Virtual Exploration of Proteins

Elke Moritz
Department of Computer Science
University of Kaiserslautern
Germany
email: moritz@informatik.uni-kl.de

Joerg Meyer
Department of Computer Science
Mississippi State University
USA
email: jmeyer@cs.msstate.edu

ABSTRACT

The rapid growth in publicly available macromolecular data sets has created the need for improved distribution and visual representation of biological data. Since the function of a molecule is directly related to its three-dimensional (3D) structure, techniques for investigating molecular structure, function and interaction are essential for microbiologists and drug designers in particular.

This prototype study explores the possibilities and potentials of surround-screen projection-based virtual reality environments for interactive visualization of multiple molecular models.

MolScript, a program for textbook-quality imaging, was enhanced with CAVE™-functionality to facilitate stereoscopic real-time rendering for displaying and manipulating realistic 3D structural models of macromolecular data (fig. 1).

An intuitive and natural user interface was added to enable exploration and side-by-side comparison as well as superimposition of scientifically meaningful and visually appealing visualizations.

KEY WORDS

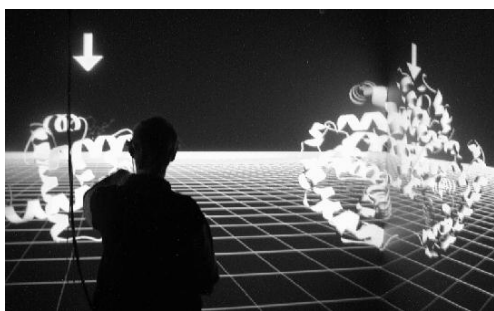Virtual Reality, Macromolecular Structures, CAVE, Scientific Visualization

Figure 1. Exploring macromolecular structures in the CAVE™.

1. Biomolecular Background

Even though the human genetic code has almost been completely decoded by participants of the Human Genome Project and commercial companies, the question of how life works can not be answered with only the mere sequencing and assembly of a genome map. The challenge of associating semantics with structural data remains.

Insights into the causes, effects, and treatment of diseases can only be gained by exploring the macromolecular structure of nucleic acids and proteins and by understanding their function.

A known amino acid sequence (1D, primary structure) permits predictions of the 3D structure of the molecule. Currently available computational methods do not allow the exact construction, only predictions of the structure through adequate and accurate pairwise or multiple sequence alignment.

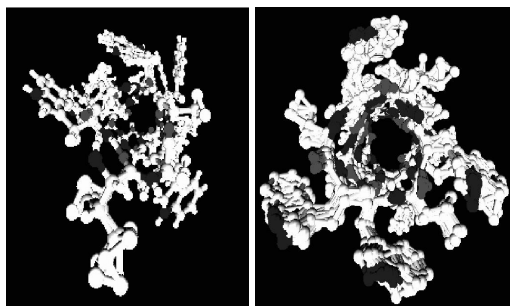Secondary structure units are often illustrated with sim-

Figure 2. Rotated side view of Gramicidin (1AL4 and 1GRM). Note that although 1AL4 was determined through X-ray Diffraction while 1GRM is a NMR data set, the characterizing structure properties can easily be compared.

plified symbols for useful overall views in so-called cartoon representations. β sheets are shown as arrows, loops as ribbons, and α helices as cylinders or helices (fig. 3). Purely chemical aspects are best represented through wireframe-like stick or ball-and-stick models, while space-filling models (*cpk*) provide information on volume, size, shape and surface structure (fig. 2). Since different models emphasize diverse aspects of the structure, the comparison of different representations is indispensable for comparing the characteristic features of the structure with those of the surface.

Simple combinations of a few secondary structures are called motifs or supersecondary structures. Domains (tertiary structures) are units of function that form compact globular structures consisting of $1..N$ secondary structures. A single peptide chain can thereby fold into one or several domains.

Figure 3. Exploration of plastocyanin (1BXU) shows the typical arrangement of β strands in barrel form.

While the amino acid sequence is responsible for the structure, the position of the amino acids in the structure is responsible for the function. The fact that the primary structure determines the tertiary structure of the molecule is commonly described as the central dogma of protein folding.

The impact of efficient visualization techniques is described in the next section.

