Looking at Molecules—An Essay on Art and Science

David S. Goodsell*[a]

Introduction

As you leaf through the pages of this journal, you will be presented with all manner of illustrations. The fields that we work in—chemistry, biochemistry, molecular biology—are particularly amenable to illustration. After all, most of our work boils down to understanding the arrangement of atoms in space. So, we use sketches, illustrations, and advanced computer graphics to examine our particular arrangement of atoms, and to present our atoms to other researchers and to the public.

Take a moment to think about how amazing this is. We are able to synthesize pictures of the arrangement of atoms in matter, both animate and inanimate. We are revealing an invisible world, where the concept of vision has no meaning, in a way that is interpretable by our senses.

Tools of the Trade

Today, manipulation and illustration of atomic structure is fast, easy, and remarkably fun. The computer forms a facile interface to atomic data (of which there is enough for a lifetime of exploration), allowing us to delve into structures, compare them to others, and ultimately create pictures of what we have discovered.

If you are a chemist, you will be dealing with hoards of chemical diagrams. Fortunately, there is a long tradition of how to represent these diagrams. Even more fortunately, these diagrams are codified to a sufficient extent to allow computers to understand the rules. By sticking to this tradition, every chemist will understand your atoms and their covalent connections. We even can use some shorthand to simplify these already parsimonious diagrams: the familiar hexagon with a circle inside represents a more complex underlying aromaticity, we don't have to put a "C" at every carbon position in a hydrocarbon, and often, we don't even have to show the hydrogens at all. Turn-key software running on personal computers allows us to churn out clear, precise versions of these diagrams.

If you are a biochemist, you will probably be dealing with larger fish, so you'll need to use a bigger hook. Protein and nucleic acid structures have hundreds or thousands of atoms—each essential—so we need advanced tools to look at them. Over the years, researchers and illustrators have tried many ways of simplifying these structures. The goal is to simplify the representation to make the picture more interpretable, throwing out information that is not needed. We have to be careful, though, not to throw out too much, or the picture will be useless. This is the artistry and pedagogy of molecular illustration.

Three basic representations, shown in Figure 1, have withstood the test of time. Covalent diagrams, such as wireframes or ball-and-stick diagrams, are direct extensions from the chemical tradition, showing the underlying chemistry of the molecule. Everything is there for the exploring, but often it can be too much, and the image becomes a sprawl of overlapping lines. Spacefilling representations (and other surfacing or solvent accessible variants) combat this sprawl by obscuring all interior detail. These diagrams look at the size and shape of molecules, and are great for thinking about interactions between different molecules. Their strength, however, is also their major limitation: they are the best way to show the shape of the molecule but all of the interesting connections inside are hidden. Ribbon diagrams round out our representational bag-of-tricks. They strip away all of the distracting atomic information and present the topology of a protein or nucleic acid chain. Ribbon diagrams are arguably the most beautiful of the representations, and thus find their way into most popular accounts of biomolecular structure. When you think of the structure of DNA, the familiar ladder diagram, first shown in that famous *Nature* paper,[1] is the picture that comes to mind. Today, you can find the curly protein ribbons and arrows on the cover of nearly any journal.

Researchers are Doing It for Themselves

Molecular graphics software is available off the shelf, so today many images of molecules are created in the laboratory where they are being studied (Figure 2). These programs range from basic molecular viewing to advanced rendering. My favorite place to start is the program RasMol (links to all of the software described in this section may be found at the Protein Data Bank: http://www.pdb.org). With two clicks of the mouse, you have a molecule on the screen and you can start exploring. Other popular programs, such as Molscript and Raster3D, allow the creation of high-quality images for publication. They require a bit more practice, but the results are well worth the effort.

