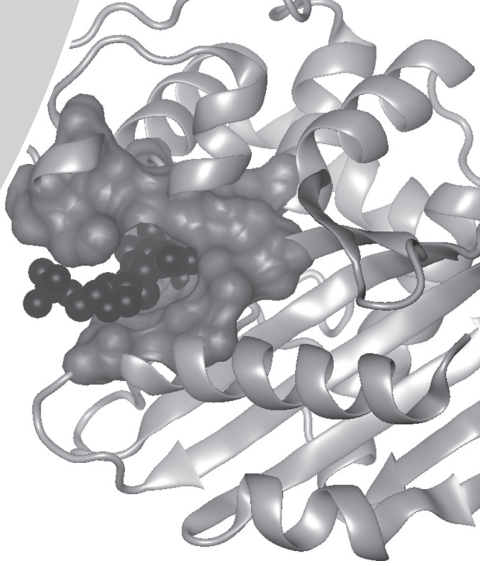
INQUIRY & INVESTIGATION

Protein Structure Analysis: Introducing Students to Rational Drug Design



AGNIESZKA SZARECKA,
CHRISTOPHER DOBSON

Abstract

We describe a series of engaging exercises in which students emulate the process that researchers use to efficiently develop new pharmaceutical drugs, that of rational drug design. The activities are taken from a three- to four-hour workshop regularly conducted with first-year college students and presented here to take place over three to four class periods. Although targeted at college students, these activities may be appropriate at the high school level as well, particularly in an AP Biology course. The exercises introduce students to the topics of bioinformatics and computer modeling, in the context of rational drug design, using free online resources such as databases and computer programs. Through the process of learning about computational drug design and drug optimization, students also learn content such as elements of protein structure and protein–ligand interactions. Based on our assessment, students enjoy the exercises, become more interested in bioinformatics and computer modeling, and demonstrate an increase in content knowledge relevant to the topics.

Key Words: *Protein structure; secondary; tertiary; ligand; binding pocket; binding affinity; interactions; therapeutic drug; drug target; ligand docking.*

○ Introduction

The use of computers and computational models in biology and biomedical research has increased dramatically in recent decades. At the same time, students at all levels use computers and the Internet in their daily lives more than ever before. The abundance of computational tools available online to support teaching and research can be used to enhance students' learning and retention through engagement in hands-on activities using devices with which they are already familiar. Here, we propose a series of lessons utilizing free online databases and computer programs to introduce bioinformatics and computer modeling. In these lessons, students more effectively learn various aspects of protein structure, cellular functions, involvement in disease processes, and protein–ligand interactions. Students are typically very interested in the process of developing new pharmaceuticals, and after the series of computer exercises proposed here they will understand the steps involved in computational drug-design pipelines, a process that uses structural data for a specific protein target to rationally design the best effector molecule – a drug candidate.

Unlike most wet-lab experiments in molecular biology and biochemistry, these computational experiments do not require expensive equipment or laboratory setups. The tools we describe here are predominantly free online resources that are accessible via standard Internet connection. Locally run software can also be used, and we will recommend free and easy-to-install programs, capable of running on the majority of common platforms. The exercises can easily be divided between pre-lab, in-class, and homework assignments. They could also be conducted in one three-hour lab with a pre-lab/post-lab assignment, or over two labs, or as some combination of lecture and lab. Students with a particular interest in this field can progress to more complex tasks that would form a month- or term-long project in drug design.

This material can be used in a variety of ways across the curriculum. The following are examples of concepts that can be taught: building blocks of protein structure, hierarchy of protein structure, protein folding, structural features in proteins that allow for interactions between proteins and small molecules (drug candidates, drugs, toxins, regulatory/signaling molecules), protein–protein or protein–nucleic acid complex formation, methods of determining three-dimensional structures of biomolecules, diversity of protein function (e.g., catalysis, signaling, transport), identification of protein targets in the mechanisms of human diseases, and rational drug design. Students will learn to visualize and analyze experimentally determined protein structures, to collect data from multiple databases, to use computer programs to calculate properties of protein surfaces and simulate protein–ligand interactions, and finally to interpret the data in the context of drug design.

