

Using Open-Source Bioinformatics and Visualization Tools to Explore the Structure and Function of SARS-CoV-2 Spike Protein

Laura L. Listenberger^{1,2}, Cassandra M. Joiner², and Cassidy R. Terrell^{3*}

¹Department of Biology, St. Olaf College

²Department of Chemistry, St. Olaf College

³Center for Learning Innovation, University of Minnesota

Abstract

The relationship between protein structure and function is a foundational concept in undergraduate biochemistry. We find this theme is best presented with assignments that encourage exploration and analysis. Here, we share a series of four assignments that use open-source, online molecular visualization and bioinformatics tools to examine the interaction between the SARS-CoV-2 spike protein and the ACE2 receptor. The interaction between these two proteins initiates SARS-CoV-2 infection of human host cells and is the cause of COVID-19. In assignment I, students identify sequences with homology to the SARS-CoV-2 spike protein and use them to build a primary sequence alignment. Students make connections to a linked primary research article as an example of how scientists use molecular and phylogenetic analysis to explore the origins of a novel virus. Assignments II through IV teach students to use an online molecular visualization tool for analysis of secondary, tertiary, and quaternary structure. Emphasis is placed on identification of noncovalent interactions that stabilize the SARS-CoV-2 spike protein and mediate its interaction with ACE2. We assigned this project to upper-level undergraduate biochemistry students at a public university and liberal arts college. Students in our courses completed the project as individual homework assignments. However, we can easily envision implementation of this project during multiple in-class sessions or in a biochemistry laboratory using in-person or remote learning. We share this project as a resource for instructors who aim to teach protein structure and function using inquiry-based molecular visualization activities.

Citation: Listenberger LL, Joiner CM, Terrell CR. 2022. Using open-source bioinformatics and visualization tools to explore the structure and function of SARS-CoV-2 spike protein. *CourseSource*. <https://doi.org/10.24918/cs.2022.5>

Editor: Charles Hauser, St. Edward's University

Received: 1/5/2021; **Accepted:** 9/9/2021; **Published:** 3/18/2022

Copyright: © 2022 Listenberger, Joiner, and Terrell. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited. The images used in supporting materials (Supporting File S3: Molecular Modeling – Summative Assessments and Data from Student Responses) are from journals that use the Creative Commons Attribution License. We cite the original source for each figure. The primary image includes student generated data and a cartoon from Pixabay, an online repository of copyright free art.

Conflict of Interest and Funding Statement: None of the authors has a financial, personal, or professional conflict of interest related to this work.

Supporting Materials: Supporting Files S1. Molecular Modeling – BioMolViz Goals and Objectives; S2. Molecular Modeling – Assignments I-IV; S3. Molecular Modeling – Summative Assessments and Data from Student Responses; S4. Molecular Modeling – Answer Key (contact corresponding author for a copy); and S5. Molecular Modeling – Teaching Resources.

***Correspondence to:** University of Minnesota, Center for Learning Innovation, 111 S. Broadway, Rochester, MN, 55904, terre031@r.umn.edu.

Learning Goals

Students will:

- understand how online databases, bioinformatics, and molecular visualization tools can be used to examine protein structure.
- understand that chemical and physical forces govern protein structure.
- understand that chemical and physical forces govern protein-protein interactions.
- understand how virus-host interactions contribute to viral infectivity

From *Vision and Change*:

- "Structure and function: Basic units of structure define the function of all living things."
- "Ability to use modeling and simulation: Biology focuses on the study of complex systems"
- "Ability to tap into the interdisciplinary nature of science: Biology is an interdisciplinary science."

From the *Threshold Concepts of Biochemistry* (1):

- "The physical basis of interactions."

Learning Objectives

Students will be able to:

- use online databases and bioinformatic tools to create a protein sequence alignment and identify conserved residues.
- use molecular visualization tools to identify secondary, tertiary, and quaternary structural features of a protein.
- predict cellular location (cytosolic vs. transmembrane) by assessing hydrophobicity of surface residues.
- identify the intermolecular (noncovalent) forces that stabilize and contribute to folding the secondary, tertiary, and quaternary structural features of a protein.
- identify the intermolecular (noncovalent) forces that mediate protein-protein interactions.
- predict how a single amino acid substitution can change protein structure and function.
- consider amino acid mutations in the context of evolution and can predict whether a specific substitution is likely to be conserved in a population.
- choose the best structural rendering (i.e., cartoon, surface, sticks, etc.) to use to convey a specific characteristic of a protein.

INTRODUCTION

Online molecular visualization and bioinformatics tools are increasingly popular for teaching protein structure. Students generally report increased understanding of protein structure and enthusiasm for assignments that engage with virtual models (2–4). Instructor assessments also show gains in student learning (5–8). Despite these advantages, development of guided molecular visualization activities is time consuming. There remain relatively few resources for instructors.

We previously developed a multi-week, inquiry-based molecular visualization project to guide students through analysis of the prostaglandin H_2 synthase (also known as cyclooxygenase-1) structure (9). We sought to (i) promote student understanding of protein structure/function, and (ii) increase student knowledge about, and confidence in using, online databases and computational tools. The skills and content included in these assignments address both core concepts (structure/function) and competencies (modeling and simulation; accessing, comprehending, and communicating science) from the ASBMB learning framework (10) and Vision and Change (11). Additionally, students practice many of the visual literacy skills outlined by Dries et al. (12) and the BioMolViz working group (Supporting File S1. Molecular Modeling – BioMolViz Goals and Objectives).

