

## Georgia Journal of Science

Volume 75 No. 2 *Scholarly Contributions from the Membership and Others*

Article 3

2017

# Exploring the Use of Free Bioinformatics Modules in an Introductory Biochemistry Course

Charlsey Dodgen  
cd179535@gordonstate.edu

Uzezi Uwerosuo  
*University of Maryland*

ChulHee Kang  
chkang@wsu.edu

Cathy Lee  
*Gordon State College, clee@gordonstate.edu*

Follow this and additional works at: <http://digitalcommons.gaacademy.org/gjs>

 Part of the [Biochemistry Commons](#), [Bioinformatics Commons](#), and the [Science and Mathematics Education Commons](#)

### Recommended Citation

Dodgen, Charlsey; Uwerosuo, Uzezi; Kang, ChulHee; and Lee, Cathy (2017) "Exploring the Use of Free Bioinformatics Modules in an Introductory Biochemistry Course," *Georgia Journal of Science*, Vol. 75, No. 2, Article 3.  
Available at: <http://digitalcommons.gaacademy.org/gjs/vol75/iss2/3>

This Research Articles is brought to you for free and open access by Digital Commons @ the Georgia Academy of Science. It has been accepted for inclusion in Georgia Journal of Science by an authorized editor of Digital Commons @ the Georgia Academy of Science.

---

# Exploring the Use of Free Bioinformatics Modules in an Introductory Biochemistry Course

## **Cover Page Footnote**

We would like to thank all of the students enrolled in the biochemistry class during Spring 2016 at Gordon State College.

## EXPLORING THE USE OF FREE BIOINFORMATICS MODULES IN AN INTRODUCTORY BIOCHEMISTRY COURSE

Charlsey Dodgen<sup>1</sup>, Vwerosuo Uzezi<sup>1</sup>, ChulHee Kang<sup>2</sup>,  
and Cathy Lee<sup>1\*</sup>

<sup>1</sup>Department of Biology and Physical Sciences  
Gordon State College

Barnesville, Georgia, 30204, USA

<sup>2</sup>Department of Chemistry

Washington State University

Pullman, Washington, 99163, USA

\*Author to whom correspondence should be addressed,  
[clee@gordonstate.edu](mailto:clee@gordonstate.edu)

### ABSTRACT

Although bioinformatics, the use of computational science to study biology, has become imperative in many areas of the biological sciences and related career paths, introductory biochemistry courses may disregard practical knowledge on bioinformatics. For this reason, we merged a hands-on activity module into an undergraduate biochemistry course in two ways. First, we incorporated bioinformatics modules for building phylogenetic trees by aligning the active sites of 10 chosen related  $\alpha$ -amylase enzymes using freely available data. Secondly, we chose three of those 10  $\alpha$ -amylase enzymes to compare the 3D structure of their active sites. This module gives the students an opportunity to understand how to access biological information from public databases such as GenBank, and analyze the information using software like MEGA, ClustalW, and RasMol. Overall, our module should provide instructors with ideas on how to develop similar modules and encourage students to develop further independence in the use of bioinformatics tools.

**Keywords:** bioinformatics module, MEGA, phylogeny, molecular data, RasMol, 3D structure of  $\alpha$ -amylase

### INTRODUCTION

Across the various branches of the life sciences over the latter part of the last century, easy accessibility of high performance computers and effective data management algorithms began to play a major role in how we decipher and interpret biological information. The relationship between computer science and biology can be traced back to the 1950s, however it was not until the 1960s when the emergence of the modern computer and networks capable of processing protein and DNA data became available (Hagen 2000). With a large expansion of the available amino acid and DNA sequences and the development of high-speed computers, the field of bioinformatics has been established and developed rapidly into what we now know today.

The term *bioinformatics* was not used until 1979 when Paulien Hogeweg chose it to describe informational processes occurring in biotic systems. Before then, early computational biologists who pioneered this field had a general idea that an integration of mathematics, computational sciences, and molecular biology would be a very effective

tool to understand the fundamental questions in life sciences (Miskowski et al. 2007). Today, bioinformatics as a multidisciplinary field is involved in various aspects of our lives from healthcare to manufacturing. For example, health informatics focuses on improving health care by using any combination of lower cost applications of bioinformatics. Hence, the importance of exposing undergraduate biology students to basic and practical knowledge of bioinformatics cannot be overlooked for their career development.

Recent studies show that an inquiry- or research-based lab course could simultaneously divulge scientific knowledge and foster students' research expertise and confidence (Gray et al. 2015). Bioinformatics and genomics education can effectively expose students to a logical way of solving complex biological problems by taking advantage of innovative and accumulated works of computer scientists and geneticists (Wightman and Hark 2012). Most laboratories including biosciences now employ routine data mining and other public database analysis through high-speed computers. As such, it has become necessary that undergraduate students of bioscience major are familiarized with bioinformatic tools and databases to keep up with the modern era (National Research Council 2002).

The National Center for Biotechnology Information (NCBI) describes bioinformatics as a practical discipline utilizing a wide range of computational techniques such as DNA sequence, protein structural alignment, and data mining (Luscombe et al. 2001). The tools for learning, research, and mining biological data are now mostly free and readily available to students at all levels (Ditty et al. 2013). Therefore, the bioinformatics module can be incorporated as a "dry lab" component of undergraduate courses using computer resources available at most colleges. Here, we present a reliable bioinformatics classroom module into an introductory biochemistry course, which identifies two critical aspects of hands-on activity: building phylogeny and 3D structure of the active site of the target enzyme,  $\alpha$ -amylase. First, the conserved amino acid sequences from  $\alpha$ -amylase active sites of 10 closely related species were used to build phylogenetic trees using Molecular Evolutionary Genetics Analysis Version 6.0 (Tamura et al., 2013). Second, three 3D structures of the 10 chosen  $\alpha$ -amylase active sites were compared using RasMol (Sayle and Milner-White 1995). This basic module can provide an effective hands-on teaching and learning experience in addition to offering an accurate assessment of where students stand on the use and understanding of bioinformatics when combined with both pre- and postmodule materials such as quizzes or worksheets.