We have designed and implemented these exercises for an approximately three- to four-hour, introductory-level session for first-year college students, assuming no prior background in molecular biology or biochemistry, thus encouraging all science majors to attend (the results of these workshop-like sessions are described

The American Biology Teacher, Vol. 81, No. 6, pp. 423–429, ISSN 0002-7685, electronic ISSN 1938-4211. © 2019 National Association of Biology Teachers. All rights reserved. Please direct all requests for permission to photocopy or reproduce article content through the University of California Press's Reprints and Permissions web page, www.ucpress.edu/journals.php?p=reprints. DOI: https://doi.org/10.1525/abt.2019.81.6.423.

Table 1. High school educational standards potentially addressed by these exercises.

Next Generation Science Standards	
HS-LS1-2	Develop and use a model to illustrate the hierarchical organization of interacting systems that provide specific functions within multicellular organisms.
AP Biology Learning Objectives	
4.1	The student is able to explain the connection between the sequence and the subcomponents of a biological polymer and its properties.
4.2	The student is able to refine representations and models to explain how the subcomponents of a biological polymer and their sequence determine the properties of that polymer.
4.3	The student is able to use models to predict and justify that changes in the subcomponents of a biological polymer affect the functionality of the molecule.
4.9	The student is able to predict the effects of a change in a component(s) of a biological system on the functionality of an organism(s).

below in the Assessment section). These exercises may be appropriate at the high school level, particularly in an AP Biology course, and the educational standards potentially addressed at this level are shown in Table 1.

○ Objective

Students collect protein structure data from an online database, visualize the structure, and collect data on structural features consistent with binding of small molecules. They also analyze the protein structure, and known ligands, to inform the process of designing a novel ligand, a potential therapeutic drug.

○ Lesson Design

In our implementation, we have worked with 15–25 students and one instructor. A teaching assistant will be necessary if the class is larger. We recommend that these exercises be conducted over four class periods (or two labs) or, alternatively, three class periods and one lab, preceded by a homework assignment and in-class discussion for students divided into groups of two to four. Table 2 lists the online databases and computer programs we recommend using in lessons 1–4. The Supplemental Material available with the online version of this article includes comments and technical tips on using these resources, as well as suggestions for written work and worksheets for the students to turn in after completing their analysis.

Table 2. Databases and computer programs that are the easiest to use and most effective for lessons 1–4.

Computer Resource	Applications & Use in the Lesson Plan
Therapeutic Target Database[a]	Drugs and targets searchable by disease (lessons 1 and 3) https://db.idrblab.org/ttd/
Protein Data Bank (PDB)[b]	Protein structures (lessons 1–4) https://www.rcsb.org/
Free program VMD[c]	To visualize and analyze protein structures (lessons 2–4) http://www.ks.uiuc.edu/Research/vmd/
DogSiteScorer server[d]	To identify cavities in the protein surface, and to calculate their properties and potential for druggability (lesson 3) https://proteins.plus/
ZINC database[e]	Known ligands and drugs (lessons 3 and 4) http://zinc.docking.org/ http://zinc15.docking.org/substances/home/
Online server ProDrg[f]	To create files containing three-dimensional structures of known and novel ligands (lessons 3 and 4) http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg
Online server One-Click Docking	To create files containing spatial coordinates of known and novel ligands (lessons 3 and 4) and to run docking simulations (lesson 4) https://mcule.com/apps/1-click-docking/
Free program Avogadro[g]	To create and visualize 3D structure files for ligands (lessons 3 and 4) https://avogadro.cc/
Online server SwissDock[h]	To run docking simulations (lesson 4) http://swissdock.ch/
Online Docking Server[i]	To run docking simulations (lesson 4) https://www.dockingserver.com/web/gettingstarted/

[a]Li et al. (2018).
[b]Additional tools are available therein: Protein Workshop and Ligand Explorer, and "Molecule of the Month" library http://pdb101.rcsb.org/motm/motm-by-title.
[c]Humphrey et al. (1996). VMD tutorial (most relevant sections are "Introduction" and "Working with a Single Molecule") is available at http://www.ks.uiuc.edu/Training/Tutorials/vmd/tutorial-html/index.html.
[d]Volkamer et al. (2012).
[e]Irwin & Shoichet (2005), Sterling & Irwin (2015).
[f]Schüttelkopf & van Aalten (2004).
[g]Hanwell et al. (2012).
[h]Grosdidier et al. (2011).
[i]Bikadi & Hazai (2009).

○ Lesson 1: What Are Therapeutic Drugs?

The timeline for Lesson 1 is as follows: group work about 20 minutes, and whole-class discussion with the instructor about 20–30 minutes (alternatively, this lesson could form a preparatory homework assignment, preferably set up for groups).

