

# Virtual Reality Modeling for Structural Biology

Jürgen Sühnel

Institut für Molekulare Biotechnologie, Postfach 100813, D-07708 Jena / Germany

jsuehnel@imb-jena.de

**Abstract.** In 1995 we have set up a virtual reality division of our Internet-based Image Library of Biological Macromolecules. This was one of the first virtual reality applications in biology. In this contribution we describe how virtual reality modeling can be applied to structural biology.

## 1. Introduction

Structural information on biological macromolecules is an essential requirement for our understanding of biological function and for a deliberate variation of this function by rational or evolutionary approaches. Progress in recombinant DNA technology and RNA synthesis, X-ray and nuclear magnetic resonance (NMR) instrumentation and computer and software technology has led to an increasing rate of accumulation of new structures. Currently (November 15, 1996), the Protein Data Bank at the Brookhaven National Laboratory has coordinate entries of 4702 proteins and of 361 nucleic acids [1, Q]. From the very beginning of structural biology visualization was essential for determining and understanding structures. Model building with metal plates and other materials has played a crucial role for the DNA double helix structure suggested by Watson and Crick in 1953 [2]. Further, it is reported that the first model of hemoglobin built by M. Perutz and colleagues in the 1960s took up 16 square feet of floor space [3]. Since the early 1980s interactive computer graphics has greatly facilitated and improved the visualization of biopolymer structures. The usual approach is to retrieve the coordinate files from a structure database and then to use one of the molecular graphics software packages. On the other hand, one would often prefer to have biopolymer images directly available without the need to spend some time for visualization or even without having access to a molecular graphics software. This is especially important for the large and heterogeneous community outside structural biology.

The recent developments in the World-Wide Web enable one to transfer images or videos over the Internet very easily. We have, therefore, started in 1993 to set up an Internet-based Image Library of Biological Macromolecules [4, F].

Currently, the Library contains more than 3500 images of about 300 structures. The images are in the public domain and can freely be retrieved from the IMB Jena WWW server [F]. The images of molecular structures are intended to

provide as much information as possible. Therefore, mixed rendering, coloring and labeling techniques are extensively used. All molecular images are available both in mono and stereo representations. In addition to the molecular images color coded distance plots are included. Distance plots relate the distances between representative atoms of amino acids and/or nucleic acids in the 3D structure to the sequence [5].

In 1995 the new VRML format was defined. The Virtual Reality Modeling Language (VRML) is essentially a three-dimensional (3D) image format supplemented by network tools. Contrary to the static images it enables one to interact with the three-dimensional image objects. Of course, for biopolymers this can be done much better using molecular graphics packages. On the other hand, VRML viewers are already becoming standard parts of current web browsers. Insofar it is immediately obvious that this new tool is of relevance for structural biology. We have therefore extended the Image Library by a Virtual Reality Division. Besides the pioneering work done at the Imperial College London [6, A] and at the Technische Hochschule Darmstadt [7, Y], this was one of the first VRML applications in biology and to the best of our knowledge the very first application which was not devoted to demonstration purposes alone. The VRML division contains now already about 650 VRML representations of biopolymer structures.

The objective of this contribution is to describe our experience with the VRML format for biopolymer structures and to speculate about possible future applications in this field.

## **2. What is VRML ?**

VRML stands for Virtual Reality Modeling Language and represents a platform independent standard language for describing 3D objects supplemented by network functionality. It is based on a subset of the Open Inventor ASCII file format [P] and it describes 3D objects or scenarios in an object oriented manner. The basic elements are various node types: shape nodes (points, lines, spheres, cylinders, ...), property nodes (color, texture maps, geometry transformation, ...), group nodes (for implementing a hierarchical structure), camera nodes, light nodes, WWWInline nodes (for loading other VRML files into the current scene) and WWWAnchor nodes (hyperlinks).

