Supplementary Information: Overview of Contents

Supp Info 1: Precise learning gains, organized by course modules (pp 1–5)
Supp Info 2: A sample PyMOL-based in-class activity – Molecular visualization & structural analysis of serine proteases (pp 1–4)
Supp Info 3: Molecular docking tutorial for the Biochemistry Lab (Chem4411/21; pp 1–12)
Supp Info 4: Sample effort report (p 1)
Supp Info 5: Sample grading rubric from the first term (pp 1–3)
Supp Info 6: Sample student assessment of their learning gains (SALG) survey questions (pp 1–4)
Supp Info 7: Sample post-course survey questions (pp 1–8)
Known Structure, Unknown Function:
An Inquiry-based Undergraduate Biochemistry Lab Course
Cynthia Gray, Carol W. Price, Christopher T. Lee, Alison H. Dewald, Matthew A. Cline, Charles E. McAnany, Linda Columbus, Cameron Mura

Supplementary Information, 1:
Precise learning gains, organized by course modules (see also Table I)

For each module in Table I (main text), we include here an outline of the learning gains that we expect for a student. Each assignment is graded in accord with these gains; that is, the questions or graded portions of any assignment are grouped according to learning gains (Fig 1, main text), such that we can use the assignment to help assess the learning gains for a particular topic or concept.

Module 1A: Literature searches; electronic resources and tools
This module includes a one-hour lecture that highlights the many resources and literature search engines used by biomolecular scientists. In addition, the UVa library offers a program wherein a subject librarian is invited for subject-based instruction on current assignments and projects, with a focus on article databases, plagiarism, and critically evaluating resources. The learning gains assessed are (i) overall understanding of the purpose of literature searches (aims and concepts), (ii) identification of relevant literature (broader context), and (iii) the ability to find and understand resources (aims and concepts).

Module 1B: Basics of pipetting with the micropipette
Students deliver water to a weigh-boat on a balance and determine the volume delivered (based on the measured mass and the density of water) from each of their pipets (a P1000, P200, and P20). They calculate standard deviations for each measurement. The learning gains assessed for this assignment are (i) pipetting skills (aims and concepts) and (ii) quantitative skills (data processing).

Module 2: Critically reading the primary literature
This module includes both an interactive instructional period and a follow-up assignment for the laboratory period. The goal of this module is to (i) define what the primary literature encompasses (peer review, publication frequency, citations); (ii) address the differences between the primary (a direct report of research results and findings), secondary (review articles), and tertiary (textbooks and references) literature sources; (iii) introduce students to active and critical reading (identify the main question being addressed in the study, the conclusions, critically evaluate the data used to support the conclusions, pinpoint any missing factors or limitations); and (iv) walk students through relevant papers (i.e., provide a
guided journal-reading experience). The last two goals are achieved using the C.R.E.A.T.E. method developed by Hoskins et al. (ref [16] in the main text). The learning gains assessed in this module are for the student to (i) understand how the data in published articles were generated (aims and concepts), (ii) critically analyze the results of each figure in the article (aims and concepts, and data processing), (iii) elucidate a hypothesis based on the results of each data figure (experimental design), and (iv) be able to propose a follow-up experiment (experimental design).

Module 3: Biochemical buffers and solutions
This Module focuses on practical aspects of buffer preparation and introduces concepts of relevance to protein solutions (e.g., factors influencing solubility and stability, Hofmeister series, etc.). Students were taught the principles of making a buffer, using the Henderson-Hasselbalch equation, in their chemistry and biochemistry lecture courses. In this module, the practical considerations for buffer selection and solution preparation are emphasized—ionic strength, buffering capacity, compounds that may act as potential interferents in the reaction/assay, etc. Students are expected to (i) choose an appropriate buffer for downstream purification and enzymatic assays for their POI, (ii) calculate how to prepare the buffer, (iii) determine a reasonable volume to prepare, and then (iv) prepare the solution. The primary learning gain addressed in this module is laboratory skills. When students revisit these concepts in Module 9 (dialysis), the choice of buffer and how it is made can be assessed in terms of experimental design.

Modules 4–5: Enzyme kinetics assay (a hands-on assay using lactate dehydrogenase or a similarly well-characterized, commercially available enzyme)
This module seeks to provide students experience with (i) performing enzyme kinetics assays, (ii) the techniques used in performing spectrophotometric enzyme assays, (iii) the process of experimental design, (iv) how to process raw data, and (v) how to analyze/interpret the resultant processed data. Students are expected to perform an extensive pre-lab assignment which requires them to think deeply about the experiment before coming to lab (the learning gain here is experimental design). While a four-hour lab session suffices for performing the experiments, completing this lab requires students to arrive well-prepared. Because of their active role in planning and executing the lab work in this module, students develop an appreciation of the practical considerations of experimental design and preparedness, and are more likely to carefully scrutinize protocols and attempts to make improvements (‘shortcuts’), versus students who simply execute a pre-prepared protocol. In addition, students gain familiarity with each step and are thus better prepared to make logical choices when troubleshooting or adapting the protocol.
In the second kinetics module (Module 5), students learn how to analyze data from kinetics assays. Using a chosen (fixed) enzyme concentrations, students perform a series of assays in which substrate concentration is varied. They convert raw data (absorbance values) to concentrations using the appropriate extinction coefficients, generate plot(s), and calculate all possible enzyme kinetics parameters (learning gain: data processing). This module includes discussion of various means of data presentation and analysis (Michaelis-Menten, Lineweaver-Burke, Hanes-Woolf plots), determination of kinetics parameters ($K_M$, $k_{cat}$, $v_{max}$), interpretation of data from inhibitor assays, and analysis of alternate substrates. We have found that special attention must be paid to (i) how to use a spreadsheet effectively (most students use Microsoft Excel, some use Origin; we do not enforce a specific program), (ii) careful calculations of concentrations, and (iii) dimensional analysis. In working-up the data, students also learn how to effectively represent quantitative data as figures, which is a skill they use extensively in the second term. When writing the lab report for these two Modules, students are expected to demonstrate their understanding of the aims of the experiment and to relate their work to a broader context. Thus, all four of our learning gains are assessed in the laboratory and assignment associated with Modules 4-5.

**Module 6: Computational biology, I: Bioinformatic tools, web/database resources**

This module introduces students to computational methods that are commonly used in modern biochemical research. For the lab portion of this module, we draw upon an extensive and up-to-date collection of ‘Education Articles’ published in *PLoS Computational Biology*, including a practical tutorial on using many different types of bioinformatic approaches to analyze protein function from 3D structure (see main text). This Module’s lecture materials touch upon the core ideas of (i) molecular evolution and phylogeny (including phylogenetic trees); (ii) sequence alignment methods (pairwise and multiple, substitution matrices, gaps, local/global alignment, E-scores); (iii) the basic idea of ‘profiles’ and functional annotation; and (iv) structural bioinformatics (pairwise structural alignment, finding evolutionarily-conserved functional ‘patches’, etc.). The last portion—3D structural analysis and an introduction to the PyMOL molecular visualization environment—supplies a natural bridge to the next computational section (Module 11).

