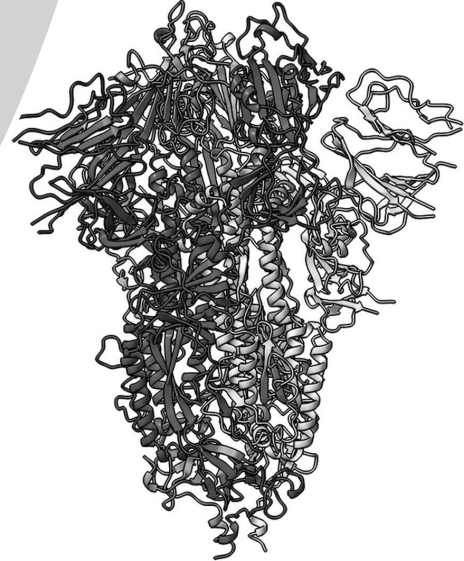


From COVID-19 to the Central Dogma: Investigating the SARS-CoV-2 Spike Protein

George P. McGuire, Camila V. Luna,
Erica M. Staehling, M. Elizabeth Stroupe



ABSTRACT

Students often struggle with visualizing protein structures when working with two-dimensional textbook and lecture materials, so introducing them to 3D visualization software developed by and for structural biologists offers them a unique opportunity to work with authentic data while furthering their spatial reasoning skills and understanding of molecular structure and function. This article presents an active learning virtual laboratory in which students use authentic structural biology data to investigate the effects of both hypothetical and real-world SARS-CoV-2 mutations on the virus's ability to bind to human ACE2 receptors and infect a host, causing COVID-19. Through this activity, introductory-level college students or advanced high school students gain a better understanding of applied biology, such as how vaccines and treatments are designed, as well as strengthening their understanding of core disciplinary concepts, such as the relationship between protein structure and function and the central dogma of molecular biology. While there were challenges during the pilot phase of activity development due to COVID-19 restrictions, students in the pilot groups came away from the activity with deeper understanding of the relationship between proteins and amino acid sequences and a new appreciation for the ways researchers design treatments for and study viruses.

Key Words: SARS-CoV-2 spike protein; COVID; structural biology; virus variants; ChimeraX.

○ Introduction

The continuing mutation of the SARS-CoV-2 virus has kept scientists on their toes and dominated news cycles, providing a rich anchoring phenomenon for students to develop and further their understanding of the central dogma of molecular biology, viral action, and ligand-receptor interaction (Windschitl et al., 2018). Members of the Coronaviridae family of viruses are known for their spike proteins that not only give their famous crown-like shape but also play a major role in their ability to infect host cells. For most known coronaviruses, including the current pandemic-causing SARS-CoV-2, the importance of the spike protein and how it mediates host cell entry is well documented (Lan et al., 2020;

Walls et al., 2020). Differences in the amino acid composition of the spike protein across strains of coronavirus and among variants of the same strain determine the host range of the virus. Specifically, the chemical nature of amino acids in the spike determines the binding affinity of the virion to specific receptors on the surface of the cells they infect (Lan et al., 2020). In the case of SARS-CoV-2, multiple variants have been identified in human patients suffering from COVID-19 in the United States, and each of these variants is proposed to have evolved from the same SARS-CoV-2 virus that originated in China (Washington et al., 2021).

In this activity, students investigate the amino acid structure of the SARS-CoV-2 spike protein and its interaction with the human ACE2 receptor using the UCSF ChimeraX 3D molecular modeling program (Pettersen, 2021). ChimeraX allows students to utilize authentic data to visualize 3D models of proteins and manipulate individual amino acids to identify how amino acid alterations can change the structure and chemical properties of a molecule. After identifying key amino acids, students investigate the corresponding codons within the SARS-CoV-2 genome, hypothesizing and discussing how potential mutations would affect the nucleotide sequence of this gene, change the amino acid sequence of the spike protein, and lead to potential changes in transmission rates and/or disease severity. Thus, modeling and uncovering the effects of viral evolution that have contributed to humanity's struggle with slowing the spread of SARS-CoV-2's increasing number of variants.