Here, we update our molecular visualization project to explore the structure of severe acute respiratory syndrome (SARS)-like coronavirus 2 (SARS-CoV-2), the virus responsible for COVID-19.

Scientists around the globe have mobilized to understand the molecular interactions that allow the virus to interact and ultimately fuse with human host cells (13). Coronaviruses, like SARS-CoV-2, include proteins that protrude out from their surfaces, forming “spikes” (14). The SARS-CoV-2 spike protein binds to the angiotensin-converting enzyme 2 (ACE2) at the surface of human cells. The interaction between SARS-CoV-2 and ACE2 triggers the host cell to take up the virus (15). In our updated molecular visualization project, students examine the SARS-CoV-2 spike protein and its interaction with ACE2. We chose exploration of the SARS-CoV-2 spike protein as part of our ongoing effort to promote inclusivity with relevant examples (16).

Instructors who aim to bring related topics into their class may also be interested in three recent teaching and learning articles. The Research Collaboratory for Structural Bioinformatics (RCSB) has published materials for exploring the SARS-CoV-2 main protease (Nsp5). Students create a sequence alignment and use Foldit to model the 3D structure of the protein (17). Bryce and colleagues guide students through analysis of genetically related viral genomes to identify small well-conserved regions of the viral replicase polyproteins PP1a and PP1ab for drug or vaccine targets (18). Lorusso and Shumskaya developed online laboratory exercises for multiple sequence alignment, phylogeny construction, and protein modeling using SARS-CoV-2 data and the PyMOL molecular visualization system (19).

Our project, referred to herein as the Molecular Modeling Project, includes four assignments that together explore the primary, secondary, tertiary, and quaternary structure of the SARS-CoV-2 spike protein. Students use the National Center for Biotechnology Information (NCBI) databases to create

a protein sequence alignment. Students use UCSF Chimera (referred to as Chimera) to visualize and analyze the structure of the SARS-CoV-2 spike protein and its interaction with ACE2, investigate how mutations within the spike protein might affect its interaction with ACE2, and hypothesize how changes to structure can alter viral infectivity. Each assignment includes instructions and embedded questions. Some questions require annotated images to display a specific feature of the protein structure. Our students complete each assignment as homework outside of class and are evaluated individually. However, consultation with the instructor, undergraduate teaching assistant, and other students is encouraged.

Intended Audience

These activities were designed for third- and fourth-year undergraduate students enrolled in an upper-level biochemistry course at one of two small, primarily undergraduate institutions (PUIs). Course sections ranged from 22–52 students. Prior to enrollment into our biochemistry courses, students have completed four semesters of college-level chemistry coursework (General Chemistry I and II, Organic Chemistry I and II) with a minimum of a C- in all prerequisite courses. Our students are mostly life science majors pursuing either a Bachelor of Arts (at St. Olaf College) or a Bachelor of Science degree (at University of Minnesota, Rochester (UMR)).

Required Learning Time

Students reported completing each assignment in 3–6 hours. These four activities were either spread across the entire semester (UMR) or completed during the first six weeks of the semester to coincide with discussion of macromolecular structure, including protein structure and function and enzymes (St. Olaf College).

Prerequisite Student Knowledge

Prior to introducing the Molecular Modeling Project, students are introduced to the amino acids and peptide structure. At both institutions, our students are required to memorize the amino acid structures, biologically relevant functional groups, and one letter codes. They are also able to draw the amino acids and short peptides in skeletal form at any pH between 0–14. Throughout the project, assignment due dates are set to occur after the corresponding content is taught in the course. Amino acids and primary protein structure, secondary protein structure, tertiary and quaternary protein structure, and protein function are covered before Assignment I, II, III, and IV are due, respectively. Within these lessons students also learn how to draw and identify noncovalent interactions between amino acid functional groups.

Prerequisite Teacher Knowledge

The instructor needs knowledge of amino acid, peptide, and protein structure and the noncovalent interactions that stabilize secondary, tertiary, and quaternary protein structure. We have structured the assignments such that anyone with basic protein knowledge can successfully use the online databases and bioinformatics and molecular visualization tools without previous experience. We suggest that instructors work through the four assignments before presenting them to the class, so that they become familiar with these tools and are able to anticipate difficulties. Links to the NCBI database, UCSF Chimera, and video tutorials of common Chimera commands can be found in Supporting File S5. Molecular Modeling – Teaching Resources.

SCIENTIFIC TEACHING THEMES

Active Learning

All assignments were completed outside of class time. Students engaged with:

- Inquiry-based learning worksheets
- Experiential learning by utilizing online databases and computer models

Assessment

Learning was measured using these methods:

- Assessment of student responses to the assignment questions
- Assessment of students' figures and figure legends using rubric analysis
- Summative assessment of student responses to in-class activity questions
- Summative assessment of student responses to exam questions

Inclusive Teaching

"Giving students opportunities to think and talk about biology" (16)

- "Allow students time to write"
 - Each student works through each assignment at their own pace, while still being able to work with teaching assistants and instructors for help. This gives the individual student the time needed to work through the different computational tools and think through the given assignments.

"Building an inclusive and fair biology classroom for all students" (16)

- "Integrate culturally diverse and relevant examples"
 - The students worked with the SARS-CoV-2 protein responsible for the COVID-19 disease and pandemic. They used molecular modeling to interrogate the structure of the main domain of the SARS-CoV-2 spike protein that is responsible for the interaction with and infectivity of humans by this virus.