Students perform extensive, in-depth bioinformatic analyses of their POI during lab time, and complete an assignment that details their findings. The learning gains assessed in this module include understanding aims and concepts, investigative skills, critical thinking, and broader context.

**Modules 7–9: Recombinant protein expression, chromatography, protein purification, SDS-PAGE, and dialysis**
This module introduces students to experimental techniques that are central to biochemical research, including two approaches deemed by the American Chemical Society (ACS) to be important general techniques: electrophoretic methodologies and chromatographic separations. This Module includes interactive lectures that cover (i) general methods for cloning recombinant proteins (so they learn how their POI plasmid was created), (ii) regulation of protein expression and induction in various plasmid vectors, (iii) the usage of chemical tags, such as (His)$_6$, for purification purposes, (iv) gel-filtration, ion-exchange, and affinity chromatography, (v) electrophoretic gel separation techniques, and (vi) dialysis. We have found that supplying student groups with novel, uncharacterized proteins, which they first research via the literature and bioinformatic methods, gives students a sense of ownership of the project and instills the excitement for discovery that only true research can bestow.

Modules 7–9 span four weeks, but are contained within one lab report. This lab report allows us to assess the students in each of the four main learning gains we have identified (Fig 1). Student lab reports should (i) demonstrate an understanding of the purpose of the experiments (aims & concepts), (ii) display a grasp of the methods (experimental design), (iii) feature clear figures of carefully processed data (data processing), and (iv) relate their work to the ultimate goal of characterizing their POI (broader context).

**Module 10: Protein concentration determination; ligand-binding assays**

In the BioLEd curriculum, students are taught how to quantify proteins by two methods: (i) UV/vis spectrophotometry (absorbance at 280nm, $A_{280}$) and (ii) a modified Bradford assay that depends on Coomassie blue binding (a BioRad assay). Also, the molecular basis of protein-ligand binding are introduced in lecture slides, and methods for analyzing such data are introduced and summarized (include equilibrium dialysis, ligand-blotting, filter-binding analysis, isothermal titration calorimetry, mobility shift assays for nucleic acid-binding proteins, and spectroscopic (notably fluorescence) measurements). In past labs, students have studied the binding of the ligand Coomassie blue to bovine serum albumin (as described in ref [18] in the main text). As for the other two lab reports in the first semester, the grading rubric for this Module 10 report also addresses the four learning gains: understanding aims and concepts, experimental design, data processing, and broader context.

**Module 11: Computational biology, II: Molecular visualization, modeling, docking**

This Module introduces the basic concepts of molecular visualization and graphics (e.g., stereoscopic viewing, different types of molecular representations, surfaces, ‘scenes’, ray-tracing, etc.), followed by an overview of some elementary ideas of molecular modeling (e.g., rotamer libraries, homology modeling). Perhaps of greatest potential utility for their POI functional studies, we introduce the students to compu-
tational methods for ligand-protein docking; in the past we have employed the PATCHDOCK server, and most recently we have begun introducing students directly to the Linux-based usage of the AUTODOCK suite (see main text). Introducing this computational Module before the second semester enables students to begin immersing themselves in the (potentially foreign) computational tools and concepts; this, in turn, leads to students (i) becoming independent practitioners of the computational methods within a matter of weeks/months, and (ii) fruitfully applying this new knowledge and computational expertise to their POI over the remainder of the year-long course. Learning gains assessed in the Module 11 assignment include data processing (to generate figures that are scientifically convincing and lucid), critical thinking (analyzing the docking results), and general biochemical knowledge (to interpret the results).
Known Structure, Unknown Function: 
An Inquiry-based Undergraduate Biochemistry Lab Course
Cynthia Gray, Carol W. Price, Christopher T. Lee, Alison H. Dewald, Matthew A. Cline, Charles E. McAnany, Linda Columbus, Cameron Mura

Supplementary Information, 2: 
A sample PyMOL-based in-class activity — Molecular visualization & structural analysis of serine proteases

Useful PDB IDs and other helpful resources (e.g., Proteopedia, http://proteopedia.org):
- **2AGI**: “The leupeptin-trypsin covalent complex at 1.14 A resolution” (2PTN, no leupeptin)
- **2CGA**: “Chymotrypsinogen A. X-ray Crystal Structure Analysis and Refinement of a New Crystal Form at 1.8 A Resolution”

Trypsin

- Open the trypsin PDB file in PyMOL and then:
  - Add hydrogen atoms (‘A’ pull-down menu → ‘Hydrogens’ → ‘Add’)
  - Issue this command: select myhelix, resi 235:245
  - Issue this command: hide everything
  - Display cartoon of myhelix
  - Double middle-click near the center of the helix (to center the molecule)
  - Zoom the view of the helix
  - Identify the N- and C-termini of the helix
  - Show main chain atoms as sticks (for the helix selection)
  - Set background color to white
  - Save the image, and label the termini and the H-bonds that stabilize the α-helix

- Now, for the same helix, select the **hydrophobic residues** and color the selection (green). Hint: consider a command such as this (all one line):
  ```
  select my_hydrophobic, resn \n  leu+val+ile+gly+pro+ala+phe+met
  ```

- Now select the **polar residues** and color the selection (red), using a command similar to this:
  ```
  select my_polar, resn \n  glu+asp+asn+gln+lys+arg+his+ser+thr
  ```

Is ‘myhelix’ amphipathic? Explain.
- Select residues 1-234 and color them gray.
- Now display the cartoon of the entire trypsin molecule.
- Print it out and label the hydrophobic regions of the helix and explain the nature of the helix with respect to the rest of the protein.

**Chymotrypsin (chymo), another serine protease**

- Open the chymo PDB file in PyMOL.
- Using the commands above, identify an amphipathic β-strand. What residues comprise the strand?

- How many disulfide bonds are there in chymotrypsin?
...The Protease Mechanism — substrate specificity!

➢ Select residues 57, 102, and 195 using this PyMOL command: __________________________?

➢ Color the entire molecule gray and the selected residues another color.

➢ Display a molecular surface.
  *Do you see a large cavity next to the colored residues?*

➢ If so...
  • What properties of the cavity do you believe to be important in binding the peptide substrate?

  • Compare the cavity to that of trypsin and elastase. Do they differ? (If so, how?)

➢ Prepare a figure that illustrates the differences amongst these proteins (highlighting active site residues, cavities, and stabilizing residues).

➢ Select residues 215-219 and color them a different color (view with and without the surface).
  *What is the function of these residues in the chymotrypsin mechanism?*

➢ The compound tosyl-L-phenylalanine chloromethyl ketone (TPCK) specifically inhibits chymotrypsin by covalently labeling His57.
Given the chemical structure, can you suggest a mechanism for the inactivation reaction? (You can consult the enzyme catalysis chapter in Voet & Voet or other standard *Biochemistry* texts.)

Why is this inhibitor specific to chymotrypsin?