### **Goals**

The exercises that are presented in our module serve three purposes:

1. We created a simple and effective bioinformatics module to test its efficacy in educating students about phylogeny as well as visualizing and locating the active site of the subject enzymes in an undergraduate one-semester biochemistry class.
2. We exposed the students to hands-on active learning research projects using bioinformatics modules. This involves online databases and programs such as GenBank to obtain molecular data, MEGA to create phylogenetic trees, and RasMol to observe the 3D structures of three closely related enzymes. Furthermore, this module can provide the participating students the ability to navigate these programs outside of an instructor's direction for their own research use.

3. We assessed and surveyed student learning and evaluated their responses to the module.

Most of the tools used are web-based and use very modest computer resources. One of the public sites used for obtaining protein sequences and other crucial information is the National Center for Biotechnology Information (NCBI). Information from this site includes nucleotide and amino acid sequences, BLAST (basic local alignment search tool) searches, and relevant publications. Another tool used is RasMol, a user-friendly graphics program for molecular visualization, which allows the students to view the 3D structure of a protein that has been determined by both X-ray crystallography and NMR.

Upon completion of the exercises in our module, the participating students should be able to perform simple tasks including multiple sequence alignments using ClustalW (a program widely used for multiple sequence alignment), data mining from the NCBI website, building a phylogenetic tree using MEGA, and also be familiar with other databases that pertain to valuable information about both proteins and nucleic acid sequences. In addition, the students will become familiar with  $\alpha$ -amylase enzymes that are a staple in undergraduate biochemistry classes due to the conservation of their active site pocket among species. Before students began the exercises, we exposed them to pre-evaluation questions by choosing three random active site structures of  $\alpha$ -amylases. Examples are as follows:

- Where is the active site for each enzyme?
- What types of amino acids form the geometry of a binding site?
- Does the chemistry of the amino acids (e.g. acidic) give you some clues as to the forces involved in binding?
- Identify the amino acids that might play an important role in facilitating the catalysis.

### **Instructor Tips**

The activity module is designed as an introductory format for exposing students to basic bioinformatics, but it can be easily modified to suit different levels of undergraduate students. The entire module can be completed in approximately three hours (depending on a few variable factors like computer speed and internet connection), but it is advised that students be given up to five hours in the event that there is any confusion. Hence, the module can be treated as an in-class group activity using the available computer lab over several days. It can also be incorporated into the laboratory section of a biochemistry course if the instructor chose to do so.

Overall, the responses from the students showed a great enthusiasm for the bioinformatics module (see the response section). The students' comments on our survey also reflected that they felt doing hands-on active research instead of passive learning was more helpful. Before the students begin this exercise, it is more effective to teach them a basic knowledge about enzymes, protein structure, phylogeny, and proteomics. Good review tutorials in those areas may aid in students' success for completing this module.

### **MATERIALS AND METHODS**

We divided the class into groups of no more than four students. Each group had a leader who was responsible for submitting and presenting the results of the completed exercise within the group. Each group was then handed instructions about the exercise and the time frame for completion of the assignment. We advise that instructors use a

peer led team learning (PLTL) method for these exercises in which students who performed well in the previous course aid the groups of current biochemistry students. This PLTL method has been shown to improve students' grades and help them in retaining the information they have been exposed to (Varma-Nelson 2006).

### Exercise I: Building a Phylogenetic Tree

The students were required to perform a brief exercise on phylogenetic tree building using MEGA. More detailed instructions on how to do this exercise can be found here [http://faculty.gordonstate.edu/clee/ABT7807\\_11\\_Lee.pdf](http://faculty.gordonstate.edu/clee/ABT7807_11_Lee.pdf) and here <https://www.youtube.com/watch?v=uMy3HB94EVw> and MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0 (Tamura et al., 2013).

To complete the activity, the students will need to access the NCBI website to obtain the data sequences, download MEGA from the Molecular Evolutionary Genetics Analysis website, and download the alignment software ClustalW from the website. The links to access these programs and software are located in Table I below.

**Table I.** Access websites and software used

Website/ Software	Web address
National Center for Biotechnology Information (NCBI)	<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>
RasMol	<a href="http://www.rasmol.org/">http://www.rasmol.org/</a>
Protein Data Bank (PCB)	<a href="http://www.rcsb.org/pdb/home/home.do">http://www.rcsb.org/pdb/home/home.do</a>
Molecular Evolutionary Genetics Analysis (MEGA)	<a href="http://www.megasoftware.net/home">http://www.megasoftware.net/home</a>
ClustalW	<a href="http://www.genome.jp/tools/clustalw/">http://www.genome.jp/tools/clustalw/</a>
RasMol Tutorial	<a href="http://openwetware.org/wiki/BISC220/S10:Mod_1_RasMol_Tutorial">http://openwetware.org/wiki/BISC220/S10:Mod_1_RasMol_Tutorial</a>

The first step of building the phylogenetic tree is to locate a conserved region in the target enzyme,  $\alpha$ -amylase, and obtain the necessary sequence information (either protein or nucleotide sequence). In this exercise, we chose the Conserved Protein Domain Family “AmyAc\_family” and obtained the protein sequence data from the following link: [http://www.ncbi.nlm.nih.gov/protein?LinkName=cdd\\_protein&from\\_uid=200451](http://www.ncbi.nlm.nih.gov/protein?LinkName=cdd_protein&from_uid=200451)

1. Students were encouraged to choose any 10 organisms in the list from the link above and record their FASTA accession numbers. Table II shows the accession numbers from one of the students as an example.
2. After selecting each organism of the 10, click on the FASTA link to display the protein sequence information.
3. Copy the highlighted FASTA text and save it. Be sure to delete all text after “>” and insert the accession number or organism name in place of the text.
4. Repeat steps 1–3 until you have done all 10 organisms.
5. Once you have chosen 10 organisms, save the file as “Amylase.fasta”. Note that one does not need to change the file type manually, and that you only need to type “Amylase.fasta” as the name.

