Bioscience Discovery, 02 (2):222 -226, June 2011 ISSN: 2229-3469 (Print)

A COMPARATIVE STUDY OF PROTEIN STRUCTURE VISUALIZATION TOOLS FOR VARIOUS DISPLAY CAPABILITIES

Shaheda N Ansari and Sayyed Iliyas[1]
Department of Computer Science
[1]Department of Botany
AKIS, Poona College of Arts, Science and Commerce, Camp, Pune-411001(MS)
Email: sayyed_iliyas@yahoo.com
Received: 25 March 2011, 28 April 2011

ABSTRACT

A molecular graphics visualization tool is required to view the structure that is encoded by atomic coordinate PDB files and to be able to manipulate the images to view the molecule 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 display capabilities such as display of sequence, finding pattern in sequence and structure, highlighting selected amino acids in the sequence in structure, highlighting selected amino acids in the structure in sequence, hydrogen bonds, hydrogen bond distance, disulfide bond, Ramchandran plot, different types of surfaces and different rendering styles of the protein structure; that will help the researchers to select the appropriate tool in their study.

Key words: Protein structure, motif etc.

INTRODUCTION

Proteins occupy a central position in the architecture and functioning of living matter. Depending upon their physical and chemical structure and location inside the cell, different proteins perform various functions. Some proteins serve as important structural elements of the body, for example as hair, wool, and collagen, and important constituent of connective tissue; other proteins may be enzymes, hormones or oxygen-carriers. Still other proteins participate in muscular contraction, and some are associated with the genes, the hereditary factors. They are, therefore, essential to cell structure and cell function (Jain *et al.* 2006).

Proteins are essential to cell structure and cell functions. Each protein has one folded shape, and consistently folds into it, usually in less than a second. The ability to visualize the 3D structures of protein is critical in many areas like drug design, protein modeling. This is because 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. 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 (Rastogi *et al.* 2006). As the nascent protein emerges from the ribosome, it rapidly folds to an energy minimum, a specific tertiary structure referred to as the proteins native state or native fold (Anfinsen 1973; Privalov and Gill 1988). Diseases such as Alzheimer's disease (Zhang *et al.* 2004), cystic fibrosis, BSE (Mad Cow disease), an inherited form of emphysema, and even many cancers are believed to result from protein misfolding. 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 2006). Analyzing the geometric structure of protein is fundamental for the study of a protein folding, docking, and interactions between proteins. One of the important geometric analyses is computing the molecular surface of protein (Ryu *et al.* 2005). Three definitions of surface are used for molecular modeling as defined by (Richards 1977). The van der Waals (VW) surface is the external surface of atoms, each represented by a spherical ball of van der Waals radius. A molecular surface calculate the volume of protein, electrostatic potential, interface surfaces in protein-protein, protein-ligand, etc (Connolly 1983a, 1983b). A solvent accessible surface is first defined by Lee and Richards in 1971 to compute the free space that a probe, a sphere enclosing a small molecule, can move around during the small molecule interacts with a protein (Lee and Richards 1971).

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 determined by X-ray crystallography and NMR. 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:

1. RasMol (http://www.umass.edu/microbio/Rasmol/),
2. Chime (http://www.mdl.com/products/framework/chime/)
3. Protein Explorer (http://www.proteinexplorer.org/)
4. Swiss-PDB viewer (http://www.expasy.org/spdbv/)
5. **WebMol** (http://www.cmpharm.ucsf.edu/~walther/webmol/)
6. MOLMOL (http://www.mol.biol.ethz.ch/wuthrich/software/molmol/)
7. 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). One of the download protein structure file (PDB ID: 1CRN) was loaded in each of the seven software's one by one and software's were observed for the following properties:

Observations were made to see whether software's display sequence of the loaded protein, locates the given pattern (motif) i.e. specific fragment of amino acids in the sequence and structure, whether selecting the amino acids in the sequence highlights them in the structure or not and whether selecting amino acids in the structure highlights them in the sequence or not. Different display capabilities in the software's were studied viz. Hydrogen bonds, Hydrogen bond distance, Ramachadran plot and disulfide bonds. Different types of surfaces viz. Van der Waals surface, Solvent-accessible surface, Solvent-excluded surface (Molecular surface) displayed by the softwares were observed.

