOX-269 Fab was recently determined and deposited with Protein Data Bank (PDB id: 7NEH). The structure of its mutant N501Y has also been determined (PDB id: 7NEG).

1. Report on the following:

- Method by which the two structures were determined and their Resolution
- Number of chains and number of residues in each chain for every structure
- Number of water molecules in the structure (for both pdb files)
- Which are the ligands (heteroatoms) that present in the structures
- Give a list of protein-protein interfaces using EMBL-EBI tools (<https://www.ebi.ac.uk>) for both structures. Would you expect to see differences between the two?

2. Determine the R.M.S.D. between receptor binding domain of SARS-COV-2 Spike glycoprotein complex and its mutant

- Over all atoms
- Over Ca atoms

3. Visualize the two protein structures of the complex using Chimera (<https://www.cgl.ucsf.edu/chimera/>) showing the secondary structure elements. Please highlight the regions of interest.

- One figure per structure
- One figure with the two structures superposed

Part II.

- Consider 50 Ca atoms of SARS-COV-2 Spike glycoprotein (PDB id: 7NEH), indexed E401 to E450. Construct the corresponding 51x51 Cayley-Menger matrix B.
- Compute $\text{rank}(B)$; explain why the obtained value is correct.
- Perturb entries of B by 2%, and 4% (maintaining symmetry, positive entries), then confirm that $\text{rank}(B) > 5$. In each case, compute Gram matrix G, apply SVD: $G = U\Sigma U^T$. Let S be the diagonal matrix containing zeros and the 3 largest singular values of G. Compute the 3D coordinates as $\sqrt{S}U^T$, and report the c-RMSD against the original structure for both perturbations.