This activity is designed for college-level and high-achieving high school biology students, following the 5E learning cycle: engage, explore, explain, elaborate, and evaluate (Bybee et al., 2006). As such, it has been field-tested with students in a for-majors introductory biology course and an upper-division elective course at a large research university in the southeastern United States, as well as with high schoolers enrolled in a summer program at that same university. The learning goals for this lesson align with the framework from the National Research Council (NRC) for K–12 science education, providing students with multiple opportunities to engage in metacognition by reflecting on what they are investigating, why they are investigating it, and the scientific practices in which they are engaged: asking questions; developing and using models; planning and carrying out investigations; analyzing and interpreting data; constructing

explanations; engaging in argument from evidence; and obtaining, evaluating, and communicating information (NRC, 2012). The lesson could be adapted to span the length of one to three class meetings and could be included within units on scientific approaches for studying organisms, basic chemistry related to living organisms, biomolecules, and cellular activity, assuming students have already been introduced to the central dogma of biology and the composition and function of proteins. Although pacing varied during the pilot phase, we found that a total of three hours was adequate for advanced high school students to achieve the learning goals.

○ Goals

Through this lesson, students will develop and further their understanding of the SARS-CoV-2 virus, mechanisms of viral entry into host cells, how mutations can affect a virus's ability to infect host cells, protein structure and function, how the sequence of DNA/RNA determines the amino acid sequence of a protein, how researchers study viruses, and the role of basic research in developing treatments for viruses. By the end of this activity, students will be able to

- Use ChimeraX and authentic structure data to examine the composition and shape of both the SARS-CoV-2 spike protein and the human ACE2 receptor protein.
- Investigate how hypothetical and real-life mutations affect a protein's structure and function.
- Explain how certain mutations could affect a virus's ability to bind to a human receptor.
- Explain how the potential for mutation impacts the design of potential treatment methods.

○ Student Material for This Lesson

Access to computers/laptops, including a three-button mouse (preferred)

- ChimeraX, a freely available 3D molecular modeling program, developed with funding from the National Institute of Health
- SARS-CoV-2 lab manual (Appendix 1)
- The ChimeraX file containing the prepared protein models (Appendix 2)
- SARS-CoV-2 presentation (Appendix 3)

○ Lesson Framework/Procedures

Prelesson Setup

Instructors should familiarize themselves with using the ChimeraX 3D molecular modeling program by watching the ChimeraX tutorial videos (<https://rbvi.github.io/chimera-tutorials/presentations/modules/chimerax-comp-structures/index.html#>) provided by UCSF and consider having students also complete tutorials before class or guiding students through using the necessary tools during class time. It is also important to confirm that students have access to computers with the ChimeraX program and the SARS-CoV-2 Spike.cxs file (Appendix 2) downloaded and functioning before beginning the lesson. The lab manual (Appendix 1) is designed to be too challenging for most students to complete comfortably on their own, and we recommend splitting students into groups of three

Table 1. Student group roles and descriptions.

Role	Description
Driver	Controls the ChimeraX program
Scribe	Fills out the lab manual (Appendix 1) with the group's data and answers
Troubleshooter	Mediates communication with the instructor or directly finds help within the UCSF user guide / Help function during the activity (https://www.rbvi.ucsf.edu/chimerax/docs/user/index.html)
Spokesperson (optional)	Communicates on behalf of the group during whole-class discussion

to four with each student assigned to a specific role (Table 1), to promote positive interdependence (Felder & Brent, 2016, p. 256).

Instructors may optionally dedicate enough class time for groups to complete the entire lab manual or may “jigsaw” and evenly split students into group types, where Group Type A works with their own hypothetical amino acid changes and mutations and Group Type B investigates the amino acid changes identified within a real-world variant of SARS-CoV-2. We present the activity assuming the jigsaw version of the implementation, and group type will come into effect during “Part 2: Investigating the Spike-ACE2 Interaction” (Appendix 1). Alternatively, if class time is limited, the instructor could simply select and modify the lab manual to focus on either hypothetical mutations or real-world variants to suit their purposes.

Engagement

Instructors should begin with a short introduction to the COVID-19 pandemic and the responsible virus, SARS-CoV-2, being sure to discuss both its structure and method of infection. This can be customized to the context of the course, but Appendix 3 provides several slides/figures an instructor might use to elicit students' prior knowledge about the targeted biology concepts and provide some brief direct instruction on the necessary background information (e.g., Figures 1 and 2). This provides students with a context to explore the central dogma and ligand-receptor interaction throughout the activity.

