

An *In Silico* Approach to Protease Discovery

Pavan Kadandale¹, Marc Sprague-Piercy² and Rachel W. Martin³

¹University of California Irvine, Department of Molecular Biology and Biochemistry, 2234 McGaugh Hall ZOT3900, Irvine CA 92697-3900 USA

²University of California Irvine, Department of Molecular Biology and Biochemistry, 3304 Natural Sciences 1 ZOT2025, Irvine CA 92697-2025 USA

³University of California Irvine, Departments of Chemistry and Molecular Biology & Biochemistry, 4136 Natural Sciences 1 ZOT2025, Irvine CA 92697-2025 USA

(pavan.k@uci.edu; piercym@uci.edu; rachel.martin@uci.edu)

A current area of research in Dr. Martin's lab is the identification of novel proteases. Based on this, we have developed a simple, 1-2 class activity that allows students to contribute to cutting-edge research as part of the regular curriculum. Using modern software, students view and analyze protein structures, and predict whether an unknown protein might be a protease with novel functionality. Based on the idea that structure and function are related and conserved, students compare structures of predicted proteases to the structure of a known protease and annotate both conserved and novel features. Through this activity, students become conversant with the main secondary structure elements of proteins and how they fold together. They will learn how to use Chimera to visualize and analyze protein structure. Data generated by the students will be analyzed in Dr. Martin's lab, potentially identifying interesting new proteases for experimental characterization.

Keywords: protein structure, biochemistry, inquiry-based learning, proteases, bioinformatics

Introduction

One of the key areas covered in almost every undergraduate Biochemistry course is our understanding of protein structure. In fact, the connection between macromolecular structure and function is one of the four foundational concepts identified by the American Society for Biochemistry and Molecular Biology (ASBMB). ASBMB has also identified that understanding how and why proteins fold is a threshold concept for mastering the fundamentals of Biochemistry (ASBMB). Distillation of numerous studies have led to the conclusion that allowing students to manipulate structures - either using physical models or computer-based representations of molecular structures - enables them to better understand the correlation between physical structure and function (Wu and Shah 2004). Curricular material based on this principle would enable students to develop a stronger understanding of the relationship between protein structure and function, and enable them to more readily grasp the part that this relationship plays in modern research.

Using Computer-Based Modeling to Teach Protein Structure

A number of different approaches have been used to incorporate modeling into the teaching and learning of macromolecular structure. Examples include the use of 3D printed models and pipe cleaners to explain protein structure and function (Herman et al. 2006; Marshall 2014). We used computer-based models in our classroom to explore protein structure and function since it offers several advantages for us:

- It is easily scalable, allowing for us to use these models in a classroom of 400 students.
- We can easily change the protein/s used for the modeling, and/or have students use different proteins in the same classroom.
- We can change the scale and representation of the model in order to explore different ideas about protein structure-function. For example, we can use space-filled models to show constraints within the active site, or overlay amino acid charges onto the structure of the protein to show the importance of differential charge on the function of a protein.

- It allows students to easily manipulate the model in three dimensions and allows them to zoom into and out of specific areas.
- We can correlate amino acid composition and/or properties with the structure and/or function of the protein.
- We can overlay different proteins onto each other to compare their structures.
- Software for protein modeling is freely available for all the major platforms and is relatively easy to install and use.
- Using computer-based models allows one to easily change the exercises to achieve different learning objectives, or to adapt them to different class contexts.
- Because of the scalable and inexpensive nature of these methods, we can create inquiry-based teaching modules in many different class contexts and get students thinking about the scientific method and discovery of new knowledge.

Leveraging all these advantages, we created an inquiry-based module that we implemented in a large undergraduate lower division Biochemistry class (~400 students) to engage the students in ongoing research at UC Irvine that is based on protein structure and function.

An *in silico* Approach to Protease Discovery

One of the areas of active research in Dr. Martin's laboratory is the discovery of novel proteases. Newly discovered proteases have the potential to be used as therapeutics to help decrease the spread of nosocomial infections or in industrial applications. Recently, Dr. Martin's lab has created an *in silico* analysis pipeline that takes the whole genome sequence of an organism and predicts the proteins that would be made by that organism based on the genome sequence (Butts et al. 2016a, 2016b). Since cysteine proteases are a very well-studied family of proteases, the analysis pipeline is then able to not only predict the cysteine proteases that the organism produces, but also predicts with very high accuracy the structure of these proteases (Butts et al. 2016a). This approach was recently used to predict the structure of all the cysteine proteases made by the insectivorous plant, *Drosera capensis* (Cape sundew) (Butts et al. 2016a, 2016b). Insectivorous plants are a particularly interesting set of organisms to study in order to discover new proteases since

they must synthesize proteases capable of breaking down the complex structural proteins of insects at ambient temperatures and physiological pH. What the current approach lacks, however, is the ability to distinguish between relatively conserved proteases and truly novel proteases. In order to achieve this goal, all of the predicted proteases from *D. capensis* need to be manually overlaid on the structure of a known protease and then compared for novel structural features that might identify protease candidates for further study. We used this aspect of the protease discovery pipeline to not only teach students about protein structure, but to also involve them in an authentic research project to predict which proteases were worthy of follow-up studies.