**Table II.** Example list of organisms and accession numbers

Organism	Accession number
<i>Streptomyces sp.</i>	WP_046263794.1
<i>Lactococcus piscium</i>	WP_047914993.1
<i>Lactococcus raffinolactis</i>	CCK19760.1
<i>Rasamsonia emersonii</i>	KKA21180.1
<i>Talaromyces stipitatus</i>	XP_002478606.1
<i>Clostridium sp.</i>	WP_042282244.1
<i>Caldithrix abyssi</i>	WP_006927553.1
<i>Lachnoclostridium phytofermentans</i>	ABX42665.1
<i>Meiothermus silvanus</i>	ADH63845.1
<i>Streptomyces sviveus</i>	EDY54032.1

```

>Lactococcus piscium
MVQQSIDKALSTFSTGENFHAQHLYLGCHRTADGYVFRVWAPHAKQVALISDKTQWRSKPIPMTLNHLGQW
EVYLLAGQVATGDYYKYLVRTRSTGQMIKLDPYAIRFGERPDTAAQIYDISPKDWQDKAWLTKRRQWDMF
TSPVSIYEVHASSWQRHDDGRLYTFSDLAKTLIPYVKS LGFTHIEFMPLMTHPLGMSWGYQLTGYFALEH
TYGTPEAFQDFVEAAHLAIGVMIDWVPGHFSQND DALAYYDGTPTYEYQDNNKARNVGVQALNFDLSKP
QVQSFLISSLKFWIEFYHLDGVRIDAVSNTLYLDYDQGEWTPNSDGSNINHGVAFKKLTSVIKLSHPD
VMSAEAEATSQTKITGMVEMGGLGDFDKWNMGWMNDVLFYQMDPLYRGQYFNLLTFSFMYMMDEHFILP
FSHDEVVHGKKS LMHKMWGDRKQQFSGLRNLLTYQMCHPGKLLFMGQEFQQLLEWKYDWQLEWETSDEL
NDQMQYFTSNLAHVYRDFPQLFELDH DYAGFEVIDADNTAQSVLSFIRK DASDNQLICVFNMAPIERHQF
TIGVPLAGTYTEVLNTEMSEFGGSWQAHNADQETQKAQWKTYEQTLSFTLPALGAVI WQVLA EETDNQTE
AVSSKSHDTS DYYVTD AQWAVLAPYFPV GARS KYEKRDLVSAVLF IKQTGRAWTKLPDRFP PYKTVYAFY
KRATKLG IWQRAMSDYETLKTDN
>Lactococcus raffinolactis
MGQTLADEALRTFTTGENFHVQHLYLGCHQTEGRFFR V WAPHAQVQLIGDMDWREHPVPMRNEKQW
EVTLPDTQASLG VYYKYLVRTRASGQVIEKMDPYAIRFGERPDTAAQIYDVSPKEWQDGDWL TARRKRKIF
SAPLNIYEVHPGSWQRHEDVRLYTFSDLTATLIPYVKS LGYTHIEFIPLMTHPLGMSWGYQLTGYFALEH
TYGTPEEFQDFVEAAHLAAGLVIVDWVPGHFSQND DALAYFDGTPTYEYQDGNRARNVGVQALNFDVSKP
QVQSFLISSLKFWIDFYHIDGVRIDAVSNMLYLDYDQGPWTPNPDGSNTNHDGV E FKKLTSVIKLAHPD
VIMAAEEASSQIKITGMLEMGGLGDFDKWNMGWMNDVLFYQMDPLYRGQHFSLLTFSFMYMMDENFILP
FSHDEVVHGKKS LMHKMWGDRYQQFSGLRNLLAYQICH PGKNLLFMGQEFQQLLEWKYDWQLEWAVSDEM
NDQIQAF TTALAHFYQTS PQLFELDH DYAGFEVIDADNTAQSVLSFIRK DASGNQLICVFNMAPIERSNF
TIGVPPFAGTYTEVLNSEMAEFDGSWLAHNADTATQKGQWKDYQNTLSFTLPALGAVI WTL SKADGA AKIK
AKKSTKSTKSTKNVTKSSRKTTS HAIASEDVKQLTQVEWES LAEYFPMGSR SKYEKRDLVNAALYIQQTG
TVWAKLPENYPYGSVYAFYKRS LKNGSWQHVL DVL

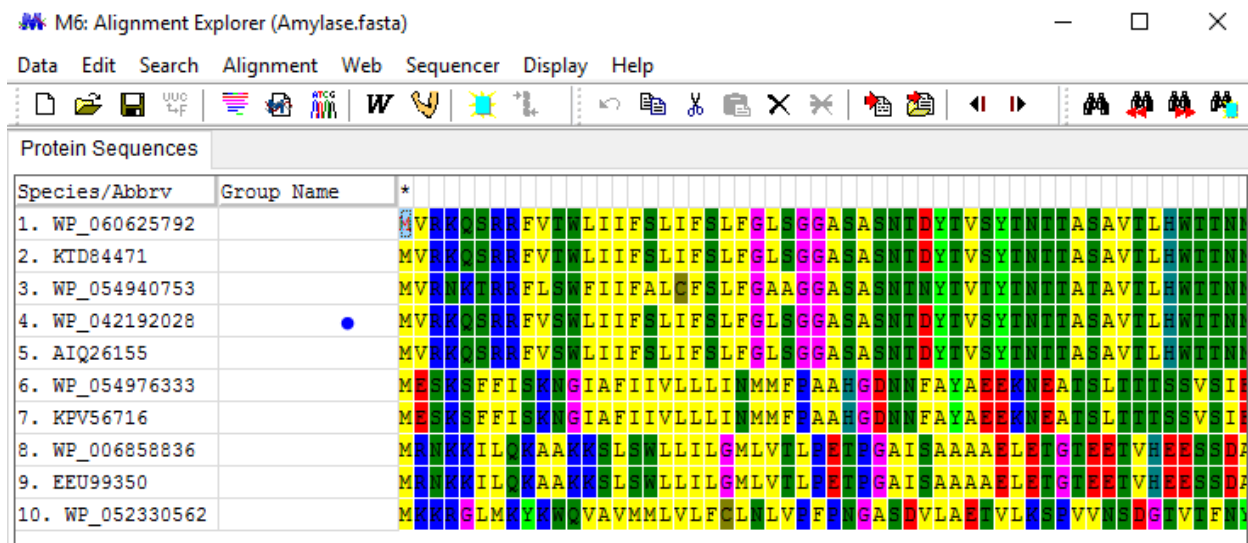
```

