

Processing of Biological Data

Prof. Dr. Volkhard Helms
Summer Semester 2017

Saarland University
Chair of Computational Biology

Exercise Sheet 4

Due: 27.06.2017 10:15

Submit your solutions (as a single PDF attachment) and code to mat@bioinformatik.uni-saarland.de. For the coding part, I encourage you to use Python.

Exercise 4.1: The Protein Data Bank (PDB) (80 points)

Familiarize yourself with the database here <http://www.rcsb.org/pdb/home/home.do>. Additionally get to know the PDB file format here http://deposit.rcsb.org/adit/docs/pdb_atom_format.html

(a) Protein-protein interface (80 points).

From the PDB homepage download the file (.pdb) for this protein: PDBID 5NI1.

- Briefly describe the protein. (5 points)
- How many chains are present in the protein. (5 points)
- Compute the center of mass of this protein. *For simplicity assume that the mass of each atom is 1.* (10 points)
- Devise a simple algorithm to identify a set of residues that contact residue(s) from another chain. A residue is in contact with other residues if the atom-atom distance between them is < 0.5 nm. Write your solution as a pseudocode. (25 points)
- Implement the algorithm. Your solution should output a text file containing the residue number, residue name and the contact residue of the other chain. (25 points)
- Do you think these residues are critical to the protein? Why? (10 points)

Exercise 4.2: Molecular Dynamics (MD) Simulation (20 points)

Familiarize yourself with MD simulation here <https://www.nature.com/nsmb/journal/v9/n9/full/nsb0902-646.html>. Also check out this rudimentary MD simulation demo <http://physics.weber.edu/schroeder/software/demos/MDv0.html>.

(a) In the demo, start the simulation and add some heat. What happens to the molecules when the temperature is high? (5 points)

(b) Time series analysis. (15 points)

From the course website, download the file `ic_500ns.csv`. The file contains inconsistency coefficient values of pocket formation from a 500 ns long simulation.

- Plot the data over time, compute and report the average for the simulation with *ion* and without *ion*. Can one compute a P-value from these two numbers? (5 points)
- Divide the data points into 20 blocks of equal length, compute and report the average on each block (block averaging). (5 points)
- Perform a Hypergeometric test on the averages. Is the difference between the two simulations statistically meaningful? (5 points)

Have fun!