2. Of Mice and Men

Proteins with similar function occuring in different species usually have similar amino acid sequences and similar corresponding DNA sequences. Differences in the sequence might have been the result of evolutionary mutations.

Mutations may, for instance, allow certain viruses to spread from animals to humans but may prevent the disease from spreading from person to person. Genetics are needed to find out which protein is responsible for the effect. Tests on mice help to find inhibitors for enzymes that stop the infection from spreading. The problem is to make, test, and safety-test enough vaccine in time, before the virus mutates and becomes drug resistant.

A virtual laboratory, which utilizes high-quality interactive visualizations, could help in finding medications that block certain binding processes and thereby prevent diseases. It enables scientists to exchange information and discuss details of molecular cell processes in a collaborative environment.

3. Related Work

Various textbooks have been published in the last few years (e.g., [1], [2], [3]) that provide a good survey on the area of bioinformatics.

A very detailed introduction to the structural specialties of proteins is given in Branden et al. [4]. Most of the example structures are illustrated through hand-drawn or computer-generated cartoon representations.

In existing computer programs, molecular structures are currently mostly visualized through standard cartoon representations (using ribbons or cylinders) to ensure high-speed rendering. While it is generally possible to display user-selected subsets of residues or atoms, the often simplified representations make it difficult to correctly interpret the structures visually.

The molecular viewer RasMol [5] is capable of rendering only one dataset at a time and can be used in conjunction with MolScript to produce high-quality output, but cannot be enhanced to support virtual environment displays. The more advanced tool VMD (Visual Molecular Dynamics, [6]) provides a complex, menu-driven user interface to control general molecular visualization as well as display and control of molecular dynamics simulations. Both programs support the standard stick, ball-and-stick, and space-filling representation modes, but color, bonds, and non-standard residues may be displayed differently.

With the introduction of the Responsive Workbench [7], a number of application fields have been explored, incl. rudimentary visualization of molecular structures [8], [9] and collaborative working environments [10]. The development of the CAVE™ (CAVE Automatic Virtual Environment, [11]) enabled further investigations of Virtual Reality tasks [12] and navigation paradigms [13] as well as the study of spatial design applied to interactivity to explore the immersive navigational experience [14].

The usability of various user interfaces, navigation strategies, and interaction techniques has been examined in a number of publications, for example those by Moritz et al. [15] and Knöpfle et al. [16].

4. Virtual Reality

This prototype study focuses on the exploration of several large molecular structures in immersive virtual environments (VEs). A large field of view is necessary to display the complex structural details of proteins.

Stereo projection and total immersion in VEs enable the viewer to explore and grab the structures and help to understand phylogenetic relations, functional meanings of structural properties, and intermolecular interactions.

This software has been designed to be usable by scientists or students with prior knowledge in molecular visualization, but not necessarily experience in 3D navigation.

Single-screen projection-based devices ensure easy accessibility for interaction with the data and allow intuitive interfaces to be shared by several users, but are only suitable for the exploration and manipulation of a single macromolecular structure, not for the comparison of various large 3D models.

Multiple-screen projection-based virtual reality systems like the CAVE™ enable collaborative exploration of several large-scale objects through manipulation and navigation in an immersive VR world.

The viewer navigates the virtual scene by naturally moving around inside the cube, while his or her field of vision is completely covered by the projected images.

In many cases, molecular exploration techniques in 3D space might reveal structures that can only be seen from a particular angle, but would be hidden otherwise. This way, VEs provide superior structural cues over traditional non-immersive visualizations.

5. New MolScriptCAVE Features

MolScript [17] is often regarded as a PDB-to-PostScript converter and is widely used for generating high-quality visualizations for print publications. It also features various other output formats, including JPEG images and OpenGL rendering.

A powerful command language enables the user to generate ribbon plots and various volume representations for atoms and molecules (including *cpk* and ball-and-stick). MolScript was chosen as basis for this project because of its modular structure and freely available source code (C with OpenGL), which allow for easy addition of new features. The project implementation followed two goals:

1.) Porting MolScript to the CAVE^{TM} by modifying existing modules and adding a new output module.