**Figure 1.** An example format containing the chosen protein sequences of two organism.

### Aligning the data

1. Open MEGA and click on “Align”, then “edit/ build alignment”, “ok”, and finally, select “Protein” on the new box.
2. A new window will open. In the new window, click on the yellow open folder icon under the edit button.
3. A folder will open. Search for the saved “Amylase.fasta” file and click on open.

4. In the tool bar, click on alignment and select align by ClustalW.
5. Once alignment is complete, click on edit, select all, and then delete gaps (an example of what this should look like can be seen in Figure 2).
6. Click on data in the toolbar then save session. Next click data again and click on export alignment and select FASTA format.
7. Close the window.



**Figure 2.** Alignment of the 10 chosen organisms from the AmyAc\_family using MEGA 6.0 (Tamura et al., 2013).

### Building the tree

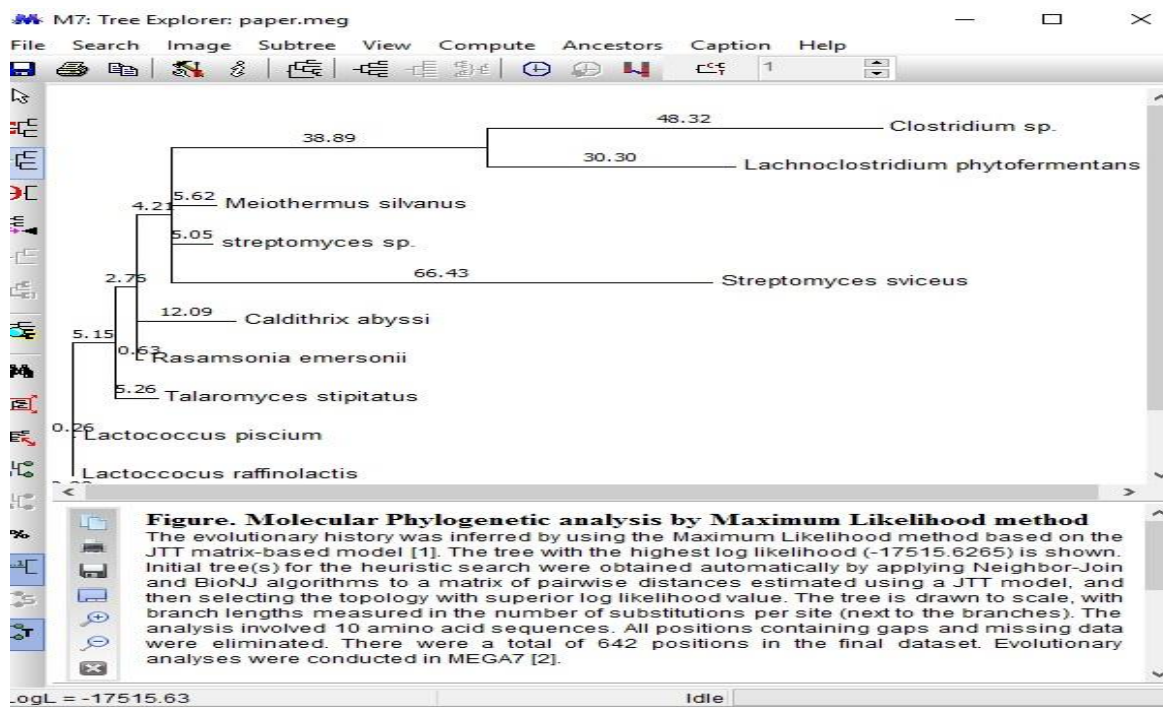
1. In the main Mega window, click on data, open a file/session, and select the .FAS file. Then, a screen will appear and choose analyze from it and then ok.
2. Once loaded, select phylogeny from the toolbar, click on construct/test maximum likelihood tree, yes, and compute.

### Exercise II: Using RasMol

For this exercise, RasMol must be downloaded and the access for the protein data sequences are found in the PDB (Protein Data Bank) website. At the PDB site, the name of the protein will be listed and a 3D structural image will be shown on the right side of the page. The links are provided in Table I. Before students start using RasMol, it is important to have a brief tutorial on how to use the various commands for manipulating the 3D image. There are different help pages for RasMol, but we found the one from [www.openware.org](http://www.openware.org) (Table I) to be the most useful.