"Monitoring (your own and students') behavior to cultivate divergent biological thinking" (16)

- "Ask open-ended questions"
 - Each assignment contained open-ended questions. Answers could vary between students. These questions were assessed based on the clarity of explanation and communication.

"Promoting engagement and self-efficacy" (20)

- It was emphasized to students that the computational and molecular visualization tools were used in academic and industrial research efforts.
- It was emphasized that the protein structures of the SARS-CoV-2 spike protein and the human ACE2 receptor that were used in this project are also used by experts to interrogate the interaction and identify possible therapeutic interventions.

LESSON PLAN

Pre-Semester Preparation

Each instructor reviewed the assignments to test the instructions with current versions of software and become familiar with the programs used in the project. We also determined when to set

due dates and how much the project contributed to the final grade. At St. Olaf College and UMR, points earned for this project contributed to 10% of the overall course grade. The assignments were posted to our respective learning management systems.

Molecular Modeling Project (Assignments I-IV) and Summative Assessments

The Molecular Modeling Project (Supporting File S2. Molecular Modeling – Assignments I-IV) and related assessments (Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses) were implemented in two undergraduate biochemistry I courses at PUIs (Table 1). Students were introduced to the project on the first day of class with an explanation of how it will help them explore and understand protein structure. The four assignments in the project were completed outside of class time. We encouraged students to begin these assignments in advance of the due dates. The amount of time needed to complete each assignment varied by student but was typically 3-6 hours per assignment. We required students to turn in their own individual assignments, although we encouraged students to talk to each other about their work. A common answer key was used to grade student responses to the assignment questions (Supporting File S4. Molecular Modeling – Answer Key (contact corresponding author for a copy)). For each assignment some questions were graded based on the answer key, with some questions awarded points for completion. The division of points was decided by individual instructors.

We supported student work on the Molecular Modeling Project through additional help sessions. These help sessions were specifically for work on the Molecular Modeling Project and were scheduled to align with assignment due dates. A typical schedule included two, two-hour help sessions scheduled for the week before each assignment due date. At St. Olaf College, help sessions were run by students who had previously completed the course. At UMR, help sessions were staffed by the course instructor. In a typical year, the help sessions were well attended with over 50% of the class choosing to participate. Moving the help sessions online during the fall of 2020 decreased attendance. We encouraged students to use the help sessions as work time to complete the assignments, rather than waiting to have a question before choosing to attend. With this model, students who got stuck while working through an assignment had easy and immediate access to help. While the teaching assistants (at St. Olaf) or course instructor (at UMR) were available to provide this help, most questions were answered by other students attending the help sessions. We found that students working in parallel on the Molecular Modeling Project in the same physical space were likely to collaborate and share ideas.

Assignment I: Primary sequence analysis of SARS-CoV-2 spike protein

Students begin the protein structure and function learning progression by analyzing the primary sequence of the SARS-CoV-2 spike protein. We encourage students to evaluate a published sequence alignment of spike proteins from coronavirus particles that infected pangolins and bats. Students read the author's conclusions from this paper and answer questions about what was learned.

Using various databases from the NCBI, students next create and analyze another protein sequence alignment of the SARS-

CoV-2 spike protein. We emphasize that these tools are a common first step when a scientist begins research with a novel or under-studied protein.

Assignment II: An introduction to UCSF Chimera and review of protein structure

In the second assignment students use computer modeling to examine the structure of a leucine zipper dimer (PDB ID: 1ZIK) that is entirely alpha helical (21). Here, students learn to use the molecular visualization program, Chimera, by testing various renderings and commands, finding hydrogen bonds, and displaying sidechains. Students create a figure to show the hydrogen bonding pattern in an alpha helix. Many of the inquiry-based learning questions guide students in using the molecular visualizations to assess aspects of protein structure and function.

Assignment III: Global appraisal of the structure and function of SARS-CoV-2 spike protein

Students return to the SARS-CoV-2 spike protein and investigate the secondary, tertiary, and quaternary protein structure and function using Chimera. In this inquiry-based, experiential assignment students use the full structure of the SARS-CoV-2 spike protein (PDB ID: 6VXX) to model quaternary structure features (22). Students are next introduced to the SARS-CoV-2 spike protein's receptor binding domain (RBD) in complex with part of the ACE2 receptor protein (PDB ID: 6M0J) (23). Students manipulate and explore secondary and tertiary protein structure and function content with this rendering. Throughout the assignment, students generate figures to represent the molecular story at hand, which include secondary structural patterns, examples of tertiary and quaternary structure, and the 3D fold of the protein.

Assignment IV: Exploring the SARS-CoV-2 spike protein and ACE2 receptor interaction interface

Students complete the project by modeling and assessing SARS-CoV-2 spike protein mutations in Chimera. Utilizing the SARS-CoV-2 spike protein RBD in complex with the ACE2 receptor (PDB ID: 6M0J), students are challenged to connect protein structure and function content to infectivity. Students begin the assignment by modeling and assessing two known mutations. In the process students learn the commands for modeling point mutations in Chimera. The questions guide students to interpret the visual results. This learning progression is appropriately scaffolded for the student to then propose a mutation that could be found in a population of infectious viral particles. Again, students represent these molecular events by creating figures in Chimera.

Summative Assessments for the Project

Summative assessments were incorporated into two in-class activities, three regular exams, and the final exam. Examples of assessment questions from UMR are shared as supplemental material (Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses).

Anticipated student challenges

After utilizing the project in our courses, we want other instructors to be aware of technological challenges students may encounter:

- Chimera does not have an “undo” or “back” feature. We remind students to save their “sessions” frequently as they work through the assignments.