2.) Development of intuitive 3D interaction methods for the CAVE^{TM}.

The main concern was to enhance the program with new features while at the same time keeping the existing functionality. Thus the changes to other MolScript modules and libraries were kept at a minimum by developing the new CAVE^{TM} module based on routines of the interactive OpenGL mode.

In the standard version of MolScript, only one PDB file can be displayed at a time. For detailed comparisons of similar structures, MolScript was enhanced with the option for loading and (initially) non-overlapping display of up to four molecular structures.
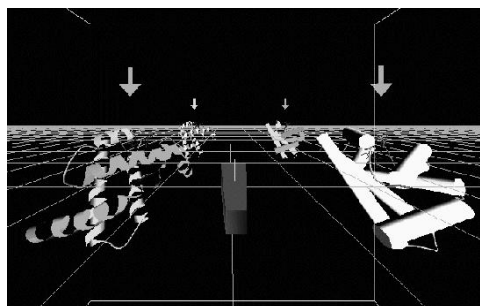

Figure 4. View of four different cartoon representations of myoglobin (1MBN).

6. Navigation Techniques and User Interface

Suitable interaction methods for easy and unencumbered exploration of molecular data have been added to the CAVE^{TM} output module.

The traditional 2D input device, the mouse, is substituted by a wand with several analog inputs, often a trigger and a jog shuttle, and a number of digital input buttons. Six-degrees-of-freedom (6dof) trackers for head and wand tracking are fully supported.

The exploration of molecular structures is possible through global and local navigation. The complete scene can be translated, and single objects can be picked, moved and rotated. A wand button toggles between these modes.

6.1 Global Navigation

Global navigation of a virtual world without any specific interaction with objects helps exploring large 3D datasets.

The user can navigate the world by using the trigger and jog shuttle on the wand to move "the entire CAVE". All objects are moved at the same time. Their relative position and orientation to each other does not change.

6.2 Local Navigation

Each molecular object can be moved separately in local navigation mode. This makes a closer examination of similarities and differences of the structures possible and even allows superimposition of the molecules.

The natural method for the selection of objects is to pick the model at a specific point and move and rotate it around this point according to the movement of the hand. The possible range of motion and displacement is bound by the length of the arm.

This grab-and-move technique was adapted from the Responsive Workbench for this study [15]. While this method may be suitable for virtual workbenches, where the operational space is very limited and the object is usually within reach of the user, this method can not be applied directly for the MolScript CAVE^{TM} module, since the models can either be very far away from the user or can be superimposed in a way that grabbing the surface directly is impossible.

Therefore, an additional rotation mode for detailed local positioning was introduced, which rotates the active molecule around its origin according to the relative wand movement.

6.3 Visual Proxies

Traditional methods for object selection such as ray or cone casting [18], [16] are not suitable for superimposed or occluded objects.

To distinguish between different molecules even if several of them are superimposed, an arrow was attached to each object. This navigational aid not only indicates the orientation of the object, but also shows which molecule is currently being moved, since the corresponding arrow of the selected object changes its color. In addition, the name of each molecule is displayed (figure 4).

Additional visual cues for better orientation in the VE are given by an optional navigational grid which supports the users' ability to more effectively perform navigational tasks if visual markers are provided.

7. Hardware Setup

The CAVE^{TM} module was implemented in C with OpenGL, using CAVELib^{TM} v2.6. An 8-processor SGI **Onyx2 Infinite Reality** system with 250MHz, R10000 processors, and 4 Gigabyte of RAM served as a rendering engine. Four channels are generated using two Infinite Reality 2 Graphics Pipelines - pipe 1 for the front and floor screens, pipe 2 for the left and right screens.
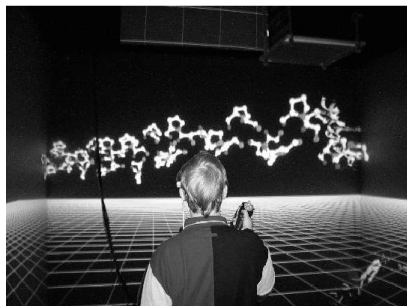

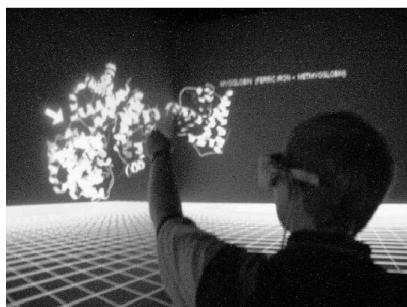Figure 5. Exploration of the ball-and-stick representation of a collagen macromolecule in the CAVE^{TM}.

