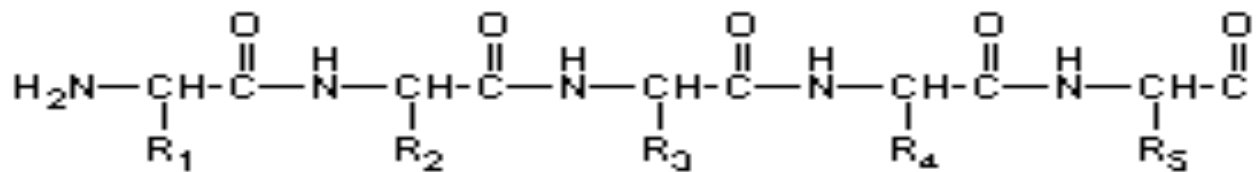


CS2110 Course Project (250 points):

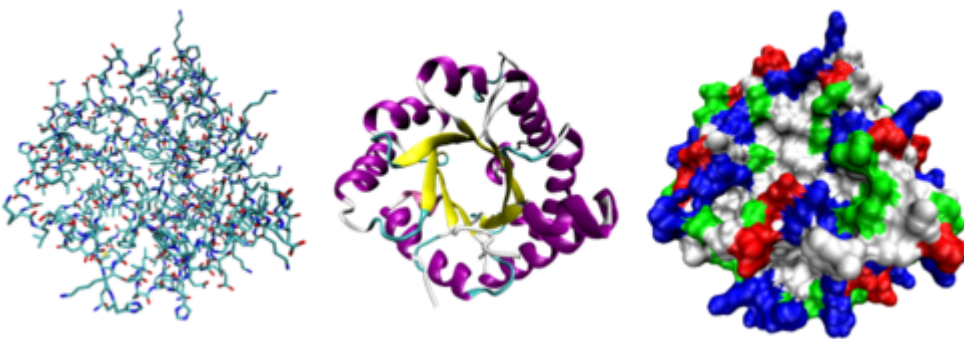
Write a C program to manipulate a protein structure file

1. Introduction

A protein is a fundamental macromolecule that carries out biological functions. It consists of a sequence of 20 different kinds of amino acids joined by peptide bond. Each amino acid is composed of a number of atoms (C, N, O, H). More information about protein is here: <http://en.wikipedia.org/wiki/Protein>



In order to carry out biological function, protein must fold into 3-D structure. So each atom of a residue has a position (x, y, z) in the 3D space.



Protein structure can be determined by X-ray Crystallography technique. The coordinates of atoms are determined and stored in a file in the Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>).

The essential fields about ATOM are: index of atom, atom name, amino acid name, index of amino acid, x, y, and z as follows.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"ATOM"	"
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Code for insertion of residues.
31 - 38	Real(8.3)	x	Orthogonal coordinates for X in Angstroms
39 - 46	Real(8.3)	y	Orthogonal coordinates for Y in Angstroms
47 - 54	Real(8.3)	z	Orthogonal coordinates for Z in Angstroms
55 - 60	Real(6.2)	occupancy	Occupancy.
61 - 66	Real(6.2)	tempFactor	Temperature factor.
77 - 78	LString(2)	element	Element symbol, right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

The detailed description about the PDB format is here:

<http://www.wwpdb.org/documentation/format23/sect9.html>

Protein structure can be visualized by Rasmol. Rasmol can be downloaded here: <http://www.umass.edu/microbio/rasmol/rasras.htm>

2. Project Description

Download a protein structure file:

