Interactive 3D Protein Structure Visualization Using Virtual Reality

Elke Moritz, Joerg Meyer
University of California, Irvine
Biomedical Engineering
644E Engineering Tower
Irvine, CA 92697-2625
{emoritz|jmeyer}@uci.edu

Abstract

Large-scale biomedical data sets of macromolecular structures such as DNA and proteins describe highly complex biomolecular entities which often consist of thousands of atoms and residues in large 3D strands of amino acids. Various types of abstract representations are used to display these data sets, ranging from traditional ball-and-stick models to feature-presenting cartoon models. Despite these abstractions, even the most simplified representations still generate a large amount of geometry data. Useful features are often difficult to recognize in the sheer amount of detail produced by the rendering software.

Comparative visualization of various structures in any type of abstract representation is helpful for understanding the relation between function and structure, and prediction of properties of newly discovered proteins which might have an impact on drug design and better treatment options for diseases.

We present a case study that uses high performance workstations and a Virtual Reality display system to process large amounts of geometry data for real-time 3D exploration, superimposition, and interactive navigational tasks in a virtual reality environment.

1. Introduction

A wide variety of different functions is fulfilled by proteins. Since their function is directly related to their 3D structure, some molecules like the structural protein Collagen are built relatively simple, while other molecules might be highly complex and show special features like binding sites if inspected closely (figure 1).

These features are difficult to recognize using traditional 2D visualizations. Due to the complexity of protein structures it is difficult to visualize the corresponding data sets on standard commodity hardware. Therefore, different levels of abstraction are used to reduce the complexity of the geometric representation to be rendered.

Virtual Reality environments allow a detailed inspection and comparison of related molecular structures and offer a different quality than standard 2D methods that are found in most computer-generated renderings and textbooks.

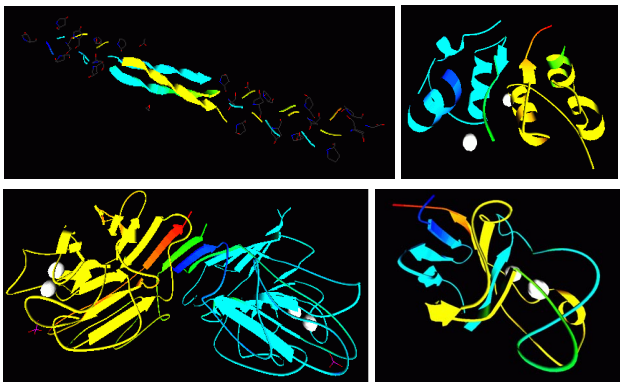

Figure 1. Proteins have different functions: structural support (Collagen, 1BKV [7]), message transport (Insulin, 4INS [2]), carbohydrate binding (Lectin, 2LAL [9]), garbage collection (Superoxid Dismutase, 2SOD [16]).

2. Data Acquisition and Abstraction

Information on complex biomolecular structures is collected in various databases for publicly available macromolecular data sets, most importantly for example the Protein Data Bank (PDB [3]).

Improved distribution and realistic visual representations of large-scale biomolecular data sets are important in biomedical research, for example in the field of drug development. Until now, computer-generated visual models have often been too simplified, making the complex interactions between large protein structures difficult to simulate. Virtual Reality can help to visually verify data through structure comparison and analysis.

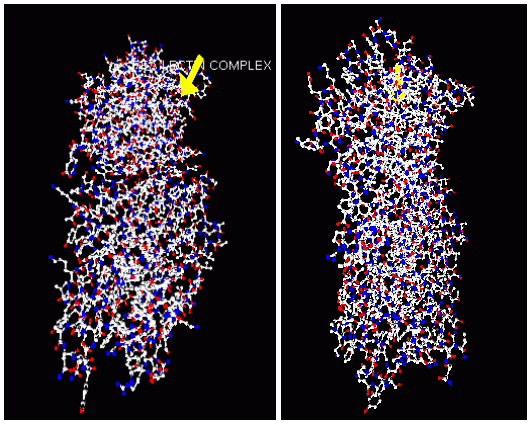

Figure 2. Ball-and-stick models of lectin from (a) peas (1RIN) and (b) lentils (2LAL).