Figure 6. Working in the CAVE^{TM}.

User movement and interaction is recorded through 6dof trackers mounted to a pair of shutterglasses for stereoscopic viewing, and a custom wand with four digital and two analog controls.

Figure 7. The CAVE^{TM} Laboratory at the ERC and the Nintendo64^{TM}-based custom wand.

PDB ID	Name	Atoms	Residues
1A3N	Hemoglobin	4997	574
1A4L	Adenosine Deaminase	11473	1396
1AL4	Gramicidin	956	36
1BXU	Plastocyanin	776	91
1GRM	Gramicidin	1360	34
1MBN	Myoglobin	1260	153
1MTM	Triosephosphate Isomerase	1825	241

Table 1. Characteristics of sample molecules.

PDB ID	Polyg.[1]	DLists[1]	Polyg[2]	DLists[2]
1A3N	8484	70	132588	11111
1A4L	28544	236	337904	28288
1AL4	512	14	17944	1718
1BXU	2518	17	21050	1765
1GRM	2400	14	115984	12873
1MBN	2330	18	36824	3083
1MTM	4986	45	55250	4625

Table 2. Number of polygons and displaylists rendered by MolScript in example cartoon representations [1] and ball-and-stick models [2].

8. Results

Various structure data sets from the PDB (Protein Data Bank, [19]) were analyzed with MolScript in the CAVE^{TM}. A representative range of features expressed in different structures (proteins, DNA, etc.) and various types of data (X-ray Diffraction, NMR) were investigated.

Detailed representations of macromolecular structures can be very complex and result in a huge number of generated polygons. Depending on the elements chosen for visualization, even a single molecule can be so large that real-time interaction is made impossible.

The number of visualization processes generated by the CAVELib^{TM} for repeated rendering of the generated polygons causes a severe bottleneck for real-time visualization in the CAVE^{TM}. The maximum number of polygons for real-time rendering depends on the size and positioning and distribution of the objects on the screens. The worst case situation would mean that one pipeline renders the complete scene.

In the end, it is up to the user to carefully choose the displayed features of the molecular structures in the data file and to ensure that the MolScript 3D representations do not generate too many polygons. In most cases, this can be guaranteed by selecting a cartoon representation for most parts of the molecule, and the more detailed ball-and-stick model only for the area of interest.

The navigational experience of exploring the detailed molecular structure in the virtual environment is far superior to 2D onscreen displays, because it allows the user to navigate through the scene, turn the object, and reveal structures that might be occluded otherwise or only become apparent when viewed from a particular angle. The possibility to display and compare up to four objects in the CAVE^{TM} allows for easy collaborative investigation of macromolecular structures.

9. Conclusions

Porting MolScript to a CAVE^{TM}-like environment provides scientists with an easy-to-use tool for high-quality visualization and exploration of macromolecular data sets.

Up to four molecules calculated from PDB structure data can be displayed simultaneously in real-time, as long as the total of generated polygons is less than approx. 100,000. Higher polygon numbers result in significantly lower frame rates.

Molecular structures can be explored through global navigation in the environment as well as through direct 6dof interaction with the objects.

Intuitive interaction methods have been explored and implemented utilizing the special features of a custom wand, allowing straightforward operations like manual rotation and positioning for side-by-side comparison and superimposing of several 3D structures.

The stereoscopic visualizations generated by the new CAVE^{TM} interface of MolScript benefit both students and educators in providing detailed interactive models in real time that are superior to traditional 2D textbook representations and static 2D projections of 3D objects.

10. Future Work

Investigation of macromolecular dynamics, especially the analysis of docking problems and macromolecule-ligand interactions is essential for drug design and discovery.

For a closer analysis of the structural data, the CAVE^{TM} module will be enhanced with additional user interface features and a set of tools for interactive exploration and manipulation.

