Tested Studies for Laboratory Teaching
Proceedings of the Association for Biology Laboratory Education
Volume 40, Article 48, 2019

Protein Characterization: A Database-Based Approach for Analyzing Protein Structure and Function

Alma E. Rodriguez Estrada

Aurora University, Biology Department, 347 S. Gladstone Ave., Aurora Illinois 60134 USA (**arodriguezestrada@aurora.edu**)

The study of proteins structure and function is essential in biochemistry courses. Once a basic understanding of the four levels of protein structure is achieved, and general information about their function is learned, an additional challenge would further cement and evaluate students' ability to understand and explain how protein structure often dictates its function and how structural changes, due to mutations or altered cellular environments, could potentially affect its function. The term paper "protein characterization" is a written assignment (mini-term paper) that students complete during several weeks in a biochemistry course. Students choose a protein of interest and explore the Protein Data Bank (PDB) in order to retrieve information related to the three-dimensional structure, amino acid sequence and protein annotation. In this mini workshop, participants become involved in the inquiry process that students experience while completing this assignment. Participants will use the PDB website and the Jmol software to examine proteins. The presenter will also share the instructions and rubrics used to asses this assignment.

Keywords: protein characterization, biochemistry, Protein Data Bank (PDB), Jmol

Introduction

Proteins are one of the most structurally and functionally diverse groups of macromolecules in living systems. Elucidating the primary structure of proteins sheds light into their physical and chemical characteristics, three-dimensional structure and biological function. The Protein Data Bank (PDB) is a worldwide repository of macromolecules' structural data established in 1971 (Berman et al. 2000, Voet et al. 2016). The number of entries available at the PDB has exponentially increased since the 1980s. Today, there are a total of 145,473 entries available in this repository (PDB statistics). The PDB is widely used by researchers, educators and students alike. Deposition of structural data in the PDB is required by most scientific journals. Biochemistry textbooks often cite the PDB accession code in the legends of protein images.

A clear understanding on how the amino acid sequence (primary structure) of proteins determines higher levels of structure (secondary and tertiary) and how the structure provides insight into a protein's function is a common learning objective in biochemistry courses. The present activity has been implemented as a written assignment, a mini-term paper named *Protein Characterization*, in a three thousand level, sixteen-week biochemistry course. The activity involves: 1) use of the Protein Data Bank, 2) reading primary research publications, and 3) use of the Jmol software. In this assignment, students select a protein, describe all levels of structure and make connections between structure and function. The students formally communicate the information in a written format as a mini-term paper.

Choosing a Protein of Interest

Proteins can be chosen from the course textbook. Ideally, the selected protein is involved in a metabolic or signaling pathway that will be covered later in the semester (e.g. glycolysis, citric acid cycle, or MAPK/ ERK pathway). However, this is not a requirement. Students might choose proteins that are somehow related to something that interests them. For example, a student interested in ophthalmology could choose to investigate rhodopsin; a student interested in cancer might choose the tumor suppressor protein p53. The length of the protein should be at least 200 amino acid residues and it cannot be an intrinsically disordered protein.

Protein Structure

The amino acid sequence of the protein is graphically represented and can be downloaded from the

PDB as a text file and/or as an image file. The amino acid sequence image is annotated. Thus, important information regarding secondary structures and relevant amino acids (e.g. amino acid residues that bind to ligands, amino acid residues essential for catalysis, etc.) can be easily identified.

The secondary, tertiary and quaternary structure of proteins in the PDB are made available from two protein viewers: NGL or Jmol/JSmol. Within each, several options allow the display of proteins in different colors and styles (cartoon, spacefill, line, etc.). It is also possible to display features such as hydrogen bonds and disulfide bridges, among many others. The viewers also allow for protein rotation. Alternatively, students can download the image file and use Jmol to further modify the structure.

Using Jmol for Visualization

Jmol is an open source software used for interactive molecular visualization. Jmol is freely accessible (http://www.jmol.org/), it is compatible with major operating system and it is relatively easy to use (Hanson 2010, Herraez 2006). Specific features of a protein can be emphasized using simple commands available in Jmol (Fig. 1).

