International Journal of Basic and Applied Chemical Sciences ISSN: 2277-2073 (Online)
An Online International Journal Available at http://www.cibtech.org/jcs.htm
2013 Vol. 3 (1) January-March, pp.1-6/Iliyas and Ansari
Research Article

A COMPARATIVE STUDY OF DIFFERENT PROPERTIES PROVIDED BY PROTEIN STRUCTURE VISUALIZATION TOOLS

[*]Sayyed Iliyas and Shaheda N. Ansari
**AKIS, Poona College of Arts, Science and Commerce,Camp, Pune-411001, India*
**Author for Correspondence*

ABSTRACT
Using information visualization tool, researchers can see experimental results more clearly than by simply viewing raw numbers. A molecular graphics visualization tool view the structure encoded by atomic coordinate PDB files from various perspectives. Without a proper tool, the PDB file will be read as a text file that lists each atom and its numerical coordinates in 3-D space. Thus researchers need tools that are capable of loading and displaying huge amount of data. Many tools have been developed to visualize a protein whose structure has been known. During present course of investigation a comparative study of seven commonly used freely available protein structure visualization tools viz. RasMol, Chime, Protein Explorer, Swiss-Pdb Viewer, WebMol, MOLMOL and Cn3D were made based on different properties such as operating systems in which software can work, program type, language used to implement graphics; can softwares print the structure, opens multiple files at a time, support command line, display contents of PDB file, display hetero atoms, identify clicked atom and residue containing clicked atom in the structure; and different coloring styles supported by the softwares; that will help the researchers to select the appropriate tool in their study.

Key Words: *Fold, Amino Acid, Secondary Structure, Domain etc*

INTRODUCTION
Proteins are essential to cell structure and cell function. They are intimately connected with all phases of chemical and physical activity that constitute the life of the cell. Each protein has one folded shape, and consistently folds into it, usually in less than a second. That complicated folded shape dictates how the protein works, and also how it interacts with other entities. Wetlaufer defined domains as stable units of protein structure that could fold autonomously. In the past domains have been described as units of: compact structure (Richardson, 1981); function and evolution (Bork, 1991); folding (Wetlaufer, 1973). Each definition is valid and will often overlap, i.e. a compact structural domain that is found amongst diverse proteins is likely to fold independently within its structural environment. Nature often brings several domains together to form multidomain and multifunctional proteins with a vast number of possibilities (Chothia, 1992).

A gene in the DNA of living cell codes the specific sequence of amino acids that make up each protein. A change in just one amino acid can change the structure and function of a protein. One of the major goals of bioinformatics is to understand the relationship between amino acid sequence and three-dimensional structure in proteins. If this relationship is known, it can be used to predict the protein structure from the amino acid sequence (Rastogi *et al.*, 2006). Based on the structure of amino acids and peptides and planar nature of the peptide bond the first elements of secondary structure, the alpha helix (α-helix) and the beta sheet (β-sheet) were suggested in 1951 by Pauling, Corey and Branson.

The arrangement of amino acids along the chain determines the structure and chemical properties of the protein. The structural and chemical relatedness of the R-groups allows classification of the twenty amino acids into chemical groups. Amino acids can be classified according to optical activity (the ability to polarize light), acidity and basicity, polarity and nonpolarity, or hydrophilicity (water-loving) and hydrophobicity (water-fearing). These categories offer clues to the function and reactivity of the amino acids in proteins (Jain *et al.*, 2006).

The most important factor governing the folding of a protein into 3D structure is the distribution of polar and non-polar side chains (Cordes *et al*., 1996). The hydrophobic effect (Silverman, 2001) plays a crucial role in protein stability and folding. The interaction of proteins with water plays a large and essential role in almost all aspects of protein folding and function (Mattos, 2002). Heat, acid, or other conditions can disturb proteins, causing them to uncoil or lose their shape and impairing their ability to function. Although certain amino acid sequences can be identified as more likely to form a particular conformation, it is still not possible to completely predict how a protein will fold based on its amino acid sequence alone, and this is an active area of biochemical research.

