

## Putting proteins in context

Scientific illustrations bring together information from diverse sources to provide an integrative view of the molecular biology of cells

David S. Goodsell

As scientists, we often focus our attention on the details. For instance, in my graduate work with Richard Dickerson, I spent years synthesizing small pieces of DNA and trying to determine the structures of their complexes with small inhibitors. These days, I use the computer to try to predict these types of macromolecule-inhibitor interactions in HIV. It's fascinating work, and work that can fill an entire lifetime of study. But it's also disconnected from the things that got me excited about science: looking at pond water through microscopes, collecting native southern California insects, trying to find Neptune with the telescope, and all of the other fascinating things that call to a curious mind. In our day-to-day research, it's easy to lose touch with the larger scientific context.

Two decades ago, I began a project to reconnect with the biological context of my main love: molecular biology. I set a goal: to create an illustration of a 100 nm cube taken from a living cell, showing every molecule at the proper size, shape, concentration, and location. There isn't any way to look at this scale level experimentally. Microscopy doesn't have the

resolution to see atomic structures in cells, and atomic methods like X-ray crystallography provide a view of the purified molecules, completely removed from their cellular context. I started with *Escherichia coli*, which was at the time (and still today) the most data-rich organism. Remarkably, there was enough data from biochemistry, structural biology, and microscopy to support such an illustration [1] (Fig. 1). I've spent the years since then creating other illustrations of cellular environments, attempting to integrate diverse data into a few images that show the cellular context of molecular biology [2–5].

The process of researching and planning this type of integrative illustration is arguably the most exciting aspect of the task. As you might imagine, there are many decisions that need to be made. Many of these decisions are based firmly on data: for instance, in the *Escherichia coli* illustrations, the number and structure of ribosomes and the exact length and sequence of the genome are well determined. Other aspects, however, are more difficult to study, so an informed decision is necessary. How is the peptidoglycan arranged?

(There are several competing models.) How compact are the proteins in the replisome? (I've shown them in a tight complex – other models have them loosely associated.) It's a humbling process, uncovering the many gray areas in our knowledge.

The Internet has revolutionized this process, allowing instant access to diverse forms of data. PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>) allows fast access to the primary research reports, uncovering information at all levels. Of particular value are the many schematic illustrations sprinkled through biochemistry research reports, showing the details of the particular biomolecular interactions being studied. When I was researching the *Escherichia coli* images in Fig. 1, these types of schematics were essential for sorting out the detailed arrangement of molecules in the cell wall and nucleoid. Interactome studies are attempting to do this in a more systematic way, but they don't yet provide the personal knowledge that is infused into the journal figures. Quite remarkably, finding structures for each of the components is the easiest part. The Protein Data Bank [6, 7] allows instant access to tens of thousands of molecular structures, including nearly all of the central molecular machines found in cells. In cases where atomic structures are not available, structures from electron microscopy are available in the EMDataBank (<http://www.emdatabank.org>), or directly in published reports.

Creation of scientific illustrations such as these presents the artist with

### Keywords:

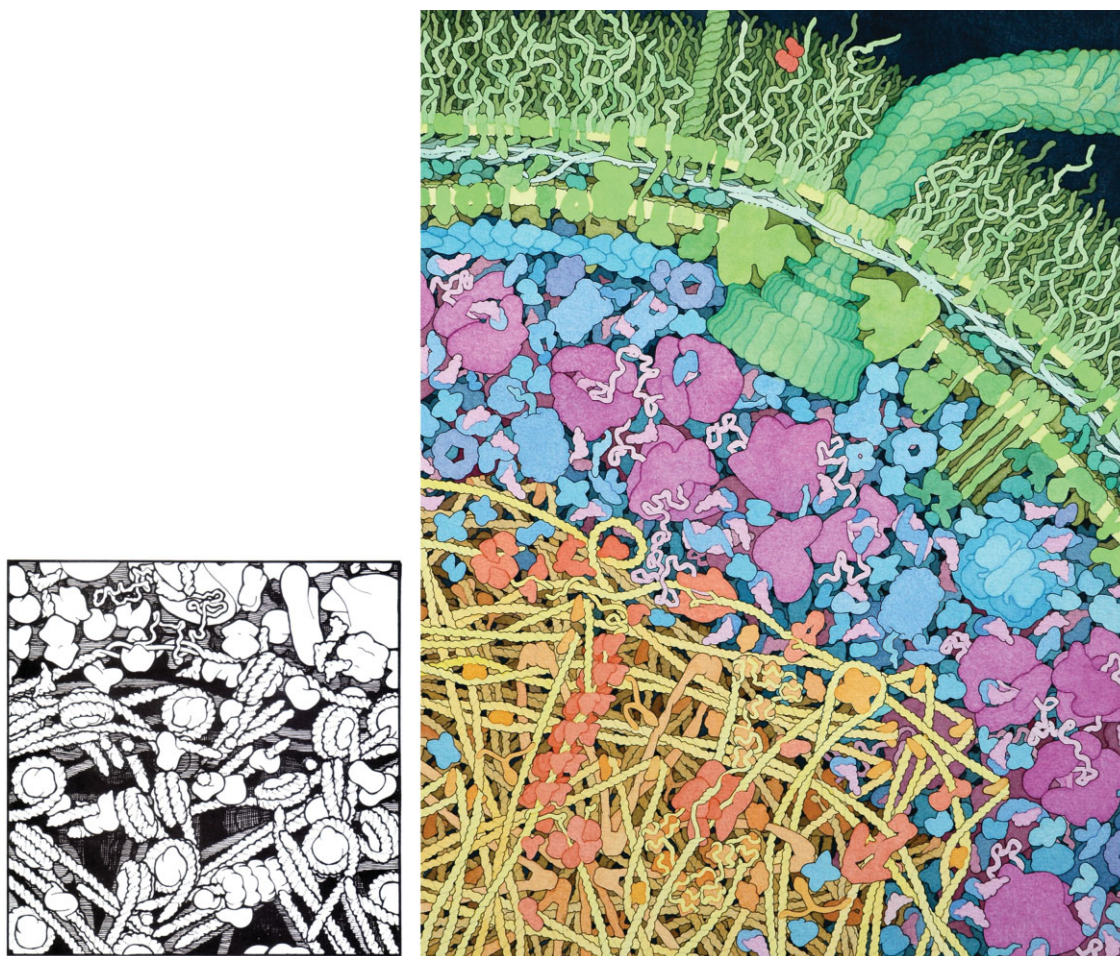
■ molecular cell biology; scientific illustration