Draw a derivative of TPCK that might inhibit trypsin, highlighting what moieties you’ve changed.
Known Structure, Unknown Function:
An Inquiry-based Undergraduate Biochemistry Lab Course
Cynthia Gray, Carol W. Price, Christopher T. Lee, Alison H. Dewald, Matthew A. Cline, Charles E. McAnany, Linda Columbus, Cameron Mura

Supplementary Information 3:
Molecular Docking Tutorial for the Biochemistry Lab (Chem4411/21)

1 Initial Setup, Introduction to Linux

In this lab, we will dock ligands to your POI using the AutoDock Vina software on the Linux operating system. This section contains some computing details that may seem superfluous at first, but the material is critical to the rest of the docking workflow, so please study it carefully.

1.1 Basic Information and Nomenclature Conventions

Here we explore the basics of using the Linux operating system, and we describe some important terminology and formatting conventions that appear throughout this tutorial. Note that new terminology is defined in context using *italics* typeface.

In the following pages, we show commands/concepts/terms in the left-hand side of the table, and matching explanations follow on the right-hand side.

| **Dolphin** | The file manager for the Linux distribution (Fedora) and window manager (KDE) that you will be using. The Dolphin system works essentially like its Windows counterpart, Windows Explorer. To start, single-click the Home icon on the Desktop. |
| **Konsole** | A graphical environment that places you in a Unix shell, which allows you to input commands as text. To open a Konsole, right click on empty space on the Desktop and select Konsole.* |
| **Home Directory** | This is the directory that is shown when you first open Dolphin. You can consider this as roughly equivalent to My Documents in Windows. It is often denoted by a ‘~’ in file-paths. |

---

*The Konsole program is, technically, a terminal emulator, which provides you with a command-line interface (CLI). There are three two fundamentally distinct modes that one uses in working in Linux: (i) GUI-based (mouse clicks, like in MS Windows or Mac OS) and (ii) text-based (in the shell, using the CLI). In reality, a hybrid of (i) and (ii) is often the most efficient approach, and for this reason we introduce you to the Unix shell in this lab course. Many commands that we will use can be run only from the CLI, or can be run far more powerfully via the CLI (this may be counterintuitive, right now!). In general, running commands and performing operations in the shell will save much effort versus other methods (and is more easily reproduced, as one can communicate to someone else a list of text commands much more easily than showing mouse clicks across the graphical desktop background).
This is how we will refer to the directory where all of your work will be done for a single job. (Think of a job as one small, self-contained unit of work; for example, it would be one replicate, if you were pipetting many solutions to repeat a wet-lab experiment in triplicate... In computational biology, you would say you performed the calculation, or job, three times.) You will create your [jobDirectory] in the next step of this tutorial, and you will need to navigate to it on several occasions.

We will use this formatting to highlight many words throughout the tutorial. This font indicates one of two things, depending on context. First, you are looking for a button, field or file called some_name. The other case is that you will be typing a text command using the keyboard. In both cases, it is text that you should find verbatim, unless...

Text with [square brackets around it] will be text that is not precisely the same for every use. This will be such things as PDB codes or ligand names, which will generally differ for each job.

GUI

This stands for Graphical User Interface, which is how you usually interact with your computer.

PGUI

This is a denotation that will be used when the command is in the small gray PyMOL box containing the File menu.

VGUI

This is a denotation that will be used when the command is in the PyMOL Viewer. Most commands given here will be on the right side panel (the graphical menu of buttons).

AGUI

This is a denotation that will be used when the command is to be issued in the AUTODock plugin. Make sure you check the tabs at the top, if you are having a hard time finding a button.
This indicates that you should type [cmd] in the PyMOL shell.

This indicates that you should type [cmd] on the Konsole command line (\[cmd\]). Note that :) is not part of the actual command, but rather it denotes the shell prompt; so, do not type a :) instead just type that text following immediately to the right of the closing parenthesis.

This command, which stands for change directory, allows you to navigate the filesystem while in a shell (in Konsole). [dir] refers to the directory that you wish to go to, so to move to a directory called foo, you would type :)cd foo. Some special symbols can occur in the place of [dir] in order to do specific things. For example...

A command that moves you up in the filesystem by one directory. So, if you are in ~/TM1689.mcline/test1, then :)cd .. would move you to ~/TM1689.mcline

In this special case of cd (when no directory is specified) you will be taken back to the Home Directory.

This extremely useful command shows a listing of all files in the current directory (analogous to seeing a list of all files graphically, in the Windows or Mac OS).

This command launches PyMOL from Konsole. Note that you can also maneuver the file-system from within PyMOL, using cd, in the same fashion as from within a shell (Konsole).

This denotes a literal TAB on the keyboard. We use this key often because <TAB> is a powerful tool when using the shell (Konsole) and from within PyMOL's command line. <TAB> triggers the computer to try and finish that which you began typing (this is known as tab completion). This means that if you have a long command name, for example autoligand, then, more often than not, it will suffice to begin typing aut and then press the <TAB> key and let the computer finish your thought. If there are multiple (reasonably few) options for completion of a command that begins aut..., then the shell will list those potential commands, and that is handy in its own right (e.g., when you know only the begining of a command, or can't remember what some file was called...).
1.2 Setting up Your Directory

In this section, you will create your `[jobDirectory]` and start PyMOL in Linux for the first time.

1. Open the home directory from the Desktop icon Home.
2. Right click in Dolphin.
3. Create New → Folder...
4. Name the Folder `[POInumber].[userID]`, where userID is your UVa computing ID and POInumber is the ID code for your protein of interest (POI). For example, I would use TM1698.mac7yx.
5. Enter the folder that you just created by single-clicking on it (a single click is often enough for this Linux environment; you don’t necessarily need to double-click like in Windows).
6. Repeat the above method to create a new job folder called `test1` (or whatever you would like) in `[POInumber].[userID]`. This is where a single job will be performed. If a new job is done then create a new folder in `[POInumber].[userID]` to work in.

2 Performing a Docking Calculation Starting with Sanitized Receptors and MOL2 ligands

Before we can begin any docking project, we must gather the necessary files into your `[jobDirectory]`.

1. Open the Home Directory and navigate to the RECEPTORS directory.
2. From the receptor directory, click once on the appropriate receptor.pdb file and use the Ctrl-c keystroke to copy the file (alternatively, right-click on the file to see the list of possible options, one of which should be to copy it).
3. Navigate to your `[jobDirectory]`, which is the directory we named `test1`.
4. Use Ctrl-v to paste the file into your `[jobDirectory]`.
7. Open PyMOL from within this Konsole by typing `:pymol`.

Now, we can set-up and run our docking calculation from within PyMOL using the following sequence of operations in the PGUI window:

```
PyMOL> load [ligand].mol2
```

For the ligand, we are using the MOL2 coordinate file downloaded from the ZINC database. If you ever need coordinates for a ligand, we advise that you start searching at the ZINC database (an online database of purchasable ligands), as your search may well end there or never be complete. Note that no pre-processing has been performed on the ligand prior to your receiving it here.
We use Protein Data Bank (PDB, .pdb) structure files for docking. The PQR format (that was also in RECEPTORS) is useful in electrostatics calculations, but that can be a topic for later analysis or discussion.