Next, focus students' attention on the SARS-CoV-2 spike protein and explain that they will have the opportunity to visualize and manipulate this structure, as well as the human ACE2 receptor, using the ChimeraX 3D molecular modeling program and the SARS-CoV-2 lab manual (Appendix 1). If a tutorial was not assigned as homework in preparation for the activity, instructors should also consider providing a short demonstration of the basic ChimeraX controls.

Exploration

The instructor should launch students into their groups to begin the lab activity by assigning and describing their roles and instructing students to read the introduction and complete page 2 of the lab manual. Students will explore a model of the entire SARS-CoV-2 spike protein while they familiarize themselves with the controls and tools available to them within ChimeraX (see Figure 3).

While exploring the model, students make guided observations of the spike protein's structure and composition. This should be followed by a quick whole-class discussion about students' initial observations and can be used as an opportunity to troubleshoot any immediate issues with using the ChimeraX software. Next, students

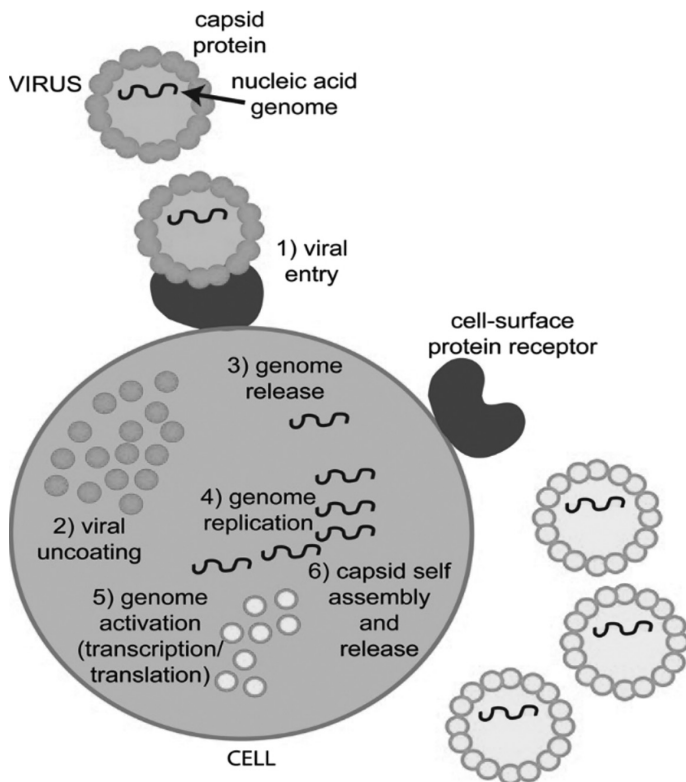


Figure 1. Slide 3 from Appendix 3. This figure depicts the basics of viral structure, entry into host cells, replication, and release of progeny. This should be referenced when introducing students to these concepts during direct instruction.

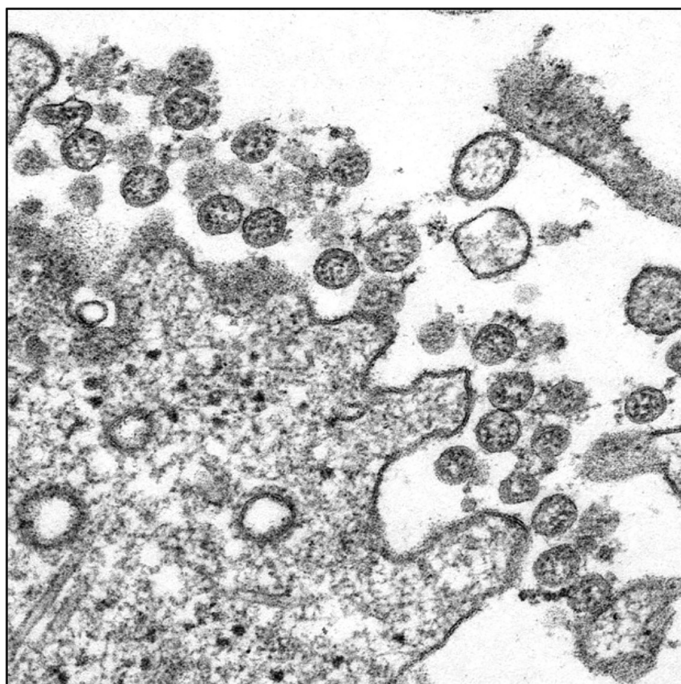


Figure 2. Image from Slide 4 from Appendix 3. This transmission electron micrograph shows actual SARS-CoV-2 virus invasion of a human cell. Source: <https://www.cdc.gov/media/subtopic/images.htm>.