Because interactivity is so important in biomolecular graphics, clever researchers have also developed a number of powerful methods to incorporate interactive figures into publications and presenta-

[a] *Prof. D. S. Goodsell*
Department of Molecular Biology
The Scripps Research Institute
10550 N. Torrey Pines Road, La Jolla
CA 92102 (USA)
Fax: (+1) 858-784-2860
E-mail: goodsell@scripps.edu

 DOI: 10.1002/cbic.200300747

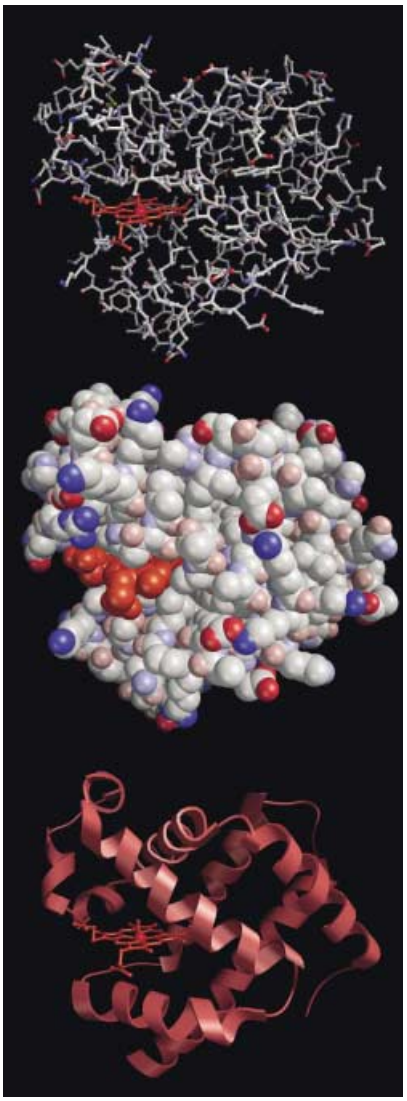


Figure 1. *Ever since the first protein structure was solved, researchers have looked for ways to display and explore these complex molecules. Today, three basic representations are commonly used. The first, shown at the top, is a modification of traditional chemical diagrams that uses lines or cylinders to show the covalent structure of the molecule. This is the workhorse of biomolecular research, and is particularly useful when viewed on an interactive computer graphics system, allowing manipulation of the three-dimensional image. The spacefilling representation, shown in the middle, was designed by Linus Pauling[5] to reveal the bulk of the molecule. If we were able, somehow, to see a molecule, we might expect it to look something like this (without, perhaps, the shiny highlights!). Ribbon diagrams, codified for proteins by Jane Richardson,[6] radically simplify the structure, showing the topology of the chain and revealing regions of specific secondary structure. These are ideal for thinking about protein folding and evolutionary relationships. These images were created with the Python Molecule Viewer (http://www.scripps.edu/~sanner/python/pmv), by using coordinates from entry* 1mbn *at the Protein Data Bank (http://www.pdb.org).*