DOI 10.1002/bies.201200072

Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA

### Corresponding author:

David S. Goodsell  
E-mail: [goodsell@scripps.edu](mailto:goodsell@scripps.edu)

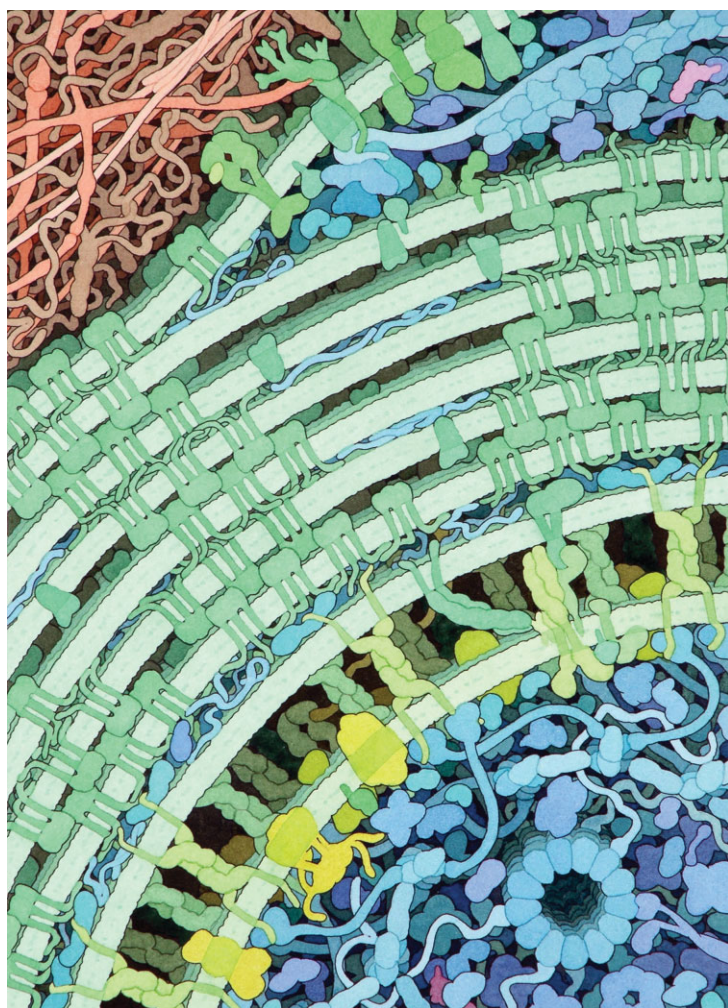


**Figure 1.** *Escherichia coli*. These two cross-sections show a portion of an *Escherichia coli* bacterium, including all macromolecules. Small molecules like ATP, glucose, and water are omitted for clarity. On the left, an illustration from 1991 showing the nucleoid, with a random arrangement of DNA helices, causing many to be clipped in the cross-section (reproduced with permission from [1]). On the right, a recent illustration employs artistic license in the arrangement of DNA and membranes, to minimize confusing artifacts. The coloring scheme is designed to highlight the functional compartments of the cell, with the cell wall and flagellar motor in green, soluble enzymes and other proteins in turquoise, ribosomes and tRNA in magenta, DNA in yellow, and proteins in the nucleoid in orange (reproduced with permission from [12]).