This command adds hydrogens based on empty valences. (As a sidenote, the PyMOL protonation tool is necessarily the ‘best’ algorithm, but its advantages are that it does not require other third-party programs or libraries, and it isn’t as picky about ligands in the PDB file (versus other methods).)

This opens the GUI plugin that we will use to set-up our docking calculations. The GUI should open to the Configuration tab, which should already be set with appropriate parameters.

In this tab, we will set-up the three-dimensional (3D) region of space where the molecular docking will occur.

This places the center of the calculation grid (last step) at the center-of-mass of your protein, which is a good starting point; we may end-up needing to adjust this in a moment (see below).

In AUTO Dock Vina, a 3D grid is laid over the protein, and the interaction energy of various atoms is computed at each grid point. By setting the spacing to 1.0, the other measures presented in the GUI will also be in units of Angstroms (Å), and thus more easily understood. (If this step is confusing, that is ok: the PyMOL plugin is smart enough to adjust your measurements to correct geometric amounts when it creates the configuration file.)

In this step, you are telling AUTO Dock Vina where to search for the potential ligand-/substrate-binding site. For faster calculations, you will want this grid to be as small as possible. You can also adjust the Grid Center Coordinates to help shrink this region.

This saves the coordinates for the grid to a file called config.txt. IMPORTANT: This is probably the first time so far that it has become crucial to have done all of the work in your [jobDirectory]. If, instead, you had been working in the Home Directory (where multiple users where working), then only one config.txt would have been saved in the Home Directory (the others would be over-written), and so odds are it isn’t yours! This can generate great confusion.
<table>
<thead>
<tr>
<th>AGUI: Receptor</th>
<th>In this tab, we will finish preparing the receptor for docking by saving <strong>AutoDock</strong>’s own special format, PDBQT, which stores some additional information (beyond the coordinates in the PDB file format). The most critical piece of additional data is the bonding information for all atoms in the system — this information defines the molecular <strong>topology</strong> and also enables us to specify which bonds we will allow to freely rotate.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGUI: Select [receptor] from the PyMOL selections.</td>
<td>Here, you are simply telling PyMOL which of the objects that it is storing is the receptor (i.e., your POI, which is to be docked to).</td>
</tr>
<tr>
<td>AGUI: Press Generate Receptor -&gt;</td>
<td>The PyMOL plugin will now go find the correct preparation script and will apply it to the receptor (your POI). So, just wait for it to finish and add your receptor to the Receptors list. While this is occurring, you should look in the Log field for any errors, because if any part of this setup was wrong then this step is likely to fail (not to worry, this is probably not your fault). Unfortunately, these error messages can be subtle and, sometimes, the program will continue on computing, but will give flawed results. If an error arises here, and if you research it a little (use Google) and do not understand it, please show your TA the error message (it may be a computer/IT problem that can be readily addressed by one of us).</td>
</tr>
<tr>
<td>AGUI: Ligands</td>
<td>This is exactly the same as the receptor (above), except that now you are choosing your ligand... So, give this a shot on your own.</td>
</tr>
<tr>
<td>AGUI: Docking</td>
<td>This is where we can print the final configuration file for <strong>AutoDock</strong> Vina. The Run Vina button seems to be broken (software is not always perfect), and so we will have to resort to the Konsole to actually run Vina.</td>
</tr>
<tr>
<td>AGUI: Press Write Vina Input File(s)</td>
<td>The program writes another file in your [jobDirectory], which is probably starting to look like a cluttered mess. That’s OK.</td>
</tr>
<tr>
<td>Open a new Konsole and :)cd [jobDirectory].</td>
<td>There is a shortcut to do this, actually: In the Konsole that is running PyMOL, double-click the free area at the bottom, located beside the current tab. This opens a new tab which provides a shell that is already in the directory of the previous tab (so you don’t have to navigate there again).</td>
</tr>
</tbody>
</table>
By executing this command — type it exactly as shown, and press Enter — your computer should happily begin computing docking conformations. When this finishes, we will begin the fun part, analysis of the docked structures of the ligands to your POI (each of these are known as docking poses). Wait for this job to run to completion, which will be apparent when the command prompt :) returns control to you (the user) rather than the program that just finished running.

This executes an in-house script that reads the log file and builds a simple table in comma-separated value (CSV) format; MS Excel or most other data-processing/math software can read/import such files.

Analysis of the docking results does not require the plugin nor any Vina software, so we can complete that stage of our work on any computer workstation with PyMOL installed (e.g., your laptops). Make sure that you save these results in a place that you can find. If you prefer to continue on the Linux platform, we have several workstations in the research lab that you can ask us about using.

### 3 Analyzing Docking Results: The Mechanics

In this section, we will load the docked ligand conformations (the poses) into PyMOL for further analysis... and that will be all that is covered in this current tutorial, because analysis of the docked poses — literally, the docking results — is your job, and is specific to your POI. (Note that by ‘analysis’ we mean visual analysis and interpretation of the locations of the ligands [on the POI], their detailed 3D structures, inter-atomic interactions, ligands···POI contacts, etc.)

1. Open log.csv in Excel or a comparable program. If you are using a Linux workstation, we suggest LibreOffice Calc. Note that there are no headings. This is because we wish for this file to be easily loaded into any program that accepts CSV, but headings may hinder such compatibility. The headings are, from right to left, ‘Ligand Identifier’, ‘Mode Number’, ‘Binding Affinity (kcal/mol)’, ‘RMSD upper bound (Å)’ and ‘RMSD lower bound (Å)’. The RMSD values are of limited value, particularly when you dock to the entire protein (known as blind docking). The binding affinity can be viewed (very roughly) as the thermodynamic binding affinity, were the ligand to bind in exactly that pose; however, these are not truly accurate $\Delta G_{bind}$ values, and are only particularly useful when internally compared across different docked conformations/ligands/etc.

2. Start PyMOL on your computer.

3. Use the PGUI: File $\rightarrow$ Open... to find and load in your receptor.[receptor].pdb and your [ligand].docked.pdbqt

4. The docked poses are now in one PyMOL object with multiple states. To switch between the states you can use the arrows at the right of the PyMOL Viewer.
Congratulations! Now that you’re familiar with the mechanics of a docking calculation, use what you already know about PyMOL and biochemistry to draw conclusions from the docked conformations; ideally, perhaps you will be able to assess the ligand-binding preferences of your POI.

Appendices

PyMOL: A Quick-start Guide

Installation

There are two major ways to install PyMOL. First, one can obtain the educational version, though that edition is some releases behind the latest production version. Nevertheless, the educational version can be installed via a relatively simple process, and if you wish to use this version, Google ‘pymol’ (follow the directions at http://www.pymol.org). The second method requires you to compile PyMOL from source-code; this considerably more complicated route does provide you with the very latest, ‘bleeding-edge’ version. For Windows, visit http://www.pymolwiki.org/index.php/Windows_Install for directions. For Apple, go to http://www.pymolwiki.org/index.php/MAC_Install. Both methods may take some tinkering and online searching in order to make sure that appropriate libraries are in-place, cross-compatibility with versions of the Python and Tcl programming languages is not a problem, and so on.