Protein visualization has become an important research topic, especially in light of the accomplishment of the Human Genome Project (Burley *et al.*, 1999). The ability to visualize the 3D structure of proteins is critical in many areas like, drug design, protein modeling. This is because that the 3D structure of a protein determines its interaction with other molecules, hence its function, and the relation of the protein to other known proteins. Modern techniques of drug development make extensive use of computer-aided visualization of molecular properties. Reliable 3D atomic coordinates of molecules are essential for the success of structure-based, rational drug design projects (Bohne *et al*., 2000). There are many well established ways of visualizing the 3D protein structures. Each way of visualization highlights a different aspect of the protein molecule (Shirky, 2000). Growing number of new structure data in Protein Data Bank open new ways for collaboration, thus emphasizes the need for visualization tools that are portable. Moreover, studying the interaction between protein molecules may also require visualizing huge number of atoms, thus researchers also need tools that are capable of loading and displaying this huge amount of data (Can *et al.*, 2003). PDB (www.rcsb.org/pdb/) is the main primary database for 3-D structures of macromolecules. The PDB entries contain the atomic coordinates, and some structural parameters connected with the atoms, or computed from the structures (secondary structure). Those who want to look at one of these 3D datasets need additional software to visualize the structure file that is stored in the database. Many tools have been developed to visualize a protein whose structure has been known. Some of these tools are: RasMol (Sayle and Milner-White, 1995), Chime (MDL Inftormation Systems, Inc.), Protein Explorer (Martz, 2002), Swiss-PDB viewer (Kaplan and Littlejohn, 2001), WebMol (Walther, 1997), MOLMOL (Koradi *et al.*, 1996), and Cn3D (Wang *et al*., 2000).

MATERIALS AND METHODS

For visualizing the 3-D structure of proteins seven free structure visualization tools were downloaded:
RasMol (http://www.umass.edu/microbio/Rasmol/),
Chime (http://www.mdl.com/products/framework/chime/)
Protein Explorer (http://www.proteinexplorer.org/)
Swiss-PDB viewer (http://www.expasy.org/spdbv/)
WebMol (http://www.cmpharm.ucsf.edu/~walther/webmol/)
MOLMOL (http://www.mol.biol.ethz.ch/wuthrich/software/molmol/)
Cn3D **(**http://www.ncbi.nlm.nih.gov/Structure/CN3D/)
Structure data files were downloaded from the Protein Data Bank (http://www.rcsb.org/pdb/) and NCBI (http://www.ncbi.nlm.nih.gov/Structure/ MMDB/mmdb.shtml).
First the study was performed for all the seven softwares to find out the operating systems in which the software works, program type and the language used to develop the software to implement graphics.
One of the download protein structure file (PDB ID: 1CRN) was loaded in each of the seven softwares one by one and softwares were observed for the following properties:
The study was carried out to see the softwares provide an option to print the structure, multiple structure files being loaded at a time or not, command-line interface support, whether softwares display the actual contents of PDB file and hetero atoms present in protein, and whether softwares identify clicked atom and residue containing the clicked atom in the structure.

The coloring styles allows different objects (such as atoms, bonds and ribbon segments) to be given a specified color. The coloring styles viz. monochrome. CPK, amino acid properties, group, chain, charge, temperature, structure, and domain supported by the softwares were observed.

RESULTS AND DISCUSSION

The results summarized in Table 1 shows that RasMol can be worked on Macintosh, Windows, UNIX and VMS, Chime on Macintosh, Windows, and Irix; Protein Explorer on Macintosh and Windows; Swiss-PDB Viewer on Macintosh, Windows, Linux and Irix; WebMol can be worked on any operating system; MOLMOL on Windows, UNIX and Irix; Cn3D on Macintosh, Windows and Unix. RasMol, Swiss-PDB Viewer and MOLMOL are stand alone programs, Chime and Protein Explorer works as plug-in for browser. WebMol can be worked as an applet or stand alone program and Cn3D can be worked as browser plug-in or stand alone program. RasMol, Chime, Protein Explorer, Swiss-PDB Viewer, MOLMOL and Cn3D were implemented using OpenGL API where as WebMol is developed using Java and main advantage of Java are its compatibility across different systems/platforms and having the ability to be run remotely through web browsers.