Currently, the most important format for the World Wide Web (WWW) is HTML (hypertext markup language). Due to the fact that VRML combines 3D with network functionality it is justified to claim that VRML extends the WWW to the third dimension or that VRML is to 3D what is HTML to 2D. It is even not unlikely that VRML will be the future of the web. This, however, will finally depend on various aspects like bandwidth and CPU requirements. The VRML 1.0 specification was a means of creating static 3D worlds. The extensions and enhancements of VRML 2.0 can be characterized by the terms enhanced static worlds, interaction,

animation and prototyping. More detailed information can be found in the VRML repository of the San Diego Computer Center [AA] and on the VRML page of Silicon Graphics [Z].

### 3. How to create VRML representations ?

We have generated the VRML files of biopolymers and of the corresponding building blocks (amino acids, nucleotides) using InsightII from Molecular Simulations [N], the MIDASPLUS molecular graphics and display system [8] and Explorer EyeChem modules [9, D]. InsightII has a direct VRML interface, whereas in the other two cases, first files in the Inventor format have to be created which then have to be converted to the VRML format. This was done using the program ivToVRML provided by Silicon Graphics, Inc. [T]. Especially with InsightII almost all of the rendering techniques developed over the years can be translated to VRML. Recently, further software of relevance to VRML has become available. The WebLab viewer developed by Molecular Simulations can read PDB structure files and generate VRML files [BB]. Contrary to the rather expensive InsightII modeling system it can be obtained for free. It is running on personal computers with Windows or Macintosh operating systems and has, of course, only a small subset of visualization tools available for the more comprehensive modeling packages. The Tripos Molecular Inventor Netscape Navigator Plugin [X] can read mol files of the Tripos modeling system SYBYL [W] and convert them to Open Inventor files. The latter tool is running on Silicon Graphics computers only and requires the molecular inventor execution environment. Finally, at the Technische Hochschule Darmstadt the program pdb2vrml for converting PDB files directly to the VRML format was developed [Y]. Again, this program has only limited visualization functionality. However, it is very useful if an automatic conversion of PDB files is intended. Finally, Silicon Graphics has recently released the Molecular Inventor development kit [M]. It extends the capabilities of Open Inventor and is, therefore, appropriate for the generation of biopolymer structures in VRML format. To the best of our knowledge, this is a complete compilation of tools currently available for generating VRML files of biopolymer structures. We expect that VRML interfaces will become standard in molecular modeling packages. According to our experience the best current tool for creating VRML files of biopolymer structures is InsightII.