Structures of related molecules from different species often show striking similarities, hinting at the fact that the functions might be nearly identical as well.

A good example is the structure of lectin from peas (1RIN [13]) and lentils (2LAL [9]). To explore the complex details and similarities as shown in figures 2 a) and b), our experiments have shown that interactive comparative visualization is very helpful.

In some cases, a structure can be so complex and rich in detail, that a ball-and-stick model of the whole structure is not of any practical use (figure 3a). Therefore, a cartoon model is chosen as an abstraction of the primary structure. An example can be seen in figure 3b.

Difficulties arise from the protein folding problem which states that the amino acid sequence of a protein determines the 3D structure of the macromolecule due to interatomic interactions, chemical characteristics and structural rules for spacing between atoms and bond lengths and angles.

If the amino acid sequence (the so-called primary structure) is known, predictions of the 3D structure of the molecule are possible. Therefore predictions of local secondary structures or small domains are easier to make than

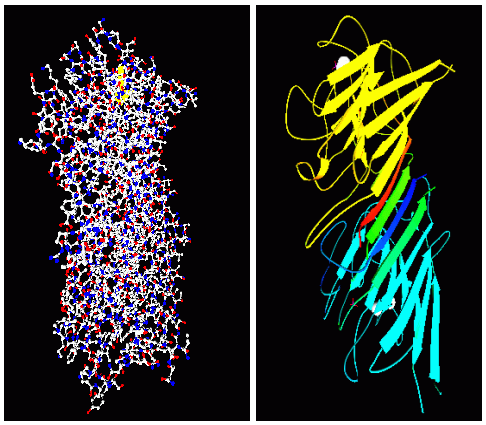

Figure 3. Comparison of (a) the ball-and-stick model and (b) a cartoon representation of lectin from lentils (2LAL).

predictions of the complete tertiary or quaternary structure of the protein (figure 4).

The ball-and-stick model (primary structure) is mostly used for displaying purely chemical aspects, while space-filling models reveal information on shape, size, volume and surface structure.

The cartoon representation however uses simplified symbols such as arrows, helices, and ribbons to represent local (secondary) structures of the object. To further simplify the cartoon model, helices can be substituted by cylinders.

These different levels of abstraction can be used to generate visual models of various complexity (tertiary and quaternary structures).

The amount of geometry data necessary to represent the individual structures is closely related to the level of abstraction.

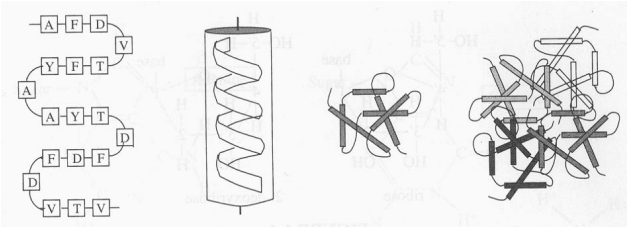

Figure 4. The primary, secondary, tertiary and quaternary structures of proteins [15].

3. Related Work

Molecular structures are usually represented by simplified models in common computer simulations and commercial software applications.

Kinemage [12] was one of the first file formats for representation of molecular structures. Mage is a general vector display program for producing static bitmap output of ”kinemages” (kinetic images). Geometric information from PDB files can be converted into this format.

RasMol [14] and VMD Visual Molecular Dynamics [5], for example, provide standard functionality for the visualization of structural protein data sets.

Visual exploration in virtual environments of rudimentary molecular structures was introduced by [11] and [1].

MolScript [8] converts protein structures as stored in the Protein Data Bank format into a geometric representation. The software package currently supports PostScript™, GIF and other formats. For our work, we have chosen MolScript due to its flexibility and extendability.

Most of these studies are based on the visualization of single, small protein structures, and do not include complex 3D structures or dynamic protein interactions.

With our system, we add interactive 3D navigation and a visual tool for the comparison and inspection of complex 3D structures and their potential for interaction with other proteins.

4. Complex Geometry Visualizations

Common ball-and-stick models comprise of many spheres which have to be transformed into the triangular representation used by OpenGL. This complex conversion is usually done by a library function.

The result is a large amount of geometry information since the triangular representation is even more complex than the ball-and-stick representation of the molecule.