Protein Structure Comparisons to Identify Interesting Proteases

The canonical cysteine protease is papain. Comparing the predicted structure of new cysteine proteases with the known structure of papain would help identify proteases with new and/or interesting structural features and would help students better understand protein structure and function. It would also expose students to modern methods in computer-based protein modeling and allow them to manipulate protein structures within the context of a research project. Based on these objectives, we developed a teaching module that is scalable, inexpensive, and adaptable to a variety of contexts. In this module, students will use the freely available software, Chimera, to view protein structures, overlay and compare two different protein structures, identify common and novel structural elements, and ultimately decide whether a protease is worthy of subsequent study. Since multiple students work independently on the same protein, compiling their results should provide a relatively accurate judgment of whether a protease is structurally novel or not. For students, the module provides a research problem that highlights the importance of protein structure to function and reinforces important concepts in protein folding and structure determination. This module engages them in research and provides one context in which the importance of protein structure to function is made apparent.

Student Outline

Objectives

- Understand how conservation of structure relates to conservation of function
- Use software to visualize protein structure in 3D
- Use software to correlate protein sequence and structure
- Use software to compare structures of two proteins and predict conserved and novel aspects of structure
- Understand how protein structure can predict protein function

Introduction

In the next two classes you will use modern software to synthesize many of the concepts that you have learned about protein structure and folding and help generate data for an ongoing research project at UC Irvine. The main points that you have covered so far, related to protein structure and folding are:

- Proteins are made up of amino acids that are joined together by covalent bonds called peptide bonds.
- The peptide bond itself is planar, but the single bonds connecting the N- C α and C α -C can rotate and bend to some extent.
- Depending on the amino acid sequence, the flexing and bending of these bonds can lead to the formation of secondary structure (α -helices and β -sheets).
- Secondary structure is primarily stabilized by hydrogen bonds.
- Hydrophobic interactions cause portions of the protein to fold inwards, in order to minimize interactions with water. This results in the formation of the tertiary structure of the protein.
- Bends, loops, and regions of disordered structure are also critical to formation of the final 3D structure of the protein.
- The structure of a protein is critical to its function.
- Many proteins share conserved structural elements called motifs and domains.
- Proteins with similar functions often have similar structures, even if their primary amino acid sequences are different.

In the next two classes, we are going to use these concepts to participate in the discovery of novel proteases - research that is ongoing in Dr. Martin's laboratory at UC Irvine.

Discovering Novel Proteases in *Drosera capensis*

Why Discover New Proteases?

Proteases have a wide variety of applications, including in food technology, medicine, industry, leather-making, and as laundry detergents. However, considerations such as their substrate specificity, sensitivity to environmental conditions, and stability currently limit their use (Feijoo-Siota and Villa 2011; Li et al. 2013; Tavano et al. 2018). Thus, finding new proteases that act on a range of substrates, in a range of environmental conditions, and that are stable, would allow us to use proteases in new and interesting applications.

*Why Use *Drosera Capensis* to Discover New Proteases?*

The Cape sundew (*Drosera capensis*) is a hardy carnivorous plant that is capable of completely digesting an insect. Thus, it must contain a wide variety of proteases that are stable, acting over a long period of time, and which are active under relatively mild environmental conditions (not requiring, for example, high heat, or extreme pH in order to be active). These proteases might potentially be useful for a variety of applications, making the Cape sundew an attractive, unexplored venue for the discovery of novel proteases.

*How to Discover Novel Proteases from *Drosera Capensis*, and What Is Your Role in this Process?*

Recently, Dr. Martin's lab has pioneered the prediction of protease occurrence and structure from whole genome sequence data (Butts et al. 2016a, 2016b). Using the whole genome sequence data from *D. capensis*, researchers could predict the proteases that were most likely encoded in the genome of the Cape sundew. Since the catalytic mechanism and structure of cysteine proteases is very well studied and knowing that key elements of catalysis and structure will likely be conserved

amongst all cysteine proteases, they could then predict the structure of a subset of proteases that were cysteine proteases, showing that this procedure could result in very accurate predictions of the structure of these novel cysteine proteases (Butts et al. 2016a, 2016b). What is lacking, however, is a means of deciding whether these proteases are worthy of further investigation.