Navigating in PyMOL

Object

This organizational unit is how PyMOL internally stores a 3D structural entity. When a protein or any other molecule is opened in PyMOL, that auto-creates one object; the next molecule that is loaded will be a new object, and so on. These objects can be edited as one group.

Object Control Panel

This is the area on the right-hand side of the PyMOL Viewer providing a list of the objects. Many of the GUI commands will be found here, and we will assume that you can explore this area on your own.

Left-click & drag

This rotates the protein representation in 3D space. Play with this for awhile to become comfortable with how this works.

Right-click & up-down drag

This zooms in and out on the protein.

Scroll wheel

This changes the clip, which is the width of a slab that dictates how much depth of the 3D space (the z-direction) is rendered at once. Most of the time, it’s not a bad idea to begin by increasing the clip until the entire protein can be seen (see also PyMOL’s closely related ‘zoom’ and ‘center’ commands). Another way to achieve this is to type zoom in the PyMOL PGUI.
PyMOL>load [file]  
This loads a structure file (e.g., in PDB format) into PyMOL, and thereby *instantiates* a new object corresponding to this structure.

PyMOL>save [file], [selection]  
This is the command for all of PyMOL save functionality, so it is a bit intricate. First, you specify the [file], which is what the file will be called. This needs the extension because that is how PyMOL determines in what file format to save. The two important types are *.pse*, which is a PyMOL session file allowing you to save your work, and *.pdb*, which simply specifies a 3D structure in PDB format.

fetch [pdbCode]  
This automatically retrieves the PDB entry from the PDB database, without your having to explicitly download it first (in fact, on Linux the PDB file will be downloaded to the local directory from which PyMOL was launched).

PyMOL>orient [object]  
This resets the view to see the [object].

PyMOL>delete [object]  
This removes the object from PyMOL.

Selections in PyMOL

Atom selections are a vital part of being able to manipulate molecules and subsets of molecules in PyMOL (or any other molecular visualization software environment). For high-quality molecular graphics, you will have to become quite familiar with *named atom selections*. Selections can be thought of as a type of object, but can contain any logical set of atoms, which can then be manipulated together as a unit (by ‘logical’ we mean in a Boolean sense). You can make selections with text commands or by clicking on the protein. The click method has seven modes for different selection *scopes*: atoms, residues, chains, segments, objects, molecules and C-alphas. To change the mode, PGUI: Mouse → Selection Mode.

PyMOL>select [selectionName], [descriptors]  
This is a command that makes a selection in PyMOL using logical descriptions. The [selectionName] is what the selection will be called in the Object Control Panel, and [descriptors] is the logic statement for whether or not an atom belongs in the selection. How to form the logic statements will be the rest of the topic of this section. If the [selectionName] is omitted, then the name will default to simply ‘sele’.

[object]  
When an object is included as part of the descriptor, then an atom must be part of that object in order to be chosen. So, if you would like to select all atoms in your receptor, the simple command would be *PyMOL>select sele, [receptor]*. This isn’t useful in and of itself, but will often be used in logic statements (when multiple objects are loaded, e.g., your POI and a homolog to be used for structural alignment in PyMOL).
**resn**

This is a descriptor that means residue name. So, if you wish to select all atoms associated with a residue that is named ‘PLP’ (in the PDB file from which the object arose), then the command would be

\[ \text{PyMOL} > \text{select sele, resn PLP} \]

**index**

This is a descriptor that means atom index number. So, if you wish to select atom 1 of the protein then the command is

\[ \text{PyMOL} > \text{select sele, index 1} \]

For early work, this descriptor is likely not as useful as others, because the mouse can achieve the same functionality (without your having to know the atomic index number(s)).

**resi**

This is a descriptor that means residue identifier. So, if you would like to select residue 1 of the protein, then

\[ \text{PyMOL} > \text{select sele, resi 1} \]

Again, this may not be as useful initially because the mouse can accomplish much the same without your needing to know the residue identifiers. However, a useful feature here is the ability to use this descriptor to select either a contiguous range of residues (e.g., ‘resi 1-10’) or a disconnected set of residues (e.g., ‘resi 1,3,5,7’)

**symbol**

This is a descriptor that means chemical symbol. So, if you seek to select all nitrogen atoms in an object, then

\[ \text{PyMOL} > \text{select sele, symbol N} \]

**chain**

This is a descriptor that chooses all atoms at the chain level. So, if you would like to select all (all atomic entities) in chain A, then issue the command

\[ \text{PyMOL} > \text{select sele, chain A} \]

This descriptor becomes a useful part of the selection logic when dealing with oligomeric (multi-chain) objects, such as is the case with many POIs.

**hetatm**

This special descriptor symbolizes every atomic entity in the object that is not part of the protein – i.e., is not proteinaceous (e.g., water molecules, bound ions, etc.). The simplest example of a command using this descriptor is

\[ \text{PyMOL} > \text{select sele, hetatm} \]

This selection *macro* gets its name from the fact that ‘hetatm’ is the starting string in these non-amino acid lines in PDB files.

**not**

This operator modifies an otherwise ‘normal’ atom selection string by (logically) negating it. An example would be

\[ \text{PyMOL} > \text{select sele, not hetatm} \]

which would select all non-hetero-atoms (i.e., the protein).

**and**

This boolean logical operator combines two descriptors by selecting only those atomic entities that satisfy both descriptors (i.e., it is the logical *intersection*). An example would be

\[ \text{PyMOL} > \text{select sele, [object] and symbol C} \]

which would select all of the carbon atoms in [object].
This boolean logical operator combines two descriptors by selecting only those atomic entities that satisfy at least one of the descriptors (i.e., it is the logical union). An example would be `PyMOL>select sele, symbol N or symbol O`, which would create an atom selection containing all of the oxygen and nitrogen atoms in the object.

