

Depicting protein structures as schematics

Jonathan Markert,^{1,*} Karolin Luger,^{2,3,*} Hayun Lee,^{4,*} Haidai Hu,^{5,*} Nicholas M.I. Taylor,^{5,*} Javier Fernandez-Martinez,^{6,7,*} Michael Rout,^{6,*} and Nathan Alder^{8,*}

As described in the last issue, there are many ways to depict protein structures. In this fourth TrendsTalk of the Special series: Scientific figure development, we focus on different approaches from authors of *TIBS* articles for creating protein structure schematic figures. They touch on questions such as: how do you decide how to represent your protein schematic? What program(s) do you prefer for generating such figures and why? What aspects do you consider when generating a protein schematic figure?

Contributing to this article are Jonathan Markert and Karolin Luger, authors of ‘Nucleosomes meet their remodeler match’ [1] (Figure 3); Hayun Lee, co-corresponding author of ‘Creating memories: molecular mechanisms of CRISPR adaptation’ [2] (Figure 2); Haidai Hu and Nicholas Taylor, the first and corresponding authors of ‘Structural basis of torque generation in the bi-directional bacterial flagellar motor’ [3] (Figures 1 and 5); Javier Fernandez-Martinez and Michael Rout, co-corresponding authors of ‘One ring to rule them all? Structural and functional diversity in the nuclear pore complex’ [4] (Figure 1); and Nathan Alder, corresponding author of ‘Mitochondrial compartmentalization: emerging themes in structure and function’ [5] (e.g., Figure 1).



Jonathan Markert and



Karolin Luger

Using simple sketches to describe complicated molecular details

When making a figure that summarizes the mechanism of a protein, it is important to maintain as much of the key structural components while also making the figure simple to understand. If the figure contains too much detailed information, such as secondary structure, then the figure becomes too complicated for the general audience. However, if over-simplified cartoons are used, such as ‘blobs’ or other stylized shapes, there is too much of a disconnect between the figure and what is really happening at the mechanistic level, in particular as it pertains to interaction between components. To find the right balance between these two, it is important to first identify what readers should be focusing on. Is the goal to show how a conformational change in a protein allows it to interact with a substrate? Or is it to show that a specific region of the protein interacts with a substrate? Or is the goal to show that the protein changes the structure of the interacting partner upon binding? All of these determine the level of detail that must be included.

The next steps are to make schematic representations of the protein (or individual domains if these are of particular interest). One approach is to project the surface representation of each protein (or domain) of interest, which allows the overall shape to be maintained, but removes complicated structural features. It is then important to identify regions within that sketch that are important to the overall model. These regions can then be highlighted by various means to focus the reader’s attention. Finally, we strongly believe that a figure must be esthetically pleasing and balanced; this affects color choices and layout.





Hayun Lee

Designing illustrations of protein structures

Protein structure is ubiquitous in modern biochemistry literature but can be complicated for non-experts. To communicate effectively with broad readers who may unfamiliar with the field, it is important to depict structures as simply as possible so that figures can be easily followed. These figures often convey key conclusions of a paper, so it is essential to consider whether the design of protein schematics is understandable and clearly summarize the concepts.

When I generate protein schematics for figures, I emphasize the main points of protein structures without getting bogged down in the details. I try to depict the overall shape of the molecule by drawing the structures as simple shapes. I also rely on other resources, such as graphical abstracts from journal articles, that can give me inspiration for designing my own graphics. Many researchers typically create graphics using Microsoft PowerPoint or Adobe Illustrator. I prefer to use Illustrator, which has great features, like the pen tool and direct select options, that are easy to use and make objects highly manipulatable. Although the program is famous for having a steep learning curve, it is worth the effort to be able to have precise control over the objects in your drawing with high-quality resolutions. Other resources (YouTube, Google, books) are available to jump-start the learning process and lots of practice will definitely help. Another potential barrier is the expensive license, but there are other good open-source options, like Inkscape, that provide similar features to Illustrator.



Haidai Hu and

How we make protein structures understandable to a broad audience

The bacterial flagellum is a complex nanomachine that is used by numerous bacteria to direct their locomotion to a favorable niche. The flagellar motor is powered by stator units. With the resolution revolution of electron cryo-microscopy, we were able to determine the atomic structure of flagellar stator unit and further in its different conformations. Our structural analysis and functional results allow us to elucidate a fundamental biological question: how the stator unit drives the bi-directional rotation of the motor. We decided to present our structural information and mechanistic insights based on the following aspects: (i) reflect our research focus; (ii) combine our and other's novel structural information and its novelty together with previous studies; (iii) be easy to be understand and be accepted by a broad audience, not only scientists in the same field; and (iv) where possible, the protein schematic should not have major distortions compared with the real structures.



Nicholas M.I. Taylor

We used Adobe Illustrator, vector graphics software, to design and generate our figures. There are several advantages when using Illustrator: (i) it is a user-friendly interface where we can customize our own workspace by integrating many useful tools in a way that we feel most comfortable, (ii) the figures generated will never lose resolution, (iii) we can easily share the figures with our collaborators and they can edit and/or correct the figures directly, and (iv) there are many good video tutorials available online from which we can learn how to use the software properly. Nevertheless, free open-source vector graphics alternatives are available as well.



