BioBrowser: A Framework for Fast Protein Visualization

Andreas Halm† and Lars Offen‡ and Dieter Fellner§
Institut für ComputerGraphik, TU Braunschweig, Germany

Abstract

This paper presents a protein visualization system called BioBrowser, which provides high quality images at interactive frame rates for molecules of extreme size and complexity. This is achieved by a shift in the tessellation approach: triangle meshes are not produced a priori on a 'just-in-case' basis. Instead, tessellation happens 'just-in-time' given a certain camera position, image size and interaction demand. Thus, our approach is based on multi-resolution meshes and on new extensions of graphics hardware. The paper shows how to reduce geometric data by using subdivision surfaces for ribbon structures and molecular surfaces and by using billboards instead of spheres consisting of triangles. It also shows how to use fragment shaders to create a three dimensional appearance and realistic sphere intersections. The combination of these approaches leads to an image quality not yet seen in interactive visualization environments for molecules of that size/complexity. All the above methods are combined to gain a high performance configurable visualization system on standard hardware.

Categories and Subject Descriptors (according to ACM CCS): H.4.3 [Information System Applications]: Communications Applications – Information browsers I.3.8 [Computing Methodologies]: Computer Graphics – Applications J.3 [Computer Applications]: Life and Medical Sciences – Biology and genetics;

Interactive rates are even more important if 3D molecular structures are being used to store, locate, and access information (in any textual or multimedia format) about the respective molecule.

Of course, there are visualization tools like Chimera [Chi], Cn3D [cn3], Rasmol [ras], or FPV [CWWS03] but they either exhibit a significant drop in rendering performance when handling very large proteins or have limited visualization styles for, e.g., ribbon structures or space fill.

In this paper we present the visualization components of a new tool called BioBrowser which aims at supporting bio scientists in the information access in a similar way web browsers support users accessing material from the web. The visualization components currently provide fast and accurate visualization routines for the following common styles:

- Balls and Sticks
- Sticks
- Space-fill
- Ribbons
- Surfaces

Due to the modular architecture of the BioBrowser it is easy to attach various other visualization styles, but so far we have limited them to the most common ones. That is to say we haven’t invented new visualization styles yet, but we applied...
various methods of modern computer graphics, like multi-resolution methods, as well as extensions of modern graphic boards like fragment shaders, to achieve interactive frame rates while preserving a high visual quality in the resulting renderings. These methods are described in detail in Section 3.

1.1. Molecule Basics

A molecule is an accumulation of atoms, where some of them are joined by so called bonds. Since the BioBrowser is in the first place a visualization tool for proteins, we will restrict the following introduction to this domain. A protein is composed of possibly more than one chain of amino acids linked together by peptide bonds. The backbone of a chain is formed by three groups of atoms in each amino acid, namely the central atom $C_\alpha$, the amino group $N-H$ and the carbonyl group $N=O$. The residue $R$, which is also bound to the $C_\alpha$ atom, characterizes the nature of each amino acid, but is not part of the backbone.

A protein can be described by a sequence of abbreviations for the different amino acids, e.g. by ‘IVNGEEAVPG’. This is known as the primary structure of a protein. Since the peptide bond is not rigid, the protein can rotate around this bond to fold itself to an energetically convenient state. This state is called tertiary structure. This structure can be revealed by Nuclear Magnetic Resonance (NMR) and X-ray crystallography and is published for example in the RCSB Protein Data bank [pdb].

Locally, the tertiary structure is built by the secondary structure which consists of alpha helices, the beta sheets and turns. The secondary and tertiary structure can be visualized using the ribbon drawing, introduced by Richardson in [Ric81]. Molecules may consist of one or more chains, which are not connected together by atomic bonding. When drawing the ribbon structure of a molecule that contains more than one chain, one must not connect the ribbons of two chains. Chains can be handled independently of each other.

![Figure 1: Structure of a protein chain: a peptide bond links amino acids in a linear way](image)

2. System Overview

The BioBrowser consists of several modules (see Figure 2). The core modules are the visualization modules, each of which implements one visualization style. The visualization modules can be used independently of each other but they can also be mixed in a single frame using alpha blending. This will allow fairly advanced molecule display when used together with additional information about the displayed molecule. Additionally, a PDB file loader module is implemented to allow easy generation for the internal data structures. All these modules together form a dynamic library, which is currently used with the BioBrowser executable, but can also be used to allow molecule display embedded in other applications, for example Internet browsers or CAVE®-applications.

3. Visualization Concepts

The general concept of all visualization modules is to reduce/avoid geometry data whenever possible. Consequently, we replace polygonal meshes by billboards and ordinary meshes by multi-resolution meshes.

