Editor: Dorée Duncan Seligmann
Avaya Labs

Molecular Visualization: A Microcosm of the E-Revolution

Susana Maria Halpine
Candle Light Productions

I remember my amused surprise when a colleague at a pharmaceutical company walked by with what looked like a small paper toilet in her hand (complete with lid). She was one of the rising stars in protein chemistry, and she was holding a paper model of the protein porin. Porin molecules are present on the surface of bacteria such as E. Coli. Like the hole in Ringo's pocket in the movie *Yellow Submarine*, the "barrel-shaped" protein forms a small channel in the membrane allowing nutrients and waste to respectively pass in and out of the cell (see http://www2.kenyon.edu/depts/biology/BMB/Chime/porins/pormast.htm). Determining the shape of the porin molecular complex was crucial to understanding its function. Similarly, knowing its form and function might guide the development of porin-targeted drugs against virulent diseases like syphilis.

Molecular visualization software programs display a 3D arrangement of molecules in an interactive format, allowing researchers to examine known molecules as well as theoretical molecules in drug development—and even recreated as paper models, if necessary.

In the last two decades, biochemistry has shifted its emphasis from bonds and connectivity to surfaces and interactions. (See the "Life's Building Blocks" sidebar for background information on biochemistry.) Molecular visualization, at the intersection of computer graphics and chemistry, accelerated this transition with interactive displays of 3D molecular models on personal computers. The free access to both molecular databases and visualization programs has enhanced our understanding of basic biological concepts. Molecular visualization also encompasses a thriving Internet community that's constantly honing and dynamically revising software implementations in response to user feedback. There's no question that computer technology has revolutionized biomedical fields such as biochemistry, genetic engineering, and molecular biology. We can now study critical biomolecules in their folded form rather than as a string of amino acids. For example, we can examine insulin in its native form, the physical shape that regulates glucose uptake in our cells (see Figure 1).

Biochemistry will undoubtedly play an increasing role in our society; cloning, DNA identification, and genetically modified foods are already standard news items. Artists will facilitate the exploration and understanding of this new world, just as European artists made Christian theology visible to the common man. Molecular visualization will provide the conceptual tools and graphic metaphors indispensable for representing this sophisticated information. Reflecting this process, artists have already begun to exploit these programs for artistic expression.

History of an e-community

The first molecular visualization program developed for widespread use on personal computers was Mage, a vector-based program created by David and Jane Richardson. It was distributed freely on a diskette with the *Protein Science* journal in 1992. Although widely used, it lacked the depth of perception needed for 3D modeling. Roger Sayle's RasMol followed, capable of generating interactive raster-based models and publication-quality images. Chime, a Netscape plug-in developed at MDL Information Systems, was derived from RasMol and brought molecular visualization to the Internet community. Protein Explorer is now being developed to make features from RasMol and Chime accessible to nontechies. The current educational technology uses Chime-based tutorials in which interactive schematic molecules, usually embedded in a page of text, demonstrate key concepts by clicking on preset JavaScript buttons.

The most widely used visualization software packages are freeware. This is due, in no small part, to Roger Sayle, the "patron saint" of molec-

1070-986X/01/$10.00 © 2001 IEEE

Life's Building Blocks

I don't know at what point in my education I realized that the 20 protein amino acids—of body-builder and health-food fame—were actually microscopic building units. For example, just as a plumber selects an elbow joint to redirect a pipe, the amino acid Proline might be used to put a kink or turn in a protein chain. Other units, through electrostatic attraction of acid-base pairs, for example, function like molecular Velcro to stabilize the folded shape. I can't say when I was struck with this understanding, but I've been hooked on biochemistry ever since. Protein chemistry became a miniature Legoland with one important difference: these building-block assemblies are the basis of life as we know it.