Students start by completing Worksheet 1 (Figure 1). The activity includes group discussion and Internet searches, and the completed worksheets are then used in a follow-up discussion with the instructor. Student learning objectives are to exchange information on (1) different prescription medications of which students have heard, (2) major developers of currently used and new drugs, and (3) the regulatory process that leads to drug approval for clinical use. They collect and discuss data on pharmaceutical companies, drug development timelines, and FDA clinical trials. Depending on the course and students' interests, the instructor can also plan a discussion of the cost of drug development and approval process, ethical and scientific aspects of clinical trial design, or marketing and pricing/patenting issues. The following resources are helpful to the instructor when preparing for the discussion or to share with students:

- European Bioinformatics Institute, https://www.ebi.ac.uk/training/online/course/functional-genomics-i-introduction-and-designing-e/functional-genomics-case-studies/drug-disc
- Food and Drug Administration, https://www.fda.gov/Drugs/DevelopmentApprovalProcess/default.htm
- Journal articles: Kaitin (2010), DiMasi et al. (2016), and Kapetanovic (2008)

It will enrich the lesson if students can revise/expand their worksheets in the course of the discussion and are then asked to turn in a final written summary of topics b–g (Figure 1). The last column in the worksheet introduces a critical component and concept of rational drug design, the molecular target. This is a specific biomolecule, most commonly a protein, that is involved in the disease mechanism and whose structure is used to design a new drug (in a process called structure-based drug design). Discussion of this concept forms a logical transition to the next lesson on proteins and the protein structure–function relationship. The instructor can choose to use Table 3 and Figure 2 to help explain the following terms: *molecular target*, *binding*

Fill out the table and answer the questions based on your own knowledge, discussion within the group, and Internet searches.

a) What prescription drugs are you familiar with?

Drug name	Disease	Treatment or cure?	*Molecular target **
Aspirin	Fever, pain	treatment	Cyclooxygenase enzyme

b) What major pharmaceutical companies are you familiar with?

c) Who approves a new drug to be used in clinical practice?

d) What's a clinical trial?

e) What is a side effect or ADR – Adverse Drug Reaction?

f) What are the main steps in the process of developing a new prescription drug?

g) *How do you understand the word "molecular target"?

Figure 1. Example worksheet introducing students to basic concepts related to the development of prescription drugs.

Table 3. Examples of common diseases, drugs used to treat them, their molecular targets, and PDB IDs of structures of the targets or related to the targets.

Disease	Drug	Molecular Target	PDB ID of Relevant Structures[a]
Hypertension	Atenolol (Antagonist)	Beta-1 adrenergic receptor	3D4S (human beta-2 receptor) 2VT4 (turkey beta-1 receptor)
Alzheimer's	Galantamine (Inhibitor)	Acetylcholinesterase	5HF9
Migraine	Sumatriptan (Agonist)	5-hydroxytryptamine 1D receptor GPCR-family 1	5V54 chain A (1B receptor)
	Dihydroergotamine	D2-dopamine receptor	3PBL (chain B) (D3 receptor)
Stroke	Ebselen	Serum albumin	1AO6
Pain/inflammation	Ibuprofen	Cyclooxygenase-2	3LN1
Pain	Vicodine/hydrocodone	Mu-type opioid receptor	5C1M chain B
Bacterial infection	Tazobactam	Beta-lactamase inhibitor	1BLC 3BLM
Anxiety	Oxazepam (Agonist)	Peripheral-type benzodiazepine receptor	2N02
Asthma	Enprofylline	CAMP-specific 3′,5′-cyclic phosphodiesterase	2QYK 3I8V 3TVX
HIV	Lamivudine	HIV reverse transcriptase	3LP0

[a]PDB ID is a unique four-character label that each structure is assigned on deposition to the Protein Data Bank (PDB). This unique identifier can be used to search for and retrieve protein structures from the PDB. In addition, a number of online modeling programs will accept the PDB ID as input and will automatically download the structure from the database.