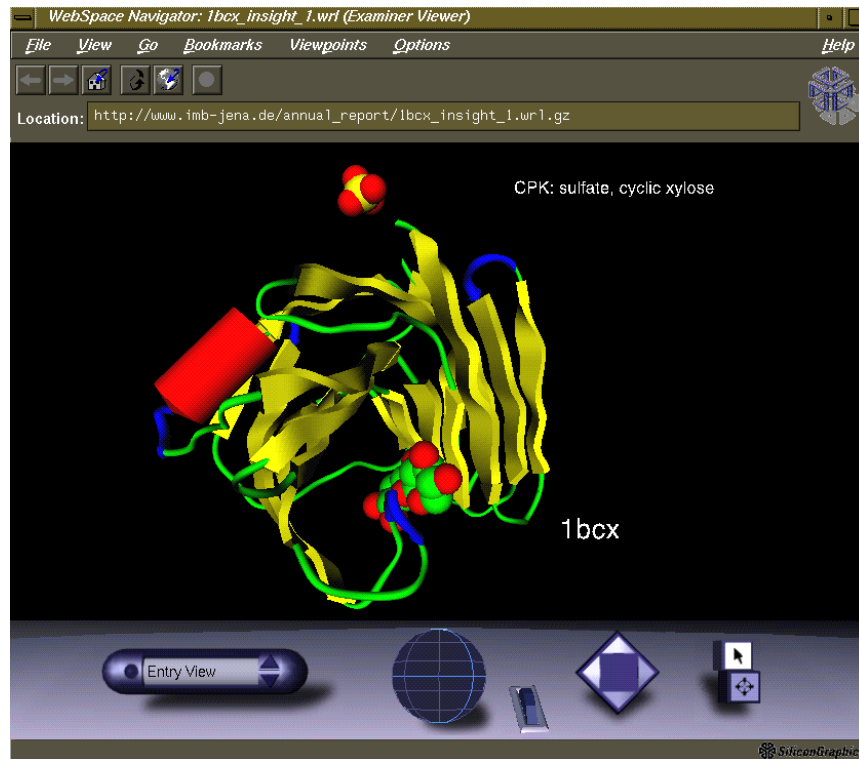
To give an impression how VRML files look like, in the following section the corresponding file for a line drawing of the amino acid alanine is shown. It was generated using the VRML interface of the molecular modeling package InsightII (version 95.0.3) by Molecular Simulations [N]:



```

}}
DEF mat_255_0_0_255 Material {
  diffuseColor 1.000 0.000 0.000
}
Coordinate3 { point [
  -0.392313 2.453175 0.253431,
  -0.059091 1.958026 0.415358,
  -0.492730 2.369304 0.203611,
  -0.159508 1.874155 0.365538,
  ]}
IndexedLineSet { coordIndex [
  0, 1,-1, 2, 3,-1,
  ]}
}

```



**Fig. 1.** Secondary structure representation of the *Bacillus circulans* xylanase (PDB code - 1bcx [10]) complexed with sulfate and cyclic xylose in the VRML viewer WebSpace (helix - tube, sheet - ribbon) [CC].

#### **4. How to view VRML files ?**

For visualizing VRML files one needs VRML viewers. The first generally available VRML viewer was WebSpace from Silicon Graphics [T, CC]. It was released for the first platforms in May 1995. The claim that VRML viewers will become standard parts of future web browsers is confirmed, for example, by the various VRML plugins of Netscape Navigator [O] and by the Microsoft Internet Explorer 3.0 [I]. Especially interesting for structural biology is the viewer i3D developed at CERN [E]. It seems to be the first viewer which supports crystal eyes stereo representations. CosmoPlayer (Silicon Graphics) is one of the first viewers which supports the upcoming VRML 2.0 specification [B]. A comprehensive collection of viewers can be found in the VRML repository at the San Diego Supercomputer Center [AA].

To give an impression, how a VRML viewer looks like Figure 1 shows the secondary structure representation of the *Bacillus circulans* xylanase complexed with sulfate and cyclic xylose in the VRML viewer WebSpace.

#### **5. Structural biology VRML web sites**

The very small number of virtual reality applications in structural biology available last year has already increased:

- Crystal Packing Visualization for PDB Files (Harvard Medical School) [C],
- Virtual Molecular Studio (Dep. of Chemistry, Imperial College, London) [A],
- Interactive Membrane Builder (Physical and Theoretical Chemistry, University of Oxford) [G],
- Image Library of Biological Macromolecules (Biocomputing, IMB Jena) [F],
- Molecular Dynamics Calculations (Johns Hopkins University Medical School) [L],
- Protein Motions Database (Stanford University Medical Center) [R],
- Solvent Excluded Surface Computation (The Scripps Research Institute) [U],
- VRML in Chemistry (Institut für Physikalische Chemie, TH Darmstadt) [Y].

Very recently, further structural biology web sites have included VRML. Examples are:

- Protein Data Bank (Brookhaven National Laboratory) [Q],
- SCOP (structural classification of proteins) Database (MRC Laboratory of Molecular Biology and Center for Protein Engineering, Cambridge) [11, S].

We expect in the future that VRML applications will become a standard tool for structural biology.