- Chimera has a “delete” option for atoms and bonds. We encourage students to use the “hide” option rather than deleting atoms/bonds. If atoms/bonds have been “hidden” they can be “shown” again, but deleted items cannot be retrieved without starting over.
- When saving an image, we suggest using the “Save Image” option under “File” to produce higher quality images. If the “Capture/Snip” tool is used the images will be pixelated.
- Students do not readily notice the lack of hydrogens and double bonds when the PDB file is retrieved. This can make recognizing functional groups and identifying amino acids challenging. Instructors can add hydrogen atoms to pdb files using the online MolProbity tool, which can then be opened in Chimera (32).
- The “Find H-bond” tool in Chimera will locate a noncovalent interaction between two polar functional groups - some of these are salt bridges and ion-dipoles, rather than hydrogen bonds. Assignment II is a good place to mention this to students either in class or in the feedback on their figure.
- In the SARS-CoV-2 spike protein - ACE2 complex (PDB ID: 6M0J), the spike protein (Chain E) and ACE2 (Chain A) have similar residue numbers. When using the “sel:” and “focus:” commands, Chimera will select the residues in both structures. This confused students, so we suggest that they hide the ACE2 complex when working on mutations in the spike protein.

TEACHING DISCUSSION

The *Vision and Change* report and the *ASBMB Foundational Concept Framework* call for learning activities that promote the use of molecular modeling and visualization in undergraduate coursework (10, 11). Integrating the exploration of protein structure and function in these activities serves to enhance student understanding of these abstract, visually complex concepts. Our previous implementation of a molecular modeling project in undergraduate biochemistry increased students' knowledge of and familiarity with the programs utilized (9). In this new version of the project, we chose to focus on the SARS-CoV-2 spike protein. This project has been tested at two undergraduate institutions in biochemistry I courses. One of the institutions shares the summative assessments (Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses) used to measure student understanding of the learning objectives. The data analysis of student responses is from UMR, where each student completed the project independently using time outside of class. In addition to the data analysis, we note observations of student behaviors and performance for each assignment separately.

Assignment I: Primary Sequence Analysis of SARS-CoV-2 Spike Protein

In general, students easily navigate the NCBI site and follow instructions to create the alignment (Supplemental Table 1 in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses). Analysis of the alignment is more difficult. Students' scores on question 11 are the lowest for the assignment (Supplemental Table 1 in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses). To answer this question, students need to identify amino acids by single

letter code and recognize similarities in amino acid size and properties. Students are then asked to use this content knowledge to judge whether aligned amino acids are conserved. Students may find this difficult because few have significant practice applying biochemical knowledge to new scenarios. Summative assessment questions were incorporated into an in-class activity to help students connect these skills and content with further practice. On exam 1, we asked similar questions about a sequence alignment that students had not seen before. Item analysis of Q1 and Q2 from exam 1 questions further indicate that many students struggle to synthesize the skills and content required to analyze a protein sequence alignment (Table 2). Despite these challenges, this assignment had the highest total score of all the assignments in this project (Supplemental Table 1 in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses). While additional work is needed, most students demonstrate the ability to correctly create a protein sequence alignment and analyze the alignment by identifying conserved residues.

Assignment II: An Introduction to UCSF Chimera and Review of Protein Structure

In the second assignment students learn a molecular modeling and visualization program to determine the factors that influence protein structure. Observationally, students ask many questions about creating the figure and concomitantly comment that this requires the most time and thought of all the questions on this assignment. We have found it helpful to talk to students about the importance of figures in clearly communicating science. We hear less frustration when we explain why we ask them to learn how to create a figure that tells a story or makes a specific point.

Some students are puzzled by Q6 as they need to reconcile that the typical length of a hydrogen bond is shorter than the observed values in Chimera. (This is because the hydrogens are not shown in the structure.) Overall, students perform well on this assignment as indicated by the total score (Supplemental Table 1 in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses). The item analysis average was at least 71.0% for each question, with the lowest and widest distribution of scores on Q3 (Supplemental Table 1 in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses). This question again requires students to synthesize cognitively complex elements. Students must see the lysine side chains oriented away from the center, recognize that lysine is a polar amino acid (the use of CPK colors could aid or complicate this for students), recognize that the center dimer interface is hydrophobic, and recall the hydrophobic effect. Student responses to other questions demonstrate the ability to use Chimera to manipulate a protein structure and identify the best visual renderings to present specific features of a structure. Students' scores on exam 1, Q3 suggest many students struggle to connect their knowledge of protein secondary structure with the visual representation of a protein ribbon diagram (Table 2).

Assignment III: Global Appraisal of the Structure and Function of SARS-CoV-2 Spike Protein

In Assignment III, students are tasked with recognizing and locating structural features using the given PDB IDs, understanding that “biological molecules are complex”, and that “structure is determined by several factors” such as the “physical basis of interactions.” (1, 10). Students comment that locating

examples of quaternary and tertiary structure is challenging and making quality figures of these interactions is time consuming. When analyzing student figures of these interactions the most common errors were mislabeling the interaction, failing to indicate the subunit each residue was in, and/or selecting residues within the same subunit when demonstrating quaternary structure. As such, the questions that correlated with the figures had lower scores. The other two assignment questions with low scores were Q1 and Q11 (Supplemental Table 1 in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses). Many students describe the secondary structure for question 1, instead of the quaternary structure. This may be because they had previous practice with this and less practice describing quaternary structure. Q11 is conceptually and visually challenging, because the unstructured regions of the protein are needed for protein dynamics - a process that Chimera (and other two-dimensional representations) does not model. We suspect our students have an under-developed mental model of protein dynamics. It's not surprising that students would struggle to understand and communicate this. In all, assignment III had the lowest total score of all the assignments in this project (Supplemental Table 1 in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses).