This is where you can contribute to this area of active research. By overlaying (and comparing) the predicted structure of the cysteine proteases from *D. capensis* with the structure of a canonical cysteine protease, papain, you can predict whether any of these new proteases have features that distinguish them from known proteases, and therefore, whether they are worth further exploration. In doing this, you will solidify many of the concepts that you have learned about protein structure and function and learn to use modern visualization software to explore protein structure in 3D.

Learning to use Chimera

Setting up Chimera

1. Download UCSF Chimera by going to the following link: <https://www.cgl.ucsf.edu/chimera/download.html>
2. Select the appropriate download file for your computer's operating system and download the newest version available for your computer's operating system. Follow the installation instructions specific to your operating system (Mac or Windows)

UCSF CHIMERA

an Extensible Molecular Modeling System

Download Chimera

- [Daily Builds](#)
- [Snapshot Releases](#)
- [Unsupported Releases](#)
- [Old Releases](#)
- [Bug Tracking System](#)
- [Licensing Information](#)
- [Experimental Chimera Features](#)
- [Plug-ins on the Web](#)
- [Graphics Driver Bugs](#)
- [Benchmark Results](#)
- [Chimera Source Code](#)
- [Cygwin Source Code](#)

Current Production Releases

- See the [release notes](#) for a list of new features and other information.
- For [more recent changes](#), use the [snapshot](#) and [daily](#) builds; they are less tested but usually reliable.

• 64-bit Releases:

Platform	Installer, Size, and Checksum	Date	Notes
Microsoft Windows 64-bit	chimera-1.12-win64.exe Size: 150984696 bytes MD5: b662f97cdacb8be458662aa638dd6694	Oct 24, 2017	Instructions Documentation Runs on Windows 7 or later.
Mac OS X 64-bit	chimera-1.12-mac64.dmg Size: 131622631 bytes MD5: 3a691632a3eba8b17286d23432bb6bbb	Oct 24, 2017	Instructions Documentation Runs on Mac OS X 10.8 or later.
Linux 64-bit	chimera-1.12-linux_x86_64.bin Size: 157795228 bytes MD5: 24b9c2560e119988a3150b47c495737b	Oct 24, 2017	Instructions Documentation Compiled on CentOS 5.11.

• 32-bit Releases (for small memory computers):

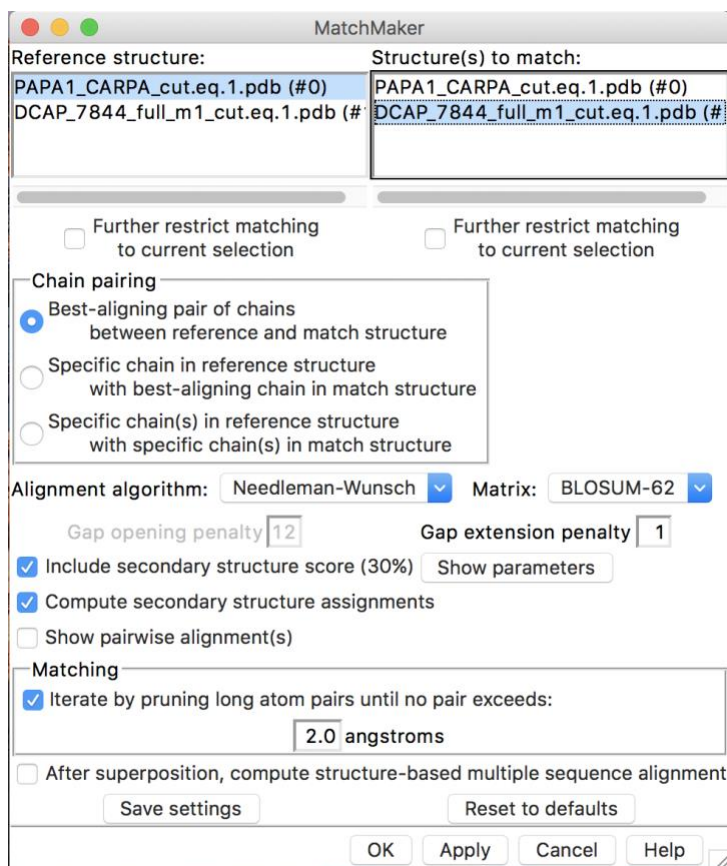
Platform	Installer, Size, and Checksum	Date	Notes
Microsoft Windows	chimera-1.11.2-win32.exe Size: 109260484 bytes MD5: c0201380b0f328f57b9d839efcf646bd	Dec 02, 2016	Instructions Documentation Runs on Windows 7 and 8 or later.
Mac OS X	chimera-1.11.2-mac.dmg Size: 103774888 bytes MD5: fa2ccd9c17c456d71088e81129c862d6	Dec 02, 2016	Instructions Documentation Runs on Mac OS X 10.8 or later.
Linux	chimera-1.11.2-linux.bin Size: 119742278 bytes MD5: 47dd12fbcfcfe01ea678599dd7001a6b	Dec 02, 2016	Instructions Documentation Compiled on Debian 4 (etch).

or Linux) and install Chimera on your computer.