## 6. Problems

It is occasionally claimed that VRML files are small as compared to the standard GIF format, for example [7]. We have found that this is only correct for simple drawings of relatively small structures. For high-quality images of complex structures the situation is completely different. Table 1 shows an example for the protein xylanase displayed in Figure 1.

**Table 1.** Filesize in kbytes for the GIF and VRML formats of the amino acid alanine and of the protein xylanase (PDB code: 1bcx, [10]) generated with the molecular modeling system InsightII from Molecular Simulations. Both structures contain no annotation and for the protein structure water and the ligands sulfate and cyclic xylose are not included.

	filesize (kbytes)			compression ratio (VRML)
	GIF (72 dpi)	VRML	VRML (gzip)	
alanine (line drawing)	3.722	2.715	0.947	0.35
alanine (high quality CPK)	42.098	4.604	1.168	0.25
xylanase (secondary structure)	62.712	2.813.342	360.599	0.13
xylanase (line drawing)	42.418	180.606	53.042	0.29
xylanase (high quality CPK)	128.412	407.482	74.540	0.18

In the alanine case indeed the VRML files are much smaller than the corresponding GIF images. On the other hand, for the xylanase, a not too large protein with 185 amino acids, the situation is completely different. For the secondary structure representation with a high-quality ribbon drawing both the

original and the compressed VRML files are much larger than the GIF file. For the line drawing and CPK representation the uncompressed VRML files are either larger or of almost the same size as the corresponding GIF images. The situation is even worse for high-quality representations of larger structures. In these case it is not unlikely that a VRML file will have a size of 50 Mbytes or even more.

Even though there is a substantial compression effect, which reduces the bandwidth demands, uncompression takes time. Therefore, it may happen that currently less powerful computers are not able to manage larger VRML files at all. One will come across such examples if one checks out the Virtual Reality Division of the Image Library. This problem poses a major challenge to the VRML tools. One should realize, however, that the performance of the viewers, like WebSpace for example, has already dramatically increased since May 1995.

## **7. Outlook**

An interesting application we expect in the near future is that electronic versions of journals will contain figures in VRML format, which will enable the reader to interact with the three-dimensional image objects. This is not only important for structure representations but for any three-dimensional figures. There are already various other tools for the usage of the VRML format. Examples are the conferencing system InPerson [H] and the IRIS Annotator tool [J] both developed by Silicon Graphics. Together with the already mentioned Molecular Inventor and with the various possibilities of creating VRML images they constitute already a not too small repertoire of software for working with VRML. In addition, the Internet programming system Java (Sun Microsystems [V]) provides a flexible environment for the integration of VRML into more complex applications [K]. It is especially important that these new developments open up the way of collaborative work on 3D objects in local or wide-area networks. The new VRML 2.0 specification will include further interesting features like the implementation of dynamic behavior and interactivity. Both current and future features of VRML add a new dimension to the current molecular modeling approaches. We are convinced that the new VRML format already has a lot to offer for a better dissemination of visual information on biological macromolecules within and outside the scientific community. It will soon become a standard tool in structural biology.

## References

### *Papers*

1. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977, 112, 535.
2. Watson, J. D. and F. H. C. Crick. A structure for deoxyribose nucleic acid. *Nature* 1953, 171, 737.
3. Hall, S. H. Protein images update natural history. *Science* 1995, 267, 620.
4. Sühnel, J. Image library of biological macromolecules. *Comput. Appl. Biosci.* 1996, 12, 227.
5. Godzik, A., Skolnick, J. and Kolinski, A. Regularities in interaction patterns of globular proteins. *Prot. Eng.* 1993, 6, 801.
6. Casher, O., Leach, C., Page, C. S., Rzepa, H. S. Advanced VRML based chemistry applications: A 3D molecular hyperglossary. *J. Mol. Struct. (THEOCHEM)* 1996, 368, 49.
7. Brickmann, J. and Vollhardt, H. Virtual reality on the world wide web: a paradigm shift in molecular modeling ? *Trends in Biotechnology* 1996, 14, 167.
8. Ferrin, T. E., Huang, C. C., Jarvis, L. E. and Langridge, R. The MIDAS display system. *J. Mol. Graphics* 6, 13-27, 36.
9. Casher, O., Rzepa, H. S. A chemical collaboratory using explorer eyechem and the common client interface. *Computer Graphics* 1995, 29, 52.
10. Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J., Yaguchi, M. Mutationally and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase. *Protein Sci.* 1994, 3, 467.
11. Murzin, A. G., Brenner, S. E., Hubbard, T., Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 1995, 247, 536.

