CURRICULUM VITAE

Name:	GURPARTAP SINGH
Date of Birth:	13. August 1986
Nationality:	Indian.
Address:	Madlamerks Veien 6, HO 232, 4041 Hafsfjord, Stavanger.
Email:	g.singh@stud.uis.no
Moblie:	0047 40385279
Telephone:	0047 51728232

PROFILE

- Full command on my subjects and quite innovative in the research projects
- Good knowledge of various techniques and procedures of wet lab as well as IT lab
- Ability to perform individually as well as in group
- Self-motivated, communicative and always ready to meet new challenges
- Organized, highly-trained individual with exceptional follow-through and a comprehensive grounding in biological as well as computational sciences
- Well confident, enthusiastic individual with a reliable sense of humour along with a keen eye for detail

SYNOPSIS

After a solid foundation in theory during my graduation in biotechnology and post graduation in bioinformatics, a brief stint as a project student at the Indian Institute of Sciences has developed my computational skills. I am experienced with a number of software packages (commercial and free), databases and laboratory techniques such as electrophoresis, chromatography, spectrophoresis, etc. I understand that research work requires discipline and perseverance to repeat tasks carefully. I am willing, to perform basic tasks and move on to solve complex problems, to learn more and adapt to new environments quickly. A strong academic record in biotechnology and computational biology along with good communication skills, team work spirit, and work efficiency further demonstrates my motivation, as well as an ability to understand and support your research goals.

EDUCATION

2007 – 2009 (2 years)
Guru Nanak Dev University, India
Department of Biotechnology
Maters of Science in Bioinformatics
Passed with aggregate of 71%

2004 – 2007 (3 years)
Guru Nanak Dev University, India
Department of Biotechnology
Bachelors of Science in Biotechnology
Passed with aggregate of 74%

PROJECTS, WORKSHOPS AND TRAININGS

(Jan – June) 2009
Masters thesis in Bioinformatics
Indian Institute of Science, Bangalore
Molecular Biophysics Unit
Prof. M.R.N Murthy, Chairman
Topic – Determination of Proteins Structures by Plotting

April 2007
Bachelors Project Report in Biotechnology
Guru Nanak Dev University, India
Department of Biotechnology
Mrs Rupinder Jit Kaur, Lecturer
Topic – Superantigens

February 2007
Workshops Attended at

- National Bureau of Animal Genetic Resources (NBARC), Karnal
- Directorate of Wheat Research (DWR), Karnal
- National Dairy Research Institute (NDRC), Karnal

(May-June) 2006
Summer Training
Verka Milk Plant, Mohali

LANGUAGES

English
Taken ILETS on 25. October 2008
Scored 7 Bands

COMPUTER SKILLS

Operating systems	Windows
Scripting Language	C, C++, Perl
Web Tools	HTML, DHTML
Database Management	SQL, MY SQL
Application package	Visual Basics

BIOINFORMATICS SKILLS

Rasmol, Swiss pdb, Pymol, Cn3d, Clustal W/X/V, BLAST2, VISTA and Database knowledge

BIOTECHNOLOGY SKILLS

DNA isolation and purification, Chromatography, Mass spectrometry (ESI, MALDI TOF), Cloning, PCR, Gel Electrophoresis, Restriction digestion, Blotting techniques, Mammalian Cell culture, Tissue culture techniques etc

RESEARCH INTEREST

- Proteomics
- Comparative Genomics (genome evolution)
- Molecular System Biology and wet lab work
- Computational Neurosciences.

DETERMINATION OF PROTEIN STRUCTURES BY PLOTTING (Master's Thesis)

Abstract

The motive of my project work on this topic was to find a simple method that can give us the overall information about the protein structures in relatively short time. As we are familiar with the two methods of protein structure determination i.e. X-ray and NMR, these are quite long procedures. I made use of bioinformatics to get to the new method. I had written computer programs in C++ that operates over the PDB (Protein Data Bank) files to do various calculations and then I generated polar plots, using sigma plot, for the different proteins. The results of this method were verified against those of other programs like pymol and were found to be accurate. This method gives us the results in few seconds. This work was totally new and any kind of this work has not published anywhere yet.