At UMR, the in-class summative assessment following the assignment III due date fell in the Lipids Module and was an opportunity to integrate the assignment content and skills with membrane structure and function content (Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses). These questions facilitated a robust discussion among the students and with the instructor. The corresponding summative assessment on the exam 2 (Q2) had a high correct response rate (Table 2); however only 64.7% of students selected the correct response to exam 2 (Q1) question - *What is the best view of protein to predict cellular location?* (Table 2). These scores indicate that students are able to understand the content information, but the visual representational skills are more challenging.

Assignment IV: Exploring the SARS-CoV-2 Spike Protein and ACE2 Receptor Interaction Interface

Students complete the project by assessing the mutations in the spike protein and predicting the impact on the interaction between the spike protein and the ACE2 receptor protein. Students are surprised and concerned that the first mutation (A348T) is not located near the ACE2 protein. There seems to be a lack of confidence in student ability to model the mutation and/or interpret what they see in the program. This spurred many conversations between the students and instructor and offered an opportunity to discuss the implications of such a mutation on the allostereics of the protein, that could not be modeled in Chimera but may occur *in vivo*. During these conversations the D614G mutation discussed in the in-class summative assessment for assignment III was used as an example of a mutation that is also distanced from the interface yet causes an allosteric change that could positively engage the spike protein with the ACE2 receptor. Student mistakes on the lowest scoring question (Q8) included mutating the ACE2 protein instead of the spike protein, mislabeling the change in noncovalent interaction, and/or not including enough information in their prediction (Supplemental Table 1 in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses). In all, the

total score on assignment IV indicates successful completion of the tasks and responses to questions (Supplemental Table 1 in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses). Additionally, students were challenged to predict whether the mutations in the spike protein would be likely to persist in a population of infectious virus, as a learning objective of the assignment was for students to understand the impact of mutations on infectivity. Students were able to predict this successfully in the assignment; however, the summative assessment questions on the third exam indicate a lack of retention and/or understanding (Table 2). Student responses to Q1 on the third exam indicate an ability to identify and rank noncovalent interactions based on representations rendered in Chimera; however, students were less successful on Q2 and Q3 from exam 3 (Table 2). Considering that a vast majority were able to identify and rank the noncovalent interactions based on a series of renderings, it is likely that students had trouble interpreting the mutation representations. The two most selected answers were the strongest noncovalent interactions, so students did correlate that a stronger noncovalent interaction would correspond with a smaller K_d. Misinterpreting how the mutations were displayed could account for low scores on exam 3 as well. In this advanced question, students needed to select only mutations that 1) occurred on the spike protein, 2) was at the receptor interface, and that 3) weakened the interaction between the spike protein and ACE2 receptor protein. Students likely would benefit from additional practice with analyzing renderings of protein mutations.

Conclusions and Future Directions

In summary, the assignments in this project provide students the opportunity to explore protein structure and function while developing visualization and modeling skills. This project and summative assessments may be incorporated into a variety of course formats. We implemented this project as homework for face-to-face (St. Olaf College) and remote (St. Olaf College and UMR) undergraduate biochemistry courses. Students had the opportunity to work with instructors or teaching assistants during optional help sessions. Other instructors may choose to assign this project to groups of students and/or engage with this work in class or lab (online or virtual). Importantly, this project brings current, relevant COVID-19 content into an undergraduate biochemistry course.

Total scores and item analysis on the assignments suggest students are successful at completing the tasks and responding to the prompts. The item analysis of student responses to the summative assessments indicates the challenge that students face in synthesizing these abstract concepts and visualizations. Assessments that measure and capture the students' mental model progression before, during, and after a series of activities, like the ones provided here, would provide interesting and useful information on how students build an understanding of these concepts.

Modifications and Extensions

The current project was designed for an upper-level biochemistry course to enhance students' understanding of protein structure and function of the SARS-CoV-2 spike protein. However, many modifications can be made to extend this project to different biology and chemistry courses. Assignment I, which focuses on SARS-CoV-2 primary sequence and its conservation among other viruses, can be used in both genetics

and immunology courses. In biophysics or other biochemistry courses, assignment II-III can be expanded to discuss the different sources used for obtaining structural data (i.e., X-ray crystallography, NMR, and cryo-electron microscopy), the advantages and limitations of those sources, and the quality of the structural data (i.e., resolution, R-value, etc.) and how those variables impact the inferences made about structures. Additionally, assignment IV, which focuses on the interaction between the SARS-CoV-2 Receptor Binding Domain (RBD) and the ACE2 receptor, can be modified for an immunology course to look at the RBD and its interactions with neutralizing antibodies that block ACE2 binding (24-26).

In assignment IV, we look at two mutations in the SARS-CoV-2 RBD found in (27) and hypothesize how these mutations affect RBD-ACE2 binding and viral infectivity. The SARS-CoV-2 spike protein is constantly mutating and several variants found in the population (i.e., B.1.1.7, B.1.351, B.1.617.2, P.1) have shown increased transmission and infectivity (28-31). This assignment can be modified to have students choose one of these variants to study in Chimera and hypothesize why these variants are transmissible based on the interaction between the SARS-CoV-2 RBD and the ACE2 receptor. For reference, we have added a table of the four variants above which includes the variant-specific mutations in the spike protein and the RBD (Supporting File S5: Molecular Modeling - Teaching Resources). In addition, this assignment could have students explore how these variants have disproportionately affected different populations.