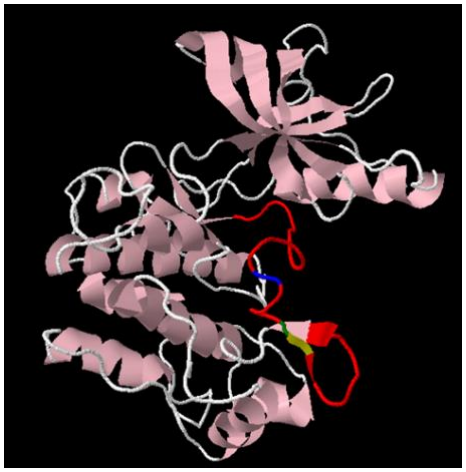

Figure 1. Tyrosine kinase domain of the human insulin receptor. The activation loop (1149 – 1170) is represented in red. The three tyrosine residues in the activation loop that become phosphorylated upon insulin stimulation are shown in blue (1158), green (1162), and yellow (1163). Autophosphorylation of the tyrosine residues results in a conformational change of the activation loop, rendering the kinase domain active (Hubbard et al. 1994, PDBid 1IRK).

From the PDB and the 3D view window, files compatible with Jmol (.pdb) can be downloaded. (the JSmol viewer has to be selected). Once the structure file is open in Jmol, right clicking on the image will open a toolbar from where the option "console" can be selected (Fig. 2A). The console window is where commands are displayed (Fig. 2B).

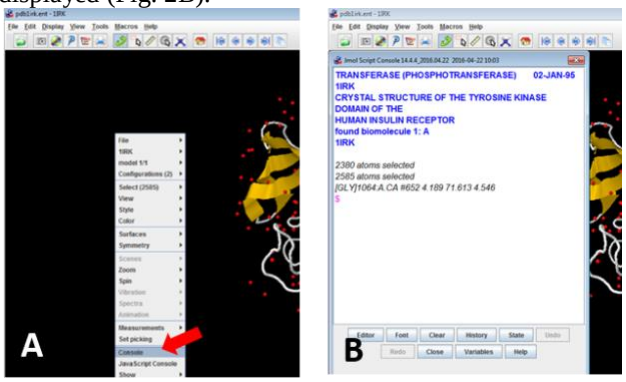

Figure 2. Structure view in Jmol. A. Right clicking on the image will open a toolbar from which "console" can be selected (red arrow). B. Open console.

The Center for Biomolecular Modeling (Milwaukee School of Engineering) has created a quick reference sheet of commonly used commands in Jmol (http://cbm.msoe.edu/includes/pdf/JmolQuickReferenceSheet.pdf). Each command is written after the $ sign. For example, in order to select and color all the helices in a structure, two commands are necessary. The first command will select all the helices and the second command will color the helices in pink:

$ select helix (click enter)

767 atoms selected (this displays automatically after pressing enter)

$ color pink (click enter) at this moment all the helices in the structure will be colored pink.

Figure 3 shows the commands used to display the activation loop and tyrosine residues displayed in Fig. 1.

Protein Function

The three-dimensional structure of proteins provides insight into its function and how a protein might interact with other macromolecules or cellular structures. Protein characterization might address multiple aspects but for the purpose of this written assignment, students select only a few aspects to focus on. For example, if the chosen protein is a protein kinase receptor such as the human insulin receptor, the focus of the assignment can be the cytosolic domain, specifically the position and structure of the activation loop or the position and structure of the catalytic site. Students could describe the composition of the activation loop, the conformational changes that occur

upon phosphorylation of the tyrosine residues, and how those changes influence the accessibility of the active site (Voet et al. 2016).

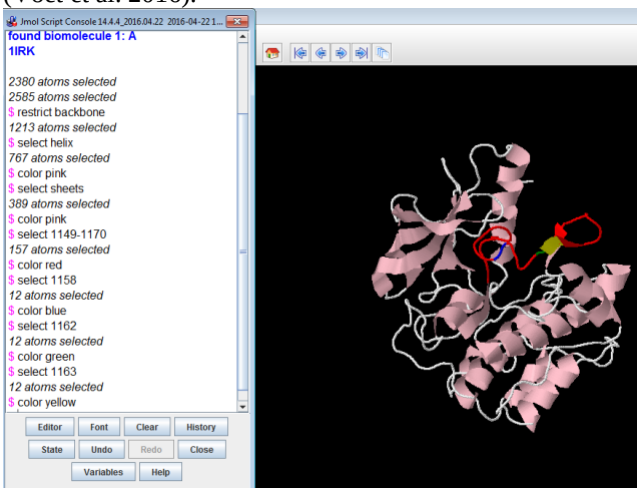

Figure 3. Jmol commands used to display the tyrosine kinase domain of the human insulin in figure 1 (PDBid 1IRK). Helices and sheets are in pink, the activation loop (1149-1170) is in red, and the three tyrosine residues involved in autophosphorylation are represented in blue (1158), green (1162) and yellow (1163) (Hubbard et al. 1994, PDBid 1IRK).