- Next, download the three .PDB files from [please add a link to where you will host the three files here].
- Store them on your computer in a location that is easy to get to.

Opening and Aligning Protein Sequences

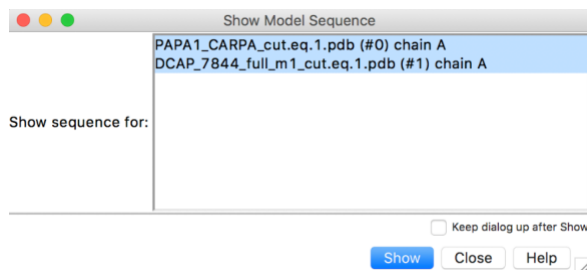
- Open Chimera.
- Open the .PDB file (*File > Open*) for the reference sequence, papain (PAPA1_CARPA_cut.eq.1.pdb) to generate a model of the protein. You should get an image like the one to the right. The reference sequence will be colored beige.
- Open a second protein (either DCAP_0624_full_m1_cut.eq.1.pdb, or DCAP_0515_full_m1_cut.eq.1.pdb – your choice!) that you will compare to papain. You will now have two proteins on top of one another in no particular orientation. Your second protein will be colored light blue.
- In order to align the two proteins with one another, scroll to “Tools”. Then select the following: *Tools > Structure Comparison > Matchmaker*. A screen like the one below will be generated. Select PAPA1_CARPA_cut.eq.1.pdb (#0) for your reference structure, and the protein of comparison (the second one you opened) as the structure to match. Keep all the other defaults the same and then hit "Apply".



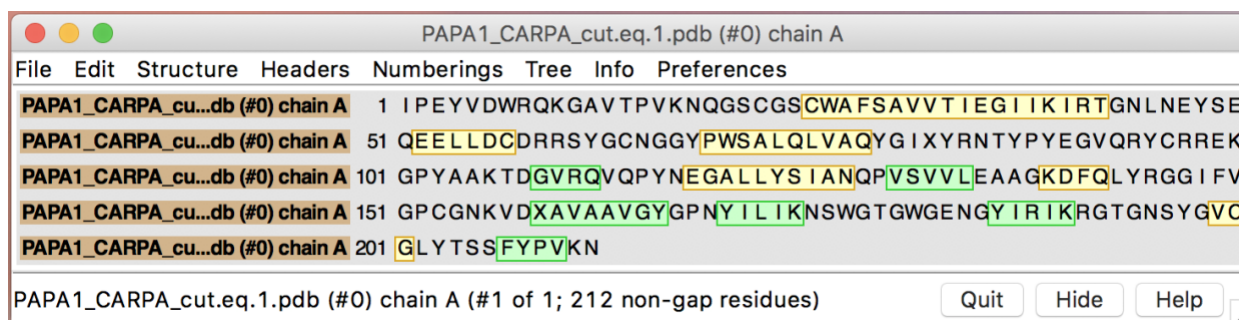
- Your protein sequences will now be aligned!

View the Protein Sequences

10. Open the amino acid sequences for the proteins. Go to *Tools > Sequence > Sequence* and you will reach the screen below. Open both protein sequences (the second protein name will correspond to the protein you chose in the steps above!).



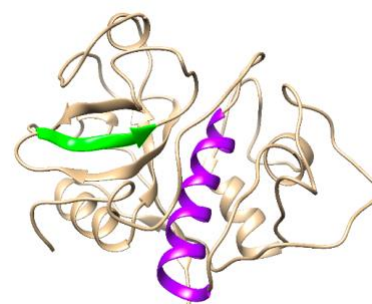
11. You should get a pop-up that looks like the image below. The yellow highlights mean that there is an alpha helix in that area and the green highlights correspond to beta strands.



Tips for Working with Chimera

13. Hover the cursor over the last amino acid in the sequence and the menu will display the residue number.

14. The tertiary structure of the protein in the image to the right contains secondary structure elements. The green colored part of the structure is a beta strand, while the purple colored part is an alpha helix. You can highlight an alpha helix in your protein by selecting the corresponding residues in your protein (in the Sequence view). Then go to *Actions > Color* and then select the color in which you want to highlight that part of your structure.