3.1. Ball-and-Stick Model

The Ball-and-Stick visualization style shows all existing bonds in the molecule as well as all atoms as equal-sized spheres (Figure 4). Since the spheres will not intersect each other it is not necessary to model the spheres as a polytope. Instead, it is sufficient to model an atom as just a quadrangle together with a texture resembling a sphere, called a billboard, whose normal will always be directed towards the viewer (Figure 3 (a)). The bonds are realized in an analogous manner (Figure 3 (b)). To create the illusion of 3-dimensional objects the bonds do not connect the mid points

---

submitted to EUROMEDIA - IEEE VGTC Symposium on Visualization (2005)
of the atoms, but are a bit shorter, so that the impression arises that the bonds intersect with the atoms. The use of billboards reduces the amount of geometric data by orders of magnitude. The drawback is that the billboards must be rotated and translated, so that the normal points to the spectator, instead of just translating the spheres. The calculation of this transformation matrix is slightly more expensive than to set up the translation matrix, but afterwards the graphic hardware must only transform the four vertices of the billboard and not every vertex of the sphere. Therefore the drawback changes to a gain when comparing the billboard to spheres of modest to high resolution. Although some of the perspective information is lost, in this case the difference is hardly noticeable.

Figure 3: Billboard for an atom (a) and a bond (b). Generating the alpha channel of the 2D texture (c)

3.2. Space-fill (van der Waals) Model

To give an overall impression of the tertiary structure the protein is visualized with spheres having the van der Waals radius of the related atom. Figure 5 demonstrates this style, called space-fill.

In this display mode the spheres do intersect each other and therefore billboards cannot be used as they won’t intersect properly. Thus, a 2D texture with alpha channel is used. The alpha channel models the relative z-depth of the unit half sphere (Fig. 3 b)). These “enhanced” billboards can be rendered by using a fragment program which translates the pixels of the billboard according to the value in the alpha channel (see Figure 6). As shown in the figure the relative z-depth of the billboard must be corrected by the projective radius of the sphere $r_p$ and the factor $\cos \alpha$, so that $\Delta z' = \Delta z \times \cos \alpha = r_p \times \Delta z \times \cos \alpha$ holds. The correction factor $r_p \times \cos \alpha$ is a simplification, because we assume a parallel projection of an atom after having projected it to the viewplane. This increases the rendering speed and in normal viewing perspective it won’t disturb the visual impression. The resulting fragment program is written as both
GL_ARB_fragment_program extension and GL_NV_fragment_program extension. This is done to allow specific optimizations for NVidia hardware. The fragment program takes two parameters: the 2D texture and the correction factor \( r_p \cos \alpha \) as a vertex attribute. The texture is evaluated per pixel and while the red, green and blue values are multiplied with the current primary color to get the effect of the \( \text{GL}\_\text{MODULATE} \) texture environment, the alpha value of the texture is multiplied with the correction factor (given as a parameter) and added to the fragment’s \( z \)-coordinate. As per-pixel depth testing takes place after the fragment program has been executed, intersection between objects are correctly rendered. This method is also known as depth sprites. A more detailed description can be found in [SGwHS98].

For older graphic hardware which does not support fragment programs or which have only slow implementations of the fragment programs the BioBrowser implements a fall-back solution. Several precalculated half-spheres with different resolutions are currently used. The correct level of detail is chosen by the projective size of the sphere on the viewplane. Then the half-spheres are rendered instead of the billboards.

3.3. Bond Model

In this model only the bonds of the molecule are drawn. The bonds are realized as precalculated meshes of half-cylinders which are rotated like the billboards. To close the gaps between such half-cylinders precomputed half-spheres are used.

![Figure 7: Stick visualization of protein 1axq](image)

3.4. Ribbon Model

The ‘ribbon drawing’ introduced by Richardson [Ric81] provides an elegant method to visualize the folding and secondary structure of proteins. There are several implementations to visualize these ribbon structures like the program from Carson which was improved over the years [CB86], [Car87], [Car91] and [Car97], or complete molecular visualization tools like Chimera [Chi], Cn3D [cn3], or Rasmol [ras].

These implementations construct either 2D structures, e.g., B-spline curves, or they construct a 3D ribbon with a fixed tessellation. This has the drawback that the visualization of huge ribbon structures is either unacceptably slow but of acceptable quality, or it is interactive but of low quality – particularly when viewed at closer distance.

Interactive speed is necessary for the user to get an immediate feedback while exploring the molecular structure. On the other hand high quality images are needed for presentations or publications. Both – up to now conflicting – requirements can only be met by the use of 3D models supporting surface representations of variable resolution: multi-resolution meshes.