### Modifying the Molecular Scene/Representation

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>PyMOL&gt;color [color], [selection]</code></td>
<td>This colors the selection to the [color]. The GUI can be used to determine which colors are available, and then this command can be used to then chose a particular color (by name).</td>
</tr>
<tr>
<td><code>PyMOL&gt;util.cbag [object]</code></td>
<td>This colors the atoms of the [object] with carbon = green, oxygen = red, and nitrogen = blue.</td>
</tr>
<tr>
<td><code>PyMOL&gt;show [representation], [selection]</code></td>
<td>This shows the representation of the selection. Note that it just adds the representation to shown representations, it does not remove representations. Use the GUI to find the different available representations then use this as a quick method to get back to that representation.</td>
</tr>
<tr>
<td><code>PyMOL&gt;hide [representation], [selection]</code></td>
<td>This hides the representation of the molecule. Note that it just removes the one representation. A common command that one might use is <code>hide everything, [selection]</code>. This removes all the representations from the active display, giving you a clean slate to work with.</td>
</tr>
<tr>
<td><code>PyMOL&gt;bg_color [representation]</code></td>
<td>This sets the background color, and is mostly used to set the background to white for making images for presentations and papers. Many people find a black background more visually appealing and simpler to work with for ‘zoomed-in’, detailed analysis of a molecular scene (better contrast); a white background is often used at a more global level (at the level of protein chains in an oligomer) and is almost always used for final rendering for purposes of a manuscript, poster, presentation, etc. (less ink used in printing a poster with molecular graphics on white backgrounds).</td>
</tr>
<tr>
<td><code>PyMOL&gt;set [name], [value], [selection]</code></td>
<td>This is a subtle and highly flexible command that is can be used to vary literally any of PyMOL’s hundreds (to thousands) of parameter settings. Some useful particularly useful settings to consider modifying/customizing are noted below.</td>
</tr>
<tr>
<td><code>PyMOL&gt;set transparency, [value], [selection]</code></td>
<td>This adjusts the transparency of any surfaces that are rendered (whether they are actively showing or hidden). The value ranges between 0 (full opacity; the default) and 1 (full transparency).</td>
</tr>
<tr>
<td><code>PyMOL&gt;set surface_color, [color], [selection]</code></td>
<td>This changes the color of the surface for the named atom selection.</td>
</tr>
</tbody>
</table>
PyMOL>set
sphere_transparency, [value], [selection]

Same as transparency (above), but adjusts the opacity of any
sphere representations, instead of surfaces.

PyMOL>ray [width]

This initiates ray-tracing of the molecular scene that is actively
visible in the viewer window, yielding high-quality, photorealistic
images. [width] specifies the width (in pixels) of the final
ray-traced output image (which is written to disk via the ‘png’
command).

For more information of molecular visualization and graphics, you can see “An Introduction to
Biomolecular Graphics” by Mura et al. [1] Note that if any of the results obtained via the
procedure described here are included in later work, then the convention is that you will need to cite
the software used — e.g., AutoDock Vina, the AutoDock PyMOL plugin, AutoDockTools-
4 (which operates behind the scenes in much of what was described above), and PyMOL. The
appropriate references are [2, 3, 4, 5].

References


with a new scoring function, efficient optimization, and multithreading. Journal of Computa-

[3] Daniel Seeliger and Bert L. Groot. Ligand docking and binding site analysis with PyMOL and

Goodsell, and Arthur J. Olson. AutoDock4 and AutoDockTools4: Automated docking with

For items 1–3, rate each group member (including yourself) on the group evaluation criteria listed below. Use the following scale:

<table>
<thead>
<tr>
<th>1 = poor</th>
<th>2 = marginal</th>
<th>3 = satisfactory/average</th>
<th>4 = good</th>
<th>5 = excellent</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Group evaluation criteria</th>
<th>self-evaluation</th>
<th>member 1 name:</th>
<th>member 2 name:</th>
<th>member 3 name:</th>
<th>member 4 name:</th>
<th>member 5 name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Participated in group meetings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Cooperated with group; supported group process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Demonstrated consistent commitment and effort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

List a skill that each student brings to the group (e.g., bioinformatics, writing, bench-work, interpersonal skills, etc.).

List a skill which is lacking from your group overall, or which could be improved.

List something specific that the group learned from you, that they may not have learned otherwise.

Overall, how effectively did your group work together on this task/assignment?

Suggest one change the group could make to improve its performance.
Known Structure, Unknown Function: An Inquiry-based Undergraduate Biochemistry Lab Course
Cynthia Gray, Carol W. Price, Christopher T. Lee, Alison H. Dewald, Matthew A. Cline, Charles E. McAnany, Linda Columbus, Cameron Mura

Supplementary Information, 5:
Sample grading rubric from the first term

Labs 7-9: Recombinant Protein Expression, Chromatography and SDS-PAGE, and Dialysis

Abstract
Identify Problem Studied  
* Isolation of protein  
2.5 pt: __________
Mention Techniques Used  
* Chromatography, SDS gels  
2.5 pt: __________
Relevant Data w/ significance  
* Information learned about protein  
2.5 pt: __________
Conciseness  
(Total: 10 points)  
2.5 pt: __________

Introduction
Student understands aims and concepts of the experiment  
Overall Clarity  
4 pt: __________
Cloning/Expression of Recombinant Protein (8 pts)
Vector and Antibiotic Selection  
2 pt: __________
Transformation  
2 pt: __________
E. coli as an expression host  
2 pt: __________
Induction with arabinose  
2 pt: __________
Chromatography (16 pts)
General explanation of chromatography  
4 pt: __________
Gel filtration  
*Separation based on size  
2 pt: __________
*Explanation of method  
2 pt: __________
Ion exchange  
*Separation based on pi  
2 pt: __________
*Explanation of method  
2 pt: __________
Affinity  
*Separation based on a specific interaction  
2 pt: __________
*Explanation of method  
2 pt: __________
SDS-PAGE (6 pts)
What does SDS-PAGE do?  
3 pt: __________
How does SDS-PAGE work?  
3 pt: __________
Protein of Interest (6 pts)
Presence and specifics of any affinity tags  
2 pt: __________
Theoretical MW 1 pt: __________
Theoretical pI 1 pt: __________
Identity 2 pt: __________
(Total: 40 points)

**M&M: Student understands experimental design**

**Cloning/Expression of Recombinant Protein (3 pts)**
- Plasmid and cell line 1 pt: __________
- Media and antibiotic 1 pt: __________
- Inducer (IPTG, arabinose, etc.) 1 pt: __________

**Chromatography (3 pts)**
- Gel Filtration 1 pt: __________
  * Sephadex G-100 resin, lysis buffer for elution
- Ion Exchange 1 pt: __________
  * DEAE (anion exchange) resin, step elution with increasing [NaCl]
- Affinity 1 pt: __________
  * Ni-NTA resin, elute with imidazole

**SDS-PAGE (2 pts)**
- *BioRad Ready SDS-PAGE (10-20% Tris-HCl) gel 2 pt: __________

**Dialysis (2 pt)** (Should be in the chromatography section, but optionally can be its own section.)
- One sentence stating that protein was dialyzed into a new buffer 2pt: __________
(Total: 10 points)

**Results: Student understands data processing**

**Overall Clarity** 2 pt: __________

**Cloning/Expression of Recombinant POI (7 pts)**
- SDS-PAGE gel image
  * Lanes labeled 1 pt: __________
  * MW marker labeled 1 pt: __________
  * Arrow/circle to indicate induction band (on each gel) 1 pt: __________
- Figure caption/legend 2 pt: __________
- Text describing results and reference to figure 2 pt: __________

**Chromatography (21 pts)**

**Gel Filtration - 7 pts**
- SDS-PAGE gel image
  * Lanes labeled (incl. where blue dextran and cyt c were observed) 1 pt: __________
  * MW marker labeled 1 pt: __________
- Figure caption/legend 2 pt: __________
- Text describing results and reference to figure 3 pt: __________