15. In Chimera there are two methods for determining residue numbers when looking at the protein structure. In the first method pictured above, you can open the sequence menu and count to where the residue is. Or, in the second, simpler method, you can hover your mouse over a certain part of the protein and the number will pop up, giving the type of amino acid, followed by the residue number.

Homework Assignment

1. Align the reference protein with one of the two proteins (either DCAP_0624_full_m1_cut.eq.1.pdb or DCAP_0515_full_m1_cut.eq.1.pdb).
2. Select the FIRST α -helix in the reference protein (Papain, the PAPA1_CARPA_cut.eq.1.pdb), and color it RED.

3. Now, select the FIRST β -strand in the protein you picked to work with (and only in this protein – not the reference) and color it GREEN.
4. Finally, select a secondary structure element that is common to both proteins, and color it YELLOW in both sequences.
5. Save the results as a .jpeg image (*File>Save image*).
6. Upload the image of your final output into the homework assignment.

Class Activity

Objective

The objective of this activity is to identify interesting structural characteristics in novel cysteine proteases. To accomplish this, we will be comparing the structure of an uncharacterized, putative cysteine protease to papain, a well-characterized cysteine protease. By comparing the differences and similarities in structure between these two proteins we hope to find interesting targets for expression and characterization.

Location of Survey:

Find the survey at [Please enter the location of your homework assignment here].

Location of Data files:

Structures of papain (PAPA1_CARPA.pdb) and the uncharacterized proteins can be downloaded from [Please enter the link to the data files here - the data files can be obtained by contacting the authors].

Protocol – Aligning Proteins:

1. Complete the initial homework assignment to understand the basics of using Chimera.
2. Download the structural data for the reference protein, PAPA1_CARPA.pdb.
3. Download the structural data for your novel protease from the link above.
4. Open UCSF Chimera.
5. Open the .pdb file for the reference protease, PAPA1_CARPA.pdb in UCSF Chimera.
6. Open the unknown protein in UCSF Chimera.
7. Align the two proteins.
8. Click on: *Tools* → *Structure Comparison* → *MatchMaker*.
9. Make sure that PAPA1_CARPA.pdb is selected as the reference structure and your uncharacterized protein is selected as “Structure(s) to match:”
10. Open the amino acid sequences for both proteins: *Tools* → *Sequence* → *Sequence*.
11. Select both proteins and click “Show”.
12. Areas on the sequence that are highlighted in yellow are alpha helices and those highlighted in green correspond to beta strands.
13. Select the file name of your uncharacterized protein from the dropdown menu.

Protocol – Comparing Structures:

1. Open up the survey from the link above. We will be working on completing this survey and collecting data on the structural comparison.
2. Select the unknown protein that you are going to be working with from the appropriate dropdown.
3. Determine the length of the protein.

- a) Hover your cursor over the last residue in the sequence, so you can find the last residue number. This corresponds to the length of the protein.
- b) Enter this length in the appropriate box in the survey.

4. Compare α -Helices

- a) In your structure there will be many α -helices. We are first going to find the helix on your *uncharacterized* protein that aligns with the **FIRST** helix on PAPA1_CARPA.
- b) Highlight the **first helix in PAPA1_CARPA** in the sequence view and change the color of this region on the structure view. Find a helix on the uncharacterized structure that **aligns in space** (by looking at the structure!) with this first helix in PAPA1_CARPA. If there is no such helix then answer “no” in the survey/homework question.
- c) Do the same for the rest of the helices in PAPA1_CARPA. For example, for the second helix in PAPA1_CARPA find a helix in your uncharacterized protein that **aligns in space**. Then repeat the same process for the third helix in PAPA1_CARPA and so on. Remember that you are looking for the helices that match up in the **structure of the proteins, not necessarily the sequences**, and that we are referring to the helix numbers based on the reference protease, PAPA1_CARPA.
- d) For each helix in your uncharacterized protein, find a residue near the center of that helix and select the number of that residue in the appropriate dropdown box. You don't have to be exact with finding the central residue - get the residue number of any amino acid within the central portion of the helix.