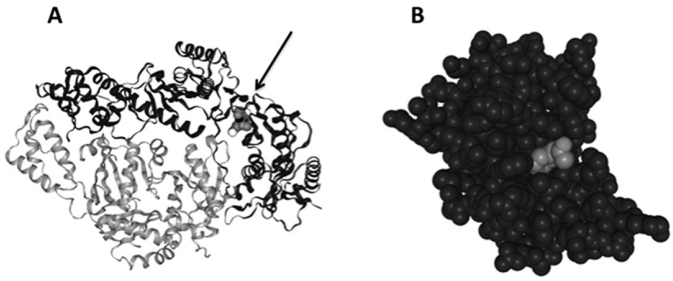


Figure 2. Examples of proteins bound to drug molecules. These images (saved from PDB Structure 3D View) illustrate the protein 3D structure and the concept of ligand binding to a cavity in the protein surface. (**A**) Ribbon model of AZT-resistant HIV reverse transcriptase (PDB ID 1RT3) bound to novel drug marked by an arrow. (**B**) Space-filled models of phospholipase A2 bound to aspirin (PDB ID 1OXR). The ligand is shown in both A and B as a space-filled model.

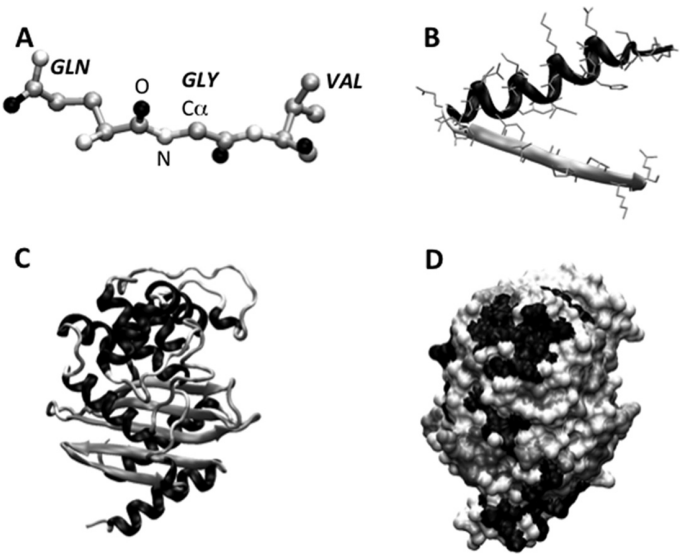


Figure 3. The hierarchical nature of protein structure. (**A**) Amino acids and peptide bonds form a protein chain. (**B**) Backbone folds to form secondary structural elements (alpha-helices and beta-strands) with side chains "hanging." (**C**) Tertiary structure showing the conformation of the backbone and secondary elements. (**D**) Space-filled model of the tertiary structure illustrating dense packing of atoms. Images created using VMD (Humphrey et al., 1996) of OXA-24 beta-lactamase (PDB ID 3PAE).

pocket, *active site*, *ligand*, *receptor*, and others. For example, aspirin is a ligand for the molecular target phospholipase A2 (Figure 2B) and functions by blocking the active site. Interestingly, aspirin also binds to cyclooxygenase (lesson 2, part 2), which can be mentioned to students as an example of a drug interacting with multiple proteins.

○ Lesson 2: Drugs Interact with Proteins

Part 1: Introduction to Protein Structure (~15 minutes)

Figure 3 can be used to explain the hierarchical structure of proteins. Alternatively, if the Protein Data Bank (PDB; Table 2) has already been mentioned, students can use a combination of Figure 3 and their own protein structure images. Concepts that can also be introduced here are native state and protein function dependence on the correct folding to the native state (i.e., adopting a unique 3D conformation of the amino acid chain). Students can complete Worksheet 2A (see Supplemental Material) as part of in-class activity or homework.

Part 2: Exploring Structures of Protein–Drug Complexes (~30 minutes)