Finding the active site requires some knowledge about the structure of the corresponding enzyme. Amylases have at least one conserved calcium-binding site, as  $\text{Ca}^{2+}$  ion is essential for the stability of the enzyme (Saboury 2002). In this exercise, we encourage the students to access the RCSB Protein Data Bank (PDB) to obtain the PDB file for three different  $\alpha$ -amylases and then to analyze the similarities and differences between their active sites using the same method as what was done for 5E6Y below. To ensure that only amylases from the conserved family are chosen, we restrict the list of amylases to only those that are in the list provided from Exercise I.





**Figure 3.** Phylogenetic tree of 10 members (in Table II) of the AmyAc\_family using MEGA 6.0 (Tamura et al., 2013).

The goal of this hands-on exercise is to help the students understand homology and common ancestry. In order to make it more understandable to the students, the teacher could explain the idea of how homology is closely linked to characters. For example, one could compare both hand bones and muscles of humans and the aye-aye (a type of lemur). Even though the aye-aye has hand bones that look similar in structure to those of humans, they evolved independently and are separate from humans due to convergent evolution.

### Step 1: Choosing the enzyme

Students should go to [www.rcsb.org](http://www.rcsb.org) and type in “1,4-alpha-glucan branching enzyme” in the search bar. From the results, pick any three of the results that falls within the list of organisms provided from exercise 1. The search results provide a concise description of the protein function, as well as the links to additional information about its role in metabolism and literature citations. Many proteins, but not all, will also have 3D structural information available (Pembroke 2000). Students can click the “view in JSmol” button below the image to view the 3D structure of the enzyme (Chen 2008). 3D representation of a molecule may be used as a teaching tool or for research (Herráez 2006). This will give the students an idea of what they will be working on using the RasMol program later on. The students should then copy the four alphanumeric accession ID and record it for future reference. To access the PDB files, click on the four-letter accession ID and a new page will open a new window containing all the information. At the top right corner of the page, a blue rectangle that reads “Download Files” will be accessible. Students should click it, and then select “PDB Format” to download the necessary files. This is done for each of the three 1,4- $\alpha$ -glucan branching enzymes that were previously chosen. Once downloaded, open the PDB files with RasMol. These PDB

files that contain the coordinates and element type for all the constituting atoms of the protein will allow the program to create visual depictions of the selected enzymes.

### Step 2: RasMol

RasMol allows the students to view the 3D structure of amylase and compare the amino acid residues between  $\alpha$ -amylases from different organisms. A quick glance at the help tutorial provided for RasMol can help the students in manipulating various parts of the image to see areas that are being studied. In our case, we were studying the active site of the enzyme from three different sources.

When you begin RasMol using a Microsoft platform, two windows will open. The first is the image, or graphics window, and the second is the command line window that is also called the terminal window. The command line window allows the user to define various parts of the structure that need to be studied. To open the PDB files downloaded, go to the file icon in the RasMol display menu and click open in the folder that you saved the document to. Then find the files that were downloaded and open them. The display page will open again with a 3D image of the enzyme (Sayle and Milner-White 1995). Only one PDB image can be opened at a time. If using the PC version, the two RasMol will be on the task bar. The command line will be used to adjust the image to our desired region of study, which is the active site.

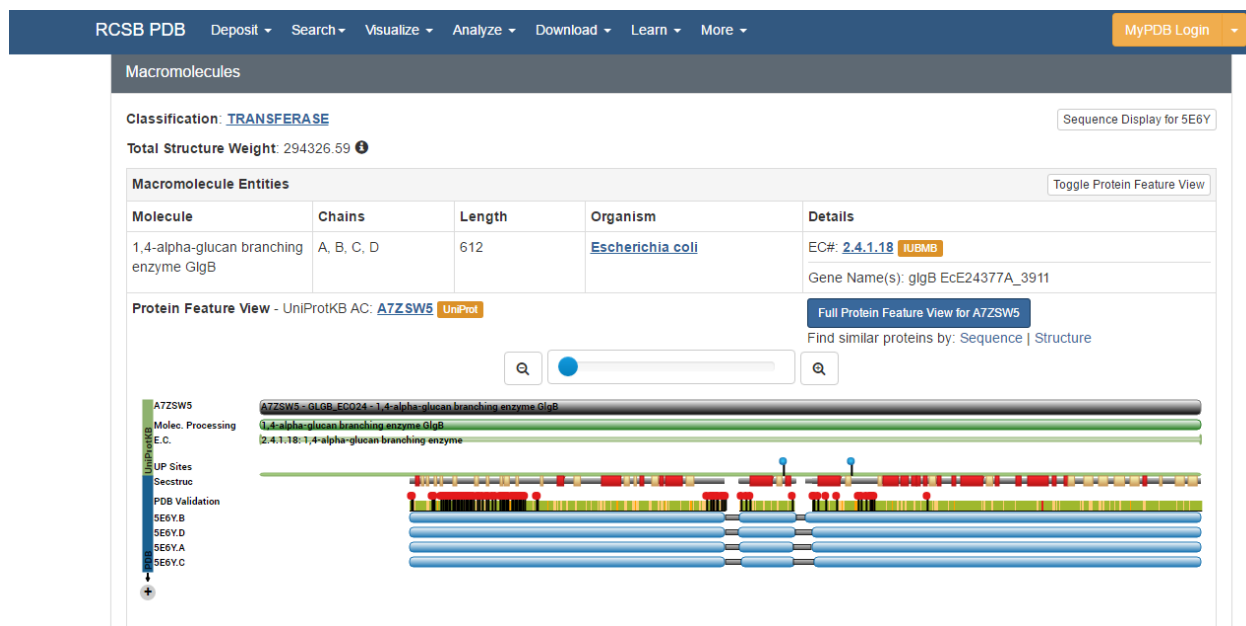
### Step 3: Adjusting the PDB image

Image adjustments can be done by typing commands in the command window. For our exercise, we use the crystal structure of 1,4- $\alpha$ -glucan branching enzyme Glgb from *E. coli* (PDBID: 5E6Y) (Feng et al. 2016). Depending on where the enzyme is derived can make the amino acid placements vary in their position (Aghajari et al. 2002). This variability allows for the flexible nature of catalytic residues of the active site.