**Ion Exchange - 7 pts**
- SDS-PAGE gel image
  * Lanes labeled 1 pt: __________
  * MW marker labeled 1 pt: __________
- Figure caption/legend 2 pt: __________
- Text describing results and reference to figure 3 pt: __________

**Affinity - 7 pts**
- SDS-PAGE gel image
*Lanes labeled 1 pt: __________
*MW marker labeled 1 pt: __________
Figure caption/legend 2 pt: __________
Text describing results and reference to figure 3 pt: __________
(Total: 30 points)

**Discussion: Student capable of interpreting data and placing in a broader context**

Cloning/Expression results 3 pt: __________
*Was transformation and induction successful? How do they know?*

Gel Filtration 8 pt: __________
*Did protein elute in expected fraction? Why or why not?*
*Oligomer in void volume*

Ion Exchange 8 pt: __________
*Did protein elute in expected fraction? Why or why not?*
*Discussion of protein pI*

Affinity 8 pt: __________
*Did protein elute in expected fraction? Why or why not?*
*Discussion of the His-Tag*

Comparison of the three chromatography methods (yield and purity) 10 pt: __________
Additional citations and outside research 5 pt: __________
*How to improve purification, what can be done after purification*

Error Analysis (That is, they attempt to explain why they did not get the results that they may have expected based on what they know about their protein (oligomeric state, pI, etc.) They don’t just say “It didn’t work as expected.” 3 pt: __________

(Total: 45 points)

**Conclusion: Summarize Results**

Place results in a broader context 4 pt: __________
No introduction of new data/information 2 pt: __________
(Total: 10 points)

**References**

5 pt: __________

Grand Total of 150 pt: __________
Known Structure, Unknown Function:
An Inquiry-based Undergraduate Biochemistry Lab Course
Cynthia Gray, Carol W. Price, Christopher T. Lee, Alison H. Dewald, Matthew A. Cline, Charles E. McAnany, Linda Columbus, Cameron Mura

Supplementary Information, 6:
Sample student assessment of their learning gains (SALG) survey questions

SALG Survey Questions
Instructions to students:
• Teachers value student feedback, which is taken into account when improving courses such as this one. Please be as precise as you can in your answers. Please choose "not applicable" for any activity you did not do. You may find one or more questions at the end of each section that invite an answer in your own words. Please comment candidly, bearing in mind that future students will benefit from your thoughtfulness. Remember that this is an anonymous survey: your teacher will never know what any individual student has written.
• You may see the following note next to some questions:
  "D" — Department question. The department head can view the responses to these questions.

Understanding
1. Presently, I understand...

<table>
<thead>
<tr>
<th>1.1 The following concepts that will be explored in this class</th>
<th>not applicable</th>
<th>not at all</th>
<th>just a little</th>
<th>somewhat</th>
<th>a lot</th>
<th>a great deal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1 Literature searches and electronic resources</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.2 Reading primary literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.3 Critiquing primary literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.4 Writing primary literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.5 Bioinformatics tools and methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.6 Molecular visualization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.7 Molecular modeling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.8 Molecular docking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.9 Buffer solutions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.10 Kinetic assays</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.11 Enzyme kinetics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.12 Data analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.13 Recombinant protein expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.14 Chromatography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.15 Protein purification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.16 SDS-PAGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.17 Dialysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.18 Protein concentration determination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.19 Protein-ligand binding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.20 Experimental design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.21 Choosing appropriate controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.22 Systematic perturbation of an experiment to test my hypothesis generated by initial data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.23 Poster preparation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.24 Poster presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2 The relationships between the concepts listed above</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3 How ideas we will explore in this class relate to ideas I have encountered in other classes within this subject area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1.4 How ideas we will explore in this class relate to ideas I have encountered in classes outside of this subject area

1.5 How studying this subject helps people address real world issues

1.6 What do you expect to understand at the end of the class that you do not know now?

**Skills**

2. Presently, I can...

2.1 Find articles relevant to a particular problem in professional journals or elsewhere

2.2 Identify patterns in data

2.3 Recognize a sound argument and appropriate use of evidence

2.4 Write documents in discipline-appropriate style and format

2.5 Work effectively with others

2.6 Prepare and give oral presentations

2.7 What do you expect to be able to do at the end of the course that you cannot do now?

2.8 Please comment on how you expect this material to integrate with your career and/or life.
Attitudes

3. Presently, I am...

<table>
<thead>
<tr>
<th></th>
<th>not applicable</th>
<th>not at all</th>
<th>just a little</th>
<th>somewhat</th>
<th>a lot</th>
<th>a great deal</th>
</tr>
</thead>
</table>

3.1 Enthusiastic about the subject

3.2 Interested in taking or planning to take additional classes in this subject

3.3 Confident that I understand the subject

3.4 Willing to seek help from others (teacher, peers, TA) when working on academic problems

3.5 Please comment on your present level of interest in this subject.

3.6 Why did you choose to take this class?

Integration of Learning

4. Presently, I am in the habit of...

<table>
<thead>
<tr>
<th></th>
<th>not applicable</th>
<th>not at all</th>
<th>just a little</th>
<th>somewhat</th>
<th>a lot</th>
<th>a great deal</th>
</tr>
</thead>
</table>

4.1 Connecting key ideas I learn in class with real research scenarios.
Known Structure, Unknown Function:
An Inquiry-based Undergraduate Biochemistry Lab Course
Cynthia Gray, Carol W. Price, Christopher T. Lee, Alisón H. Dewald, Matthew A. Cline,
Charles E. McAnany, Linda Columbus, Cameron Mura

Supplementary Information, 7:
Sample post–course survey questions

Hello: You are invited to participate in this post-course survey to help assess and evaluate Chem4411/21,
Biochemistry Labs I & II. The questionnaire should take approximately 20 minutes to complete. Your par-
ticipation in this study is completely voluntary. There are no foreseeable risks associated with this pro-
ject. However, if you feel uncomfortable answering any particular questions (unless marked as required),
you may skip the question. You can withdraw from the survey at any point. It is very important for us to
learn your opinions. Your survey responses will be strictly confidential, and data from this research will
be reported only in aggregate. Your information will be coded and will remain confidential. If you have
questions at any time about the survey or the procedures, you may contact Cindy Gray by email
cg4eq@virginia.edu). Participation in this survey will enter you into the lottery system described in the
preliminary announcement email (we will contact you if you have won a prize from our lottery!). Thank
you very much for your time and support. Please start with the survey now by clicking the ‘Continue’ but-
ton below.

Current Occupation

For Chem4421 (second semester), who was your Teaching Assistant (TA)?
1. Abelin, Sarah
2. Dawidowski, Alison
3. Ebmeier, Jennifer
4. Fox, Donald
5. Kabzinski, Joseph
6. Kroncke, Ryan
7. Lo, Brett
8. Malaker, Tracy
9. Oliver, Ronald
10. Patterson, Peter
11. Randolph, Jennifer
12. I don't remember
Prior to these courses, did you have any experience in a research-based laboratory?