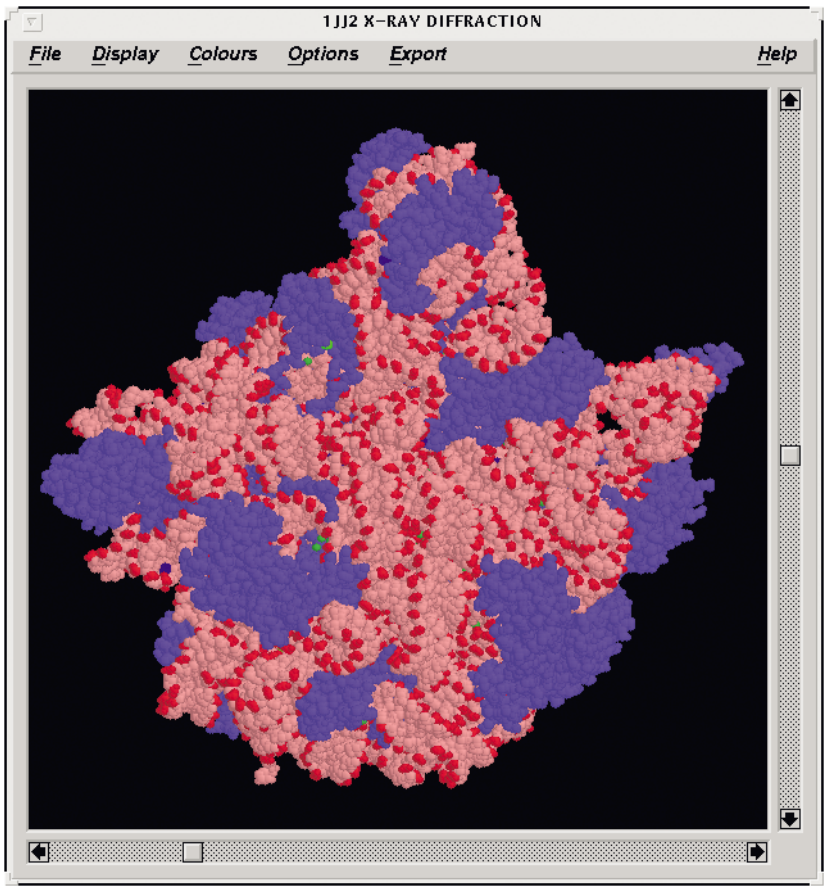


Figure 2. *Amazingly powerful molecular graphics programs are available on personal computers. A snapshot of RasMol (http://www.bernstein-plus-sons.com/software/rasmol) in action is shown here. It allows interactive manipulation of this enormous ribosome structure, which includes nearly a hundred thousand atoms, and a variety of options for coloring and representation. With RasMol and other similar programs, molecular structures are at anybody's fingertips. Coordinates were taken from entry* 1jj2 *at the Protein Data Bank.*

tions. David and Jane Richardson pioneered the approach with the program Mage. It allows authors to create a "kinemage", an animated, interactive figure. The reader is given some basic freedom to interact with the figure, but not—and this is important—total freedom. The author designs the kinemage to pave the way for the reader, picking the best representations to display the particular topic at hand, and removing options that might cause the reader to get lost.

More recently, the Chime plug-in for Netscape has moved interactive molecular graphics into the world wide web. With Chime, authors can place interactive windows into web pages, allowing perfect integration between interactive exploration of the structure and any explanatory text or links to other sites.

State-of-the-Art

Molecular graphics is an active field, undergoing significant changes. A major thrust in current development is to improve the modularity and reusability of graphics methods. The idea is to create a collection of modular tools that can be connected together to perform custom functions. We no longer create ponderous, monolithic programs that do everything, instead, we take an atomic coordinate manager and have it feed data to a molecular dynamics tool, and hook it up to a molecular viewer so we can watch what is happening.

For example, Michel Sanner is using the Python programming language as the glue to connect these diverse modules.[2] He has created a visual programming

Figure 3. *A visual programming environment is used to link services on the WWW with a molecular viewer. Here, a WWW-based resource, ConSurf, provides values for evolutionary conservation of amino acids in a protein. They are then used to modulate the color of the chain representation in the Molecular Viewer. The Network Builder allows many different options to be explored: pulling in data from a variety of resources and using it to modulate any of the parameters of rendering. Figure provided by Michel Sanner, the Scripps Research Institute.*

Figure 4. *Rapid prototyping methods are being used to create physical models of large molecules. This model of chymotrypsin (notice the deep, dark specificity pocket) was created by using machinery from Z-Corporation. The machinery lays down thin layers of gypsum powder and sprays on colored glue with an ink-jet printer, building the model up layer-by-layer from bottom to top. The hand-sized model allows direct, tactile exploration of the protein structure.*

environment that allows the networking of different modules to form custom applications. The use of Python is the key advantage over previous data network methods (such as AVS), because it allows, with modest coding effort, the "wrapping" of other applications to allow them to communicate their results to the network. The center of this network is PMV, the Python Molecule Viewer, which takes care of all of the graphics tasks. Figure 3 shows an example of using a WWW-based service to direct the rendering of molecules. In other applications, symmetry information, molecular dynamics, electrostatics, quantum mechanics, and many other sources of data have been linked into the networks.