Essentially, life is a self-sustaining chain of chemical reactions. The day-to-day cellular functions are largely staffed and maintained by strings of proteins folded up into interdependent shapes. These proteinaceous shapes operate within a pinball-like environment: one protein interacts with another altering its shape, which causes it to interact with a third molecule, and so on. This chain reaction eventually results in a critical molecule becoming active and triggering a biological activity. Each protein string, hence its innate folded form, is determined by a line of DNA code in our genetic makeup.[1] A mutation in the DNA dictates the insertion of a different amino acid, or construction unit, in the protein. Replacing any unit changes the protein's shape and thereby changes its function. For example, straight hair becomes wavy, or in more ominous cases, an enzyme looses its lock-and-key activity and blocks a healthy biological function.

The importance of mapping the human genome has been somewhat overstated.[2] To a large extent, the new map gives only the sequence of DNA bases without showing where instructions for one protein end and another begin. Having information on DNA without its correlation to protein synthesis is like having an imprint of a roadmap without knowing your starting point, where you want to go, or what the map is describing along the way. The human genome map is a good start, but the resulting protein structures will be the real clues to health and disease.

References

1. S.M. Halpine, "Peptide Synthesis Animation," *Life, The Science of Biology*, sixth ed., W.K. Purves et al., eds., W.H. Freeman and Co., New York, 2000, http://www.whfreeman.com/purves6edemo/con_index.htm?99xex.
2. E.K. Wilson, "Gearing up for Genomic's Protein Avalanche," *Chemical and Eng. News*, vol. 78, no. 39, Sept. 2000, pp. 41-44.

ular visualization. Sayle developed RasMol as part of his PhD in computer science, using ray-tracing algorithms to render shadowed space-filled models in record time (see http://www.umass.edu/microbio/rasmol/pershist.txt). After he posted the software and its source code on the Internet free of charge, RasMol became one of the most widely used programs around the world. It was immediately adopted by commercial enterprises and research universities, including universities in developing countries that couldn't afford high-priced commercial software packages. Sayle, however, remained an impoverished research assistant at the University of Edinburgh.

Eventually GlaxoWellcome UK—realizing that some recompense was due him because of the widespread use of his program within their own pharmaceutical company—provided him with a temporary job so he could continue to improve his software. Sayle's spirit of selflessness—sharing the fruits of his research with the world for the greater good—continues to imbue the molecular visualization field. His gift of RasMol to the scientific community is held high as an example toward which molecular visualization devotees endeavor. In fact, part of the etiquette within these highly competitive biomedical fields involves maintaining a free exchange of information.

According to Sayle, several software companies adopted the RasMol code for commercial purposes (see http://www.umass.edu/microbio/rasmol/faq_ras.htm). But his spirit of cooperation infiltrated even commercial developers: MDLI reworked RasMol's code and released the Chime

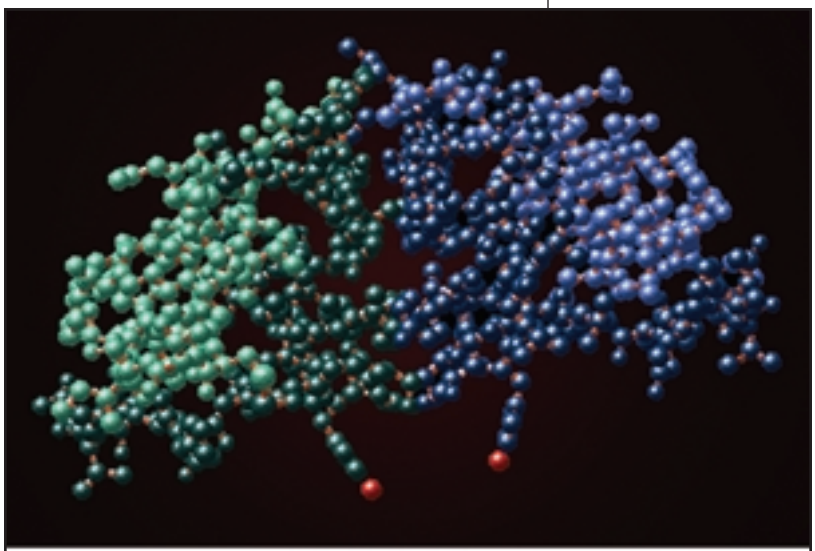

Figure 1. Insulin molecule as a 3D model (PDB# 4INS) and a sequence of letter-coded amino acids, by artist-biochemist Susana Maria Halpine.