Highly complex biomolecular structures, however, generate ball-and-stick models which are often too complex to investigate, so that the substitution of the overall model with a cartoon model, which uses a ball-and-stick representation potentially only at a designated region of interest, is preferred.

The complexity of the geometry representation of ball-and-stick models is greatly reduced by introducing several levels of abstraction based on the secondary and tertiary structures, enabling the visualization and exploration of various complex models at interactive frame rates.

5. MolVis

Our prototype study [10] enhances the features of the software package MolScript [8] in the following way: It not only features comparative visualization and rendering of multiple structures on commodity graphics hardware, but also the exploration of up to four complex macromolecular structures on surround-screen projection-based virtual environments like the CAVE™ [4]. It also adds interactive navigation using a wand-type controller, electromagnetic head and hand tracking, and stereoscopic vision.

While this allows the collaborative exploration and detailed analysis of the structure and therefore the functions, interactions, and dynamics of multiple biomolecular structures, it also requires the generation of a large amount of geometry data. Several views of the objects must be calculated to accomodate multiple screens, and at the same time real-time interactive navigation, for example in a CAVE™, is enabled, requiring updates of up to eight images (4-sided CAVE with stereo projection) at interactive frame rates.

A high performance rendering engine, an 8-processor SGI Onyx2 Infinite reality system with 250MHz, R10000 processors and 4 Gigabytes of RAM was used, providing four channels for the four screens using two Infinite Reality 2 Graphics Pipelines for each screen pair.

Using multiple levels of abstraction as described in section 2, we have successfully integrated comparative visualization of biomolecular structures in a virtual environment, as shown in figure 5.

A six-degrees-of-freedom navigation interface enables the user to compare complex structures side-by-side, to superimpose them, and to position them so that details that are otherwise occluded are revealed.

This way, useful information about the biomolecular structures can be gained, allowing new insights into functional constraints of structural properties, intermolecular interactions, and phylogenetic relations.

Virtual exploration of ball-and-stick models (see figure 2) of data sets from the Protein Data Bank [3] in a CAVE™ can be seen in figure 5a, while figure 5b shows the investigation of complex cartoon models.

6. Case Study: Pyruvate Kinase

Our experimental setup consisted of a four-sided, room-sized, multi-person, high-resolution, three-dimensional video and audio theater, commonly referred to as CAVE™ (or Cave Automatic Virtual Environment). Each of the four stereo-ready video projectors has a resolution of $1024 * 768$ pixels and a refresh rate of 57Hz.

A test group consisting of six participants, including biologists and computer scientists, was asked to explore up to four macromolecular structures at a time, wearing shutter glasses. The position of one of the users was tracked using an electromagnetic tracking system known as ”Flock of Birds”. The perspective was constantly updated for the tracked user, and only this user was able to interact with the

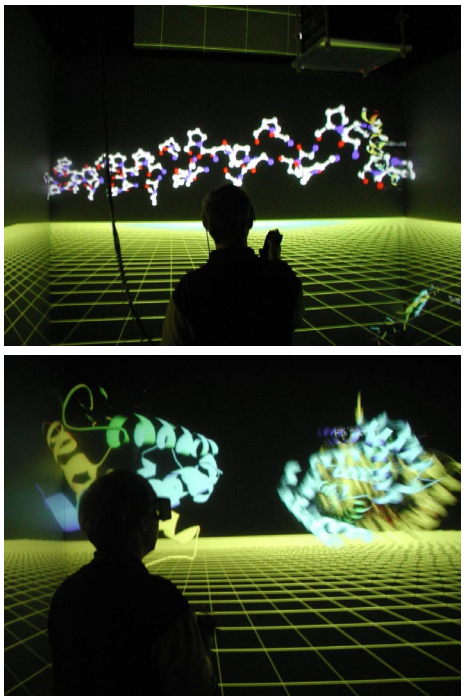

Figure 5. Comparative visualization and exploration of various biomolecular structures in the CAVE™ Laboratory at the Engineering Research Center (ERC, Mississippi State University): (a) ball-and-stick model (top), (b) cartoon model (bottom).

data through a custom wand (a modified Nintendo64 controller) and navigate through the scene. The other users saw a slightly distorted image, but were still able to see the same spatial features as the person who was navigating. The number of users was limited by the size of the projection space.