This exercise can be completed in a few different ways depending on the students' level of interest and preparation. The easiest way is to use PDB and its Molecule of the Month library (http://pdb101.rcsb.org/motm/motm-by-category). We recommend choosing the "You and Your Health" or "Drug Action" categories. However, the instructor or students can choose any other category depending on the overall context in which this material is taught. Each protein–drug pair in the Molecule of the Month library is described in the context of disease symptoms and mechanism of the protein's native and disease-causing function, and a URL link (showing as PDB ID) is provided directly to the protein structure record in the PDB (for an explanation of PDB ID, see Table 3 footnotes). In this way, students learn the necessary background before accessing the structure file. We typically use the example of cyclooxygenase from "Drug Action" (http://pdb101.rcsb.org/motm/17). This particular example is advantageous because it pertains to a well-known drug (aspirin), allows for an introduction to pain killers in general, and explains the off-target interactions and side effects that drugs most often cause. After reading the information provided, students can follow the link to the structure of the COX-ligand complex, 4COX, or 1PTH. Figure 4 shows the summary page of the PDB record for 1PTH, from which students can access the structure file (via Download tab), literature reference for this protein structure, and "3D View" programs ("Structure" or "Ligand Interaction"). "3DView" Structure/Ligand can be used to analyze the protein's secondary and tertiary structures as well as the ligand binding site. The exact location of the ligand binding pocket can be determined along with the amino acid residues that form the pocket. More detailed analysis of the structure of the binding pocket (using "Ligand Interaction"; Figure 5) will help students understand the mechanism of drug action (blocking the enzyme active site near the heme group), and identify which amino acids bind the drug and through what types of interactions (hydrogen bond, hydrophobic, metal ion). Students learn how the ligand "fits" the binding pocket in terms of space (ligand size) and interactions with neighboring amino acids. This knowledge helps inform the design of an alternative/improved drug. We recommend Worksheet 2B (see Supplemental Material) to guide students through this part of lesson 2 and to help them write meaningful observations that should be used in lessons 3 and 4.

Alternatively, a more difficult and more advanced form of this exercise would be to search PDB directly for (1) all available structures of a specific drug target (e.g., adrenergic receptors) to identify structures that contain a ligand or different ligands, or (2) all available protein structures that are bound to a specific ligand (e.g., ibuprofen).