Table 1: Operating Systems, Program type and Language used for Graphics for protein structure visualization tools

Software	Operating Systems	Program type	Language used for Graphics
RasMol	Macintosh Windows Unix VMS	Stand alone Program	OpenGL
Chime	Macintosh Windows Irix	Browser plug-in	OpenGL
Protein Explorer	Macintosh Windows	Browser plug-in	OpenGL
Swiss-Pdb Viewer	Macintosh Windows Linux Irix	Stand alone Program	OpenGL
WebMol	All	Applet or Stand alone Program	Java
MOLMOL	Windows Unix Irix	Stand alone Program	OpenGL
Cn3D	Macintosh Windows Unix	Browser plug-in or Stand alone Program	OpenGL

Results summarized in Table 2 indicate that RasMol, Chime, Protein Explorer and WebMol contain an option to print the structure where as Swiss-PDB Viewer, MOLMOL and Cn3D do not have print option. Multiple structure files can be opened at a time in Swiss-PDB Viewer and MOLMOL where as other software can open only one file at a time. RasMol, Protein Explorer and MOLMOL also supports command-line interface which is not provided by Chime, Swiss-PDB Viewer WebMol and Cn3D. Protein Explorer, Swiss-PDB Viewer and WebMol also displays the actual contents of the PDB file where as rest of the softwares can not display. Hetero atoms can be displayed in all the seven softwares. WebMol do

not display the clicked atom but only can identify residue containing the clicked atom in the structure where as other softwares can identify the clicked atom as well as residue containing the clicked atom in the structure.

Table 2: Different properties of protein structure visualization tools

Software	Structure can be Printed	Multiple files can be opened	Command Line	Display Contents of PDB File	Display Hetero atoms	Identify clicked atom in structure	Identify residue containing clicked atom in structure
RasMol	+	-	+	-	+	+	+
Chime	+	-	-	-	+	+	+
Protein Explorer	+	-	+	+	+	+	+
Swiss-Pdb Viewer	-	+	-	+	+	+	+
WebMol	+	-	-	+	+	-	+
MOLMOL	-	+	+	-	+	+	+
Cn3D	-	-	-	-	+	+	+

+ = YES, - = NO

Table 3: Coloring Styles used for protein structure visualization tools

Software	Mo-chrome	CPK	Ami-Acid properties	Group	Chain	Charge	Temperature	Structure	Domain
RasMol	+	+	+	+	+	+	+	+	-
Chime	+	+	+	+	+	+	+	+	-
Protein Explorer	+	+	+	+	+	+	+	+	-
Swiss-Pdb Viewer	+	+	+	-	+	-	+	+	-
WebMol	-	+	-	-	-	-	+	+	-
MOLMOL	+	-	-	-	-	-	-	-	-
Cn3D	+	+	-	+	+	+	+	+	+

+ = YES, - = NO

The study of coloring styles supported by the softwares summarized in Table 3 shows that RasMol, Chime and Protein Explorer have all the color styles except domain. Swiss-PDB Viewer does not have group and domain color schemes. WebMol supports CPK, Temperature and structure color styles.

MOLMOL have only monochrome color style. Cn3D supports all the coloring styles except color style according to amino acids properties. The domain color style is provided by only Cn3D.

Conclusion

Information visualization techniques have become an attractive option for the field of bioinformatics; researchers can see experimental results more clearly than by simply viewing raw numbers due to the difficulties inherent in understanding large quantities of data. One of the important sub-disciplines within bioinformatics is the development and implementation of tools that enable saving, retrieving and analysis of Biological data and extracting the meaningful information from the mass of molecular biology or biological databases and to carry out sequence or structural analysis. The visualization tool should be developed using Java 3D API. The main advantages of Java are its compatibility across different systems/platforms and having the ability to be run remotely through web browsers. Java 3D also promises high performance; because it is capable of taking advantage of the graphics hardware in a system. The visualization too should also display the contents of the PDB file. It should have command window to support its functionality but all the commands should also be provided by pull-down menus and/or buttons so that novice users can use the tool easily without expending time to study the syntax of the commands.