One of the macromolecular structures we explored in our case study was Pyruvate Kinase (molecule 1A3X [6] in the PDB), an enzyme that can be found in red blood cells. In case of an anaerobic metabolism, this protein helps convert glucose to energy. The PDB data set included coordinate information on 7,472 atoms. The 3D models we generated using our software contained 225,780 polygons in the standard ball-and-stick model and 22,604 polygons in the cartoon representation of the complete tetramer.

The molecule in figure 6 shows a complex 3D structure consisting of two chains. As can be seen in figure 6a, the standard ball-and-stick representation is too complex and overloaded with detail to allow any substantial visual analysis.

For a more comprehensive exploration, a concentration on just one polypeptide chain in the cartoon representation

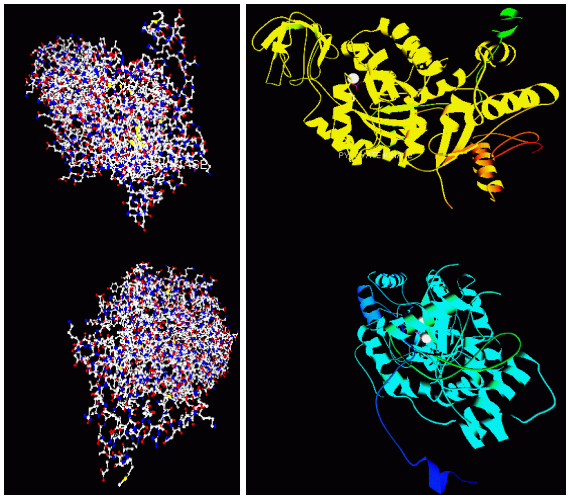

Figure 6. 3D structure visualization of Pyruvate Kinase (1A3X): a) ball-and-stick model (left), b) cartoon model (right).

(figure 6b) exposes more details on how the chain folds into several domains.

As a result of this study, it turned out that particular perspectives or the initial position of the molecule were not optimal for recognition of the barrel structures.

The examples in figure 7 demonstrate the importance of enhanced exploration features, since only specific orientations clearly reveal the different domain structures including an α/β barrel, antiparallel β strands and an open twisted α/β structure.

Figure 7a (top left) clearly shows the side of the barrel structure on the left side of the image, and 7b (top right) shows the front of the same domain at the top, while in 7d (lower right) this detail is completely occluded. This was observed by all members of the test group.

Our intuitive interaction and navigation interface allowed easy exploration in the 3D Virtual Reality test environment so that the test group could closely inspect details in different orientations of the molecule which appeared to be hidden otherwise.

7. Conclusion

The case study has shown that interactive navigation tools and stereoscopic vision in a virtual environment add a new quality to structural protein analysis. By visual inspection of a large molecule, the test group was able to detect features that were included in a dynamic 3D environment, but occluded by the overwhelming amount of other features in a 2D representation. Since the function of a pro-

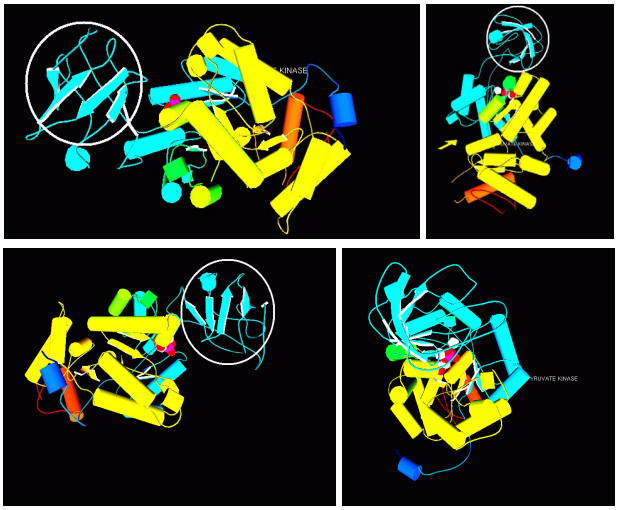

Figure 7. Various visualizations of one chain of Pyruvate Kinase (1A3X). For illustration purposes, the barrel structure has been marked by a circle.

tein is determined by its structure, the results of this study are an indicator for the fact that a functional analysis can be performed much easier in a virtual environment than in commonly used single-perspective 2D projections found in other simulation software or textbooks.