Student Outline
Protein Characterization : Mini-Term Paper – 50 points

Objectives:

- Become familiar with the Protein Data Bank (PDB): https://www.rcsb.org/
- Interpret protein characterization data
- Relate protein structure and function
- Read, understand and synthesize primary research articles

There are three major deadlines associated with this task (Table 1). Two deadlines are optional.

Table 1. Due dates for the three major tasks related to the term paper.

Activity	Due Date	Description	Points	Instructions
Select your topic	Week 6	Provide the protein's PDB ID code, common name and one sentence describing the protein function. A sign-in will be available during class.	5	- Visit the Protein Data Bank website: http://www.rcsb.org/pdb/home/home.do - Choose the protein you want to work with and record the accession code (#XXX). Alternatively, pick a protein from your textbook. But make sure that the code is also in the PDB. - Your protein has to have a minimum length of 200 amino acid residues. Anything smaller than that has to be approved by the instructor. - Write down the title of the research article associated with this structure - Retrieve the publication (it might be available through google scholar or through the University library. Avoid any structure which work has not been published. - As you choose your protein, you have to make sure you explore the primary literature associated with the protein. Read the article abstract and skim other sections of the article. Please make sure you are able to understand the purpose of the research work and the outcomes. This will require some effort.
Protein characteri-zation form and list of references **OPTIONAL**	Week 9 - 12	1. Complete the "Protein Characteristics Form" 2. Provide a list of at least 3 references: - One reference is the publication associated with your protein ID. - Two references should correspond to scientific literature (research papers found in specialized	- -	- Schedule an appointment with me (preferably during office hours). Completion of this step will not be done by e-mail. - Print and bring me your protein characterization form (COMPLETED) and the articles you found so far related to the selected protein. - The protein characterization form is available on Moodle. - Please, make sure to take notes during the

		journals such as J*ournal of Molecular Biology, Biochemistry*, *Nature*, *Science*, etc.). - An additional reference could be your book. But this is not part of the minimum of 3 references.		meeting.
Term paper (final version)	Week 14	The final version of the term paper should fulfill all the requirements regarding format and content. Notice that the minimum number of references for your final paper is 6!	45	- Upload this assignment as a word file in the dropbox "term paper"
Resubmis-sion (optional)	Week 15			
Total			50	

Paper Structure

A. Title (20 words max): should accurately reflect the content of your paper. Make sure your title is interesting and make sure that the title does not match any of your references. As you further investigate your protein, you will come up with a title that reflects the content of the paper.

B. Protein Characterization (4 pages, 2 of those are images):

B.1. Origin (organism and title of the research publication)

B.2. Structure: length, molecular weight, amino acid sequence, secondary, tertiary and quaternary structure, ligands, substrate, post-translational modifications, etc.

B.3. Relevant images (3D view and sequence view from the PDB site, one page for each maximum). Please include one or two additional and relevant images.

B.4. Relationship between structure and function. For example, if this is an enzyme, describe the binding site and what amino acids are involved in the catalytic reaction; describe the domains involved in binding the substrates and describe the catalytic reaction. If this is transmembrane protein (channel protein for example), describe the role of relevant amino acids in the protein specificity for the ions or molecules that cross the membrane. If this is a carrier protein such as hemoglobin, identify relevant amino acids involved in the interactions with the heme group, amino acids involved in the subunits interactions, amino acids involved in the shifting between the T and R state, etc. In this section, you should cite at least 3 research publications.

B.5. Protein localization in the cell, organ, organ system, etc. Make sure you describe the function of the protein within the organism.

B.6. Why are researchers interested in studying this protein? Why is this protein relevant? In this section, you should **tell the story about the protein**. When, how and why was this protein first discovered? etc. In this section, you should cite at least 3 research publications.

C. Future Work (1 page): Several years of research work have been done about that protein. Describe how this information will be used in the future. Discuss the authors' perspective and your own.