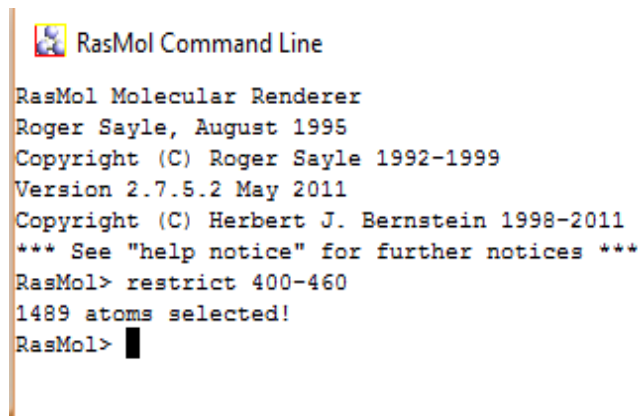
To easily identify the active site, the students need to go back to the area where they downloaded the PDB file and scroll down to the section titled “Macromolecule” that is near the middle of the page. An example image (utilizing 5E6Y) of what the macromolecule section looks like can be seen in Figure 4. The student needs to place his or her cursor over the blue dot above the domain of UPSites SecStruc line. A pop up will display the nucleophilic active site and the next blue dot will display the proton donor active site. The numbers next to the word *active site* illustrate the positions of the amino acids that makes up the active site. This gives us the information needed to restrict the region that contains the active site of that particular  $\alpha$ -amylase enzyme (5E6Y).

Once students identify the range in which the active site is located for their enzyme, they can begin restricting the active site area. For instance, 5E6Y has a nucleophile active site at 405 and a proton active site at 458, and so we can instruct RasMol to restrict everything except 400–480 by typing “restrict 400–480” in the command window (as seen in Figure 5). This allows students to see an image of where the active site is located (Figure 6).

The following (Figure 7 and Figure 8) are two other examples (in a bigger view) of active site images from  $\alpha$ -amylase from *Lactococcus raffinolactis* and *Streptomyces sviveus*.



**Figure 4.** Identifying the active site of 1,4-alpha-glucan branching enzyme, GlgB from *E. coli* (5E6Y; Feng et al. 2016) from Protein Data Bank.



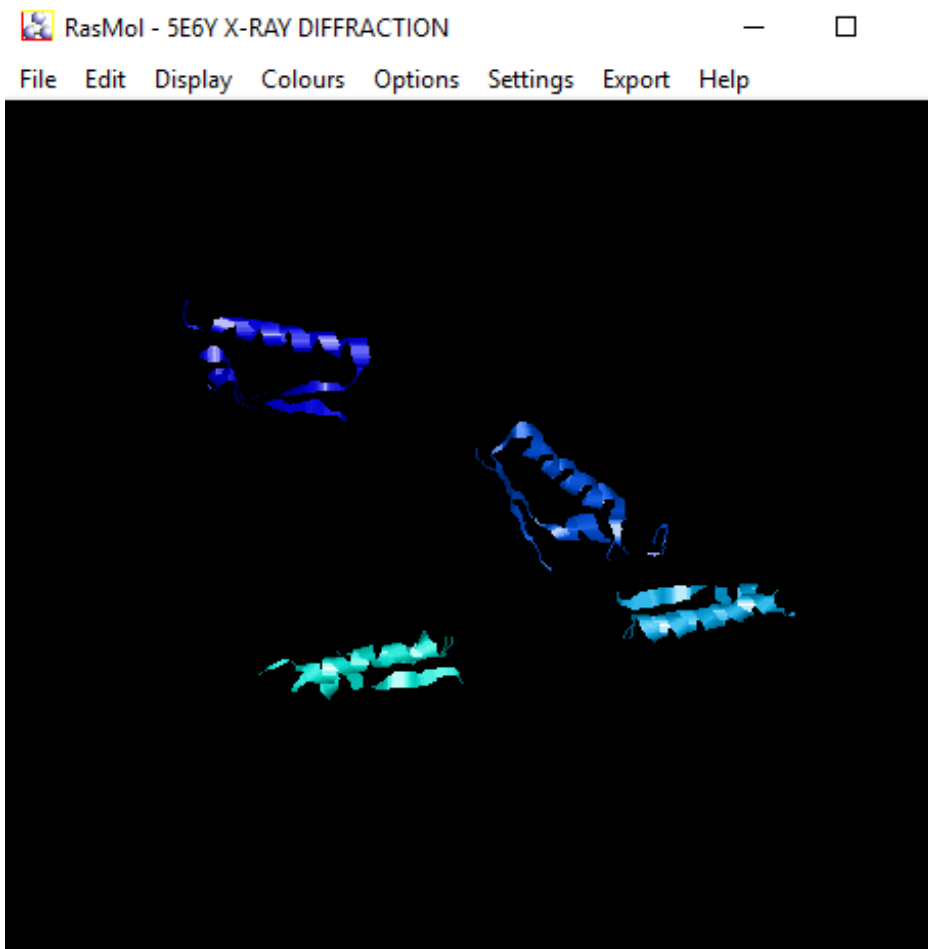
**Figure 5.** Commands of RasMol (Sayle and Milner-White 1995) for visualizing the active site of the 1,4-alpha-glucan branching enzyme, GlgB from *E. coli* (5E6Y; Feng et al. 2016) using RasMol (Sayle and Milner-White 1995).