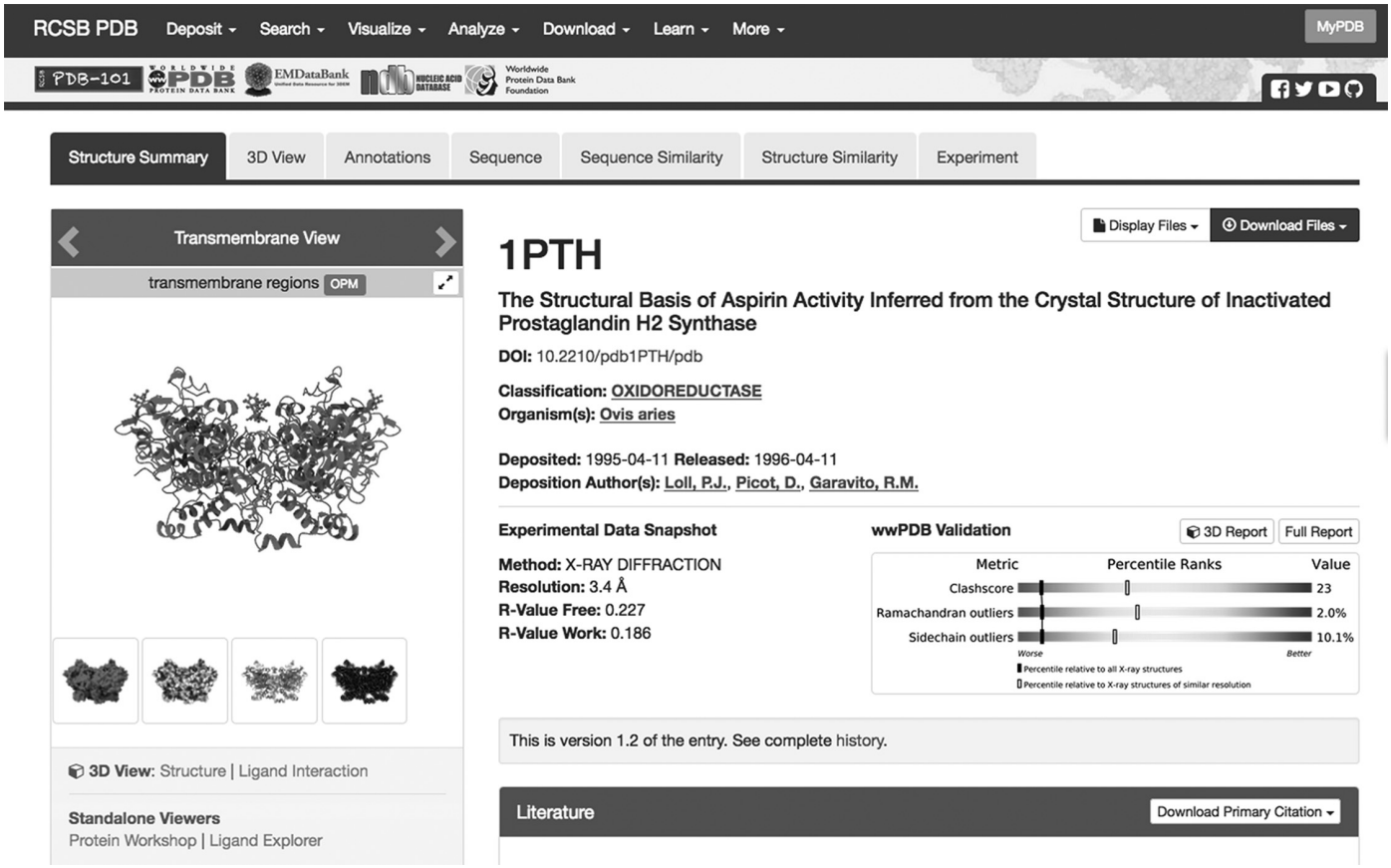


Figure 4. Protein Data Bank summary page for cyclooxygenase (PDB ID 1PTH).

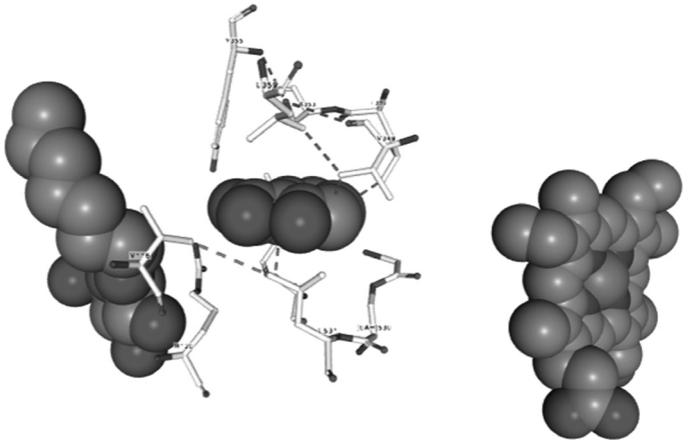


Figure 5. Example of the Ligand View program output that shows the protein residues forming interactions with the bound ligand.

Other tools to visualize the complex structures are Protein Workshop and Ligand Explorer (Figure 4 bottom left); standalone Java-based viewers available from the PDB Summary page; or a separate program, VMD (Humphrey et al., 1996). To use VMD most effectively, we recommend that students download structure files directly from the PDB (in PDB format; see Figure 4, top right, "Download Files") and complete the program tutorial to learn the basic functions (see Table 2 footnotes). In our experience, learning the basic operations takes about 15 minutes with instruction and one to two hours if students work independently through the first part of the tutorial ("Single Molecule"). Learning VMD is strongly recommended for highly motivated students or students who will work on an inquiry-based, longer-term project.

○ Lesson 3: Ligand Design

We recommend that students complete a homework assignment before lesson 3 that includes online registrations for servers or download and installations for the locally run programs to ensure they complete lesson 3 within 50 minutes. The homework can also include short practice with molecular editors such as ProDrg or Avogadro (Table 2).

After completion of lesson 2, students have sufficient background in protein structure, protein involvement in disease, and the concept of a small drug molecule binding to its protein target. The goals of lesson 3 are (1) to choose their own target disease and target protein for rational drug design and (2) to design novel or modified ligand molecules. The easiest tool to search for a target or target-drug pair is the Therapeutic Target Database (Table 2), which students can query by disease, by target, or by drug. Students should note the results of their searches, including search criteria, database hits, and their final selections. They can also be asked to write a short explanation of why they selected a particular disease/drug and a summary of information provided by the database for this disease/drug. Table 3 can also be used to initiate the

search or to direct students to specific therapeutic targets. It is important to note that rational drug design relies on availability of target structures. Structures for many targets are available in the PDB, but for others only homologous structures may be available (e.g., the mouse protein rather than human, or a different receptor type). Homologs can be used for the simplest purpose of lessons 3 and 4 (to illustrate drug design as the process of designing a ligand that has the highest affinity to the target), but students engaged in an inquiry project should understand the caveats. Most importantly, the active site of a nonhuman target may differ in amino acid sequence and conformation, which may then affect ligand binding. In principle, students could try to build a homology model of a human protein based on the structure of a homolog, but that would be a long and complex extension of the lesson we are proposing here. We refer readers interested in extending the project and including homology modeling to Schmidt et al. (2014). At the end of this part of the lesson, students should write a brief report on what structures are available in the PDB and whether the structure available is a homolog.