Molecular Models

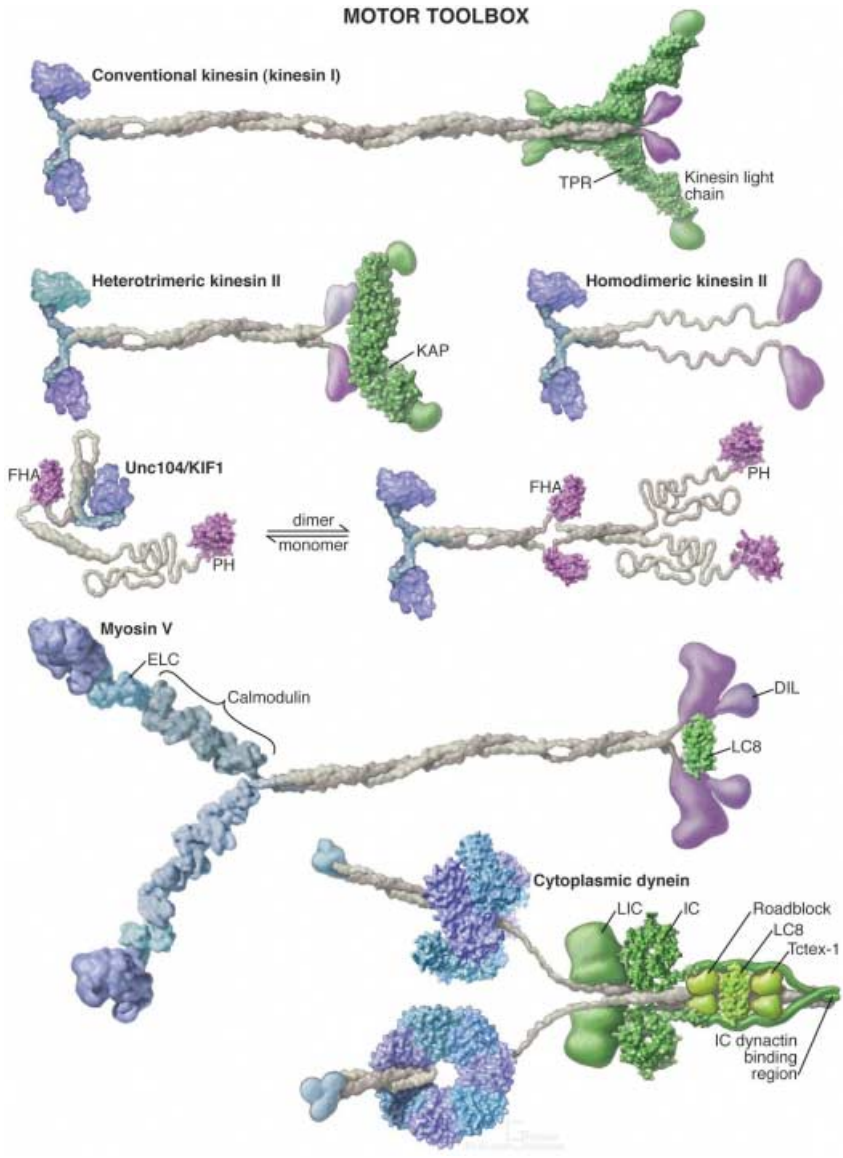
Wooden or plastic ball-and-stick models play an indispensable role when teaching about covalent bonding and stereochemistry. These models—an adult form of Tinkertoys—provide hours of fun and insight. But until recently, physical models were limited to the world of chemistry. Snap-together models get too unwieldy when you try to build trypsin or a ribosome.

Researchers are now borrowing technologies from engineering to create physical models of proteins and large molecular assemblies. These technologies were designed for rapid prototyping, to build and test car parts and the like. They build up a model one layer at a time by squirting on tiny dots of molten plastic, by cutting and gluing together layer after layer of paper, or by gluing down thin layers of gypsum powder with an ink-jet printer. After a little programming, three-dimensional printers are now being used by biologists to provide a tangible alternative to computer graphics. The result is a perfect three-dimensional model of any desired molecule (Figure 4). These models are irresistible: researchers and students alike are finding that it is impossible not to handle and explore them.

Collaboration with an Artist

Now that these exciting tools are available on our desktop, why would a scientist need to go to an artist? Desktop molecular graphics are superb for the representations that they are designed to create, but only for those. Artists are essential in cases where the subject is just too complex for routine graphics. Artists figured prominently in the first few decades of protein structure, when scientists and readers struggled to understand for the first time the complex three-dimensional arrangement of atoms of myoglobin and lysozyme. The collaboration of Irving Geis and Richard Dickerson is a milestone: working as one, they created illustrations that brought this new world to life.[3]

In other cases, the artist/scientist collaboration can benefit both the image being created and the science being presented. A perfect example is the illustration by Graham Johnson, shown in Figure 5. He was approached by Ron Vale to create a picture showing the current state of knowledge of motor proteins. Graham was able to synthesize structural information from many sources: from atomic structures, ultrastructural information from microscopy, and, for some pieces, simple molecular weights. These are combined into a coherent, interpretable picture, with very little fabrication. The animations by Drew Berry (Figure 6), or my own paintings of mole-



cules in cells (Figure 7) are other examples of the utility of this approach.