Javier Fernandez-Martinez
and



Michael Rout



Nathan Alder

Illustrating massive assemblies such as nuclear pore complexes

As proteins are the workhorses of the cell, their structures and interactions convey an enormous amount of information about their functionality and mechanism of action. However, this information resides at different spatial levels, from the atomic scale to quaternary structures of interacting protein assemblies. Often, these assemblies can be extremely large, with structures such as the nuclear pore complex (NPC) being composed of millions of atoms and hundreds of protein copies. Informatively representing this detail in a single figure is of course impractical. Instead, we favor deciding in advance what are the minimal spatial levels we need in order to represent the particular message that a given figure is intended to convey. We strive, though certainly don't always succeed, to follow the guidance of arguably the master of information design and data visualization, Edward Tufte, who states that 'graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space' and, moreover, 'clutter and confusion are failures of design, not attributes of information'.

Thus, we try to limit ourselves to illustrating at the scale at which we wish to convey the desired information. For example, in figures that illustrate the functional consequences of a mutation in a protein, we would usually focus on atomic scale representation of how that mutation affects the protein's structure and neighboring interactions and discard other features that are not also immediately relevant; and if this can be conveyed schematically (perhaps in parallel), all the better. For larger assemblies, the same Tufte guidelines hold: discard all spatial levels and features that do not provide the information intended to be conveyed. At the level of the very largest assemblies, such as the NPC, we have sought to illustrate only the kinds of proteins that comprise the full assembly, how they are arranged overall, and how this arrangement varies between different organisms, resulting in alternative NPC architectures. We generally use ChimeraX for the initial representations, which are then arranged and labeled in Adobe Illustrator.

Tuning complexity to match the text

Einstein's philosophy that the irreducible basic elements of any theory should be 'as simple as possible but no simpler' applies equally well to the task of preparing protein schematics. With the availability of high-resolution structures (or homology models from programs like DeepMind's AlphaFold), there is a temptation to represent all protein illustrations in the greatest detail possible. But I strive to tune the complexity of a figure to match the detail in the corresponding text. Too little detail will dilute the information content of a figure, to be sure. But too much detail can be distracting, perhaps leaving readers wondering if they are missing an important point of the narrative.

I therefore prefer illustration software that allows for a wide range of detail in preparing protein schematics. For figures that do not require explicit protein structure, I use Adobe Illustrator. As a vector graphics-based program, Illustrator offers high geometric precision, scalability without resolution loss, and an almost infinite palette of artistic options to target the right level of complexity. For figures that refer to specific structural elements, there are many excellent molecular graphics programs available but my go-to is the Schrödinger suite (Maestro and PyMol). When deciding how to translate a pdb file into an illustration, again the most effective figures convey just enough detail to complement the text: for instance,

surface representation for basic domain architecture, ribbon diagrams when secondary structure is discussed, and atomic detail to illustrate residue rotamers or side-chain interactions.

But regardless of the level of detail one uses for individual figures, the value of consistent iconography throughout the document cannot be overstated. Uniformity of style (for instance, red always means ‘phosphorylated’, atoms have consistent color scheme, etc.) provides cohesiveness and clarity, making it much easier for the reader to follow.

¹Department of Cell Biology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA

²Department of Biochemistry, University of Colorado Boulder, Boulder, CO, USA

³Howard Hughes Medical Institute, Chevy Chase, MD, USA

⁴Department of Biological and Chemical Engineering, Dongyang Mirae University, Seoul, South Korea

⁵Structural Biology of Molecular Machines Group, Protein Structure & Function Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

⁶Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY, USA

⁷Ikerbasque, Basque Foundation for Science and Instituto Biofisika (UPV/EHU, CSIC), University of the Basque Country, Leioa, Spain

⁸Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA

*Correspondence:

Jonathan_Markert@hms.harvard.edu (J. Markert), Karolin.Luger@Colorado.edu (K. Luger), hayunlee@dongyang.ac.kr (H. Lee), haidai.hu@cpr.ku.dk (H. Hu), nicholas.taylor@cpr.ku.dk (N.M.I. Taylor), jfernandez@mail.rockefeller.edu (J. Fernandez-Martinez), rout@mail.rockefeller.edu (M. Rout), and nathan.alder@uconn.edu (N. Alder).

<https://doi.org/10.1016/j.tibs.2023.01.006>

© 2023 Elsevier Ltd. All rights reserved.

References

1. Markert, J. and Luger, K. (2021) Nucleosomes meet their remodeler match. *Trends Biochem. Sci.* 46, 41–50
2. Lee, H. and Sashital, D.G. (2022) Creating memories: molecular mechanisms of CRISPR adaptation. *Trends Biochem. Sci.* 47, 464–476
3. Hu, H. *et al.* (2022) Structural basis of torque generation in the bi-directional bacterial flagellar motor. *Trends Biochem. Sci.* 47, 160–172
4. Fernandez-Martinez, J. and Rout, M.P. (2021) One ring to rule them all? Structural and functional diversity in the nuclear pore complex. *Trends Biochem. Sci.* 46, 595–607
5. Iovine, J.C. *et al.* (2021) Mitochondrial compartmentalization: emerging themes in structure and function. *Trends Biochem. Sci.* 46, 902–917