Finally, this project was designed to have students learn to use Chimera, a free open-source molecular visualization system. Students can also use the next-generation system, ChimeraX, but instructors will need to check the command language between Chimera and ChimeraX and modify the assignments as needed. Additionally, instructors can use the online MolProbity tool to add hydrogen atoms to structures in PDB files, which can be viewed in Chimera (32).

SUPPORTING MATERIALS

- S1. Molecular Modeling – BioMolViz Goals and Objectives
- S2. Molecular Modeling – Assignments I-IV
- S3. Molecular Modeling – Summative Assessments and Data from Student Responses
- S4. Molecular Modeling – Answer Key **Please contact corresponding author for a copy of the answer key.*
- S5. Molecular Modeling – Teaching Resources

ACKNOWLEDGMENTS

We thank our students for their feedback and participation in this study. This study was approved by the IRB at the University of Minnesota under IRB Protocol # 0908S71602 titled "Investigating Student Learning in an Integrated Curriculum" and per IRB guidelines all identifying information was removed for data analysis and dissemination.

REFERENCES

1. Loertscher J, Green D, Lewis JE, Lin S, Minderhout V. 2014. Identification of threshold concepts for biochemistry. *CBE Life Sci Educ* 13:516–528. doi: 10.1187/cbe.14-04-0066
2. Rigsby RE, Parker AB. 2016. Using the PyMOL application to reinforce visual