D. References (6 references minimum, 1 page): Please, make sure you utilize reliable bibliographic resources including

books and research articles. Three of your references must correspond to research articles, one can be your book, one is the PDB website. Using the web is fine as long as the sources are reliable. For example, use research institutions and Universities web sites, electronic journals, etc. Other web sites are a good starting point, but should be used mainly for your own background research and for finding further references such as reliable books/journal articles. See the document "reference guidelines" posted in Moodle. References must be listed in alphabetical order by author's last name. Refer to the reference guidelines posted on Moodle.

Format:

Font: Times New Roman No. 12

Line Space: 1.5

Margin: normal 1'

Length: 6 complete pages (including the references). Papers with less pages will have a % deducted from the rubric grade, equivalent to the amount missing. For example, if your paper is 3.5 pages long your grade will be 3.5/6 = 58.3 % of the grade calculated by the rubric. Likewise, if your paper has more than 6 pages, a percent equivalent to the extra number of pages will be deducted (1/6 x 100 = 16.6%).

Page Numbers: include page numbers/total page number

File Extension: word files ONLY

File Name: Last name first name protein common name 4 letter ID. Example: RodriguezAlma_Insulin_XXXX

In-Text Citations: It is essential to use in-text citation to validate the content and accredit others research. After your present facts, ideas, data, etc. you must cite in parenthesis the source. In-text citations should contain the author (s) last name and the year of publication. When more than one reference is used, the citations need to be arranged in alphabetical order. The paragraph below was extracted from a scientific publication (accelerated telomere shortening in response to life stress by Epel et al. 2004) and it serves as an example.

> "People who are stressed over long periods tend to look haggard, and it is commonly thought that psychological stress leads to premature aging and the earlier onset of diseases of aging. Numerous studies demonstrate links between chronic stress and indices of poor health, including risk factors for cardiovascular disease and poorer immune function (McEwen 1998; Segerstrom and Miller 2004)"

Remember that copying content word by word is plagiarism. Please, become familiar with Aurora University's Code of Academic Integrity (http://www.aurora.edu/academics/resources/code.html#axzz2ZuhXzqJL).

Submission: **Late submissions** of the topic selection will not grant you any points. **For the final version,** 10 points will be deducted for each day of delay (**24 hours)**. Remember that your topic must be approved by the instructor. Thus, **if you complete a term paper for which approval was never granted, your paper will not have any value**. A paper submitted **three days after** the due date **will not be graded**.

RUBRICS:

Component	P	Above Standards	Meets Standards	Approaching Standards	Below Standards
Title (20 words maximum)	5	The title clearly reflects the content of your paper and is attractive. The title is not the same title as any of the references used. Report the PDB ID in the title. 5 pts.	The title reflects the content of your paper but it is not attractive. The title is not the same title as any of the references used. The PDB ID might or might not appear in the title. 3 pts	The title does not reflect the content of your paper. The title might or might not be identical to one of the references. The PDB ID might or might not appear in the title. 2 pts.	The title is broad and uninformative. The title is the same as one of the references used. The PDB ID might or might not be present in the title. 1 pts.
Protein character-ization (3 pages of text)	15	All the components are included (B.1 to B.6). The information provided is congruent. Item B.6 is thorough, thoughtful, clear and relevant. Written in paragraph form. 15 pts.	All the components are included (B.1 to B.6). The information provided might or might not be congruent. Item B.6 is not thorough, thoughtful, clear or/and relevant. Written in paragraph form. 12 pts.	Some components (B.1 to B.6) are missing. If all are present, they are not congruent. If all present, not written in paragraph form. Item B.6 is vague. 9 pts.	Some components (B.1 to B.6) are missing. If all are present, they are not congruent. If all are present, the information is not written in paragraph form. Item B.6 is poorly stated. 5 pts.
Protein character-ization: (images, 1 page)	8	Two images have been included (3D structure and sequence). An additional relevant image is also included. The images are labelled, the images have captions (including figure number) and citations. Images are referred through the text. 8	Two images have been included (3D structure and sequence). The images are labelled, the images have captions (figure numbers might and might not be present) and citations. Images might or might not be referred through the text. 6	Two images are included but the images are not labelled. Captions, figure numbers, citations might be missing. Images might or might not be referred through the text. 4	Only one image is included. If both are present, they might not have labels, figure numbers, citation, captions. Images might or might not be referred through the text. 2
Future work (1 page)	5	The information in this section is thorough, thoughtful, clear and relevant. Researchers and owns views are discussed. 5 pts.	The information in this section is present but it is not thorough, thoughtful, clear or relevant. Researchers or owns views might be absent. 4 pts.	The information in this section is unclear and not thoughtful. Researchers or owns views might be absent. If present, they are irrelevant. 2.5 pts.	The information in this section is unclear and not thoughtful. Researchers or owns views might be absent. 1 pts.