Note that this process is a two-way street. The collaboration is an opportunity for the scientist to compile an exhaustive set of information: after all, Graham needed to know what information was available for every part in each structure, and Drew needed to know where and how fast the parts of the polymerase were acting. The artist/scientist collaboration forces us to reveal both the parts that are well understood, and the parts that require further scientific scrutiny.

One Size Does Not Fit All

Today, we can easily turn out colorful, accurate illustrations to support our research projects. These are perfect to accompany our journal articles, but we may run into trouble when we move to other audiences. We often make the mistake of using the same imagery when faced with less technical audiences, and lose their interest in the process. The

Figure 5. *Ron Vale went to Graham Johnson to create this atlas of motor proteins for an article in* Cell.[7] *This is a challenging task, since parts of these proteins are known in great detail, and parts are less well defined. After digging up all currently-available structural data, Graham developed a style that shows high detail where warranted and smoother representations for domains and coiled-coils where only amino acid sequences and molecular weights are known.*

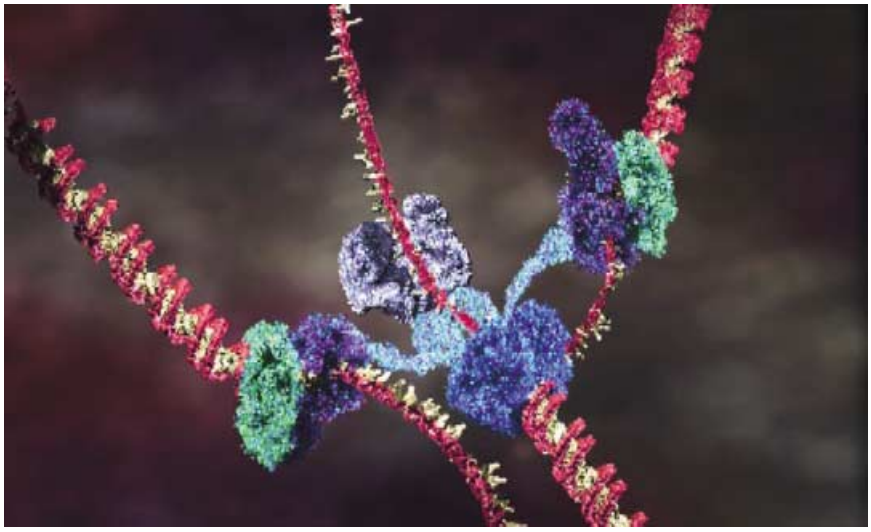


Figure 6. *Drew Berry tackled the challenge of animating a DNA replication fork in action. Based on the many atomic structures of the players, he assembled this model of the DNA replisome. As with all complex models, parts are based firmly in data, and parts, such as the geometry of the polymerases, primases, and clamps in the overall complex, are still the subject of speculation and study. Since the overall process – -most notably, the discontinuous priming and replication of the lagging strand—is modeled accurately, the animation is an excellent teaching tool. It is also a boon to science, forcing researchers to look at* all *of the steps in this process, and develop tests to decide if Drew's particular model is completely correct.*