### Post Activity/Follow Ups

After the exercise, follow up exercises were presented to the students to help test and reinforce their understanding of the exercise. Students that worked on this activity were evaluated using self-evaluating questionnaires, assessment questions, and verbal response.

Examples of post-assessment questions are as follows:

- What is the name of the enzyme? What species of organism did it come from? What is the percentage of homology?



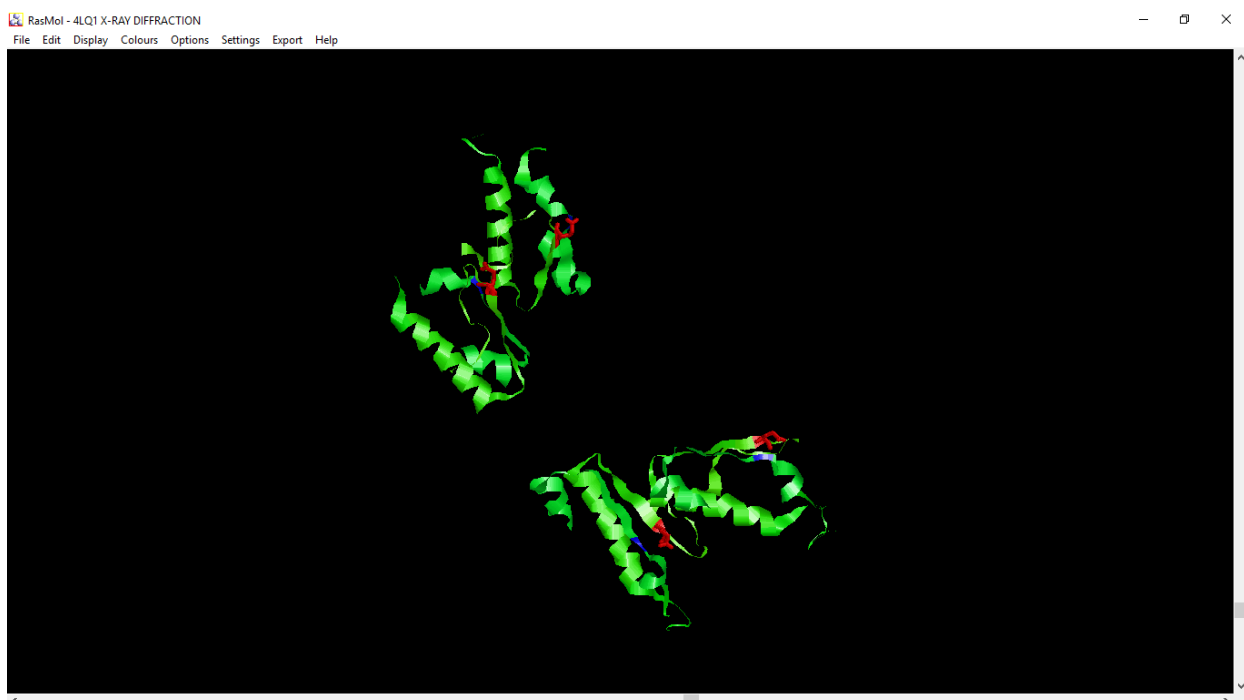
**Figure 6.** RasMol image of restricted 1,4- $\alpha$ -glucan branching enzyme, Glgb from *E. coli* (PDBID: 5E6Y, Feng et al. 2016).

- Why does it appear that there are multiple results for the same enzyme in your BLAST search?
- Is there a similarity among the enzymes that you chose (in the amino acids or structure)?
- What are the catalytic residues in the active site of all three enzymes?
- How does the location of the residue affect substrate binding?
- In humans, there are two types of  $\alpha$ -amylase; pancreatic and salivary. Why is it important to have this specialization of the enzyme in different locations?
- Are these examples of convergent or divergent evolution?

Students responded to exercises by submitting their findings for grading and evaluations. A class discussion concerning the students' learning experience can be used to better fine-tune the module for later applications. Additionally, pre- and postmodule quizzes will be very helpful in the assessment of students' knowledge gains.



**Figure 7.** Ribbon diagram of  $\alpha$ -amylase from *Lactococcus raffinolactis* showing the overall structure and active site using RasMol (Sayle and Milner-White 1995, Bernstein 2000).



**Figure 8.** Ribbon diagram of  $\alpha$ -amylase from *Streptomyces sviveus* showing the active site in red and blue using RasMol (Sayle and Milner-White 1995, Bernstein 2000).

## RESULTS

A total of 22 students of the 30 in the biochemistry class (73.3%) were able to give their feedback and rating of the activity after completion of module. No incentives were given for the students' responses to the modules. Surveying was done using the Likert scale from 1–5 with 1 being strongly disagree and 5 being strongly agree. Table III shows the questions that were asked for the survey and the students' responses.