The study was performed to see the different rendering styles supported by the softwares viz. Wireframe, Stick, Ball and Stick, Spacefill, Backbone, Trace, Ribbons, Strands, and Cartoons.

RESULTS AND DISCUSSION

Results summarized in Table 1 shows that all the studied softwares except chime display the sequence. Protein Explorer, Swiss-PDB viewer, WebMol and Cn3D have separate window for sequence. RasMol and MOLMOL have no separate sequence window. RasMol shows sequence in command window by executing command "show sequence" and MOLMOL shows sequence when "ribbon" button is pressed to display structure in ribbon form by calculating the secondary structure. Given pattern (motif) in the sequence and structure can be found in Swiss-PDB Viewer, WebMol and Cn3D. If the given pattern occurs multiple times then Swiss-PDB viewer highlights the found pattern one by one i.e. one at a time where as WebMol and Cn3D highlights all at a time. Protein Explorer finds only sequence motif. In Protein Explorer, Swiss-PDB Viewer, WebMol and Cn3D the selected part of a sequence were highlighted in the structure. Only in Swiss-PDB viewer and Cn3D selected part of a structure were highlighted in the sequence.

Results summarized in Table 2 shows that hydrogen bonds can not be seen using MOLMOL and Cn3D where as can be displayed by other five software's and only Swiss-PDB Viewer can also display the hydrogen bond distance. Ramachandran plot can be displayed in

Swiss-PDB Viewer, WebMol and MOLMOL but not in other four software's. Disulphide bridges can be seen by all the software's except MOLMOL.

The study of different types of surfaces displayed by the softwares summarized in Table 3 shows that RasMol and WebMol display Van der Waals and Solvent-accessible surfaces. Chime, Protein Explorer,

Table 1: Display Sequence, Find pattern and Correlation between Structure and Sequence for Protein structure visualization tools.

Software	Display Sequence	Find pattern (motif) in sequence and structure	Selecting ami- acid in the sequence highlight it in the structure	Selecting ami-acid in the structure highlight it in the sequence
RasMol	+	-	-	-
Chime	-	-	-	-
Protein Explorer	+	-	+	-
Swiss-Pdb Viewer	+	+	+	+
WebMol	+	+	+	-
MOLMOL	+	-	-	-
Cn3D	+	+	+	+

+ = YES, - = NO

Table 2: Different displays used for protein structure visualization tools.

Software	Display			
	Hydrogen bonds	Hydrogen bond distance	Ramachandran Plot	Disulfide Bonds
RasMol	+	-	-	+
Chime	+	-	-	+
Protein Explorer	+	-	-	+
Swiss-Pdb Viewer	+	+	+	+
WebMol	+	-	+	+
MOLMOL	-	-	+	-
Cn3D	-	-	-	+

+ = YES, - = NO

Table 3: Display different type of surfaces for protein structure visualization tools.

Software	Van der Waals surface	Solvent-accessible surface	Molecular surface
RasMol	+	+	-
Chime	+	+	+
Protein Explorer	+	+	+
Swiss-Pdb Viewer	+	+	+
WebMol	+	+	-
MOLMOL	+	+	+
Cn3D	-	-	-

+ = YES, - = NO

Swiss-PDB Viewer and MOLMOL displays Van der Waals, Solvent-accessible and Molecular surfaces where as Cn3D do not display any type of surface.

The study of rendering styles provided by software's summarized in Table 4 shows that RasMol and Protein Explorer supports all the rendering styles that were observed viz. wireframe, stick, ball and stick, spacefill, backbone, trace, ribbon, strands and cartoons. Trace representation is not provided by Chime. Swiss-PDB viewer do not have ball and stick and trace representations. WebMol provides only wireframe and backbone representations. MOLMOL do not support trace and strands representations. And Cn3D do not have stick, ribbon, strands and cartoons representations.

Table 4: Rendering Styles used for protein structure visualization tools.

Software	Wireframe	Stick	Ball and Stick	Spacefill	Backbone	Trace	Ribbons	Strands	Cartoons
RasMol	+	+	+	+	+	+	+	+	+
Chime	+	+	+	+	+	-	+	+	+
Protein Explorer	+	+	+	+	+	+	+	+	+
Swiss-Pdb Viewer	+	+	-	+	+	-	+	+	+
WebMol	+	-	-	-	+	-	-	-	-
MOLMOL	+	+	+	+	+	-	+	-	+
Cn3D	+	-	+	+	+	+	-	-	-

+ = YES, - = NO

CONCLUSION

As the genome sequencing projects proceed, scientists have gained access to tremendous amounts of biological information. 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. The need is to develop methods and algorithms to model, simulate and analyze three-dimensional protein structures and their molecular properties, to apply these techniques to understand biological processes at a molecular level, and to make these methods available to the general biomedical research community.