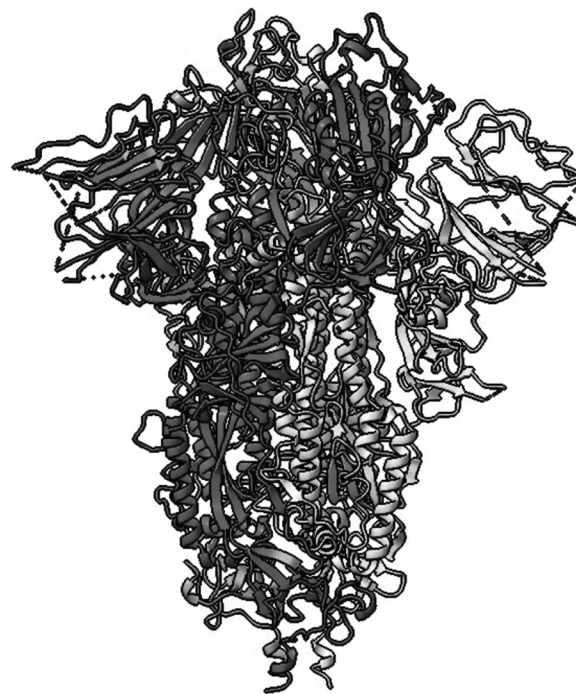


Figure 3. Image of the spike protein model that students will use during Part 1: Exploring the SARS-CoV-2 Spike Protein (in Appendix 2), generated with ChimeraX.

resume group work by moving on to page 3 of the lab manual. To promote metacognition, instructors can encourage students to revisit these initial observations after completing the lesson.

Explanation

Once students have familiarized themselves with the general structure of the SARS-CoV-2 spike protein and have completed page 3 of the lab manual, the instructor should bring the class back together to initiate a whole-class discussion of students' analysis and predictions, providing direct instruction about ligand-receptor interactions, their function in viral infection, and introduction to the human ACE2 receptor that mediates entry of SARS-CoV-2 into human cells as needed. This instruction can also include more specific information about the difference between the receptor binding domain (RBD) and the stalk portions of the spike protein, both in composition and function, using slide 5 of the SARS-CoV-2 presentation (Appendix 3).

Elaboration

After unpacking their initial analysis and predictions and receiving just-in-time instruction, students return to their groups to engage in deeper analysis of the spike protein and its interaction with ACE2 (pages 4–14 of Appendix 3). If following the jigsaw structure, remind groups of their group type (A or B), as each will be working with a different dataset. All groups will complete pages 4–7 and 14, but Group Type A will work with data on pages 8–10, and Group Type B will work with data on pages 11–13. This is the most complicated part of the activity, both in practical ability to use the ChimeraX program and student understanding of how to find the necessary data. Thus, instructors should do their best to circulate between groups and be available to answer questions to provide clarification. Completion of the recommended ChimeraX tutorial will mitigate the potential difficulty of this part of the activity.

During this stage, students will focus on identifying key spike amino acids within the RBD (including the amino acid side chain properties, location within the spike protein gene, and associated nucleotide sequence within the gene), as well as gather data related to their interaction with ACE2, such as associated receptor amino acid and bond distance (in Ångstroms, where $1 \text{ Å} = 1 \times 10^{-10}$ meter, about the length of the bonds found in protein structures).

Using their data, students will hypothesize about and experiment with potential amino acid changes that could result from mutations in the spike protein gene. Depending on their Group Type (A or B), students will experiment with their own hypothetical mutations (A) or will analyze real mutations that have been identified in the alpha (B.1.1.7) variant of SARS-CoV-2 (B) (Washington et al., 2021). Both groups gather data related to the effects of these amino acid changes, such as potential mutations that could cause that amino acid change, changes in side chain properties, changes in shape of the amino acid, and changes in bond distance.

Using data amassed in Tables 2 and 3 of the lab manual (Appendix 1), students next apply their new knowledge and understanding about the Spike-ACE2 interaction to answer questions regarding how SARS-CoV-2 variants evolve and how knowledge of their differing viral properties (such as host binding affinity) relates to researchers' abilities to understand the mechanisms of this virus and develop potential prevention/treatment methods against COVID-19.