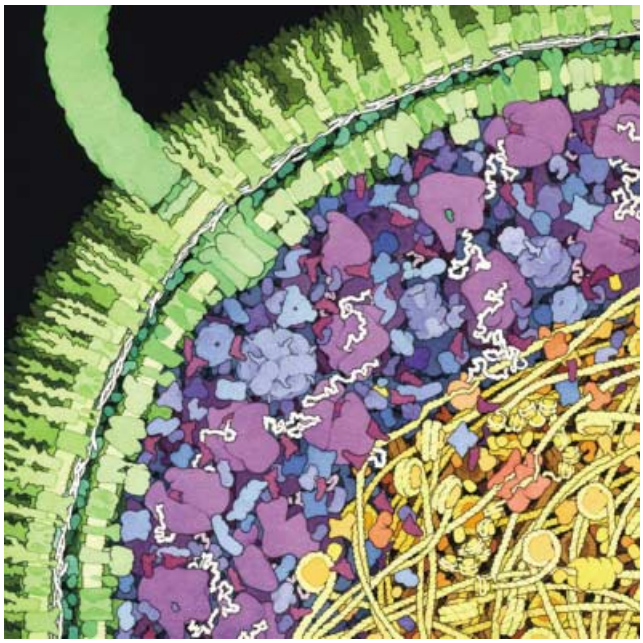


Figure 7. *In my own paintings, I use art to reveal a world that is difficult to explore directly by experiment.[8] Microscopy reveals the rich world of cellular ultrastructure, but falls short of resolving individual molecules. X-ray crystallography and NMR spectroscopy, on the other hand, reveal individual molecules in splendid detail, but taken completely out of their biological context. The painting of* Escherichia coli *shown here combines ultrastructural data with molecular structure, synthesizing a picture of the many molecules inside this living cell. The two-layered cell wall is shown in green, with a large embedded flagellar motor complex. Inside is the cytoplasm, dominated by ribosomes, shown in purple, and enzymes, shown in blue. In the lower right corner is a tangle of DNA, shown in yellow, and the many proteins involved in its replication and transcription. This painting was the introductory figure in an article about prokaryotes by Hoppert and Mayer.[9]*

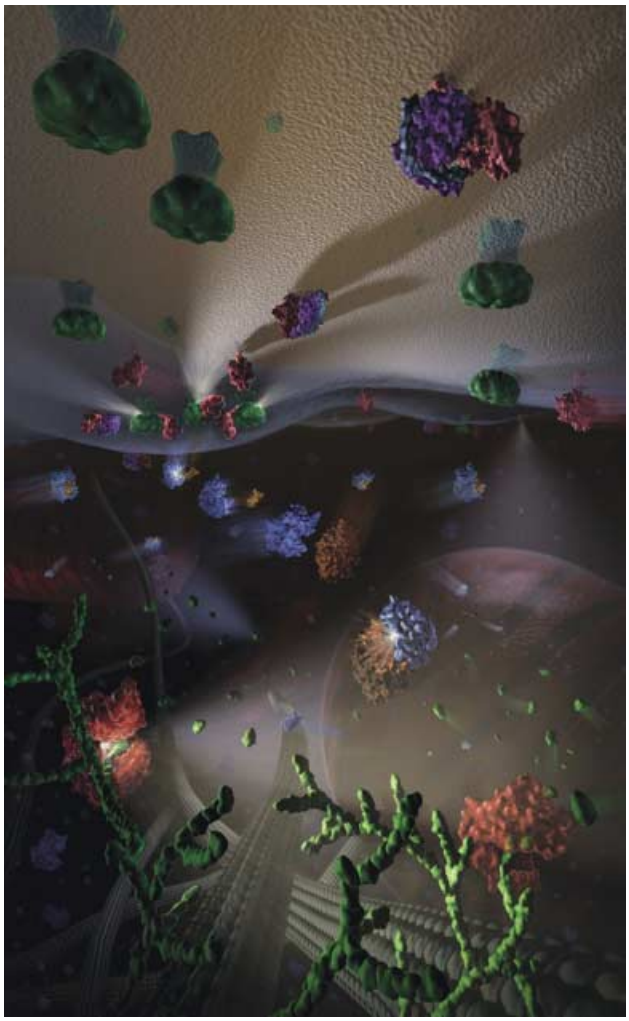


Figure 8. *Artists are often called upon to create a dynamic image for the cover of a publication or other high-profile applications. In these, the artist can pull out all the stops: panoramic views, dramatic lighting, and careful highlighting combine to draw the viewer in. This image by Graham Johnson, created for the cover of June 2003* Accounts of Chemical Research*, depicts the β-adrenergic signaling pathway leading to glycogen breakdown.*

entire field of scientific illustration has emerged to fill this need: the need to create illustrations for textbooks, science museums, and science programming. These illustrators perform a careful balancing act: paring down the scientific data to its essential core, and then creating compelling imagery to present it.