5. Compare Beta Strands (β -Strands)

- a) In your structure there will be many β -strands. We are first going to find the β -strand on your *uncharacterized* protein that aligns with the **FIRST** β -strand on PAPA1_CARPA.
- b) Highlight the **first β -strand in PAPA1_CARPA** in the sequence view and change the color of this region on the structure view. Find a β -strand on the uncharacterized structure that **aligns in space** with this first β -strand in PAPA1_CARPA. If there is no such helix then answer “no” in the survey/homework question.
- c) Do the same for the rest of the β -strands in PAPA1_CARPA. For example, for the second β -strand in PAPA1_CARPA find a β -strand in your uncharacterized protein that **aligns in space**. Then repeat the same process for the third β -strand in PAPA1_CARPA and so on. Remember that you are looking for the β -strands that match up in the **structure of the proteins, not necessarily the sequences**, and that we are referring to the β -strand numbers based on the reference protease, PAPA1_CARPA.
- d) For each β -strand in your uncharacterized protein, find a residue near the center of that β -strand and select the number of that residue in the appropriate dropdown box. Again, remember that you don't have to be exact with finding the central residue - get the residue number of any amino acid within the central portion of the β -strand.

6. Compare Beta Sheets (B-Sheets)

- a) There are two beta sheets in PAPA1_CARPA. One has 4 strands and the other has 2. Find each of these sheets on PAPA1_CARPA. On your uncharacterized protein find any beta sheet structures that align with these PAPA1_CARPA β -sheet structures. If there is no such structure then note that fact. If the structure is there but has a different number of β -strands then note the difference.

7. Active Site Analysis

- a) You now want to select the “Catalytic” triad of residues in papain. These are the three amino acids that directly catalyze the hydrolysis of the peptide bond, which allows papain to function as a protease. These residues are Cys-25, Asn-175 and His-159 (NOTE: His might be labeled as “X” in these models!) (Figure 1 and 2). Use the numbering corresponding to papain (not the top numbering, which might be misleading) to select these three residues (remember use Shift+Click+Drag to select multiple residues!).
- b) Once you have the three catalytic residues in papain selected, use *Actions* > *Atoms/bonds* > *Side chain/base* > *Show* to display the side chains of the three catalytic residues. Then use *Actions* > *Color* to make these bases white (or any other color of your choosing!). You should end up with something like Figure 2.

- c) You now know where the active site of papain (PAPA1_CARPA) is. The active site of PAPA1_CARPA does not have any loops or helices that could fully or partially block the active site, but your protein might. Find the V-shaped cleft that leads to the active site pocket and note if there are any loops or other structures that could possible block this pocket in the uncharacterized protein.

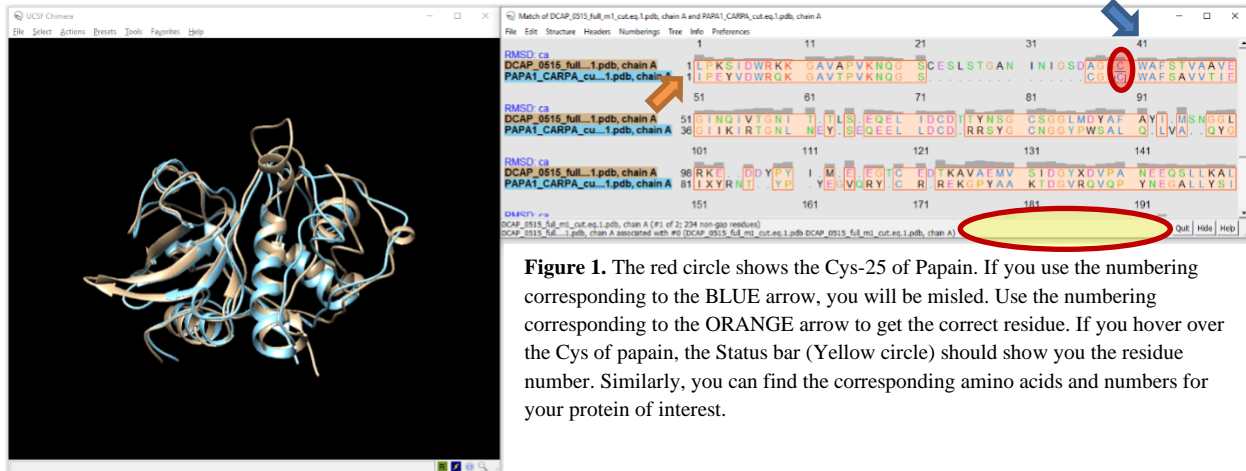


Figure 1. The red circle shows the Cys-25 of Papain. If you use the numbering corresponding to the BLUE arrow, you will be misled. Use the numbering corresponding to the ORANGE arrow to get the correct residue. If you hover over the Cys of papain, the Status bar (Yellow circle) should show you the residue number. Similarly, you can find the corresponding amino acids and numbers for your protein of interest.