When I present these illustrations, both in educational settings and research settings, viewers are typically surprised by the density and complexity of the scenes. One of my major goals is to provide a bridge, linking molecular biology with cell biology. The illustrations are designed to remind us that molecules are performing their jobs in a complex environment, which has properties that often differ from purified molecules. The most obvious feature is the crowding, a property that is only recently being studied and appreciated in molecular biology laboratories [9]. For instance, the crowded environment effects the way proteins interact, favoring complexes relative to separate molecules. Proteins must also be highly specific in their interactions, to avoid unproductive interactions with their many neighbors [10]. For me, the illustrations also make me think about the role of copy number. There are many ribosomes busily doing their jobs, but for a repressor, there may be only a

some engaging boundaries – boundaries that are not often present for other forms of art. Most importantly, a scientific illustration must be an accurate representation of the science. The term “accurate”, however, is slippery, and in practice, scientific illustrators must employ all manner of approximations and artistic license to create a useful rendering [8]. For instance, cells contain many fibrous molecules, such as DNA and actin, and large planar membranes. A random cross-section through a cell would intersect with many of these elements, creating clipped views that

would be visually confusing. The artist, however, can arrange the molecules in orientations that reflect their actual locations, but that will minimize these types of visual artifacts. Compare the two illustrations in Fig. 1. The small black-and-white illustration is one of the first that I created, and includes many clipped DNA strands. This gives the impression of many small, disconnected pieces. The more recent illustration in color artificially arranges all the DNA strands so that nothing is clipped, and giving a better impression of a long, continuous DNA strand.



**Figure 2.** Myelin sheath. This cross-section shows a nerve axon at the bottom, surrounded by a Schwann cell, which wraps around the axon and forms multiple layers of insulation. The illustration highlights the many infrastructural proteins that support the complex interaction of these two cells (reproduced with permission from [12]).

small number. As described in a thought experiment by Halling [11], looking across a population of cells, a molecule with low copy number may be much more prevalent in some cells, or even absent in others, as a result of statistical variation.

The process of creating this type of illustration forces us to look at the larger context of our subject, and to think more deeply about the many aspects that don't normally enter into our day-to-day research. When planning the illustrations for the second edition of the *Machinery of Life* [12], I had several epiphanies. First, I came to appreciate the role of unstructured proteins in cell function. The traditional view has proteins folding into perfect globular structures, but it is becoming apparent

from many lines of research that proteins often act as unstructured chains [13]. Second, I came to appreciate the role of infrastructure in cell structure and function. Reading through the many papers on neuronal structure, I found that if a protein is in a particular place, there will be an infrastructure to hold it there, and a separate regulatory network to make sure it's put there at the proper time. Both of these features are included in the myelin illustration in Fig. 2. Many unstructured chains form the infrastructure of the extracellular matrix, between the myelin membranes and in the axon cytoplasm, and flexible chains are important in the function of the voltage-gated channels. As for infrastructure, it can be argued that all of the molecules in this painting

serve as infrastructure to ensure that the three membrane-bound proteins shown in yellow (an ion pump and two ion channels) are in the right place and environment to propagate a nerve signal.

It is my hope that these illustrations will inspire researchers and educators to create similar studies. This is a great way to organize what is known, and identify areas that need more study. For instance, I've had the opportunity to work with Dan Klionsky, a researcher in the field of autophagy, in this capacity. Over the past decade or so, we have worked together to create a series of illustrations encapsulating the data that is known at the particular time [14]. It has been an exciting project, watching the picture develop as more pieces of the puzzle are added each year.

### Acknowledgments

This work was supported in part by the RCSB Protein Data Bank (NSF DBI 0829586). The author has no conflicts of interest in this work.

### References

1. Goodsell DS. 1991. Inside a living cell. *Trends Biochem Sci* **16**: 203–6.
2. Goodsell DS. 2009. Neuromuscular synapse. *Biochem Mol Biol Educ* **37**: 204–10.
3. Goodsell DS. 2009. *Escherichia coli*. *Biochem Mol Biol Educ* **37**: 325–32.
4. Goodsell DS. 2010. Mitochondrion. *Biochem Mol Biol Educ* **38**: 134–40.
5. Goodsell DS. 2011. Eukaryotic cell panorama. *Biochem Mol Biol Educ* **39**: 91–101.
6. Berman HM, Westbrook J, Feng Z, Gilliland G, et al. 2000. The Protein Data Bank. *Nucleic Acids Res* **28**: 235–42.
7. Berman H, Henrick K, Nakamura H. 2003. Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **10**: 980.
8. Goodsell DS, Johnson GT. 2007. Filling in the gaps: artistic license in education and outreach. *Plos Biol* **5**: 2759–62.
9. Ellis RJ. 2001. Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci* **26**: 597–604.
10. Sear RP. 2004. Specific protein-protein binding in many-component mixtures of proteins. *Phys Biol* **1**: 53–60.
11. Halling PJ. 1989. Do the laws of chemistry apply to living cells? *Trends Biochem Sci* **14**: 317–8.
12. Goodsell DS. 2009. *The Machinery of Life*. New York: Springer.
13. Wright PE, Dyson HJ. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **293**: 321–1.
14. Goodsell DS, Klionsky DJ. 2010. Autophagy: the art of autophagy – the Cvt pathway. *Autophagy* **6**: 3–6.