

## ARTICLE

# Teaching protein structure and function through molecular visualization

 Elisabeth Baland<sup>1</sup> | Lucía Pérez Jimenez<sup>1,2</sup> | André Mateus<sup>1,3</sup> 
<sup>1</sup>Department of Chemistry, Umeå University, Umeå, Sweden

<sup>2</sup>Department of Molecular Biology, Umeå University, Umeå, Sweden

<sup>3</sup>The Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå, Sweden

## Correspondence

 André Mateus, Department of Chemistry, Umeå University, Umeå, Sweden.  
Email: [andre.mateus@umu.se](mailto:andre.mateus@umu.se)

## Funding information

European Research Council, Grant/Award Number: ProFITGut-101076015; Swedish Research Council (Vetenskapsrådet), Grant/Award Numbers: 2021-06602, 2022-02958

## Abstract

The function of proteins is governed by their three-dimensional structure. This structure is determined by the chemical characteristics and atomic interactions of amino acids. Students of biochemistry, with a particular focus on protein chemistry, benefit from looking at protein structures and understanding how proteins are built and fold. Due to their three-dimensional nature, static two-dimensional representations in textbooks can be limiting to student learning. Here, we developed a series of tutorials that introduce students to molecular graphics software. The students are challenged to apply the software to look at proteins and to get a deeper understanding of how amino acid properties are linked to structure. We also familiarize students with some of the latest tools in computational structural biology. Students performed the tutorials with visual enthusiasm and reported general satisfaction in being able to visualize theoretical concepts learned during lectures. We further stimulated student engagement by allowing space for self-exploration. We share the tutorial instructions for other teachers to build on them, and we also offer suggestions for further improvement based on student feedback. In summary, we present a series of tutorials aimed at students of an advanced course in protein biochemistry to enable them to explore the universe of protein structures and how those relate to function.

## KEYWORDS

biochemistry, molecular graphics, motifs, protein structure and function, tutorials

## 1 | INTRODUCTION

The three-dimensional (3D) structure of proteins defines their function. This 3D structure is determined by the physicochemical characteristics of the sequence of amino acids and interactions between their atoms. The structure of a protein can be experimentally resolved by using methods

such as X-ray crystallography, nuclear magnetic resonance (NMR), or cryogenic electron microscopy (cryoEM). Over 200,000 protein structures have been deposited in the protein data bank (PDB).<sup>1</sup> These data have been used to train deep learning methods to predict protein structures,<sup>2–4</sup> which has resulted in virtually all known protein sequences now having a predicted structure.<sup>5</sup>

Most students of biochemistry encounter protein structures only as static two-dimensional illustrations in

Elisabeth Baland and Lucía Pérez Jimenez contributed equally.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Biochemistry and Molecular Biology Education* published by Wiley Periodicals LLC on behalf of International Union of Biochemistry and Molecular Biology.

textbooks. However, there are multiple tools available for visualizing protein structures.<sup>6</sup> One example is ChimeraX,<sup>7</sup> a free software for non-commercial use. The development team has compiled multiple tutorials for specific tasks,<sup>8</sup> but it can be challenging for a student to know where to start.

