Visualization of Molecules with Positional Uncertainty

Penny Rheingans[1] and Shrikant Joshi[2]

[1] Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County
Baltimore, MD 21250, USA
rheingan@cs.umbc.edu
http://www.cs.umbc.edu/ rheingan/index.html
[2] Celcore Inc.
3800 Forest Hill Irene Rd.
Memphis, TN 38125, USA
sjoshi@celcore.com

Abstract. Designing new and better chemotherapeutic compounds requires an understanding of the mechanism by which the drugs exert their biological effects. This involves consideration of the geometry of the active site, determination of the geometry of the drug, and analysis of the fit between them. This problem of drug-substrate fit, often called the docking problem, can be greatly influenced by uncertainty in the position of drug side chains. Traditional molecular graphics techniques fail to capture the distribution of likely atom positions. This paper describes a range of techniques for showing atom positions as probability distributions that more completely describe parameters which determine fit.
. . .

1 Introduction

The field of data visualization offers graphical solutions to interpreting large amount of data. Molecular graphics is one such area where a few good images can greatly enhance insight into the problem. Designing new and better chemotherapeutic compounds (drugs) requires an understanding of the mechanisms by which the drugs exert their biological effects. The structure of the drug molecule and that of the substrate molecule influence the binding property and the effectiveness of the drug docking process.

The docking problem involves determining the position and orientation of the drug molecule with respect to the substrate such that the energy of interaction of the two is minimized. The interactions are very specific in nature. Even a slight change in the side chain of a drug molecule can inhibit or enhance the interaction. Unfortunately, side chain position is not always well determined. The exact locations of atoms in a side chain can be subject to a substantial amount of uncertainty. The ability to visualize not only the best estimate of atom locations but also the range of likely atom locations would be a valuable
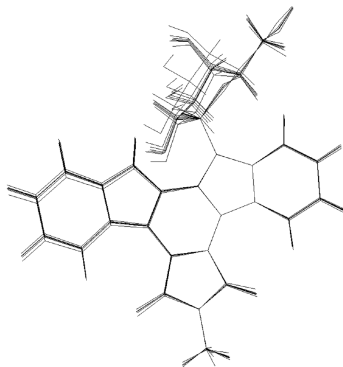
Fig. 1. Line drawing representation of eight family members.

tool in the study of structure and interactions between drug and substrate in the docking problem.

1.1   Drug Design

The drug docking problem is crucial in enzyme catalysis, antigen-antibody interactions, drug design, and advanced materials development. The particular driving focus addressed here is the mechanism by which anti-tumor drugs exert their biological effects [11]. Synthetic olegonucleotides are used as model DNA systems to study the drug binding phenomenon. High field NMR is used to examine each solution (drug and substrate mixture) at the molecular level. Initial geometries of molecules are constructed and these geometries are optimized using energy minimization calculations.

The drug molecule under study has a side chain which orients differently for each complex. Knowledge of the physical extent of the uncertainty of side chain position can be of great help in modifying, re-designing, and understanding the interaction of this drug with the substrate. Different versions of the drug molecule with identical atoms and connectivity but potentially different atom positions form a family of molecules. Each family member shares much of the molecule structure, differing mainly in side chain position and orientation. Previously a family of molecules was shown in wire-frame form in one image, with each bond shown by a line. Figure 1 shows an example of this technique. In sections where the members of the family have identical structure, multiple lines lie in the same place forming a heavy line. In sections where family members differ, a fan of options shows for the side chain. This representation is fairly successful at showing which sections have positional certainty and which do not, but is less successful at showing structural details. In sections with high positional uncertainty the display gets too cluttered to convey structure clearly. Even in

the more stable sections the 3D structure can be difficult to see due to the lack of depth cues.

2   Related Work

Common discrete and static techniques used to display molecular structure include wire frame depictions of molecule bonds, ball and stick representations showing both atoms and bonds, ribbons to trace out the backbones of complex molecules, and space-filling models showing each atom's van der Waals radius [4]. Molecular dynamics are commonly displayed using animations. In situations of limited dynamics, such as in crystal structures, line drawings showing atoms as thermal-motion probability ellipsoids have been used [2]. A few researchers have created more continuous representations of molecules using volume techniques to represent electron density [3]. One focus of the research described was to combine discrete and continuous techniques into a single unified view.