Protein structure visualization software should locate the given pattern in the sequence and structure. Locating the given pattern can be used to look for specific sites such as active sites, glycosylation sites, etc. This might also be useful to compare the conformation of a specific motif in different structures to draw conclusions about its function. Selecting residues in the structure should also be highlighted in the sequence window, and selecting resides in the sequence window should also be highlighted in the structure window like in case of Cn3D and Swiss-PDB viewer to see the sequence or structure of the selected portion respectively.

Structure visualization tools should display hydrogen bonds; hydrogen bond distance and Ramachandran plot because this information is useful in assessing the protein's secondary structure. It should display disulfide bonds. Disulfide bonds might join two peptide chains (an inter-chain disulfide bond) or two regions of the same chain (an intra-chain disulfide bridge). This bond is very important to the folding, structure and function of proteins. The presence of disulfide bond helps to maintain the tertiary structure of the protein. Softwares should also display molecular surface of protein because function of protein depends on its molecular surface.

Softwares should display structure of loaded protein molecule in the different rendering styles. Each representation of a protein molecule highlights a different aspect of the structure and have advantages and disadvantages compared to each other. Like Wireframe represents the bonding information in the protein molecule. Spacefill models each atom using its van der Waals radius, so that the viewer gets an idea of the relative sizes of the atoms making up the protein molecule and it is useful in visualizing the volume a protein molecule occupies, it gives an overall view of the molecule and thus provides a good view of the tertiary structure but it lacks information about how amino acids are connected to each other, i.e. how the chain is formed. The cartoons, ribbons, and strands display options are useful for viewing protein secondary structure (alpha helices and beta pleated sheets).

LITERATURE CITED

Anfinsen C B. 1973. Principles that govern the folding of protein chains. *Science* **181**:223-230.

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

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

Connolly ML. 1983a. Analytical molecular surface calculation. *Journal of Applied Crystallography* **16**:548-558.

Connolly ML. 1983b. Solvent-accessible surfaces of proteins and nucleic acids', *Science*, **221**: 709-713.

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

Kaplan W and Littlejohn TG. 2001. Swiss-PDB Viewer (Deep View)', *Briefings in Bioinformatics.* **2**:195–197.

Koradi R, Billeter M and Wüthrich K. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graphics,* **14**:51-55.

Lee B, and Richards FM. 1971. The interpretation of protein structures: estimation of static accessibility. *Journal of Molecular Biology* **55**: 379-400.

Martz E. 2002. Protein Explorer: easy yet powerful macromolecular visualization. *Trends Biochem. Sci.* **27**:107–109.

Privalov PL and Gill SJ. 1988. Stability of protein structure and hydrophobic interaction. *Adv Protein Chem.* **39**:191-234.

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 Limited.

Richards FM. 1977. Areas, volumes, packing, and protein structures. *Annu. Rev. Biophys. Bioeng*, **6**:151-176.

Ryu J, Kim D, Cho Y, Park R and Kim DS. 2005. Computation of Molecular Surface Using Euclidean Voronoi Diagram. *Computer-Aided Design & Applications*, **2** (1-4):439-488.

Sayle RA and Milner-White EJ. 1995.RASMOL: biomolecular graphics for all. *Trends Biochem Sci.*, **20**:374-376.

Shirky C. 2000. SevenWays of Looking at a Protein. *FEED Magazin.*

Walther D. 1997. WebMol – a Java-based PDB viewer. *Trends Biochem Sci.* **22**:274–275.

Wang Y, Geer LY, Chappey C, Kans JA, and Bryant SH. 2000 Cn3D: sequence and structure views for Entrez. *Trends Biochem. Sci.* **25**:300–302.

Zhang Q, Powers ET, Nieva J, Huff ME, Dendle MA, Bieschke J, Glabe CG, Eschenmoser A, Wentworth P, Lerner RA and Kelly JW. 2004. Metabolite-initiated protein misfolding may trigger Alzheimer's disease. *Proc Nat. Acad Sci U S A* **101**: 4752-4757.