We developed a new method ([HOF04]) to visualize these ribbon structures using a combination of simple cubic polynomial B-spline curves, which build the skeleton of the resulting ribbons and a powerful boundary representation (BRep) called Combined BReps (CBRep for short) [Hav02, HF03b]. The CBRep combines a polygonal boundary representation with Catmull/Clark subdivision surfaces [CC78] by attaching a sharpness flag to each edge in the BRep mesh. The control points of a B-spline curve defining the skeleton of the ribbons can be determined by an interpolation process ([Far90]) so that the resulting curve will pass through the positions of the \( C\alpha \) atoms.

![Figure 8: Computation of the base mesh. The stars show the positions of the \( C\alpha \) atoms. While the generated base mesh does not need to include the \( C\alpha \) positions, the spline curve that is generated at run time always does.](image)

By using the CBRep structure it is now quite easy to generate the ribbon structures: each control vertex is extended to a quad (representing the cross section) and two quads are joined to an element of the ribbon (see Figure 8). The method used is able to generate images with very high detail. Additionally, setting the visible detail dynamically according to the parameters...
allows us to reach interactive frame rates on even slowest computers. Figure 9 displays some tessellations for different levels of detail. Table 1 shows the resulting frame rates for some selected molecules. The whole molecule was always visible, so no view frustum culling took place. All the tests were made on an Athlon 1200 with a GeForce 5200 graphics card. It is worth noting that our approach also significantly

speeds up the culling process as a single test for the complete (in)visibility of one CBRep patch eliminates the many culling tests which were necessary if the ribbons had been tessellated during preprocessing resulting in tons of individual triangles, each of which had to be tested separately.

The columns High detail/Low detail give the frame rate resulting from always running the highest/lowest available detail setting. The column Adaptive detail lists the frame rates with our adaptive method that switches the level-of-detail according to the current environment determined by the graphics and computing hardware, the scene complexity, and the viewing position taking the parameters mentioned above into account. The number of triangles rendered in the different modes is also shown, whereas the number for adaptive level of detail is an average value.

3.5. Molecular Surface Model

In addition to the van der Waals surface described in Section 3.2 two other molecular surfaces can be defined by using a sphere with radius \( r_p \), called probe, which is rolled over the van der Waals surface. The first one is the Solvent Accessible Surface (SAS) defined by Lee and Richard in [LR71] as the topological boundary of the molecule spheres with their radius increased by \( r_p \).

The other is called Solvent Excluded Surface (SES) after
Table 1: Display speed of the generated ribbons

<table>
<thead>
<tr>
<th>Name</th>
<th>PDB ID</th>
<th>Atoms</th>
<th>Amino Acids</th>
<th>Secondary Structures</th>
<th>fps/#tris rendered</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV-1 Capsid</td>
<td>1A43</td>
<td>573</td>
<td>72</td>
<td>13</td>
<td>29/163,000 &gt;30/1,600 30/160,000</td>
</tr>
<tr>
<td>Calcium-Bound Triple Mutant Of Carp Parvalbumin</td>
<td>1B8L</td>
<td>806</td>
<td>108</td>
<td>13</td>
<td>21/236,500 &gt;30/2,300 22/145,000</td>
</tr>
<tr>
<td>Oxidized Dsba Crystal Form II</td>
<td>1A2J</td>
<td>1,449</td>
<td>188</td>
<td>21</td>
<td>12/413,000 &gt;30/4,045 22/145,000</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>1IDR</td>
<td>1,867</td>
<td>253</td>
<td>33</td>
<td>9/565,000 &gt;30/5,520 24/33,000</td>
</tr>
<tr>
<td>Ribonuclease</td>
<td>1IBA</td>
<td>1,925</td>
<td>248</td>
<td>38</td>
<td>9/580,000 &gt;30/5,725 24/31,000</td>
</tr>
<tr>
<td>Antigen-Antibody Complex</td>
<td>2HRP</td>
<td>6,852</td>
<td>875</td>
<td>178</td>
<td>3/2,200,000 25/22,000 25/25,000</td>
</tr>
<tr>
<td>Ribosomal Subunit</td>
<td>1FGJ</td>
<td>51,770</td>
<td>3,897</td>
<td>346</td>
<td>1/5,500,000 10/54,000 10/54,000</td>
</tr>
<tr>
<td>Ribosomal Subunit</td>
<td>1FFK</td>
<td>64,268</td>
<td>6,484</td>
<td>29</td>
<td>1/7,200,000 7/70,000 7/70,000</td>
</tr>
</tbody>
</table>

The reduced surface is generated by virtually rolling the probe along the intersections between the atoms on the surface of the molecule. In general this will create some concave spherical triangles, where three spheres intersect and some torus sweeps at the intersection between two spheres. These created faces must be converted into Bézier patches.

Figure 12: Reduced surface for the protein 1a3j. On the right the SAS edges are shown, too.