The representation of data with uncertainty is an active area of research in the field of Geographic Information Systems (GIS), using such approaches as the use of color, transparency, line width, blur, haze, interactive probing, and animation [[1], [5], [10], [12]]. Within the field of visualization, researchers have also recognized the importance of uncertainty data to accurate visualization and have been active in developing new representation techniques. Their techniques have included iterated function system fractal interpolation, fat surfaces, and sonification of uncertainty [14], [6], [7]].

3   Techniques for Showing Positional Uncertainty

Both drug and substrate are sufficiently complex that judging the quality of the fit is difficult. In order to more clearly illustrate the potential fit between a drug molecule and the target site, a visualization should show the range of possible drug configurations.

3.1   Discrete Representations

A common molecular graphics technique which addresses the problem of insufficient shape cues in the wire frame form is the ball and stick representation. In this representation, atoms are represented as spheres of appropriate color and size, while bonds are represented as shaded rods or tubes. Normally the ball size is a function of the van der Waals radius for the atom and color is chosen to conform to the standard molecular graphics conventions. Since solid objects are used to represent the atoms and bonds, one gets better depth cues from the rendered image. The solid objects and the accompanying shading of the atoms and bonds create an object boundary for the atoms. These occlusion boundaries convey a strong relative depth information about the superimposed surfaces.

Unfortunately, simply putting ball and stick representations of all family members in the same image produces a very cluttered image. See Figure 2a.
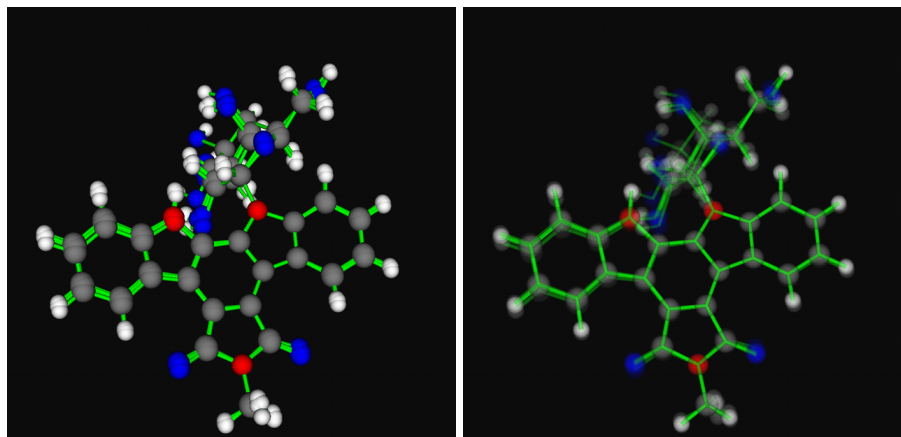
Fig. 2. Eight family members shown, a) each with full opacity, b) each with an opacity of 0.125. Carbon atoms are grey, nitrogen red, oxygen blue, and hydrogen white.

Worse yet, the least certain parts of the molecule produce the most solidly packed areas of the images. A slight modification of the ball and stick technique would assign to each family member a fraction of available energy corresponding to its likelihood. See Figure 2b.

In addition to making the image less crowded, this technique is effective in separating the image into two regions. The opaque part represents the more static portion of the drug molecule, whereas the more transparent part shows portion of the molecule that is dynamic. This gives sections of the molecule a visual weight corresponding to the likelihood that the section is actually at the displayed location. It also gives a sense of motion blur to this part of the image.

3.2   Likelihood Volume Representation

In the previous techniques each member of the molecule family is displayed as a separate and discrete possibility. Such representations could imply that these are the only configurations possible. It would be more accurate to consider molecule family members as samples of a presumably continuous possibility space. In such a space, the position of each atom is a probability distribution. In order to create this sort of continuous visualization, the first step is to create the likelihood volume that describes the probability of an atom being at each location. The original data set is discrete and unstructured; the likelihood volume is continuous and structured.