<http://www.cs.missouri.edu/~chengji/cs2110.pdb>

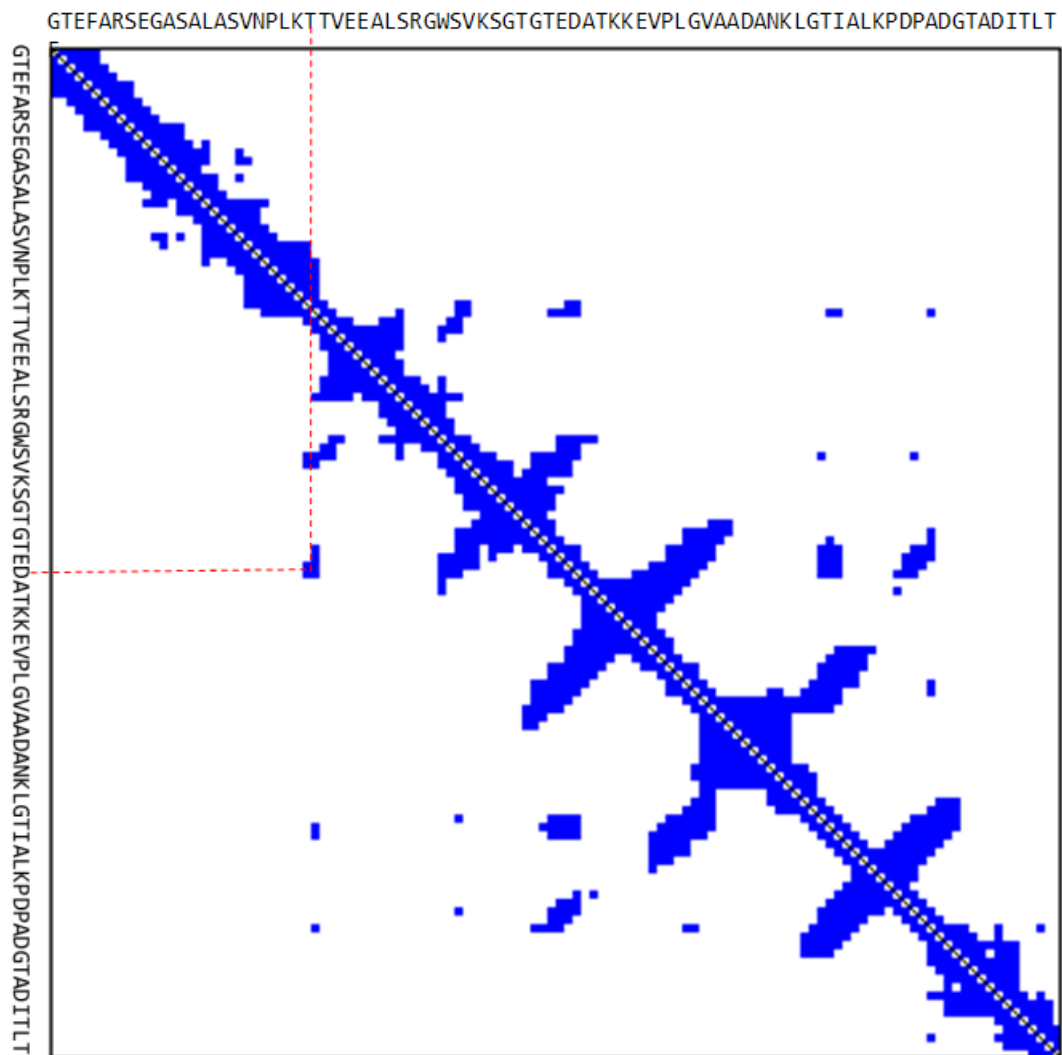
Write a C program to manipulate the structure in the file as follows (**140 points**):

1. Read a protein structure file into the program by a file name (10 points)
2. Print out the file to screen (10 points)
3. Print out the x, y, z coordinates of an atom according to an index typed in by a user (15 points)
4. Print out the x, y, z, coordinates of all the atoms of an amino acid according to the index of the amino acid (15 points)
5. Calculate the distance between any two atoms according to indices of atoms (i.e. atom serial number) (15 points)
6. Calculate the distance between two CA atoms of two amino acids according to indices of amino acids (i.e. residue sequence number) (15 points)
7. Rotate a protein structure with respect to Z axis by an angle θ in degrees. (20 points)
 - a) Calculate the x, y, z coordinates of the centroid of structure. The coordinates are just average values of all the coordinates of all the atoms.
 - b) Update the coordinates of each atom by subtracting them by the coordinates of the centroid.
 - c) Do transformation. The new x' , y' coordinates after rotation can be calculated from x, y coordinates before rotation as follows.

$$x' = x \cos \theta + y \sin \theta$$
$$y' = -x \sin \theta + y \cos \theta$$

8. Save the structure into a PDB file. (10 points)
9. Print out the amino acid contact map (20 points)

A contact map is a visual representation of distance between two amino acids. If the distance between the CA atoms of two amino acids is less than a threshold (let's say 8), the two amino acids are in contact, otherwise not in contact. Your task is to print a 2-Dimensional map for the protein as in the following figure.



An example of contact map of a protein. Each small square represents a contact where the distance of two corresponding amino acids is less than 8.

In this project, you can use a “*” to represent a contact.

10.Exit

Finally, create a command line menu to integrate the entire programs (step 1-8) together (**10 points**).

The menu may look like this:

Please enter a number to select an operation:

1. Read a protein structure from a PDB file
2. Print out the structure
3. Print the coordinate of an atom
4. Print the coordinates of atoms of an amino acid
5. Calculate the distance between two atoms
6. Calculate the distance between CA atoms of two amino acids
7. Rotate the structure with respect to Z axis
8. Save the structure into a file
9. Print the amino acid contact map to the screen
10. exit

3. Project Report, Source Code and Submission

Report (60 points)

After you finish your project, write a report to describe what you have done.

The report must contain the following sections.

1. Abstract: briefly summarize the problem and solution
2. Introduction: describe the problem in details
3. Methods: describe the function and structure of your program
4. Results: print out the output of each step. For the step 7, please include the pictures of a structure before rotation and after rotation. You can visualize a PDB structure by Rasmol
5. Compilation and Run: list the number of source code files and describe how to compile and run the programs.

Source Code Requirements (50 points)

Put all your source code files in one directory and zip it into a zip file.

New requirement about program organization: divide your programs into multiple .h and .c files. Ideally, each task has a pair of .h file and .c file. The main

program (e.g. prot.c) will call the functions defined in other .h and .c files. Create a make file to compile all the files into one program.

The make file is worth 10 points.

The organization of program is worth 20 points.

All the functions need to have comments to explain inputs, outputs, and pre-condition and post-condition. These comments are worth 20 points.

Submission

Submit your report and the zipped source code file to the TA's email box.

The project is due on May 14, 2010