The study has also shown that different levels of abstraction can help to see macroscopic structures, such as binding sites, more clearly than in a full-detail representation.

Further studies are necessary to consolidate the results and to find out which levels of abstraction are the most suitable for certain features in a molecule.

8. Acknowledgements

This work was sponsored in part by the National Partnership for Advanced Computational Infrastructure (NPACI) under award no. 10195430 00120410, and the National Institute of Mental Health (NIHM) under award no. 5 P20 MH60975.

We would like to thank the members of the Creative Interactive Visualization Laboratory (CIVL) at the University of California, Irvine, for their cooperation.

References

[1] N. Akkiraju, H. Edelsbrunner, P. Fu, and J. Qian. Viewing geometric protein structures from inside a cave. *IEEE Computer Graphics and Applications*, 16(4):58–61, 1996.

[2] E. Baker, T. Blundell, J. Cutfield, S. Cutfield, E. Dodson, G. Dodson, D. Hodgkin, R. Hubbard, N. Isaacs, C. Reynolds, and et al. The structure of 2zn pig insulin crystals at 1.5a resolution. *Philos Trans R. Soc. Lond. B. Biol. Sci.*, 319:369, 1988.

[3] H. M. Berman, J. Westbrook, Z. Feng, G. G. T. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000. http://www.rcsb.org/pdb/.

[4] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the cave. In *SIGGRAPH 93 Computer Graphics Conference*, pages 135–142, Anaheim, Aug. 1993. ACM SIGGRAPH.

[5] W. Humphrey, A. Dalke, and K. Schulten. VMD - visual molecular dynamics. *J. Molecular Graphics*, 14, 1996.

[6] M. Jurica, A. Mesecar, P. Heath, W. Shi, T. Nowak, and B. Stoddard. The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. *Structure (London)*, 6:195, 1998.

[7] P. Kramer, J. Bella, P. Mayville, B. Brodsky, and H. Berman. Sequence dependant conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.*, 6:454, 1999.

[8] P. J. Kraulis. Molscript: A program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography*, 24:946–950, 1991.

[9] R. Loris, J. Steyaert, D. Maes, J. Lisgarten, R. Pickergill, and L. Wyns. Crystal structure determination and refinement at 2.3 angstroms resolution of the lentil lectin. *Biochemistry*, 32:8772, 1993.

[10] E. Moritz and J. Meyer. Virtual exploration of proteins. In *Proceedings of the Second IASTED International Conference on Visualization, Imaging and Image Processing (VIIP)*, pages 757–762. IASTED, Sept. 2002.

[11] U. Obeysekare, C. Williams, J. Durbin, L. Rosenblum, R. Rosenberg, F. Grinstein, R. Ramamurthi, A. Landsberg, and W. Sandberg. Virtual workbench: A non-immersive virtual environment for visualizing and interacting with 3D objects for scientific visualization. In R. Yagel and G. M. Nielson, editors, *Proceedings of the Conference on Visualization*, pages 345–349, Los Alamitos, Oct. 27–Nov. 1, 1996. IEEE.

[12] D. C. Richardson and J. S. Richardson. The kinemage: A tool for scientific communication. *Protein Science*, 1:3–9, 1992.

[13] J. Rini, K. Hardman, H. Einspahr, F. Suddath, and J. Carver. X-ray crystal structure of a pea lectin-trimannoside complex at 2.6 angstroms resolution. *J. Biol. Chem.*, 268:10126, 1993.

[14] R. A. Sayle and E. J. Milner-White. Rasmol: Biomolecular graphics for all. *Biochem. Science*, 20:374–376, 1995.

[15] J. Setubal and J. Meidanis. *Introduction To Computational Molecular Biology*. PWS Publishing Company, Boston, 1997.

[16] J. Tainer, E. Getzoff, K. Beem, J. Richardson, and D. Richardson. Determination and analysis of the 2 angstrom structure of copper, zinc superoxide dismutase. *J. Mol. Biol.*, 160:181, 1982.