**Table III.** Mean score of students' survey of the module

	Statement	Mean
1	Before the semester began, I was very comfortable with using NCBI to get genetic sequences or information.	1.3
2	Overall, the phylogeny tree building exercises improved my comfort level with basic GenBank data acquisition.	3.3
3	Before the semester began, I was familiar with the basic bioinformatics tools for examining DNA sequences, studying protein structure, and identifying active sites.	1.6
4	Overall, this exercise improved my understanding and appreciation for the processes and techniques used to analyze genetic sequences and protein structure.	3.9
5	I have a better understanding of the process of bioinformatics after performing these exercises than if I had just heard about them in lecture or read a textbook.	4.5
6	Bioinformatics was easier than I initially thought it would be.	2.9
7	I feel more comfortable using the NCBI and PDB websites now than I did before starting this exercise.	4.3
8	This exercise gave me a better appreciation of bioinformatics and how computers are used to analyze protein structure.	4.2
9	I would recommend having students carry out these exercises in this course in the future.	4.4
10	I would be interested in doing more of this exercises in future biology courses.	4.8

## DISCUSSION

The average grade of students who participated in the first activity was a B. Most of the students were able to follow the instructions found here [http://faculty.gordonstate.edu/clee/ABT7807\\_11\\_Lee.pdf](http://faculty.gordonstate.edu/clee/ABT7807_11_Lee.pdf) and produce the same result as the manuscript. Among the students who submitted their results from the activity, only about 5 % could not produce the phylogenetic tree. The most common problem was deleting the gaps from the sequence after alignment. The majority of the students agreed that both activities were relatively straightforward and took about two hours to complete the first activity, and an hour to complete the second exercise.

For the module of the protein structure of  $\alpha$ -amylase, the students commented that they had little difficulty following the instructions and getting the image. The only problem that some commented was that they did not know what the active site looks like, and so when they got to it they were not too sure if they had found the right part of the enzyme.

Additionally, an instructor may choose to separate this module to better illustrate the relatedness between bioinformatics and biochemistry. For instance, Exercise I could be given to the class during a certain part of the semester, and Exercise II could be given out at a later date. Depending on the order in which the biochemistry instructor lectures, it may be more ideal to make this into two instructional modules rather than a combined assignment.

In conclusion, the field of bioinformatics is complex, growing rapidly, and is becoming a very important part of all biological sciences. The incorporation of a simple introductory exercise can easily expose the students to the basic concepts in bioinformatics. A simple module like the one we developed goes a long way in providing undergraduate students with basic knowledge of the field with little financial cost.

### ACKNOWLEDGEMENTS

We would like to thank all of the students enrolled in the biochemistry class during spring of 2016 at Gordon State College.

### REFERENCES

- Aghajari, N., G. Feller, C. Gerday, and R. Haser. 2002. Structural basis of alpha-amylase activation by chloride. *Protein Sci*, 11(6), 1435–1441.
- Bernstein, H.J. 2000. Recent changes to RasMol, recombining the variants. *Trends in Biochemical Sciences (TIBS)*, 25(9), 453–455.
- Chen, J.X. 2008. *Guide to Graphics Software Tools*. Springer Science and Business Media.
- Ditty, J.L., K.M. Williams, M.M. Keller, G.Y. Chen, X. Liu, and R.E. Parales. 2013. Integrating grant-funded research into the undergraduate biology curriculum using IMG-ACT. *Biochemistry and Molecular Biology Education*, 41(1), 16–23.
- Feng, L., R. Fawaz, S. Hovde, F. Sheng, M. Nosrati, and J.H. Geiger. 2016. Crystal structures of *Escherichia coli* branching enzyme in complex with cyclodextrins. *Acta Crystallogr D Struct Biol*, 72, 641–647. doi:10.1107/S2059798316003272.
- Gray C, C.W. Price, C.T. Lee, A.H. Dewald, M.A. Cline, C.E. McAnany, L. Columbus, and C. Mura. 2015. Known structure, unknown function: An inquiry-based undergraduate biochemistry laboratory course. *Biochemistry and Molecular Biology Education*, 43(4), 245–262. doi:10.1002/bmb.20873.
- Hagen, J.B. 2000. The origins of bioinformatics. *Nature Review Genetics*, 1, 231–236. doi:10.1038/35042090.
- Herráez, A. 2006. Biomolecules in the computer: Jmol to the rescue. *Biochemistry and Molecular Biology Education*, 34(4), 7.
- Luscombe, N.M., D. Greenbaum, and M. Gerstein. 2001. What is bioinformatics? A proposed definition and overview of the field. *National Center for Biotechnology Information*, 40(4), 346–358.
- National Research Council. 2002. *Undergraduate Education to Prepare Biomedical Research Scientists*. Washington, D.C.
- Miskowski, J.A., D.R. Howard, M.L. Abler, and S.K. Grunwald. 2007. Design and implementation of an interdepartmental bioinformatics program across life science curricula. *Biochemistry and Molecular Biology Education*, 35(1), 9–15.

- Pembroke, T.J. 2000. Bio-molecular modelling utilizing Rasmol and PDB resources: a tutorial with HEW lysozyme. *Biochemistry and Molecular Biology Education*, 8, 297–300.
- Saboury, A.A. 2002. Stability, activity and binding properties study of  $\alpha$ -amylase upon interaction with  $\text{Ca}^{2+}$  and  $\text{Co}^{2+}$ . *Biologia, Bratislava*, 57/Suppl., 11, 221–228.
- Sayle, R. and E.J. Milner-White. 1995. RasMol: Biomolecular graphics for all. *Trends in Biochemical Sciences (TIBS)*, 20(9), 374.
- Tamura K, G. Stecher, D. Peterson, A. Filipski, and S. Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, 30, 2725–2729.
- Varma-Nelson, P. 2006. Peer-led team learning. *Metropolitan Universities*, 17(4), 19–29.
- Wightman, B and A.T. Hark. 2012. Integration of bioinformatics into an undergraduate biology curriculum and the impact on development of mathematical skills. *Biochem Mol Biol Educ*, 40(5), 310–319. doi: 10.1002/bmb.20637.