This transformation from an unstructured data set to a structured one can be accomplished using a variation of splatting. Westover introduced splatting as an object-order method for directly rendering structured volumes [13]. Using this method, 2D footprints of voxels of a 3D volume are composited onto a 2D image. Vtk includes a generalization of Westover's splatting technique which
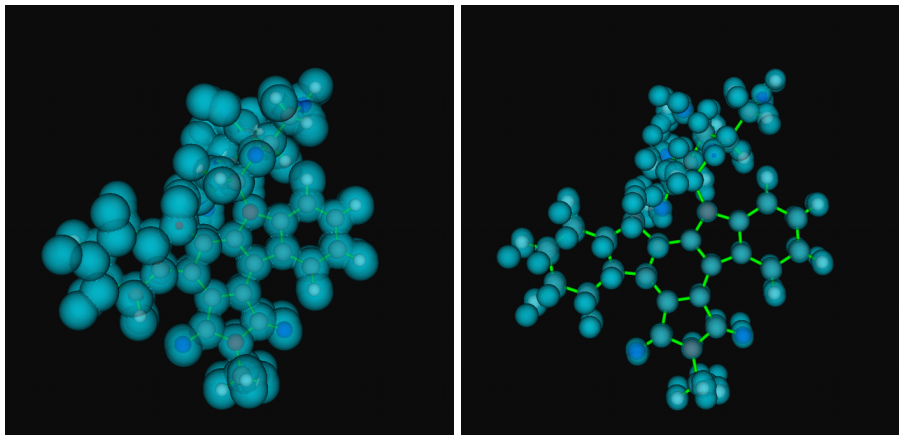
Fig. 3. Gaussian splatting and isosurface extraction. a) isolevel value of 0.9, b) isolevel value of 0.95.

can be used to sample unstructured points into a structured point set. Vtk's GaussianSplatter composites 3D footprints into a 3D volume by accumulating the maximum splat value at each voxel. The splatting function used here is the uniform Gaussian distribution. The function can be cast into the form

$$SF(x, y, z) = se^{-f(r/R)^2} \qquad (1)$$

where s is a scale factor, f is the exponent factor controlling decay rate, r is the distance between the point and the Gaussian center point, and R is the radius of influence of the Gaussian.

Using this technique, an overall function that predicts the position of the drug molecule can be constructed. Each atom from each of the molecule family members is splatted into a 3D structured points object to produce aggregate information. This resultant aggregate is the value of the function that indicates the likelihood of an atom occupying that location.

An isosurface can be extracted from this volume, using the marching cubes algorithm [8]. This isosurface contains the volume in which atoms are most expected to be located. See Figure 3a. The blobby portion in the top center part of the image is the area having high uncertainty (the dynamic side chain of the molecule). The choice of isolevel employed to extract the isosurface has a great influence on the characteristics of the final image. A good isolevel is one that can extract the structure and geometry of the static part of the molecule, keeping the blobby representation of the dynamic part of the molecule unchanged. Alternatively, the isolevel can be chosen to give atoms isosurfaces which approximately match the van der Waals radius of the atom. Atoms in stable sections appear to be spheres. Atoms in more dynamic sections become less regular isosurfaces which reflect the likely range of positions. Figure 3b shows this representation.
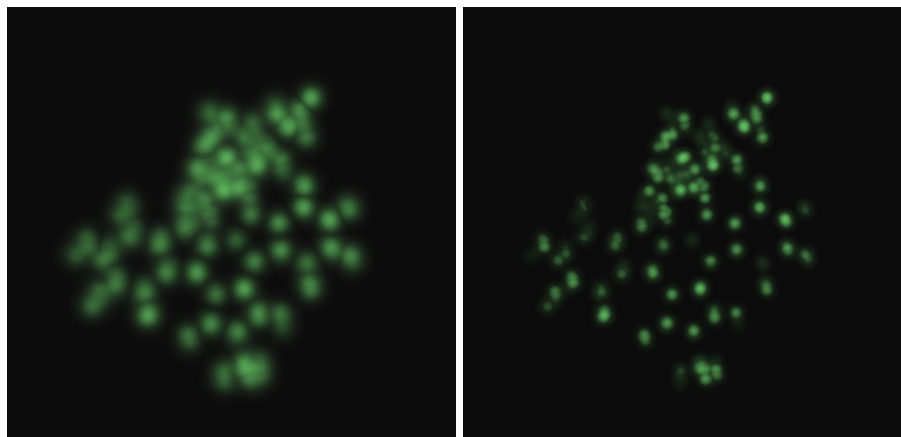
Fig. 4. Volume rendering of likelihood volume. a) emphasizing areas of greatest likelihood, b) greater contribution from less likely areas.

