

Institute of Theoretical Computer Science

Lecturer: Prof. Jack Snoeyink

Assistant: Yves Brise

April 4, 2007

Geometric Computations in Molecular Biology SS07

Exercise Set 3

Course Webpage: <http://www.ti.inf.ethz.ch/ew/courses/GCMB07/>

Due date: April 18, 2007

Exercise 1 (Programming assignment RMSD)

Task summary:

- Implement the subroutine $RMSD(P, Q)$ that takes two sets of points in \mathbb{R}^3 , $P = \{p_1, \dots, p_n\}$ and $Q = \{q_1, \dots, q_n\}$, and computes the root mean squared deviation (RMSD) between P and Q after they have been optimally aligned.
- Extract the points of the backbone atoms (N , CA , C) of all residues (1-326) from two PDB data files provided, TAM.pdb and LF_A.pdb.
- Plot a chart of the RMSD values for the fragments starting with residue i and ending with residue j , for $i, j = 3, 6, 9, \dots, 324$.
- Hand in source code and plot picture.

Details:

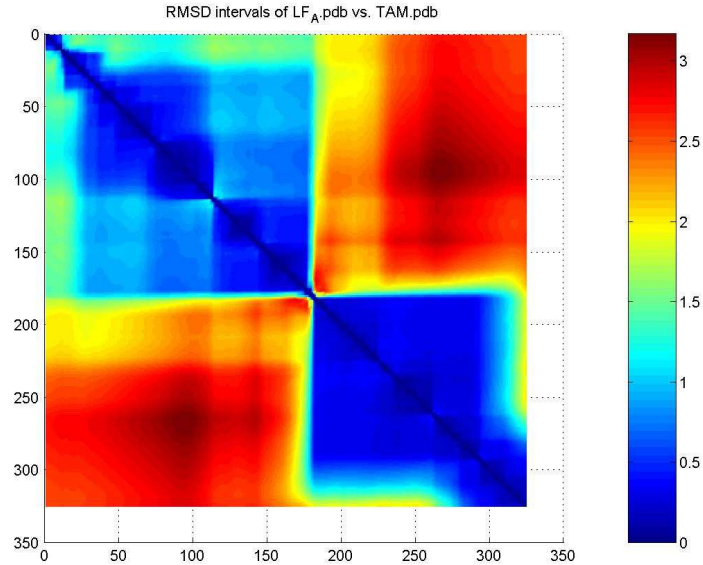
Idea:

You are given two conformations of the same moveable protein. Read in the positions of the atoms and then try to figure out where the motion occurs. Do this by computing the RMSD for corresponding fragments of the two protein versions. You can then plot the RMSD values by coloring a point (i, j) for the two fragments that have all residues in the range $[i, j]$ on the corresponding proteins.

To cut down on the number of intervals to measure, just do this for intervals where i, j are multiples of three. That is, starting with every 9th atom, since there are three atoms in the backbone of each residue.

Example: $i = 3, j = 6$, means to compare the backbone of residues 3 to 6 of the two proteins. That again means computing the RMSD value for 12 pairs of atoms.

As a final result you should get something like the following picture.



Data:

Tryptophanyl-tRNA Synthetase, or TrpRS, is a molecule that moves as it performs its function. It actually behaves a little like a pacman: when it binds an ATP nucleotide it can close down and snip off a fragment from another molecule.

We have made available two PDB files from the lab of Charlie Carter, in Biochemistry at UNC Chapel Hill.

LF_A.pdb	open, unliganded conformation
TAM.pdb	closed, product state

TAM.pdb is also available from www.rcsb.org with id 1I6L, if you wish to see more references, or images. TrpRS actually functions as a dimer (two copies of the molecule); these PDB files contain only one copy, sometimes called the asymmetric unit.