Figure 2. **VRML Gallery of Molecular Landscapes and Sculpture***, created by artist Ken Eward (http://www2.kenyon.edu/depts/biology/eward/gallery/galleryintro.html).*

plug-in free of charge. (Chime isn't open source, however, and much effort goes into finding work arounds to make it compatible with browsers other than Netscape 4.)

If Sayle is molecular visualization's patronsaint, then Eric Martz is its leading evangelist. Collaborating with an international network of devotees, Martz maintains several educational Web sites, hosts a dedicated email list, and coordinates the development of new visualization software—all while teaching immunology at the University of Massachusetts, Amherst (see http://www.umass.edu/molvis/martz). His aim is to draw the uninitiated into the fold. To this end, Martz guides bumbling e-list novices with the utmost patience and is hard at work coordinating the development of Protein Explorer for the technically challenged. The child of artist-parents, Martz is clearly enthralled by molecular visualization's amalgam of artistic and scientific sensibilities. As he states on his Web site, "I have always longed to create beauty as well as truth. RasMol and Chime enable me to do both."

Atomic coordinates

Molecular visualization software programs create an interactive display of any molecule for which a 3D structure is available. The software packages require an atomic coordinate file specifying the XYZ position of every atom in a given molecule. Researchers can usually predict the structures of small molecules from their elemental components. However, predicting the 3D structure of large molecules isn't reliable because of the large number of possible angle combinations (see http://www.stanford.edu/group/pandegroup/Cosm). To use a plumbing analogy, it would be necessary to predict the direction of dozens of elbow joints, in addition to hundreds of iterations of the other 19 construction units.[1] Although theoretical molecular models can be useful, true 3D macromolecular representations must be based on empirical data, usually from X-ray diffraction analysis. Once we know the atomic coordinates, we can vividly conceptualize the macromolecule using molecular visualization programs. Through interactive and command-line interfaces, the visualization programs allow shading and rotation of the models as well as access to more in-depth molecular analysis.

Visualization programs aren't restricted to viewing proteins. They can also display inorganic salt crystals and organic molecules including fullerenes and DNA (see http://dragon.klte.hu/~gundat/povraya.htm). One researcher has adapted a program to plot 3D data sets (see http://www.bio.umass.edu/biology/kunkel/pub/kin_files). However, because of these macromolecules' size and complexity, proteins are the predominant focus of molecular visualization. In fact, it would be nearly impossible to study large 3D protein structures without these programs.[1] A chemist at a leading university, for example, admitted to a sense of frustration at describing biochemical reactions in traditional text journals. Changes in conformation, angle of rotation, and lock-and-key docking of molecules are cumbersome to define in words, but we can readily convey these reactions with animated graphics.

Molecular visualization, therefore, has revolutionized the teaching and understanding of biochemistry by using interactive motion graphics. Researchers from around the world submit the atomic coordinates of known proteins to databases such as the Brookhaven Protein Databank (stored in PDB format). Having your protein structure accepted in a databank provides the same sense of accomplishment as discovering a plant variety or a new star; the databanks display names of the research team members beside the molecular model. This public forum lets other researchers corroborate, enhance, and use the scientific findings. Anyone can access the files and view the molecular models using free, downloadable visualization programs including Kinemage, RasMol, Chime or Protein Explorer as well as demos of commercial programs such as MacMolecule, PovChem, and ChemDraw (see http://molvis.sdsc.edu/visres). By

manipulating a 3D molecular model on a computer screen, researchers and students can begin to visualize the Lego-like nature of protein chemistry.