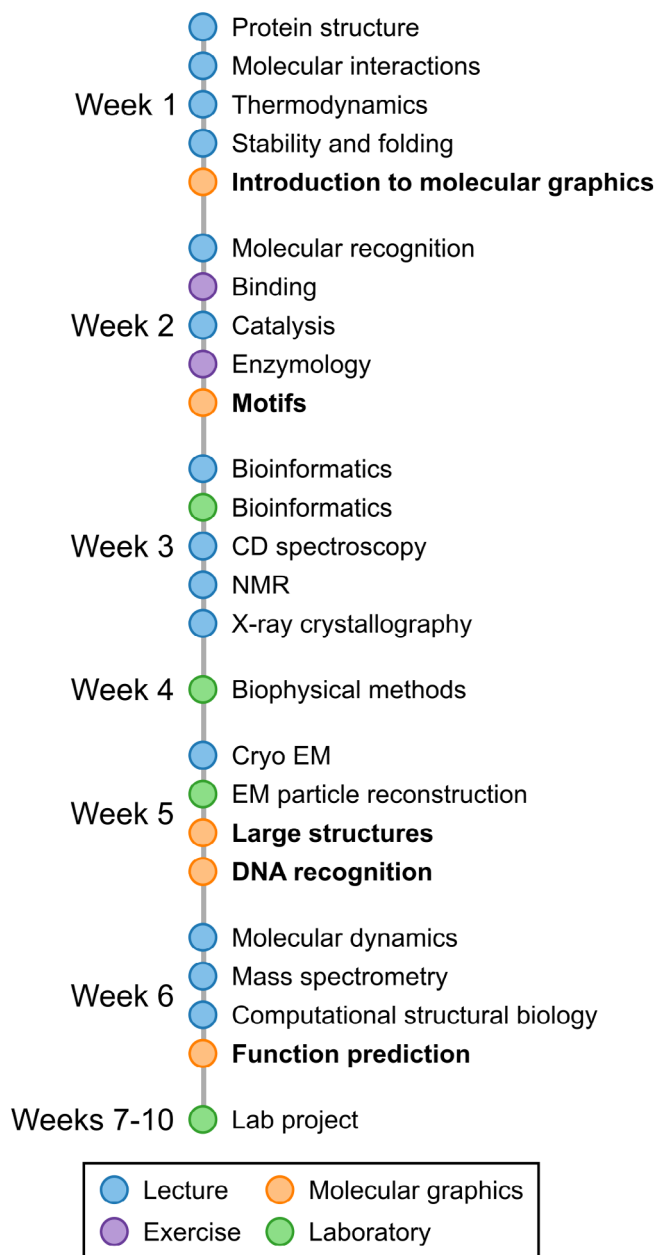
For that reason, as part of an advanced course in protein chemistry, we have developed a series of five teaching moments focused on molecular graphics. These consist of structured step-by-step tutorials that introduce students to molecular graphics software, and lead them to explore the physicochemical characteristics of proteins, the interactions between their atoms, motifs and domains, and how they can use other computational tools to get hints on protein function. By allowing students to manipulate 3D representations of protein structures, we aimed at deepening the students understanding of protein chemistry,<sup>9–12</sup> and prepare them to use the latest tools in computational structural biology.

## 2 | MATERIALS AND METHODS

### 2.1 | Course structure

The teaching moments described in this manuscript are part of the course “Biochemistry: protein structure and function course” at Umeå University in Sweden. This is a 2.5-month course given at the master level and requiring students to have at least a 2.5-month education in basic biochemistry. The syllabus can be found in Data S1 and the intended learning outcomes (ILOs) are that by the end of the course the students should be able to:

- Identify different motifs and domains in protein structures and how these relate to function.
- Understand and use the thermodynamic principles that determine the structure and folding of proteins, how and these impact function.
- Understand and describe the methodologies to determine protein structure, and suggest which one is more adequate to study a particular protein.
- Use the principles of molecular recognition to understand interactions with proteins.
- Relate the theoretical framework of catalysis to its application to study enzyme function.
- Use computational tools to visualize protein structure and to predict their function.
- Design and perform experiments to study which aspects of the structure are important for the stability, structure, and function of proteins.
- Critically discuss the results of experiments designed to study protein stability, structure, and function.



**FIGURE 1** Timeline for the course “Biochemistry: protein structure and function course” at Umeå University. The course comprises lectures and laboratory moments on different aspects of protein chemistry and methodologies to determine protein structure, interactions, and function. To help students visualize protein structures, we have designed a series of five hands-on tutorials with molecular graphics software (highlighted in the figure).

To achieve the ILOs, this course comprises a series of lectures on basic principles of physicochemical properties of proteins, protein stability and folding, protein interactions and catalysis, and introduction to different methods to determine protein structure, interactions, and function (Figure 1). These concepts are applied during laboratory moments on bioinformatics, biophysical methods, and particle reconstruction from cryoEM data, as well as a

final project. Throughout the course, the students participate in a series of five hands-on step-by-step tutorials using the molecular graphics software ChimeraX,<sup>7</sup> and are encouraged to use the software in other teaching moments.