3.5.1. Generating Bézier Patches

The solvent excluded surface can be built with only two different kinds of patch types: a spherical triangle and a torus sweep. These two patch types can easily be generated by subdivision surfaces, as shown in detail below.
3.5.1.1. Sphere Triangle The spherical triangle is a triangle bounded by three vertices and three circle segments \( s_0, s_1, s_2 \) connecting these vertices. Each of these circle segments \( s_j = (c, p_0, p_2) \) (Figure 13 a) can be converted into a Bézier curve through the following dependencies ([SR04]):

- \( p_0 \) and \( p_2 \) have Bézier weights \( w_0 = w_2 = 1 \).
- The control point \( p_1 \) lies on the line \( m \) with \( m = \frac{1}{2} (p_0 + p_2) \), distance \( d = \frac{1}{2m} \) from \( c \) and weight \( w_1 = \cos \alpha \).

From these three Bézier curves a patch generating a spherical triangle can be obtained by connecting the three control points as shown in Figure 13 b).

3.5.1.2. Torus Sweep The torus sweep is a quad bounded by four circle segments \( s_0, s_1, s_2, s_3 \) as shown in Figure 13 c). From these four segments only two are needed for the sweep, namely \( s_0 \) and \( s_1 \). From the generation process we know that:

\[
(p_3 - p_0, (p_0 - p_1) \times (p_2 - p_1)) = 0
\]

Therefore it is possible to rotate and translate the segments so that \( s_1 \) lies in the \( xz \)-plane, \( s_0 \) lies in the \( xy \)-plane and the mid point of \( s_0 \) lies in the point of origin. Now it is very easy to generate \( p_3, p_4 \) and \( p_5, p_8 \) by just scaling \( s_0 \) to the desired radius given by \( p_1 \) and \( p_2, p_0, p_1, p_2, p_0, p_1 \) and \( p_8 \) all have weight one. \( p_3, p_4 \) and \( p_5 \) have a weight according to the formula above. The patch can be obtained by building four quads like in the Figure 13 c).

The generated patches can, for example, be subdivided by the de Casteljau algorithm ([dCS9, dC63]).

\[ \text{Figure 13: (a) Calculating a Bézier Curve (b) Generating a triangular Bézier patch (c) Generating a quadrangular Bézier patch} \]

3.5.2. Results for Surfaces

As shown in Table 2 the complete running time of the surface calculation is more or less linear to the number of atoms. This is optimal since the number of faces in the reduced surface is in \( O(n) \) as shown by Sanner in [San92]. The number of patches created for the surface is in \( O(n) \), too. The most time-consuming part of the surface calculation is the generation of the reduced surface, followed by the handling of special cases and the generation of the subdivision mesh. But the overall computation time is short enough for even the biggest molecules.

\[ \text{Figure 14: Generated solvent excluded surface for protein 1a3i. On the left with adaptive tessellation and on the right with full resolution.} \]

4. Conclusion and Future Work

In this paper we presented fast visualization concepts for molecular structures. These concepts are based on multi-resolution meshes, subdivision surfaces and modern graphic boards extensions. Therefore, these concepts are not limited to visualization problems in biology but can also be used in other fields. For example, the billboard idea might be used in astronomy, when visualizing thousands of stars or in other particle systems ([BDST04, HE03]). Also, the combined BReps together with a sophisticated method to generate the base mesh is extremely efficient and can easily be adapted for other purposes.

By using these techniques, we developed the tool Bio-Browser, which can be used for interactive navigation through the proteins as well as for the production of high quality images for publications or presentations. In the past this was a two step procedure: first navigate to the desired position with one tool, then saving this position and load it to another tool for generating the high quality image.

\[ \text{Figure 15: First impression of a semantic lens} \]

The next steps will implement new visualization modules. For example, the bond model will be replaced by a new model using billboards and fragment shaders, similar to the space-fill model. Another useful feature will be a semantic lens which will help the user to switch between the different
visualization styles and to combine two or more styles in one molecule, e.g., combining a half-transparent surface with the ribbon and stick style. Figure 15 gives an idea of what it will look like. With all the necessary visualization features available at interactive rates, even for extremely complex proteins, we will be able to 'upgrade' the bare visualization tool into a tool which allows a researcher to embed and to retrieve additional bio-information, typically textual, as well as hyperlinks directly into the 3D molecular structures – similar to what current web browsers offer in the textual domain. Speaking of web browsers: of course, the BioBrowser is also available as a browser plug-in for visualizing molecules over the Internet.

Acknowledgment

The support of this project by the German Research Foundation (DFG) under grants Fe431/4-3 and Fe431/6-1 is gratefully acknowledged.

We also want to thank Sven Havemann for his CBrep Library 'Mesh-Lib' and the support he gave us, and Lars Reusche, who helped us with the subdivision-surfaces in the surface model.

References


submitted to EUROGRAPHICS - IEEE VGTC Symposium on Visualization (2005)