Once a target structure has been identified, students proceed to calculating all potential druggable cavities in the protein surface using the DogSiteScorer server (Table 2). To input the target structure to the server, students can (1) specify the PDB ID, (2) download the selected protein structure prior to using DogSiteScorer and then upload it from disk, or (3) upload any PDB protein file that has been user-modified. The server offers multiple different programs; for this lesson, the DogSiteScorer module should be selected after uploading the protein file. Students should write a report in which they record the number and locations of cavities that may be potential binding sites as well as their physical and chemical properties (e.g., volume and hydrophobicity) directly from the DogSiteScorer output. If the target PDB ID corresponds to a protein–ligand complex, they identify and analyze the binding site they have already seen in the previous lesson and evaluate its "druggability" compared to other cavities in the protein surface. Students then use these data to design their own ligands, namely small organic molecules that are to be tested as the drug candidates. For example, if the pocket is highly hydrophobic and roughly spherical, students can design a drug ligand whose structure corresponds to these characteristics.

The second goal of lesson 3 is to create 3D input files for the candidate drugs. Students are familiar with drawing two-dimensional chemical formulae of molecules, but they need to use a program to convert each such formula to spatial coordinates for all atoms that are then saved in a text file. Many programs are available to do so. We recommend that students design their own novel ligand molecules and create the 3D input files using ProDrg or Avogadro (Table 2). One-Click Docking is also very convenient in that it allows the user to streamline the ligand structure drawing, choice of docking target, and docking simulations. When saving the ligand's 3D coordinates, we recommend choosing mol2 format (for more information on file formats, see Supplemental Material, section II).

The best form of this exercise is for students to identify at least one ligand that is known to bind to their target protein and use this ligand structure as the basis for chemical modifications to match the amino acid structure of the binding pocket and/or the shape and physical properties of the binding cavity. This would correspond to drug optimization (e.g., improving binding affinity). Alternatively, novel structures can be designed following Lipinski's rule of five to discover novel binders, compounds with higher affinity, and/or potentially better pharmacokinetic parameters (Lipinski et al., 1997). Lipinski's rule of five is an empirically based predictor of absorption or permeation of a drug and includes, for example, the number of hydrogen bond donors and acceptors in the ligand molecule. Students should write a report in which they show the ligands they have designed and explain (1) how the new structure relates to other known ligands/drugs and (2) why they think the ligand will "fit" one of the previously identified pockets.

We recommend that students submit their input files (target protein and ligand structure files) for docking simulations (Table 2) as homework after lesson 3. This will ensure that the simulation results are returned by servers in time for analysis during lesson 4. Alternatively, part of lesson 4 could be devoted to starting additional simulations or data analysis, and the report could be completed as homework after lesson 4.

○ Lesson 4: Docking Simulations

There are multiple free and licensed programs to carry out protein–ligand simulations, but for this activity we have chosen to use two online-based programs that are the easiest to learn and fastest to execute. We suggest SwissDock or One-Click Docking servers (Table 2) to carry out the docking simulations (for additional comments and tips, see Supplemental Material, section II). Both servers will also display the results (docked ligand poses and estimated binding energies) at the end of the simulation for students to analyze their results. We recommend that in their written reports, students should (1) record the estimated binding energies for different poses or different ligands, (2) paste in and/or describe the location of the binding pocket within the protein structure, and (3) identify the most favorable orientation of the ligand in the pocket. The students should try to predict the most promising ligand and suggest further modifications that can be tested. More advanced students can use VMD to list all the amino acids forming contacts with the ligand and determine the type of contacts (i.e., hydrogen bonding, hydrophobic, charge-charge, etc.). Several possible versions of this exercise can be implemented:

- Students design a few chemical modifications to the known drug molecule and dock them to a single target protein (drug optimization).
- Students dock several known drugs to proteins other than their designed targets (identifying potential off-targets).
- Students design a set of completely novel molecules and dock them to a selected single target to explore possible regulatory sites (looking for binding to non-orthosteric cavities).
- Students conduct a series of consecutive docking simulations, each with a ligand progressively modified according to the properties of the binding pocket.
- Students can evaluate the accuracy of the docking method by "re-docking" a known ligand to reproduce the crystal conformation of the complex (several different docking methods could be used for comparison if this variant is selected).