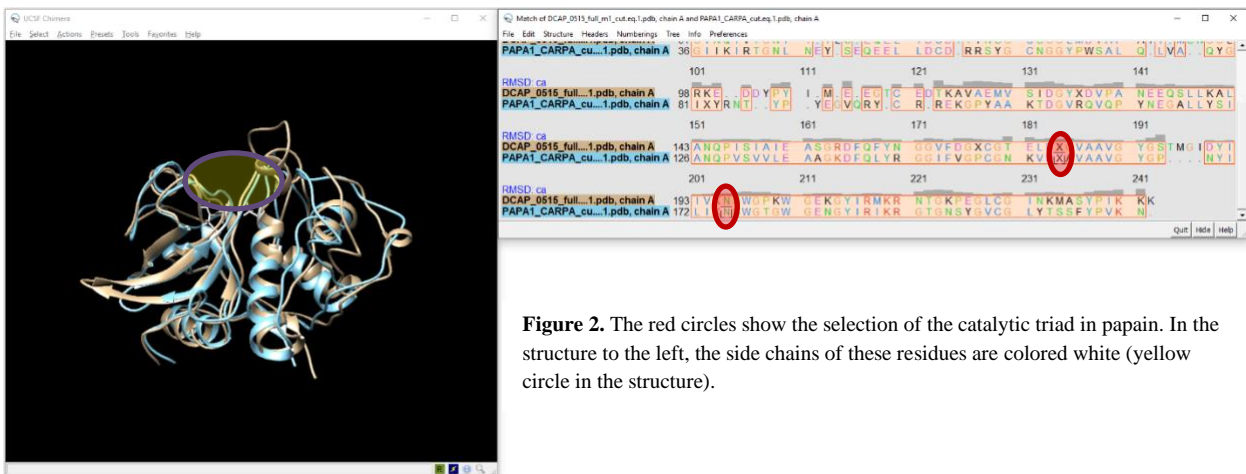


Figure 2. The red circles show the selection of the catalytic triad in papain. In the structure to the left, the side chains of these residues are colored white (yellow circle in the structure).

8. N-Terminus and C-Terminus

- Find the N-Terminus (beginning) and C-Terminus (end) of the PAPA1_CARPA. Note if your protein has an N and C-Terminus that is longer, shorter, or the same as PAPA1_CARPA.

9. Important Features.

- Compare the structure of the novel protease with PAPA1_CARPA. Does the novel protease have any extra helices, β -strands, loops, or other structures that are not found in PAPA1_CARPA? If so, note it and describe the feature.

Cited References

Butts CT, Bierma JC, Martin RW. 2016a. Novel proteases from the genome of the carnivorous plant *Drosera capensis*: structural prediction and comparative analysis. *Proteins*. 84:1517–1533.

- Butts CT, Zhang X, Kelly JE, Roskamp KW, Unhelkar MH, Freitas JA, Tahir S, Martin RW. 2016b. Sequence comparison, molecular modeling, and network analysis predict structural diversity in cysteine proteases from the Cape sundew, *Drosera capensis*. *Comput Struct Biotechnol J*. 14:271–282.
- Feijoo-Siota L, Villa TG. 2011. Native and biotechnologically engineered plant proteases with industrial applications. *Food Bioprocess Tech*. 4:1066–1088.
- Li Q, Yi L, Marek P, Iverson BL. 2013. Commercial proteases: present and future. *FEBS Lett*. 587:1155–1163.
- Tavano OL, Berenguer-Murcia A, Secundo F, Fernandez-Lafuente R. 2018. Biotechnological applications of proteases in food technology. *Compr Rev Food Sci F*. 17:412–436.

Materials

A computer with Internet access and the Chimera program (freely available for all three operating systems at <https://www.cgl.ucsf.edu/chimera/download.html>) is required for each student.

You will also need the .pdb files for papain, example proteases, and the predicted proteases from *Drosera capensis* that we are studying. To obtain these files, please contact the lead author (Pavan Kadandale, pavan.k@uci.edu). Since these proteases are part of current, unpublished research in Dr. Martin's lab, we cannot post the data on a publicly accessible location, and hence ask you to contact us directly to get access to the data.

You will need to create a survey activity for the students - again, we are happy to provide you with the one that we use (please contact the lead author) - or you can create a custom activity for your students. We are happy to help with the creation of new activities.

Notes for the Instructor

We spent two 50-minute periods on this activity, to allow for students to really think about the structures and manipulate them in 3D. However, it is possible to have them do more homework activities, and spend only one 50-minute period on the exercise.