1. no
2. yes, but only as a course
3. yes, but only as an occupation/internship
4. yes, both as a course and an occupation/internship

What was your overall grade in the courses?

1. A
2. B+
3. B
4. B-
5. C+
6. C
7. C-
8. D+
9. D
10. D-
11. F
12. Prefer Not to Answer

What are your plans, if any, for science education beyond your undergraduate degree?

1. Ph.D. in biology–related field
2. Ph.D. in chemistry–related field
3. Ph.D. in physical science*
4. M.A. in life science*
5. M.A. in physical science*
6. Advanced degree in field other than sciences
7. Medical School (MD)
8. MD/PhD*
9. Other health profession
10. Law or business degree
11. Teaching
12. Peace Corps or similar
13. Work first
14. No school after college, science career
15. No school after college, non-science related career
16. Other

How did the research experience in these courses influence your postgraduate plans?

1. I had a plan for postgraduate education that has not changed.
2. It helped confirm of my postgraduate education consideration.
3. It changed my prior plan in the direction toward a postgraduate education.
4. It changed my prior plan in the direction away from a postgraduate education.
5. I still do not have plans for postgraduate education.
The following statements refer to the poster presentation portion of the courses. Please rate how much you agree/disagree with the following statements:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composing the poster helped me prioritize the data of my research.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presenting the poster developed my oral scientific communication.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The poster presentations made me more confident in my research.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall, the introduction of poster presentation to the courses gave me a deeper understanding of biochemistry.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What specific elements of poster presentation did you find useful? What should be improved?

The next few statements refer to the computational aspects of the courses (bioinformatics, databases, literature searches, docking, etc.). Please rate how much you agree/disagree with the following statements:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>The computational aspects of the courses helped me become more independent in my research.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The computational aspects of the courses made me more confident in my research.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The computational aspects of the courses provided tools for me to be an active participant in discovery.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The computational aspects of the course made my research more tangible.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall, I have a deeper understanding of biochemistry due to the computational aspects of these courses.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What specific computational aspects of these courses did you find useful? What could be improved?

The next few statements refer to the course lectures/class-times. Please rate how much you agree/disagree with the following statements:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>I attended lectures regularly.</td>
<td>☐</td>
<td>☒</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The lectures worked well in conjunction with the lab.</td>
<td>☐</td>
<td>☒</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The lectures were clear and coherent.</td>
<td>☐</td>
<td>☒</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The lectures helped me to start thinking independently.</td>
<td>☐</td>
<td>☒</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Overall, I have a deeper understanding of biochemistry due to the course lectures.</td>
<td>☐</td>
<td>☒</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

What specific elements of the lectures in the courses did you find useful? What could be improved?

The next few statements will refer to the course labs. Please rate how much you agree/disagree with the following statements:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>The labs increased my factual knowledge.</td>
<td>☐</td>
<td>☒</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The labs increased my critical thinking.</td>
<td>☐</td>
<td>☒</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I have retained skills in experimental design because of the labs.</td>
<td>☐</td>
<td>☒</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>These labs have given me confidence in my research.</td>
<td>☐</td>
<td>☒</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I have retained skills in data analysis because of the labs.</td>
<td>☐</td>
<td>☒</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I have retained skills in group work because of the lab.</td>
<td>☐</td>
<td>☒</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Overall, I have a deeper understanding of biochemistry due to the course labs.</td>
<td>☐</td>
<td>☒</td>
<td>☒</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
What specific elements of the course laboratories did you find useful? What could be improved?

The next few statements refer to manuscript writing. Please rate how much you agree/disagree with the following statements:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>The courses improved my scientific writing skills.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I learned how to organize my research in a scientific manuscript.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing the manuscript gave me more confidence in my research.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall, I have a deeper understanding of biochemistry due to the introduction of manuscript writing to these courses.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What specific elements of the manuscript writing in the courses did you find useful? What could be improved?

The next few statements refer to group meetings. Please rate how much you agree/disagree with the following statements:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>I believe there were a sufficient amount of group meetings.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The group meetings gave me constructive feedback to improve my research.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The group meetings gave me constructive feedback for the final projects of the course (poster and manuscript).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall, I have a deeper understanding of biochemistry due to my participation in group meetings.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What specific elements of the group meetings in the courses did you find useful? What could be improved?

Please rate how much you agree/disagree with the following statements:

“Compared to other undergraduate laboratory classes I have taken, this class……”

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>…encourages more independent thinking.</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>…teaches more skills in time management.</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>…teaches more skills in scientific communication.</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>…better prepares students to present scientific information.</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>…encourages greater confidence in a student’s scientific knowledge.</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
</tbody>
</table>

Please rate how much you agree or disagree with these statements:

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>I learned to communicate well with my group.</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>I learned to work professionally with my group.</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>My group was able to delegate tasks well.</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>My group met a sufficient amount of time outside of class.</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>I have a deeper understanding of biochemistry due to working in groups.</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
</tbody>
</table>

For the purposes of these courses, I believe the ideal group size would be ______ students.

1. two
2. three
3. four
4. five
5. six
6. seven, or more
Please rank the following items in relation to each other from 1 to 9, with a ‘1’ indicating the greatest contribution to your understanding of biochemistry and a ‘9’ indicating the least contribution.

<table>
<thead>
<tr>
<th>Item</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor</td>
<td></td>
</tr>
<tr>
<td>TA</td>
<td></td>
</tr>
<tr>
<td>Group members</td>
<td></td>
</tr>
<tr>
<td>Assigned readings</td>
<td></td>
</tr>
<tr>
<td>Course Lecture</td>
<td></td>
</tr>
<tr>
<td>Course Lab</td>
<td></td>
</tr>
<tr>
<td>Group meetings</td>
<td></td>
</tr>
<tr>
<td>Poster presentation</td>
<td></td>
</tr>
<tr>
<td>Writing a manuscript</td>
<td></td>
</tr>
</tbody>
</table>

Please briefly elaborate on what you ranked in the above list as having the greatest contribution (‘1’) and what factor you ranked as having the least contribution (‘9’) to your understanding of biochemistry.

Do you feel that these Chem4411/21 courses adequately prepared you to work more independently in a laboratory setting? If so, how? If not, what would have helped you feel more prepared?

Do you feel that these courses helped you think like a scientist rather than a student? If so, how? If not, how can these courses be improved in that respect?
What are the most positive and most negative differences that you saw in these labs, compared to previous lab courses that you have taken? Please elaborate.

These next few statements will refer to your TA in the Chem4421 course (second semester). Please rate how much you agree/disagree with the following statements:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Disagree</th>
<th>Not Sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>My TA was able to answer my questions in lab.</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
</tr>
<tr>
<td>My TA was available for questions during their specified office hours.</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
</tr>
<tr>
<td>My TA gave constructive feedback.</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
</tr>
<tr>
<td>My TA was approachable.</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
</tr>
<tr>
<td>Overall, I have a deeper understanding of biochemistry due to the contributions of my TA.</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
<td>❑</td>
</tr>
</tbody>
</table>

What do you think your TA did well? What could your TA improve upon?

Do you have any other comments about these courses?