## 2.2 | Designing the molecular graphics moments

We designed tutorials with the major goal of introducing students to molecular graphics software and allowing them to directly interact with protein structures. The tutorials can be found in Data S2. These incorporated aspects of previously published approaches,<sup>10–12</sup> and of an older version of these tutorials. That version had received some criticism from the students during the course evaluations. For example, one of the students reported “The molecular graphics class could be more structured [...] it was easy to feel lost, especially at the beginning,” while another student suggested “Change the exercises so that they bring something else than just look at a protein in 3D [...] The purpose of those sessions was not clear”, and another student offered “The molecular graphics part could be improved, sometimes it was really confusing what we had to do.”

With this in mind, we decided to structure our tutorials in five sessions spread throughout the course and linked to the rest of the curriculum (Figure 1). The sessions also included recently developed computational tools that are now routinely used in structural biology.<sup>2,13,14</sup> The first four sessions were scheduled for 2 h each, and the last session was scheduled for 4 h. The topics of each of the sessions were:

1. Introduction to molecular graphics.
2. Motifs in protein structures.
3. Larger structures.
4. DNA recognition.
5. Getting hints on protein function using computational tools.

During session 1, the students learn to open PDB<sup>1</sup> or AlphaFold models<sup>2</sup> and get familiar with using the mouse actions, menus, and the command line. However, we attempted to have exercises that required minimum interaction with the command line to not overwhelm students that are not familiar with programming. The students also explore different representations of proteins (atoms/cartoons/surfaces), how to specifically color different aspects of the protein or how to measure distances. Finally, the students are trained to compare structures and to visualize possible movements in proteins. Students are asked to put these skills to practice by comparing the

quality of AlphaFold models in relation to experimentally determined structures, and by visualizing atomic interactions in secondary structures of proteins, for example, how hydrogen bonding stabilizes alpha helices (Figure 2a) and beta sheets (Figure 2b).

During session 2, the students connect how the side chains of the different amino acids contribute to interactions (e.g., electrostatic interactions) and structure (e.g., the rigid structure of proline introduces kinks in alpha helices). The students also explore simple structural motifs such as the EF-hand, hairpins or the Greek key, and how sequence motifs connect to structure.

During session 3, the students engage with larger structures, such as membrane proteins, and link the physicochemical properties of amino acids with their location on the structure, for example, hydrophobic side chains promote membrane insertion in transmembrane regions of proteins, while the hydrophilic interior contributes to the passage of polar solutes (Figure 2c).

During session 4, the students look at how proteins interact with nucleic acids, and how DNA can acquire different conformations.