Coordinates:

Suppose that we take the coordinates of just the backbone atoms (N, CA, C) with residue numbers 1 – 326. That is, we take the first three floating point numbers (which are in units of Ångströms) from lines like the following ones. The integer number preceding the first floating point number is the residue number. So, the three lines below make up the backbone of residue number 276. You can discard all except the backbone atoms. Note that the spacing between the entries is fixed (see <http://www.wwpdb.org/documentation/format23/sect9.html#ATOM>).

```
ATOM 2185 N VAL A 276 58.493 24.954 20.318 1.00 15.79 N
ATOM 2186 CA VAL A 276 57.832 24.003 21.217 1.00 14.04 C
ATOM 2187 C VAL A 276 58.005 22.557 20.768 1.00 16.58 C
```

Algorithm:

To compute the RMSD, we encourage you to use the quaternion version of Horn's algorithm, that was presented in the lecture. If you feel reluctant though, there is a version of the very same algorithm using orthonormal matrices. Here, we will state the summary of the quaternion version.

input : Two finite sets of points $P, Q \subset \mathbb{R}^3$, $|P| = |Q|$
output: RMSD value for P and Q optimally aligned

1. Compute coordinates of points relative to centroids;
2. Compute 3×3 matrix M ;
3. From the entries of M , construct 4×4 matrix N ;
4. Compute largest eigenvalue and corresponding eigenvector \dot{e}_m of N ;
5. Normalize eigenvector to obtain rotation quaternion $\dot{q} = \dot{e}_m / \|\dot{e}_m\|$;
6. Compute $val = \sqrt{\frac{1}{n} \sum_{i=1}^n \|\dot{r}'_{r,i} - \dot{q} \dot{r}'_{l,i} \dot{q}^*\|^2}$;
7. Compute translation $\dot{r}_0 = \dot{r}_r - \dot{q} \dot{r}_l \dot{q}^*$;

return val

The definition of all variables and symbols was given in the lecture. The notation is also consistent with Berthold Horn's paper "Closed-form solution of absolute orientation using unit quaternions", which is available on the course webpage.

Programming language:

You are free to choose your favourite programming language for this task. You should take some time to consider though, because it will be best if you also do the subsequent programming assignments in that language, in order to perhaps reuse some of the code. We should be able to give you support in most of the better-known languages. In case of doubt, ask us first.

One word of advice regarding Matlab and Maple. Even though it might seem advantageous to use them because of the easily accessible algebra and plotting functionality, they have at least one huge disadvantage: they're slow!

Visualizing optimal alignment (not required):

If we optimally align LF_A and TAM and measure the RMSD, we should get 2.5485 Å.

We can actually get Pymol to tell us this and show the alignment using its command "fit", which takes two sets of selected atoms. It does not do this quite as well, though, because it is using an iterative alignment algorithm. It may also skip some atoms where the amino acid types are not identical in the two structures.

To do this, load TAM.pdb with Pymol and then

```
PyMOL>sel tbb, (name n,ca,c and resi 1-326)
Selector: selection "tbb" defined with 978 atoms.
PyMOL>load LF_A.pdb
ObjectMolecule: Read crystal symmetry information.
Symmetry-Error: Missing space group symbol
CmdLoad: "LF_A.pdb" loaded as "LF_A".
```

```
PyMOL>sel lbb, (name n,ca,c and resi 1-326 and not tbb)
Selector: selection "lbb" defined with 978 atoms.
PyMOL>fit (lbb),(tbb)
Executive: RMS = 2.550 (936 to 936 atoms)
```

Change to ribbon display for both molecules to focus on the backbone. You can see that the blue is a little wider than the green.

Not all parts move equally:

```
PyMOL>fit (lbb and resi 1-100), (tbb and resi 1-100)
Executive: RMS = 1.398 (294 to 294 atoms)
PyMOL>fit (lbb and resi 100-200), (tbb and resi 100-200)
Executive: RMS = 2.384 (279 to 279 atoms)
PyMOL>fit (lbb and resi 200-300), (tbb and resi 200-300)
Executive: RMS = 0.673 (300 to 300 atoms)
```

If we look at this last one, we see the the right side of the picture has good alignment and the left is a little out of alignment — there is some sort of hinge in the middle.