Format, sequencing, page numbering, grammar and spelling, references	5	The format is followed, including the references' format, page numbers, font size, etc. The material is presented in a logical sequence. The author makes no errors in grammar or spelling that distracts the reader from the content. The scientific names are in italics; acronyms are first spelled. Pages are numbered. The file's extension is compatible with Microsoft Word (.doc or .docx). 5 pts.	Very few problems were identified. 3 pts.	Some problems were identified. 2 pts.	Many problems were identified. 0 pts.
In text citations	5	Citations consistently appear through the paper. The sources of information can be easily traced back to the original reference. Figures also contain citations. 5 pts.	Citations appear often through the paper. However, there are some sections where the sources of information cannot be traced back to the original reference. Figures might or might not contain citations. 3.5 pts.	Citations appear sporadically through the paper. There are many sections where the source of information cannot be traced back to the original reference. Figures do not contain citations. 2.5 pts.	Citations appear rarely through the paper. There are many sections where the source of information cannot be traced back to the original reference. Figures do not have citations. 1.5 pts.
File name	2	Saved as indicated in the instructions. 2 pts.	The name is partially correct. 1 pt.	Incorrect or random name. 0 pts.	

Total points: 45
Protein selection: 5

Total points: 50

Materials

A computer with Internet access and the Jmol software installed.

Notes for the Instructor

The instructions and rubrics for the protein characterization mini-term paper are distributed to students during the fourth or fifth week of the semester, right after the chapter on proteins has been covered. The final version is due the last week of the semester (week 15). Two additional deadlines are in place to assure progress in this investigation. Two weeks after the instructions and rubrics are distributed, students are required to submit the following information: 1) PDB identification code, 2) common name of the protein, 3) one or two sentences describing the protein's function, and 4) number of amino acid residues (minimum 200). At this point, students are required to have an electronic or a hardcopy of the research publication (primary citation) associated with the PDB ID code. Two students in the class cannot have the same identification code.

The second deadline is optional and involves the period of time between the point when students selected a protein and two weeks before the final paper is due (approximately seven to eight weeks total). During this time, students complete a "protein characterization form" that includes additional information such as the reported (in the PDB) and the estimated (amino acids x 110 Da) molecular weight, organism where the protein is expressed (native and recombinant protein if applicable) and images of the protein's primary and tertiary structures. Students meet in person with the instructor to review the protein characterization form and are requested to bring a hardcopy of the primary article. During the meeting, the instructor performs a quick data check, answers questions, clarifies content and provides guidance.

The third deadline is to submit the final version of the paper. Students who wish, could address the instructor's feedback and re-submit the assignment few days later.

Cited References

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN. 2000. The Protein Data Bank. Nucleic Acid Research. 28(1):235-242. http://www.rcsb.org/

MSOE Center for Molecular Modeling. Jmol Quick Reference Sheet. http://cbm.msoe.edu/includes/pdf/JmolQuickReferenceSheet.pdf. Accessed October 2018.

Hanson RM. 2010. Jmol – a paradigm shift in crystallographic visualization. Journal of Applied Crystallography. 43:1250-1260

Herraez A. 2006. Biomolecules in the Computer. Biochemistry and Molecular Biology Education 34(4):255-261.

Hubbard SR, Wei L, Ellis L, Hendrickson WA. 1994. Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature 372:746-754. PDBid 1IRK.

Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/

Voet D, Voet JG, Pratt CW. 2016. Fundamentals of Biochemistry. Fifth edition. Wiley.

About the Authors

Alma Rodriguez is Associate Professor in the Biology Department at Aurora University since 2013. Alma teachers a variety of lower (e.g. biology of cells) and upper level courses (e.g. biochemistry, genetics, research in biology) designed for science, non-science majors and pre-nursing students.

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/.**

Papers published in *Tested Studies for Laboratory Teaching: Peer-Reviewed Proceedings of the Conference of the Association for Biology Laboratory Education* 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.

Citing This Article

Rodriguez Estrada AE. 2019. Protein characterization – a database-based approach for analyzing protein structure and function. Article 48 In: McMahon K, editor. Tested studies for laboratory teaching. Volume 40. Proceedings of the 40th Conference of the Association for Biology Laboratory Education (ABLE). **http://www.ableweb.org/volumes/vol-40/?art=48**