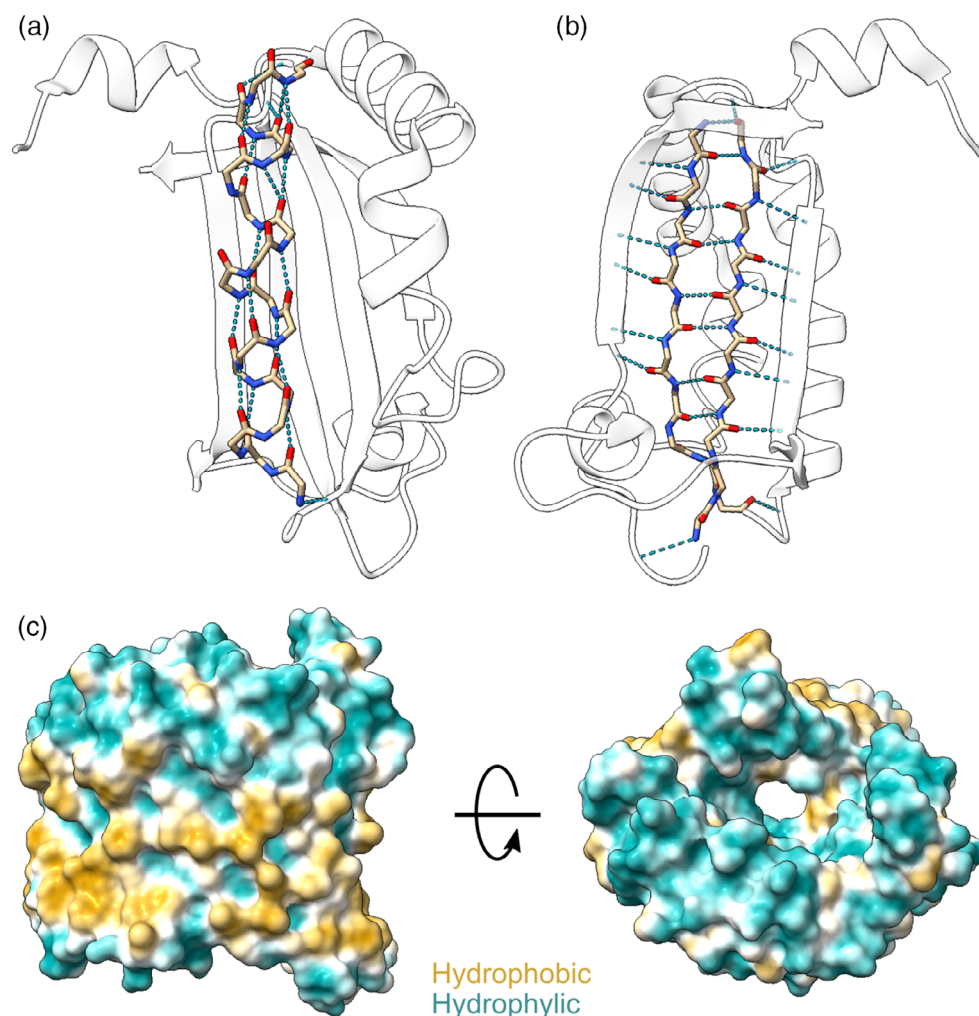
During session 5, groups with approximately four students each are given a UniProt entry<sup>15</sup> of an uncharacterized protein, and use traditional tools to identify functional motifs (such as BLAST<sup>16</sup> or InterPro<sup>17</sup>), and recently developed tools to compare protein structures to the whole protein universe (e.g., Foldseek<sup>13</sup>) or to predict protein function (e.g., DeepFRI<sup>14</sup>). Their goal is to create the first steps to study the function of proteins. At the end of the session, the students share their strategies and findings with the rest of the class in the form of a short oral presentation.

## 2.3 | Course evaluation

During the first implementation of these teaching moments, the students were asked to rank the sessions during the final course evaluation with the following prompt “This year, the molecular graphics sessions were greatly redesigned. How much did you like the format of the molecular graphics sessions? Tell your opinion if something could be improved.” They were given the chance to answer: “Not at all,” “Bad,” “OK,” “Good,” or “Very good,” as well as, input any free text in an online platform. All information was collected anonymously, and no personal data was collected that could identify the students, in accordance to the Swedish Ethical Review Act (2003:460).

## 3 | RESULTS

We designed a series of tutorials to introduce students to molecular graphics software and to give them the

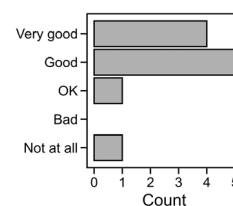


**FIGURE 2** Physicochemical properties and interactions of amino acids determine protein structure. Hydrogen bonds are essential for providing stability to (a) alpha helices and (b) beta sheets (PDB: 2OKQ). (c) Hydrophobic side chains promote the insertion of transmembrane proteins in the membrane, while the hydrophilic interior allows the passage of polar molecules (PDB: 1A0S).

opportunity to use it to explore aspects of protein chemistry. Further, the tutorials introduced students to recently developed computational tools that are now used routinely by structural biologists to understand protein function from structure.

The sessions were attended by all 17 students registered for the course. The students followed the tutorials individually, but were encouraged to ask their classmates how certain tasks were performed. To activate the students, the instructions omitted how certain tasks should be performed and instead pushed the students to look for solutions on their own, either by trial-and-error or by using a search engine or a large language model. The role of the teachers was to support students when they could not complete one of the assignments and to summarize the key concepts learned in each session. While the sessions were implemented in a small student cohort, their nature makes them scalable to a large number of students.