### *Internet addresses*

- |   |   |
|---|---|
| A. Virtual Molecular Studio                       | <a href="http://chemcomm.clic.ac.uk/VRML/">http://chemcomm.clic.ac.uk/VRML/</a>               |
| B. CosmoPlayer                                    | <a href="http://vrm1.sgi.com/cosmoplayer/">http://vrm1.sgi.com/cosmoplayer/</a>               |
| C. Crystal Packing Visualization<br>for PDB Files | <a href="http://dta.med.harvard.edu/ubc/ubc.html">http://dta.med.harvard.edu/ubc/ubc.html</a> |
| D. EyeChem Explorer Modules                       | <a href="http://www.ch.ic.ac.uk/EyeChem/">http://www.ch.ic.ac.uk/EyeChem/</a>                 |
| E. i3d Viewer                                     | <a href="http://www-venus.cern.ch/i3d/">http://www-venus.cern.ch/i3d/</a>                     |
| F. Image Library of Biological<br>Macromolecules  | <a href="http://www.imb-jena.de/IMAGE.html">http://www.imb-jena.de/IMAGE.html</a>             |

- G. Interactive Membrane Builder <http://bellatrix.pcl.ox.ac.uk/people/alan/WebSpace/builder/form.html>
- H. InPerson [http://www.sgi.com/Products/inperson\\_main.html](http://www.sgi.com/Products/inperson_main.html)
- I. Internet Explorer <http://www.microsoft.com/ie/default.asp>
- J. IRIS Annotator <http://www.sgi.com/Products/Annotator.html>
- K. Java <http://java.sun.com/>
- L. Molecular Dynamics Calculations <http://physiology.med.jhu.edu/tom/tbw.html>
- M. Molecular Inventor <http://www.sgi.com/ChemBio/MolecularInventor/>
- N. Molecular Simulations <http://www.msi.com/>
- O. Netscape <http://www.netscape.com/>
- P. Open Inventor <http://www.sgi.com/Technology/Inventor.html>
- Q. Protein Data Bank <http://www.pdb.bnl.gov/>
- R. Protein Motions Database <http://hyper.stanford.edu/~mbg/ProtMotDB/vrml/>
- S. SCOP Database <http://scop.mrc-lmb.cam.ac.uk/scop/>
- T. Silicon Graphics <http://www.sgi.com/>
- U. Solvent Excluded Surface Computation [http://www.scripps.edu/pub/olson-web/people/sanner/html/msms\\_server.html](http://www.scripps.edu/pub/olson-web/people/sanner/html/msms_server.html)
- V. Sun Microsystems <http://www.sun.com/>
- W. Tripos, Inc. <http://www.tripos.com/>
- X. Tripos Molecular Inventor Plugin <http://www.tripos.com/MolecularInventor/>
- Y. VRML in Chemistry <http://ws05.pc.chemie.th-darmstadt.de/vrml/>
- Z. VRML Page of Silicon Graphics <http://vrml.sgi.com/>
- AA. VRML Repository (San Diego Supercomputer Center) <http://www.sdsc.edu/vrml/>
- BB. WebLab Viewer <http://www.msi.com/weblab/>
- CC. WebSpace Viewer <http://vrml.sgi.com/>