Questions that students should answer after the exercise include:

- Which ligand binds with greater affinity?
- Is this ligand likely to be selective?

- What interactions are formed between the protein and the ligand?
- Does the ligand bind in a competitive or noncompetitive site?

The analysis can be completed within a class period, but depending on the complexity of the selected research question, it may also be extended to a homework assignment followed by a written report on the docking results. Note that due to space limitation, we provide little background on the computational methods behind protein–ligand docking, but for interested students we recommend Kitchen et al. (2004).

○ Assessment

We administered a survey following our most recent workshop that asked whether students (1) found the material useful, (2) enjoyed the exercises, (3) understood the process of drug discovery, and (4) became more interested in bioinformatics and computer modeling in general. The survey was given to 35 students, and 90% responded positively to all four of these questions. When asked what they appreciated most about the entire event, they responded with positive comments about (1) being able to explore and learn through hands-on activities; (2) learning what resources exist for studies of drug interactions and discovery, including online databases; (3) being able to visualize structures and simulations of biomolecules; (4) the structure of the activity (building up from basic background to more complex concepts); (5) the combination of discussion, analysis, and computational work; and (6) the good pace of the workshop.

We also assessed student understanding of the learning objectives of the workshop in a pre- and post-workshop quiz. The questions on the quiz covered the concepts of proteins folding to a biologically active native state, methods for obtaining new therapeutic drugs, and types of biomolecules with which therapeutics interact. Among the 35 attendees, 34% answered all questions correctly on the pre-workshop quiz, while 60% did so on the post-workshop quiz. Given that the workshops are attended by students from different majors, we conclude that covering at least some aspects of protein structure and function in the context of drug design, and in the format of four consecutive hands-on computer exercises, is highly effective and engaging.

○ Conclusions

We are proposing a set of easy and versatile computer-based exercises that introduce students to protein structure, function, involvement in human disease, and interactions with therapeutic drugs. Students become acquainted with concepts behind computer-aided, rational drug design and computer modeling of protein function. Tools recommended here can be used as an easy and fast illustration of the concepts behind drug design or they can be used in more advanced, inquiry-based, dry-lab experiments.

References

Bikadi, Z. & Hazai, E. (2009). Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *Journal of Cheminformatics*, *1*, 15.

DiMasi, J.A., Grabowski, H.G. & Hansen, R.W. (2016). Innovation in the pharmaceutical industry: new estimates of R&D costs. *Journal of Health Economics*, *47*, 20–33.

Grosdidier, A., Zoete, V. & Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, *39*, W270–W277.

Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E. & Hutchison, G.R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, *4*, 17.

Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, *14*, 33–38.

Irwin, J.J. & Shoichet, B.K. (2005). ZINC – a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*, *45*, 177–182.

Kaitin, K.I. (2010). Deconstructing the drug development process: the new face of innovation. *Clinical Pharmacology & Therapeutics*, *87*, 356–361.

Kapetanovic, I.M. (2008). Computer-aided drug discovery and development (CADD): in silico-chemico-biological approach. *Chemico-Biological Interactions*, *171*, 165–176.

Kitchen, D.W., Decornez, H., Furr, J.R. & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery*, *3*, 935–949.

Li, Y.H., Yu, C.Y., Li, X.X., Zhang, P., Tang, J., Yang, Q., et al. (2018). Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Research*, *46*, D1121–D1127.

Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, *23*, 3–25.

Schmidt, T., Bergner, A. & Schwede, T. (2014). Modelling three-dimensional protein structures for applications in drug design. *Drug Discovery Today*, *19*, 890–987.

Schüttelkopf, A.W. & van Aalten, D.M. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica D*, *60*, 1355–1363.

Sterling, T. & Irwin, J.J. (2015). ZINC 15 – ligand discovery for everyone. *Journal of Chemical Information and Modeling*, *55*, 2324–2337.

Volkamer, A., Kuhn, D., Rippmann, F. & Rarey, M. (2012). DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics*, *28*, 2074–2075.

AGNIESZKA SZARECKA (szarecka@gvsu.edu) is an Associate Professor in the Department of Cell and Molecular Biology and CHRISTOPHER DOBSON (dobsonc@gvsu.edu) is a Professor in the Department of Biology at Grand Valley State University, Allendale, MI 49401.