- understanding of protein structure. *Biochem Mol Biol Educ* 44:433–437. doi: 10.1002/bmb.20966
3. Hark AT. 2017. Understanding protein domains: A modular approach. CourseSource. doi: 10.24918/cs.2017.21
 4. Sung R-J, Wilson AT, Lo SM, Crowl LM, Nardi J, St. Clair K, Liu JM. 2020. BiochemAR: An augmented reality educational tool for teaching macromolecular structure and function. *J Chem Educ* 97:147–153. doi: 10.1021/acs.jchemed.8b00691
 5. White B, Kim S, Sherman K, Weber N. 2002. Evaluation of molecular visualization software for teaching protein structure differing outcomes from lecture and lab: Differing outcomes from lecture and lab. *Biochem Mol Biol Educ* 30:130–136. doi: 10.1002/bmb.2002.494030020026
 6. Loertscher J, Villafañe SM, Lewis JE, Minderhout V. 2014. Probing and improving student's understanding of protein α -helix structure using targeted assessment and classroom interventions in collaboration with a faculty community of practice. *Biochem Mol Biol Educ* 42:213–223. doi: 10.1002/bmb.20787
 7. Allred ZDR, Tai H, Bretz SL, Page RC. 2017. Using PyMOL to explore the effects of pH on noncovalent interactions between immunoglobulin G and protein A: A guided-inquiry biochemistry activity. *Biochem Mol Biol Educ* 45:528–536. doi: 10.1002/bmb.21066
 8. Lineback JE, Jansma AL. 2019. PyMOL as an instructional tool to represent and manipulate the myoglobin/hemoglobin protein system. *J Chem Educ* 96:2540–2544. doi: 10.1021/acs.jchemed.9b00143
 9. Terrell CR, Listenberger LL. 2017. Using molecular visualization to explore protein structure and function and enhance student facility with computational tools. *Biochem Mol Biol Educ* 45:318–328. doi: 10.1002/bmb.21040
 10. Tansey JT, Baird T, Cox MM, Fox KM, Knight J, Sears D, Bell E. 2013. Foundational concepts and underlying theories for majors in “biochemistry and molecular biology.” *Biochem Mol Biol Educ* 41:289–296. doi: 10.1002/bmb.20727
 11. Brownell SE, Freeman S, Wenderoth MP, Crowe AJ. 2014. BioCore guide: A tool for interpreting the core concepts of Vision and Change for biology majors. *CBE Life Sci Educ* 13:200–211. doi: 10.1187/cbe.13-12-0233
 12. Dries DR, Dean DM, Listenberger LL, Novak WRP, Franzen MA, Craig PA. 2017. An expanded framework for biomolecular visualization in the classroom: Learning goals and competencies. *Biochem Mol Biol Educ* 45:69–75. doi: 10.1002/bmb.20991
 13. Abdelrahman Z, Li M, Wang X. 2020. Comparative Review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A respiratory viruses. *Front Immunol* 11:552909. doi: 10.3389/fimmu.2020.552909
 14. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. 2021. Coronavirus biology and replication: Implications for SARS-CoV-2. *Nat Rev Microbiol* 19:155–170. doi: 10.1038/s41579-020-00468-6
 15. Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, Li F. 2020. Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A* 117:11727–11734. doi: 10.1073/pnas.2003138117
 16. Tanner KD. 2013. Structure matters: Twenty-one teaching strategies to promote student engagement and cultivate classroom equity. *CBE Life Sci Educ* 12:322–331. doi: 10.1187/cbe.13-06-0115
 17. Burley SK, Bromberg Y, Craig P, Duffy S, Dutta S, Hall BL, Hudson BP, Jiang J, Khare SD, Koeppe JR, Lubin JH, Mills SA, Pikaart MJ, Roberts R, Sarma V, Singh J, Tischfield JA, Xie L, Zardecki C. 2020. Virtual boot camp: COVID-19 evolution and structural biology. *Biochem Mol Biol Educ* 48:511–513. doi: 10.1002/bmb.21428
 18. Bryce S, Heath KN, Issi L, Ryder EF, Rao RP. 2020. Using COVID-19 as a teaching tool in a time of remote learning: A workflow for bioinformatic approaches to identifying candidates for therapeutic and vaccine development. *Biochem Mol Biol Educ* 48:492–498. doi: 10.1002/bmb.21413
 19. Lorusso NS, Shumskaya M. 2020. Online laboratory exercise on computational biology: Phylogenetic analyses and protein modeling based on SARS-CoV-2 data during COVID-19 remote instruction. *Biochem Mol Biol Educ* 48:526–527. doi: 10.1002/bmb.21438
 20. Trujillo G, Tanner KD. 2014. Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity. *CBE Life Sci Educ* 13:6–15. doi: 10.1187/cbe.13-12-0241
 21. Gonzalez L, Woolfson DN, Alber T. 1996. Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat Struct Biol* 3:1011–1018. doi: 10.1038/nsb1296-1011
 22. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181:281–292.e6. doi: 10.1016/j.cell.2020.02.058
 23. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X. 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. 7807. *Nature* 581:215–220. doi: 10.1038/s41586-020-2180-5
 24. Barnes C, Jette C, Abernathy M, Dam K-M, Esswein S, Gristick H, Malyutin A, Sharaf N, Huey-Tubman K, Lee Y, Robbiani D, Nussenzweig M, West Jr. A, Bjorkman P. 2020. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* 588: 682–687. doi: 10.1038/s41586-020-2852-1
 25. Barnes C, West Jr. A., Huey-Tubman K, Hoffmann M, Sharaf N, Hoffman P, Koranda N, Gristick H, Gaebler C, Muecksch F, Lorenzi J, Fink S, Hägglöf T, Hurley A, Millard K, Weisblum Y, Schmidt F, Hatzioannou T, Bieniasz P, Caskey M, Robbiani D, Nussenzweig M, Bjorkman P. 2020. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. *Cell* 184: 828–842. doi: 10.1016/j.cell.2020.06.025
 26. Fu D, Zhang G, Wang Y, Zhang Z, Hu H, Shen S, Wu J, Li B, Li X, Fang Y, Liu J, Wang Q, Zhou Y, Wang W, Li Y, Lu Z, Wang X, Nie C, Tian Y, Chen D, Wang Y, Zhou X, Wang Q, Yu F, Zhang C, Deng C, Zhou L, Guan G, Shao N, Lou Z, Deng F, Zhang H, Chen X, Wang M, Liu L, Rao Z, Guo Y. 2021. Structural basis for SARS-CoV-2 neutralizing antibodies with novel binding epitopes. *PLoS Biol* 19: e3001209. doi: 10.1371/journal.pbio.3001209
 27. Lokman S, Rasheduzzaman Md, Salauddin A, Barua R, Tanzina A, Rumi M, Hossain Md, Siddiki A, Mannan A, Hasan Md. 2020. Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: a computational biology approach. *Infect Genet Evol* 84: 104389. doi: 10.1016/j.meegid.2020.104389
 28. Davies N, Abbot S, Barnard R, Jarvis C, Kucharski A, Munday J, Pearson C, Russell T, Tully D, Washburne A, Wensekeers T, Gimma A, Waites W, Wong K, van Zandvoort K, Silverman, J, CMMID COVID-19 Working Group, The COVID-19 Genomics UK Consortium, Diaz-Ordaz, K, Keogh R, Eggo R, Funk S, Jit M, Atkins K, Edmunds W. 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372: eabg3005. doi: 10.1126/science.abg3055
 29. Wang P, Nair M, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong P, Graham B, Mascola J, Chang J, Yin M, Sobieszczyk M, Kyratsous C, Shapiro L, Sheng Z, Huang Y, Ho D. 2021. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* 593: 130–135. doi: 10.1038/s41586-021-03398-2
 30. Yadav P, Sapkal G, Abraham P, Ella R, Deshpande G, Patil D, Nyayanit D, Gupta N, Sahay R, Shete A, Panda S, Bhargava B, Mohan V. 2021. Neutralization of variant investigation B.1.617 with sera of BBV152 vaccines. *Clinical Infectious Diseases* ciab411. doi: 10.1093/cid/ciab411
 31. Wang P, Gasner R, Nair M, Wang M, Yu J, Gerutti G, Liu L, Kwong P, Huang Y, Shapiro L, Ho Do. 2021. Increased resistance to SARS-CoV-2 variant P.1 to antibody neutralization. *Cell Host & Microbe* 29: 747–751. doi: 10.1016/j.chom.2021.04.007
 32. Williams C, Headd J, Moriarty N, Prisant M, Videau L, Deis L, Verma V, Keedy D, Hintze B, Chen V, Jain S, Lewis S, Arendall B, Snoeyink J, Adams P, Lovell S, Richardson J, Richardson D. 2018. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* 27: 293–315. doi: 10.1002/pro.3330

Table 1. Lesson Plan for the Molecular Modeling Project.