The sessions were generally well received by the students. The teachers noticed engagement from most students during the tutorials. The students generally



**FIGURE 3** Answers from the students to “How much did you like the format of the molecular graphics sessions?”  $n = 11$  students.

communicated how visualizing aspects of protein structure in 3D helped them in understanding concepts they had learned during the lectures. The last session also provided an opportunity for students to experience the role of a researcher in structural biology. Although challenging, since for most proteins the students could not obtain much information, the students appreciated this moment.

During the course evaluation, 11 students (65%) answered the question on what they thought of the sessions (Figure 3). The vast majority answered that they



thought they were “Good” (45%) or “Very good” (36%). One student selected the lowest score, but unfortunately only commented “I didn’t find it very useful.”, which does not provide a chance for improvement. The student that thought the sessions were “OK” reported “I think they were very clear, but sometimes in one session there was a lot of repetition and I think it was pretty clear to everyone, so this wasn’t really necessary. [...] However, it was very clear! I worked with it a lot before, but this was the first time it was actually explained.” Related to this, one of the other students wrote “It was a good exercise, but some sessions can be finished in one hour” (“Good”). Encouragingly, the other students stated that this was a “Great intro to molecular graphics! Maybe more “real-life” examples, what this tool is used for in research” (“Very good”), “I think the way the molecular graphics sessions were made, made it easy to understand the lecture and helped to have a visual representation” (“Good”), “I really liked it, it was fun to identify and know more about the protein using different tools” (“Very good”), and “It was nice to experiment on the computer for the future” (“Good”).

In general, the students thought that these sessions were useful and increased their knowledge of how proteins obtain their structures.

## 4 | DISCUSSION

Protein structure visualization is a common aspect in advanced courses in protein chemistry.<sup>10,11,18</sup> In this work, we developed a new series of tutorials taking the students from the basic aspects of using molecular graphics software to applying it to understand how the structure of proteins is related to amino acid composition. We also introduced the students to novel computational tools that are now part of the toolkit of structural biology. We share these tutorials (Data S2) to inspire other teachers to build on them.

Despite providing step-by-step instructions on how to perform specific tests, we included some steps in which the students are stimulated to look on their own how to perform certain tasks, either by using a search engine or a large language model, such as chatGPT. This not only ensures that the tutorial stays relevant as novel computational structural biology tools become available and the students can find their way to such tools on their own, but it also promotes student engagement.<sup>19</sup>

Based on the student feedback, it is possible to shorten some of the sessions by reducing some steps that appear redundant. Alternatively, it is possible to include extra assignments for students that grasp concepts faster and want to further explore the capabilities of the

molecular graphics software. One option is to make it clearer that students can explore the tutorials created by the developers of the software<sup>8</sup>—although this is already explicitly written in the instructions.

Another aspect that could be included is for students to predict their own protein structures by using cloud-based services, such as ColabFold<sup>20</sup> or AlphaFold server.<sup>4</sup> Even though structures are available for all protein sequences present on UniProt,<sup>5</sup> scientists routinely encounter novel sequences for which it is useful to predict a structure<sup>21</sup> or want to model multimeric protein complexes. We are also considering including an extra session in which students would fit a molecular model to an electron microscopy map. This would be linked to the laboratory moment in which students perform particle reconstruction from cryoEM data following a tutorial created by the developers of CryoSPARC.<sup>22,23</sup> Finally, since protein structures are inherently three-dimensional, and virtual/augmented reality becomes more accessible, some of the exercises could be performed in such an environment.<sup>24</sup>

## 5 | CONCLUSION

In conclusion, visualizing protein structures can help students better understand why proteins fold into specific conformations and how some of those are linked to function. As part of an advanced course in protein chemistry, the series of tutorials presented here functions not only as an introduction to molecular graphics, but also in preparing the students to be able to use computational tools that are frequently used by structural biologists.

## AUTHOR CONTRIBUTIONS

All authors participated in the design of the teaching moment, performed the analysis, and wrote the manuscript.

## ACKNOWLEDGMENTS

We thank Prof. Magnus Wolf-Watz who had developed a previous version of this tutorial that served as an inspiration for this version. A.M. is supported by the Swedish Research Council (Vetenskapsrådet; grant number 2022-02958), and the European Research Council (grant number ProFITGut-101076015). The Laboratory for Molecular Infection Medicine Sweden is supported by the Swedish Research Council (Vetenskapsrådet; grant number 2021-06602).