Key to the exercise is the creation of an online survey (or homework!) where students enter their answers to the questions in the class activity. The survey that we used for our activity can be accessed at:

https://uci.co1.qualtrics.com/jfe/form/SV_8dGERkcQONcivjv

When designing the activity, we found it best to provide drop-downs for answers wherever possible, rather than allowing students to enter their answers in text. This makes analysis of the answers much easier later on. Also, we had to validate student entries. For example, when we asked students to enter the number of the amino acid at the center of a helix, we expected them to enter numbers in the form "54", or "32". Instead, we saw a range of answers, including "fifty-four", "the central amino acid was 54," and "tyrosine is the central amino acid in helix #2". To avoid this diversity and to standardize answers we added a validation step to make sure that all the answers were a number.

You might want to create a different homework assignment/survey based on what you want to accomplish with your students. We are happy to help you with this. We only gave students participation points for the exercise, and as long as they answered all questions, they received points for completion.

You can also use any of the novel proteases in other structure-related activities. For example, you could have them use the amino acid sequence of the novel protease to identify other related proteases using BLAST searches. If a novel protease was found to have additional domains, you could BLAST only the sequence of that domain to find other proteins with similar structures. The number of activities that can be designed around this central theme are fairly diverse, and we invite you to create a rich set of activities tailored to your context, using our module as a starting point.

Acknowledgments

We thank the students of the Biochemistry classes in the Winter quarter of 2018 and 2019 for actively participating in this research-based activity. This work was supported in part by NSF grant DMS-1361425. Marc Sprague-Piercy is supported by the HHMI Gilliam Fellows Program. The authors thank Carter Butts for helpful discussions and Suhan Kim for assistance with the pilot survey.

Cited References

- American Society for Biochemistry and Molecular Biology (ASBMB). Concept-driven teaching, assessment tools and course materials. [accessed 2019 Oct 22]. <https://www.asbmb.org/education/teachingstrategies/>
- Butts CT, Bierma JC, Martin RW. 2016a. 2016. Novel proteases from the genome of the carnivorous plant *Drosera capensis*: structural prediction and comparative analysis. *Proteins*. 84:1517–1533.
- Butts CT, Zhang X, Kelly JE, Roskamp KW, Unhelkar MH, Freitas JA, Tahir S, Martin RW. 2016b. Sequence comparison, molecular modeling, and network analysis predict structural diversity in cysteine proteases from the Cape sundew, *Drosera capensis*. *Comput Struct Biotechnol J*.14:271–282.
- Herman T, Morris J, Colton S, Batiza A, Patrick M, Franzen M, Goodsell DS. 2006. Tactile teaching: exploring protein structure/function using physical models. *Biochem Mol Biol Educ*. 34:247–254.
- Marshall PA. 2014. From pipe cleaners and pony beads to apps and 3D glasses: teaching protein structure. *J Microbiol Biol Educ*. 15:304–306.
- Wu H-K, Shah P. 2004. Exploring visuospatial thinking in chemistry learning. *Sci Educ*. 88:465–492.

About the Authors

Pavan Kadandale works on understanding how to better teach critical thinking in large classes, and also develops better assessment strategies to test for critical thinking in these large classes.

Marc Sprague-Piercy is a graduate student in Dr. Martin's lab, currently studying protein structure.

Dr. Rachel Martin is an experimental biophysical chemist whose research is focused on protein structure-function relationships and enzyme discovery.

Mission, Review Process & Disclaimer

The Association for Biology Laboratory Education (ABLE) was founded in 1979 to promote information exchange among university and college educators actively concerned with teaching biology in a laboratory setting. The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises. For more information about ABLE, please visit <http://www.ableweb.org/>.

Advances in Biology Laboratory Education is the peer-reviewed publication of the conference of the Association for Biology Laboratory Education. Published articles and extended abstracts are evaluated and selected by a committee prior to presentation at the conference, peer-reviewed by participants at the conference, and edited by members of the ABLE Editorial Board. Published abstracts are evaluated and selected by a committee prior to presentation at the conference.

Citing This Article

Kadandale P, Sprague-Piercy M, Martin RW. 2020. An *in silico* approach to protease discovery. Article 10 In: McMahon K, editor. *Advances in biology laboratory education*. Volume 41. Publication of the 41st Conference of the Association for Biology Laboratory Education (ABLE). <https://doi.org/10.37590/able.v41.art10>

Compilation © 2020 by the Association for Biology Laboratory Education, ISBN 1-890444-17-0. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright owner.

ABLE strongly encourages individuals to use the exercises in this volume in their teaching program. If this exercise is used solely at one's own institution with no intent for profit, it is excluded from the preceding copyright restriction, unless otherwise noted on the copyright notice of the individual chapter in this volume. Proper credit to this publication must be included in your laboratory outline for each use; a sample citation is given above.