It is easy to make the mistake of judging all scientific illustration according to one standard. In a recent Commentary in *Nature*,[4] Julio Ottino strongly criticized the many colorful illustrations currently in use for science outreach. What was not addressed in his commentary, however, is the context of illustrations. When we create an illustration for a journal article, factual accuracy and lack of distortion are essential, since we are using the illustration to support our results. But when we decide to create illustrations for a more general audience, the goals, and the rules, change. We are no longer creating an illustration to support a body of data. Instead, we are creating an illustration to pique a reader's interest, or to present an entire scientific concept in one easy-to-swallow bite, or simply to sell magazines. The artist is given far greater leeway and control, and the results, like Graham Johnson's image in Figure 8, can be exciting and engaging.

Conversely, I occasionally find myself acting as a harsh critic of illustrations in journal articles (for instance, there are colors other than saturated red, yellow, blue, and green). However, this criticism is misplaced. Journal illustrations are meant for one function, and one function only: to support the findings of the authors. If the illustrations succeed in this task, they are a complete success, even if the color combination doesn't fit the current cultural aesthetic.

Molecules for Everyone

Perhaps the most exciting aspect of the revolution in molecular graphics is the general accessibility of advanced structural research. The Protein Data Bank is a perfect example. There, researchers, students and teachers, and the general public can explore the latest results in anthrax structure, prion structure, molecular motors (Figure 9) and a host of other topical subjects. Molecular structure is no longer the exclusive domain of mainframe

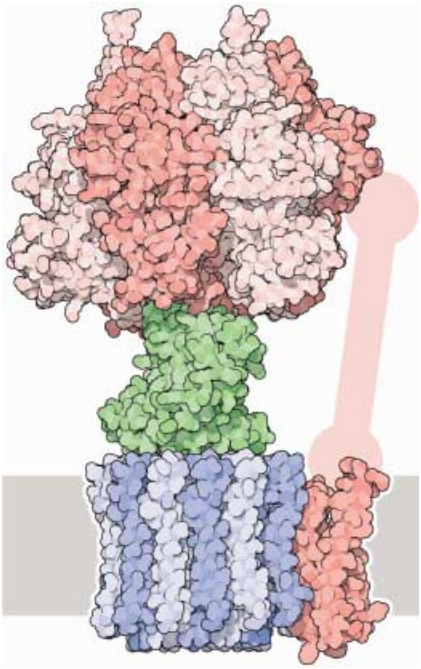


Figure 9. The molecular world is full of surprises, and always rewards the curious explorer. ATP synthase is a perfect example. It has two separate rotary nanomotors, each powered by a different fuel. The motor at the top, colored red, is powered by ATP, and the motor at the bottom, colored blue, is powered by an electrochemical gradient. Since they are tethered together, the cell can use the electrochemically-powered motor to drive the upper one, forcing it to act as a generator instead of a motor, so that it builds new ATP fuel. Coordinates were taken from entries 1c17 *and* 1e79 *at the Protein Data Bank.*

hardware, costly graphics engines, and specialist researchers. The latest structures are just a few clicks of the mouse away. So start browsing!

Keywords: molecular graphics · proteins · protein structures · scientific illustration

[1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737 – 738.
[2] M. F. Sanner, *J. Mol. Graphics Mod.* **1999**, *17*, 57 – 61.
[3] B. P. Gaber, D. S. Goodsell, *J. Mol. Graphics Mod.* **1997**, *15*, 57 – 59.
[4] J. M. Ottino, *Nature* **2003**, *421*, 474 – 476.
[5] W. L. Koltun, *Biopolymers* **1965**, *3*, 665 – 679.
[6] J. S. Richardson, *Adv. Protein Chem.* **1981**, *34*, 167 – 339.
[7] R. D. Vale, *Cell* **2003**, *112*, 467 – 480.
[8] D. S. Goodsell, *Trends Biochem. Sci.* **1991**, *16*, 203 – 206.
[9] M. Hoppert, F. Mayer, *Am. Sci.* **1999**, *87*, 518 – 525.