Protein structures are determined by one of the two methods: X-ray crystallography and NMR. These methods are very powerful ab-initio methods for protein structure determination. Those proteins whose structures have been solved by any of the above methods are deposited in the Protein Data Bank (PDB).

To study these structures various visualization tools are available. These tools display three dimensional structures in 3-D space and provide an efficient environment for studying them. Some of these tools are Rasmol, Pymol, Jmol, SPDB viewer etc.

The main motive of my project was to find a simple method which can provide us the overall information about the protein structures and folding in relatively short time. To develop such a method for studying protein structures I used 'Graph methodology'. I wrote appropriate computer program to generate graphs that reveal various aspects of protein structure. Programs written in C++ helped me to do various calculations over the protein sequences.

To start with, I used those proteins whose structures are already known and present in PDB. I started with the small proteins choosing single domain proteins. The proteins were selected from the different classes of proteins viz alpha, beta, alpha + beta and alpha/beta. Six proteins from each of these classes were studied.

The first step in this method is to find the centroid of a polypeptide chain. Centroid represents a centre point of the protein around which atoms of residues of a polypeptide are distributed.

To find centroid of a protein we need all the three coordinates, i.e. X, Y and Z, of all C-alpha atoms of all residues. Only C-alpha atoms are selected because it is bearing the side chain and the most important atom of protein backbone. A program in C++ is written to find the centroid. This program is able to calculate centroid of any protein.

Formula to calculate centroid is:

Centroid = $\sum$ X coordinates/ N, $\sum$ Y coordinates/ N, $\sum$ Z coordinates/ N

Where: $\sum$ X, $\sum$ Y and $\sum$ Z are the sum of all the X, Y and Z coordinates of all the C-alpha atoms, respectively.

'N' is the total number of C-alpha atoms in the protein.

After this, we have to find distance from the centroid to the C-alpha atom for all the residues. For this also a program in C++ is written that calculates this distance for all the residues.

Formula for finding the Centroid to C-alpha distance:

Distance= $\sqrt{(C_x - X_i)^2 + (C_y - Y_i)^2 + (C_z - Z_i)^2}$

Where: C_x, C_y and C_z are the X, Y and Z coordinates of Centroid, respectively.

X_i, Y_i, and Z_i are the X, Y and Z coordinates of C-alpha atom.

Then we have to perform certain other calculations for finding the Mean, Deviation, Square deviation, Mean square deviation and Root mean square deviation (RMSD). These calculations are also done by making use of a program.

Formulae for calculating mean, deviations, square deviations, mean square deviations, RMSD:
Mean = (Sum of distances) /N;
Deviation = Distance [i] – mean;
Square deviations = $(\text{Deviation})^2$;
Mean square deviation = (Sum square deviations)/N;
Root mean square deviation (RMSD) = √ Mean Square Deviation;
Next we have to generate plots for the proteins. These plots provide us the required information about the protein structure. They are of great importance to us in this study. The information provided by the graphs includes:

- Structural components of the protein, i.e. whether it contains helices or sheets or both.
- If there occurs any gene duplication in the protein sequence.
- Class of a protein, i.e. whether it belongs to
    1. Alpha protein or
    2. Beta protein or
    3. Alpha + beta protein or
    4. Alpha / beta protein.
- Position of amino terminal group and carboxy terminal group.
- Number of domains the protein contains, i.e. whether the protein is single domain or multidomain.
- Position of structures other than helices and sheets which includes, folds, loops etc.

On the plots we can mark the atoms which form helices and sheets. The information regarding position of loops and helices can be taken from a PDB file. We have to mark the sector on the circumference of the plot that corresponds to every helix and sheet depending upon the number and position of atoms giving rise to that particular structure. This procedure is done for all the protein. Besides this we can highlight the mean distance of each helix and sheet on the plot. This will give us more significant picture about the position of these structures in the protein. For finding out the mean of helices and sheets another program is developed.
This method of studying the structures of proteins is based on some facts which include:

- Helices are made of hydrophilic residues and sheets are made up of hydrophobic residues.
- Hydrophilic residues are present on the outer surface of a protein and hydrophobic residues are concentrated at the core of the protein.
- Loops are present at the surface regions of the proteins.

The method takes into account these conditions and results are studied on the basis of these facts.