When all groups have completed data collection and answered the "Analysis and Discussion" questions on page 14 of the lab manual (Appendix 1), ensure that all students have access to a copy of their group's completed lab manual and bring the students together for whole-class discussion.

Evaluation

The evaluation section of this lesson is focused on building students' metacognition skills by evaluating what they have learned and how this information applies to their understanding of real-world phenomena, as well as key biological concepts and practices. This is accomplished through whole-class discussion in review of student answers to the activity questions.

Whole-class discussion can be instructor-led by sequentially moving through the questions, asking students to share their answers, and eliciting further evidence of student understanding using "talk moves" (Grinath & Southerland, 2018). Alternatively, whole-class discussion can be orchestrated and sequenced based on varying group responses (Cartier et al., 2013). In either case, the instructor should facilitate students' metacognition skills by encouraging them to think deeply about how the concepts and scientific practices they engaged in apply to real-world problem solving, as well as challenging them to consider how their understanding at the end of this activity compares to their prior knowledge of these concepts/practices.

Students from different group types (A or B) should be instructed to compare their experiences and conclusions during this discussion. Alternatively, instructors can pair groups from each type or shuffle students into new groups (some from A and some from B) for small-group discussion before whole-class discussion.

○ Assessment of Learning Outcomes

The assessments included with this lesson were designed to address the desired learning goals, including a low-stakes graded assessment through having students submit their groups' completed lab manual and grading for accuracy using the SARS-CoV-2 lab manual

grading rubric (Appendix 4). However, student understanding is also demonstrated during whole-class discussion, as discourse with the instructor and their peers evokes more profound explanation and understanding than is typically demonstrated in written text within a lab manual.

This activity was field-tested in the fall of 2020, spring of 2021, and summer of 2021 at a large research university in the southeastern United States, in the midst of the COVID-19 global pandemic. When piloting this activity, we asked students to reflect explicitly on their change in understanding for each of the learning objectives by completing a short survey. Student responses to this survey showed that the majority of students strongly agreed or agreed that this activity improved their understanding of the SARS-CoV-2 virus, mechanisms of viral entry into host cells, how mutations can affect a virus's ability to infect host cells, protein structure and function, how researchers/scientists study viruses, and the role of basic research in developing treatments.

Additionally, our postactivity survey provided an opportunity for students to provide feedback to the instructor about which parts of the lesson were beneficial, and which were not, toward improving their understanding for the identified learning goals. This feedback was considered iteratively in the development of this activity and while writing this article. For instance, most of the issues students faced were related to the learning curve associated with using ChimeraX. We found that students who completed the ChimeraX tutorials before class readily overcame these challenges. The developers of ChimeraX are responsive to questions from the community through a listserv (<https://www.rbvi.ucsf.edu/mailman/listinfo/chimerax-users/index.html>) where questions can be posed.

Here are a few exemplary quotes highlighting authentic student experiences with this activity:

- "The most interesting thing I learned was how much work goes into understanding how viruses work. To understand how COVID-19 functions, you need to have an understanding of the properties of different amino acid chains and what differences a mutation in that region would cause."
- "The most interesting thing I learned was probably how a mutation in an amino acid sequence of a virus allows it to evolve from certain medications and preventative measures, since they would no longer have complementary shapes to one another (the virus and human proteins)."
- "Being able to break down the proteins and be able to get information about the virus through the program was super interesting. It also helped my understanding of how they are built significantly."
- "I really liked how the activity allowed for you to rotate and pinpoint parts of the protein. I also found changing the amino acids to be informative."
- "I think having an interactive model definitely made me feel more confident in my understanding of proteins in general, not just in the case of viruses."

○ Special Considerations

This activity was field-tested in the midst of the COVID-19 global pandemic at a large research university in the southeastern United States. During field-testing with biology majors enrolled in the university, the entire activity spanned two consecutive class periods of 1.25 hours each. During field-testing with high school students, the

entire activity spanned two consecutive class periods of 1.5 hours each. From these experiences, we recommend that instructors dedicate about 3 hours to this activity to ensure that students can fully engage, with time for elaboration and evaluation during whole-class discussion.