The time needed for collision detection for manipulations and modifications of the data can be dramatically reduced by applying mesh decimation algorithms that will decrease the number of triangles in the grid, thus making the rendering process more efficient and enabling the exploration of even larger data sets in real time.

11. Acknowledgments

12. PDB Example Data Sets

1AL4 B.M. Burkhart, D.A. Langs, W.A. Pangborn, W.L. Duax: Peptide Folding and Aggregation in Mixed Dimers of Gramicin D; To Be Published

1BXU T. Inoue, H. Sugawara, S. Hamanaka, H. Tsukui, E. Suzuki, T. Kohzuma, Y. Kai: Crystal Structure Determinations of Oxidized and Reduced Plastocyanin from the Cyanobacterium Synechococcus Sp. PCC 7942; Biochemistry, V.38, p. 6063 (1999)

1GRM A.L. Lomize, V.Yu. Orekhov, A.S. Arseniev: Refinement of the Spatial Structure of the Gramicidin A Transmembrane Ion-Channel; Biol. Membr. (USSR), V.18, p. 182 (1992)

1MBN H.C. Watson: The Stereochemistry of the Protein Myoglobin; Prog. Stereochem., V.4, p. 299 (1969)

References

[1] A. D. Baxevanis and B. Ouellette, eds., *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Wiley-Liss, Inc., 1998.

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

[3] C. Gibas and P. Jambeck, *Developing Bioinformatics Computer Skills*, O'Reilly & Associates, Sebastopol, CA, Apr. 2001.

[4] C. Branden and J. Tooze, *Introduction to Protein Structure*, Garland Publishing, Inc., New York, 1991.

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

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

[7] W. Krüger and B. Fröhlich, "Visualization blackboard: The Responsive Workbench (virtual work environment)," *IEEE Computer Graphics and Applications* **14**, pp. 12–15, May 1994.

[8] 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 *Proceedings of the Conference on Visualization*, R. Yagel and G. M. Nielson, eds., pp. 345–349, IEEE, (Los Alamitos), Oct. 27–Nov. 1, 1996.

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

[10] L. D. Cutler, B. Fröhlich, and P. Hanrahan, "Two-handed direct manipulation on the responsive workbench," in *Proceedings of the Symposium on Interactive 3D Graphics*, pp. 107–114, ACM Press, (New York), Apr. 27–30, 1997.

[11] 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*, pp. 135–142, ACM SIGGRAPH, (Anaheim), Aug. 1993.

[12] S. Volbracht and G. Domik, "Developing effective navigation techniques in virtual 3d environments," in *Virtual Environments 2000 - Proceedings of the Eurographics Workshop in Amsterdam, The Netherlands*, J. D. Mulder and R. van Liere, eds., pp. 55–64, Springer, (Wien), June 1–2, 2000.

[13] D. A. Bowman, D. Koller, and L. F. Hodges, "Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques," in *Proceedings of the Virtual Reality Annual International Symposium (VRAIS)*, pp. 45–52, (Albuquerque, NM), 1997.

[14] M. Watson, "Constructing virtual space," in *3rd International Immersive Projection Technology Workshop*, pp. 201–209, Springer, (Heidelberg), May 1999.

[15] E. Moritz, F. Kuester, B. Hamann, K. I. Joy, and H. Hagen, "Towards immersive clay modeling: Interactive modeling with octrees," in *Stereoscopic Displays and Virtual Reality Systems VII, Proceedings of SPIE Vol. 3957*, J. O. Merritt, S. A. Benton, A. J. Woods, and M. T. Bolas, eds., pp. 414–422, SPIE, Jan. 2000.

[16] C. Knöpfle, "Interacting with simulation data in an immersive environment," in *Virtual Environments 2000 - Proceedings of the Eurographics Workshop in Amsterdam, The Netherlands*, J. D. Mulder and R. van Liere, eds., pp. 157–166, Springer, (Wien), June 1–2, 2000.

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

[18] J. L. Gabbard, "A taxonomy of usability characteristics in virtual environments," Master's thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1997.

[19] 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), pp. 235–242, 2000. http://www.rcsb.org/pdb/.