Future hybridizations

My personal goal is to merge molecular visualization with Macromedia Flash Web animation. This would let users manipulate molecular models on a time line and export them in 3D Swift format for interactive viewing in Flash players. This cross-platform player is free, open source, and installed on millions of computers worldwide where it has been thoroughly field tested. Visualizing molecular models on a Flash player will allow access for a greater audience and might avoid many of the programming issues still plaguing Chime tutorial developers. To achieve this integration, visualization programs could potentially export in Swift format. Alternatively, 3D Web animation packages, such as Electric Rain Swift 3D, might have import capabilities for the Drawing Exchange Format (DXF) in the near future, allowing animation of DXF molecular models converted from the PDB format (see http://www.biology.washington.edu/molecular and http://mcn.simplenet.com/graphics/moldxfw). Regardless of the technical approach, however, animating molecular reactions on the Web will greatly expand the impact of molecular visualization.

The molecular visualization phenomenon has launched an exciting cross-fertilization within education. Many artists already integrate visually complex molecular models within their work (see Figure 2). Conversely, some biochemistry professors are accepting test answers "explained in words and/or drawn as pictures."

Molecular visualization has even begun to infiltrate even K–12 education (see http://www.nyu.edu/pages/mathmol/K_12.html). Introducing 3D models at early stages of intellectual development, in combination with physical models, might make the spatial aspects of molecular biology second nature to future high school students (see Figure 3). Certainly, the interactive, visual nature central to biochemical modeling will begin to attract students with abilities usually assigned to more conventional artistic pursuits.

Sayle freely admits to little formal training in molecular biology: "Much of what I know [of biochemistry] has been a result of conversations ... with people much brighter than I." In this regard, Sayle is an unprecedented example of how someone with a completely different skill set can make a tremendous contribution in another field. The technical assets intrinsic to molecular visualization—namely interactive graphics and electronic delivery—make it accessible to a vast audience in science education. It's now time to use this technology to instruct scientists and nonscientists alike in the fundamental aspects of biological life. With biochemistry influencing everyday decisions in grocery stores, jury boxes, and voting booths, we should no longer keep the excitement of molecular visualization to ourselves. **MM**

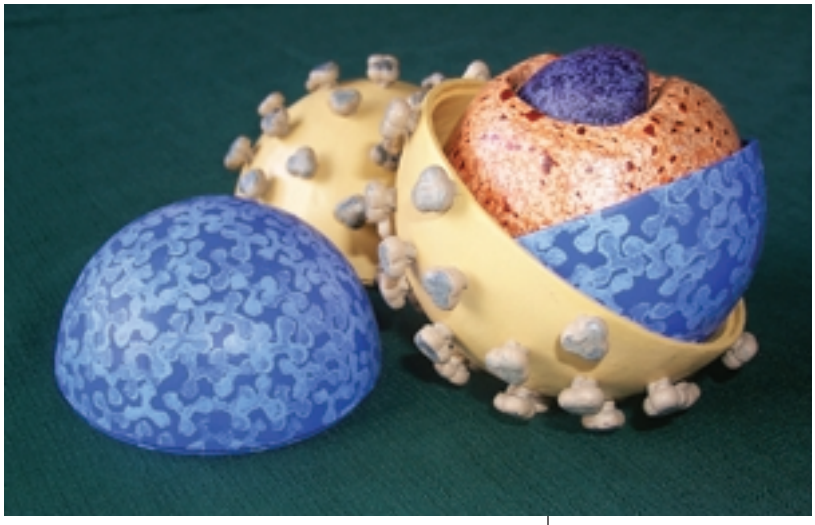

Figure 3. Three-dimensional model of HIV virus, made by biochemist Teresa Larsen of the The Foundation for Scientific Literacy (http://www.scientificliteracy.org).

Acknowledgments

I thank Linda Roux and Stacy Marsella for their comments on this article.

References

1. E.K. Wilson, "Gearing up for Genomic's Protein Avalanche," *Chemical and Eng. News*, vol. 78, no. 39, Sept. 2000, pp. 41-44.

Readers may contact Halpine at Candle Light Productions, 8640 Gulana Ave. J-1016, Playa del Rey, CA 90293, email shalpine@earthlink.net, or visit http://home.earthlink.net/~shalpine for related URLs.

Contact Artful Media editor Dorée Duncan Seligmann at Avaya Labs, Room 2B-315, 600-700 Murray Hill, NJ 07974-0636, email doree@avaya.com.