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

Data available in article [Supporting Information](#).

## ORCID

André Mateus  <https://orcid.org/0000-0001-6870-0677>

## REFERENCES

1. PDB Consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* 2019; 47:D520–8. <https://doi.org/10.1093/nar/gky949>
2. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596:583–9. <https://doi.org/10.1038/s41586-021-03819-2>
3. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science.* 2021; 373:871–6. <https://doi.org/10.1126/science.abj8754>
4. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature.* 2024;630:493–500. <https://doi.org/10.1038/s41586-024-07487-w>
5. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Židek A, et al. Highly accurate protein structure prediction for the human proteome. *Nature.* 2021;596:590–6. <https://doi.org/10.1038/s41586-021-03828-1>
6. Martinez X, Chavent M, Baaden M. Visualizing protein structures—tools and trends. *Biochem Soc Trans.* 2020;48:499–506. <https://doi.org/10.1042/BST20190621>
7. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* 2021;30:70–82. <https://doi.org/10.1002/pro.3943>
8. ChimeraX Tutorials. <https://www.rbvi.ucsf.edu/chimerax/tutorials.html>
9. Wu H-K, Shah P. Exploring visuospatial thinking in chemistry learning. *Sci Educ.* 2004;88:465–92. <https://doi.org/10.1002/sce.10126>
10. Abreu PA, Carvalho KL, Rabelo VW, Castro HC. Computational strategy for visualizing structures and teaching biochemistry. *Biochem Mol Biol Educ.* 2019;47:76–84. <https://doi.org/10.1002/bmb.21199>
11. Canning DR, Cox JR. Teaching the structural nature of biological molecules: molecular visualization in the classroom and in the hands of students. *Chem Educ Res Pract.* 2001;2:109–22. <https://doi.org/10.1039/B1RP90013G>
12. Trivedi VD. Protein structure-function exploration initiative in undergraduate biochemistry and independent research courses. *Biochem Mol Biol Educ.* 2022;50:473–5. <https://doi.org/10.1002/bmb.21664>
13. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol.* 2023;42:243–6. <https://doi.org/10.1038/s41587-023-01773-0>
14. Gligorijevic V, Renfrew PD, Kosciolk T, Leman JK, Berenberg D, Vatanen T, et al. Structure-based protein function prediction using graph convolutional networks. *Nat Commun.* 2021;12:3168. <https://doi.org/10.1038/s41467-021-23303-9>
15. UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 2023;51:D523–31. <https://doi.org/10.1093/nar/gkac1052>
16. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
17. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. *Nucleic Acids Res.* 2023;51: D418–27. <https://doi.org/10.1093/nar/gkac993>
18. Goodsell DS, Jenkinson J. Molecular illustration in research and education: past, present, and future. *J Mol Biol.* 2018;430: 3969–81. <https://doi.org/10.1016/j.jmb.2018.04.043>
19. Struyven K, Dochy F, Janssens S. ‘Teach as you preach’: the effects of student-centred versus lecture-based teaching on student teachers’ approaches to teaching. *Eur J Teacher Educ.* 2010;33:43–64. <https://doi.org/10.1080/02619760903457818>
20. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods.* 2022;19:679–82. <https://doi.org/10.1038/s41592-022-01488-1>
21. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science.* 2023;379:1123–30. <https://doi.org/10.1126/science.ade2574>
22. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods.* 2017;14:290–6. <https://doi.org/10.1038/nmeth.4169>
23. Get Started with CryoSPARC: Introductory Tutorial. <https://guide.cryosparc.com/processing-data/get-started-with-cryo-sparc-introductory-tutorial>
24. Goddard TD, Brilliant AA, Skillman TL, Vergenz S, Tyrwhitt-Drake J, Meng EC, et al. Molecular visualization on the holo-deck. *J Mol Biol.* 2018;430:3982–96. <https://doi.org/10.1016/j.jmb.2018.06.040>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Baland E, Pérez Jimenez L, Mateus A. Teaching protein structure and function through molecular visualization. *Biochem Mol Biol Educ.* 2025;53(1):15–20. <https://doi.org/10.1002/bmb.21860>