We conducted field-testing with class sizes of 30–100 students, with a wide range of backgrounds in biological content and scope of interest. The main considerations for class size are ensuring that all groups have at least three students, and that there are enough instructors or assistants present to provide timely assistance for groups that encounter issues during the activity. The largest class size (about 100 biology majors) was sufficiently facilitated by one instructor, with two assistants that were familiar with the activity. In classes larger than 100 students, special consideration should be made toward facilitating whole-class or large-group discussion, assuring that all groups have the opportunity to deliberate their findings with other groups. Additionally, the structure of the lab manual utilizes guiding questions that are accessible to less interested students, as well as providing the opportunity for high-achieving students to rigorously investigate the concepts addressed.

While the pandemic necessitated remote implementation, this lesson would likely be simpler to implement in person, especially with access to a computer lab with the necessary software preinstalled. For field-testing, the instructors utilized Canvas, Zoom, and students' own computing resources during two synchronous class meetings. One key struggle was finding enough students with a three-button mouse to create reasonable group sizes. During field-testing, some groups were as large as five to seven students to facilitate access to a three-button mouse. Based on this experience, we recommend limiting group size to three to four students per group, if possible. However, ensuring that all groups have access to a three-button mouse improves the efficiency of manipulating the ChimeraX program and may necessitate larger groups.

○ Expanding the Scope

The general outline of the activity described is adaptable to address other ligand-receptor interactions, or individual protein mutations, as the anchoring phenomenon. As such, this lab activity could be applied to a wide range of biological systems and could evolve as the scientific community uncovers more about the structure and action of influential proteins. As we learned in our field test with high school students, motivated students can use the research methods introduced in this activity to further investigate SARS-CoV-2 mutations, as well as related topics, such as environmental bioremediation. Several high school students in our summer program used this activity as a starting place for a larger term-long independent research project in summer 2021, resulting in original submissions to a journal aimed at high school student researchers.

○ Acknowledgment

This work was partially supported by National Science Foundation grant MCB1856502 to MES.

○ Supplementary Material

Appendices are available with the online version of this article.

- Appendix 1: SARS-CoV-2 lab manual
- Appendix 2: SARS-CoV-2 Spike.cxs (file to be opened in ChimeraX)
- Appendix 3: SARS-CoV-2 presentation
- Appendix 4: SARS-CoV-2 lab manual grading rubric
- ChimeraX tutorial videos: <https://rbvi.github.io/chimera-tutorials/presentations/modules/chimerax-comp-structures/index.html#>

References

- Bybee, R., Taylor, J., Gardner, A., Scotter, P., Powell, J., et al. (2006). *The BSCS 5E Instructional Model: Origins, Effectiveness and Applications*. A report prepared for the Office of Science Education, National Institutes of Health.
- Cartier, J.L., Smith, M.S., Stein, M.K. & Ross, D.K. (2013). *5 Practices for Orchestrating Productive Task-Based Discussions in Science*. National Council of Teachers of Mathematics.
- Felder, R.M., & Brent, R. (2016). *Teaching and Learning STEM: A Practical Guide*. John Wiley & Sons.
- Grinath, A.S. & Southerland, S.A. (2018). Applying the ambitious science teaching framework undergraduate biology: Responsive talk moves that support explanatory rigor. *Science Education*, 103(1), 92–122. <http://dx.doi.org/10.1002/sce.21484>.
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581, 215–21.
- National Research Council. (2012). *A Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., et al. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1), 70–82.
- Walls, A.C., Park, Y., Tortorici, M.A., Wall, A., McGuire, A.T. & Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, 180, 281–92.
- Washington, N.L., Gangavarapu, K., Zeller, M., Bolze, A., Cirulli, E.T., et al. (2021). Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *Cell*, 184, 1–4.
- Windschitl, M., Thompson, J.J. & Braaten, M.L. (2018). *Ambitious Science Teaching*. Harvard Education Press.

GEORGE P. MCGUIRE (mcggeorge@gmail.com) is a recent graduate of the College STEM Teaching program at Florida State University (FSU) and an adjunct professor of biology at Tallahassee Community College (TCC) in Tallahassee, FL 32306. CAMILA V. LUNA (cvluna@fsu.edu) is a recent graduate of the College STEM Teaching program at FSU, former adjunct professor at TCC, and current laboratory technician at Quest Diagnostics, Seattle, WA 98134. ERICA M. STAEHLING (estaehling@fsu.edu) is the director of the STEM Teaching program and associate director of the Office of STEM Teaching Activities at FSU. M. ELIZABETH STROUPE (mestroupe@bio.fsu.edu) is an associate professor in the Department of Biological Science at FSU.