A different approach is to render the likelihood volume using direct volume rendering. This approach can show the whole distribution of the volume, rather than just those portions exceeding a threshold likelihood. Techniques to volume render molecular properties expressed as a structured point dataset have been previously used by Goodsell [3]. By volume rendering, we can see the interior of the volume associated with the object. This can give a better insight as to the distribution of the position and orientation of the molecule along with the uncertain region in which the side chain may exist. The choice of transfer function to convert voxel values to resultant color and opacity values can greatly change the characteristics of the image produced. Figure 4 shows volume renderings of the likelihood volume using different transfer functions. Figure 4a emphasizes the areas where atom occupation is probable. Figure 4b includes a greater contribution of areas where atom occupation is much less likely.

4   Implementation

The techniques described here are implemented using the Visualization Toolkit (vtk) [9]. Vtk is a freely available 3D graphics and data visualization toolkit which can be used for a wide variety of computer graphics and data visualization problems. This software employs an object-oriented approach using C++ with an optional Tcl/Tk interface. The package can be extended with the construction of additional classes. To facilitate the exchange of information between Tcl and C++ methods, automatically generated wrapper code written in C is used. Applications using this toolkit are easy to implement and portable across UNIX and PC platforms.

5   Summary and Conclusions

The field of data visualization offers graphical solutions to interpreting large amounts of data. Molecular graphics is one such area where a few good images can greatly enhance insight into the problem. The goal of this project was to create these few good images. The wire frame and the ball and stick model give a good starting point for visualizing the complex. By putting all the copies of the drug molecule into one image and controlling the opacity, the resulting image can be divided into stable and dynamic regions. Splatting provides an straightforward and useful way of converting the unstructured point dataset of atom positions into a structured point set describing a likelihood volume. The choice of isolevel for the isosurface from the splatted data can be altered to create visualizations with slightly different goals. Alternatively, the likelihood volume can be directly volume rendered, creating a more holistic representation of the entire likelihood distribution.

6   Acknowledgements

References

1. Kate Beard and William Mackaness: Visual Access to Data Quality in Geographic Information Systems. Cartographica, vol. 30, no. 2-3. (1993) 37-45
2. Michael Burnett and Carroll Johnson: ORTEP-III: Oak Ridge Thermal Ellipsoid Plot Programs for Crystal Structure Illustrations. Oak Ridge National Laboratories Report ORNL-5138. (1996)
3. S. D. Goodsell, S. Mian, and A. J. Olson: Rendering volumetric data in molecular systems. Journal of Molecular Graphics, vol. 7. (1989) 41-47
4. S. D. Goodsell: Visualizing Biological Molecules. Scientific American, vol. 267, no. 5. (1992)
5. Adrian Herzog: Modeling reliability on statistical surfaces by polygon filtering. In The Accuracy of Spatial Databases, Michael Goodchild and Sucharita Gopal, eds.. Taylor&Francis, London, (1989) 209-218
6. Suresh Lodha, Alex Pang, and Robert Sheehan, and Craig Wittenbrink: UFLOW: Visualizing Uncertainty in Fluid Flow. Proceedings of IEEE Visualization '96. IEEE Computer Society Press (1996) 249-254
7. Suresh Lodha, C.M. Wilson, and Robert Sheehan: LISTEN: sounding uncertainty visualization. Proceedings of IEEE Visualization '96. IEEE Computer Society Press, (1996) 189-196
8. W. E. Lorensen and H. E. Cline: Marching Cubes: A High Resolution 3D Surface Construction Algorithm. Computer Graphics, vol. 21. (1987) 163-169

9. W. Schroeder, K. Martin,. and W. E. Lorensen: The Design and Implementation of an Object-oriented Toolkit for 3D Graphics and Visualization. Proceedings of IEEE Visualization '96 IEEE Computer Society Press, (1996) 93-100
10. David Theobald: Accuracy and bias in surface representation. In The Accuracy of Spatial Databases, Michael Goodchild and Sucharita Gopal, eds. Taylor&Francis, London, (1989) 99-106
11. R. M. Wadkins and D. E. Graves: Interactions of Anilinoacridines with Nucleic Acids: Effects of Substituent Modifications of DNA-Binding properties. Biochemistry, vol. 30, no. 17. (1991) 4277-4283
12. Matthew Ward and Junwen Zheng: Visualization of Spatio-Temporal Data Quality. Proceedings of GIS/LIS '93. (1993) 727-737.
13. L. Westover: Footprint Evaluation for Volume Rendering. Computer Graphics, vol. 24. (1990) 367-376
14. Craig Wittenbrink: IFS Fractal Interpolation of 2D and 3D Visualization. Proceedings of IEEE Visualization '95. IEEE Computer Society Press, (1995) 77-84