Activity	Description	Notes
Pre-semester preparation		
Instructor preparation	<ol style="list-style-type: none"> 1. Complete the assignment and work through questions to anticipate student difficulties 2. Determine due dates 3. Add assignments to learning management system (optional) 4. Add assignments to syllabus 	<ul style="list-style-type: none"> The assignments are provided in Supporting File S2. Molecular Modeling – Assignments I-IV. In our courses, this project (Assignment I-IV) was 10% of the overall course grade.
First day of class		
Mini-lecture introducing the project	Introduce the project and tie to course goals	
Assignment I: Primary sequence analysis of SARS-CoV-2 spike protein		
Content to cover before assignment due date	Cover content related to primary protein structure	
While the assignment is open	<ol style="list-style-type: none"> 1. Remind students of assignment due date. 2. (Optional) Hold help sessions. 3. (Optional) Add the summative assessment questions to an in-class activity and/or an exam. 	<ul style="list-style-type: none"> The assignment is provided in Supporting File S2. Molecular Modeling – Assignments I-IV. The in-class activity and exam summative assessments questions are in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses.
After the assignment closes	Grade assignment and return feedback before the Assignment 2 is due.	The answer key is provided in Supporting File S4. Molecular Modeling – Answer Key (contact corresponding author for a copy).
Assignment II: An introduction to UCSF Chimera and review of protein structure		
Content to cover before assignment due date	Cover content related to secondary protein structure.	This is also a good time frame to cover what the various renderings (surface, ball and stick, spheres, cartoon, etc.) mean and when to use them.
While assignment is open	<ol style="list-style-type: none"> 1. Remind students of assignment due date 2. (Optional) Hold help sessions. 3. (Optional) Add the summative assessment questions to the next exam. 	<ul style="list-style-type: none"> The assignment is provided in Supporting File S2. Molecular Modeling – Assignments I-IV. The exam summative assessment questions are in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses.
After assignment closes	Grade assignment and return feedback before the Assignment 3 is due.	<ul style="list-style-type: none"> The answer key is provided in Supporting File S4. Molecular Modeling – Answer Key (contact corresponding author for a copy). Returning feedback to the students before the next assignment is due will help them improve their figures.
Assignment III: Global appraisal of the structure and function of SARS-CoV-2 spike protein		
Content to cover before assignment due date	Cover content related to tertiary and quaternary protein structure.	
While assignment is open	<ol style="list-style-type: none"> 1. Remind students of assignment due date 2. (Optional) Hold help sessions 3. (Optional) Add the summative assessment questions to an in-class activity and/or exam. 	<ul style="list-style-type: none"> The assignment is provided in Supporting File S2. Molecular Modeling – Assignments I-IV. The in-class activity and exam summative assessments questions are in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses.
After assignment closes	Grade assignment and return feedback before the Assignment 4 is due.	<ul style="list-style-type: none"> The answer key is provided in Supporting File S4. Molecular Modeling – Answer Key (contact corresponding author for a copy). Returning feedback to the students before the next assignment is due will help them improve their figures.

Activity	Description	Notes
Assignment IV: Exploring the SARS-CoV-2 spike protein and ACE2 receptor interaction interface		
Content to cover before assignment due date	Cover content related to protein function.	
While assignment is open	<ol style="list-style-type: none"> 1. Remind students of assignment due date 2. (Optional) Hold help sessions 3. (Optional) Add the summative assessment questions to next exam. 	<ul style="list-style-type: none"> • The assignment is provided in Supporting File S2. Molecular Modeling – Assignments I-IV. • The exam summative assessment questions are in Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses.
After assignment closes	Grade assignment before exam assessment.	<ul style="list-style-type: none"> • The answer key is provided in Supporting File S4. Molecular Modeling – Answer Key. • If using the exam summative assessment question return grading feedback to the student before the exam.

Table 2. Questions from each summative exam were scored and converted to a percentage. The point value used for scoring is in parentheses after the question. The sample count indicates the total number of student responses scored. The mean \pm standard error are represented as percentages, as is the median. Refer to Supporting File S3. Molecular Modeling – Summative Assessments and Data from Student Responses for the summative assessments. All responses are from UMR and students completed each exam individually.

Summative Assessment Question	Sample Count	Mean % (\pm standard error)	Median %	Percent of students earning full credit
Questions corresponding to assignment I content				
Exam 1. (Q1) Which of the following residues are MOST likely to be located in the hydrophobic core? (5 pts)	51	66.7 (5.4)	80.0	45.1
Exam 1. (Q2) Which of the following residues are MOST likely to form favorable noncovalent interactions with the cannabidiolic acid A (CBDA) molecule? (3 pts)	51	52.9 (6.2)	33.3	43.1
Questions corresponding to assignment II				
Exam 1. (Q3) In the protein image here, the H-bonds in the structure pointed to with arrow C are between_____. (2 pts)	51	65.7 (6.6)	100.0	64.7
Questions corresponding to assignment III				
Exam 2. (Q1) What is the best view of protein to predict cellular location? (2 pts)	51	64.7 (6.8)	100.0	64.7
Exam 2. (Q2) We have covered the biochemistry of the coronavirus in a few different contexts (protein structure + function, enzyme inhibitors, membranes, etc.). Select all true statements related to the biochemistry of the coronavirus. (5 pts)	51	81.6 (3.2)	100.0	54.9
Questions corresponding to assignment IV				
Exam 3. (Q1) Using these figures depicting interaction between the SARS-CoV-2 spike protein and the ACE2 receptor place them in order from strongest to weakest noncovalent interaction. (4 pts)	50	80.0 (4.1)	100.0	66.0
Exam 3. (Q2) Using these figures select the mutation that would result in the smallest Kd between the SARS-CoV-2 spike protein and the ACE2 receptor. (2 pts)	50	42.0 (7.1)	0.0	42.0
Exam 3. (Q3) Using these figures select the mutations would you anticipate to find in an infections population of coronavirus. (3 pts)	50	59.3 (5.1)	50.0	38.0