X3DPROT: A Tool for Distributed 3D Protein Structure Visualization and Manipulation

Marcos Bonfadini
Universidade Anhembi Morumbi, Diretoria de Pesquisa e Extensão
pbonfa@anhembi.br

Abstract

In recent years Proteomics has gained strength and has also delivered a huge amount of data, which must be interpreted. Those who work on the field welcome any effort to speed up the whole research process. A 3D protein structure determines its interaction with other molecules. On the 3D protein structure visualization aspect, several applications have been used with a certain degree of success, although a few of then have the capability of collaborative visualization and model construction via Internet. The visualization system here presented will allow the examination of a given protein molecule through different visual representations, such as: space fill, backbone, ball-stick and ribbon. Another feature designed is an annotation system that will make possible the sharing of particular observations of various researchers geographically distributed on specific regions of the protein model. These observations would be made on audio, video and text formats.

1. Introduction

The gene sequencing of an organism delivers a huge amount of information that must be interpreted. Some DNA regions encode information about protein fabrication [1]. Proteins are made of amino acids linearly chained together folded into various shapes. The problem of finding theoretically the shape of a protein from its sequence of amino acids is known as the *protein-folding problem* [2], and is one of the hardest tasks in science today. Another way to find the 3D structures of proteins is through experimental methods such as X-Ray Crystallography [3] and Nuclear Magnetic Resonance (NMR), both providing the relative 3D atom positions for the model. Those relative coordinates and several other important information can be deposited in a world-class data bank, the Protein Data Bank (PDB) [4].

The construction of a 3D protein structure model and the understanding of how this model would describe its interactions with other proteins are made necessarily with the aid of visualization softwares, indicating that protein visualization is of extreme importance.

Collaborative software environments would allow a group of distributed researchers to work on a 3D protein structure model simultaneously, providing synchronized views and controlled access through a protein visualization software.

The system here presented is intended to have manipulation and annotation features based on the X3D [5] new standard from W3 consortium using Java3D [6] for visualization.

2. Visualization

From the PDB id of the protein provided by the user the molecule can be visualized where the loader will create a Molecule composed of several Java classes. The 3D model will be generated and represented by a Java3D scene graph.

PDB files are previously obtained from the Protein Data Bank by the user or from local lab files. There are several forms of representation of a protein structure, each one more indicated for particular aspects one wants to observe from the structure: 1. *Cα*: this is a wire-frame representation of the model that shows only the alpha carbons bonded between them. This type of bond does not exist in nature, being just an artifact to visualize the amino acids pathway folding; 2. *Backbone:* using the alpha carbon, carbon and nitrogen atoms from the amino acids in the molecule, the backbone model is created. Those atoms laid on their spatial atomic positions are represented as spheres. The bonds between the atoms from an amino acid and between the peptides are also shown in this model; 3. *Ball and Stick:* This model shows all the existing bonds in the molecule as sticks and all the atoms as equal sized spheres; 4. *Space Fill or van der Waals:* This model gives an overall view of the molecule that provides a good sense of the volume occupied by the

molecule and its tertiary structure. The atoms are represented as spheres which radius is proportional to their van der Waals radius; 5. *Ribbon:* is used to display the secondary structure in the protein, as proposed by Kabsch and Sander [7]. These forms of representation are modeled as Java3D classes.

3. Architecture

X3DPROT is written in Java, and as previously said, requires the Java3D extension to the Java language that adds high-level support for 3D graphics, allowing easily the construction and manipulation of a scene graph description of three-dimensional geometry. It will run as an applet or as a stand alone application on a Java-enabled web browser such as Netscape or Internet Explorer. In the former, run way signing the applets to certify their trusted precedence would circumvent secure restrictions. The customization of a given scene will be made through Python scripts and loaded as a standard X3D scene. Some standalone applications, like Molscript [8], InSight [9], Ribbons [10] and Pymol [11] can generate on the local machine VRML representations of a protein molecule that will be actualized to X3D.

The system is designed to operate as peer-to-peer and as master-slave. As the former multiple users will be allowed to visualize and collaborate on a 3D model, which one requesting this privilege at a time. The master-slave mode is intended for class teaching and remote learning where a single master user is allowed to change the model.

4. Conclusions

X3DPROT is being developed to be used by a single user or in a shared collaborative environment for protein structure modeling. By using a Java-enabled WWW browser, the program runs platform independently, and high-quality 3D molecular graphics can be performed easily in a distributed manner with relatively low software maintenance requirements. Using Java 3D as a graphics engine has also the additional advantage of rapid application development. The use of the X3D new standard for virtual reality puts this scientific visualization tool updated to the specs of the latest streaming or rendering extensions and close tied to XML.

5. References

[1] P. Hieter, M. Boguski, "Functional Genomics: It´s All How You Read It", *Science*, 278, 1997, pp. 601-602.

[2] A. Sali, K.M. Shakhnovich, "How Does a Protein Fold", *Nature*, 369, 1994, pp. 248-251.

[3] N. Chayen, J. Helliwell, "Protein Crystallography: The Human Genome in 3D", *Physics World*, 1998, pp.43-48.

[4] H.M.Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, "The Protein Data Bank", *Nucleic Acids Res*., 28, 2000, pp.235-242.

[5] W3 Consortium, http://www.web3d.org.

[6] Java3D Application Programming Interface, http://java.sun.com/products/java-media/3D/

[7] W. Kabisch, C.Sander, "Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-bonded and Geometrical Feature", *Biopolymers*, 22, 1983, pp.2577-2637.

[8] P.J.Kraulis, "Molscript: A Program to Produce Both Detailed and Schematic plots of Protein Structures", *J.Appl. Crystallogr.*, 24, 1991, pp.946-950.

[9] Accelrys Inc. San Diego, CA, http://www.accelrys.com

[10] M. Carson, "Ribbons 2.0", *J.Appl. Crystallogr.*, 24, 1991, pp.958-961.

[11] W.L. DeLano "The PyMOL Molecular Graphics System", DeLano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org.