It should display 3D molecular structure in different coloring styles. Monochrome in which each structure object - one PDB entry - is assigned a single color. CPK color scheme color `atom' objects by the atom (element) type. Amino acid properties colors amino acids according to traditional amino acid properties, the purpose of coloring is to identify amino acids in an unusual or surprising environment. Group color scheme color codes residues by their position in a macromolecular chain; each chain is drawn as a smooth spectrum from blue through green, yellow and orange to red. Chain color scheme assigns a different color to each chain of protein; it is particularly useful for distinguishing the parts of multimeric structure. Charge color scheme color codes each atom according to the charge value stored in the input file (or beta factor field of PDB files), high values are colored in blue (positive) and lower values colored in red (negative). Temperature color scheme color codes each atom according to the anisotropic temperature (beta) value stored in the PDB file, typically this gives a measure of the mobility/uncertainty of a given atom's position, high values are colored in warmer (red) colors and lower values in colder (blue) colors. Structure color scheme colors the molecule by protein secondary structure. Domain color scheme assigns a different color to each domain in a chain.

REFERENCES

Bohne A, Lang E and Von der Lieth CW (2000). Molecular visualization programs on the web. *Drugs Fut* **25**(5) 489.

Bork P (1991). Shuffled domains in extracellular proteins. *FEBS Letter* **286** 47-54.

Burley SK, Almo SC, Bonanno JB, Capel M, Chance MR, Gaasterland T, Lin DW, Sali A, Studier FW and Swaminathan S (1999). Structural genomics: beyond the human genome project. *Nature Genetics* **23** 151-157.

Can T, Wang Y, Wang YF and Su J (2003). FPV: fast protein visualization using Java 3D. *Bioinformatics* **19**(8) 913-922.

Chothia C (1992). Proteins. One thousand families for the molecular biologist. *Nature* **357** 543-544.

Cordes MH, Davidson AR and Sauer RT (1996). Sequence space, folding and protein design. *Current Opinion in Structural Biology* **6** 3-10.

Jain JL, Jain S and Jain N (2006). *Fundamentals of Biochemistry*. New Delhi S Chand and Company Ltd.

Kaplan W and Little john TG (2001). Swiss-PDB Viewer (Deep View). *Briefings in Bioinformatics* **2** 195–197.

Koradi R, Billeter M and Wuthrich K (1996). MOLMOL: a program for display and analysis of macromolecular structures. *Journal of Molecular Graphics* **14** 51-55.

Martz E (2002). Protein Explorer: easy yet powerful macromolecular visualization. *Trends Biochemical Sciences* **27** 107-109.
Mattos C (2002). Protein-water interactions in a dynamic world. *Trends Biochemical Sciences* **27** 203-208.
Pauling L, Corey RB and Branson HR (1951). The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci USA* **37**(4) 205-211.
Rastogi SC, Mendiratta N and Rastogi P (2006). *Bioinformatics Methods and Applications Genomics, Proteomics and Drug Discovery Second Edition.* New Delhi. Prentice Hall of India private Ltd.
Richardson JS (1981). The anatomy and taxonomy of protein structure. *Advances in Protein Chemistry* **34** 167-339.
Sayle RA and Milner-White EJ (1995). RASMOL: biomolecular graphics for all. *Trends Biochemical Sciences* **20** 374-376.
Shirky C (2000). SevenWays of Looking at a Protein. *FEED Magazine.*
Silverman BD (2001). Hydrophobic moments of protein structures: spatially profiling the distribution. *Proc Natl Acad Sci USA* **98** 4996-5001.
Walther D (**1997)**. WebMol – a Java-based PDB viewer. *Trends Biochemical Sciences* **22** 274-275.
Wang Y, Geer LY, Chappey C, Kans JA and Bryant SH (2000). Cn3D: sequence and structure views for Entrez. *Trends Biochemical Sciences*. **25** 300-302.
Wetlaufer DB (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci USA* **70** 697-701.