Molecular Docking with Open Access Software: Development of an Online Laboratory Handbook and Remote Workflow for Chemistry and Pharmacy Master's Students to Undertake Computer-Aided Drug Design

Bethanie A. Clent, Yuhang Wang, Hugh C. Britton, Frank Otto, Christopher J. Swain, Matthew H. Todd,* Jonathan D. Wilden,* and Alethea B. Tabor*
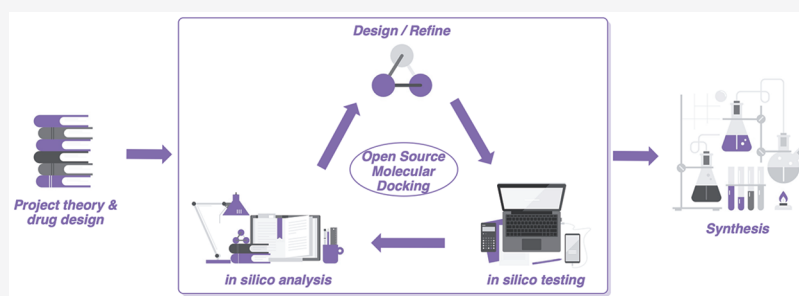
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: In response to the closure of many university laboratories due to the Covid-19 pandemic in 2020, a handbook and remote webinar approach designed to support students in the use of software tools for computer-aided drug design has been developed. Specifically, the course has been designed for chemistry and pharmacy students who have little or no experience of computational techniques and can use open-source software on their own machines. In this way a flexible and relevant course, giving a rigorous academic experience, could be delivered even in the most challenging of circumstances. We believe that this laboratory protocol will help to "democratize" the scientific process in this field.

KEYWORDS: *Graduate Education/Research, Chemoinformatics, Interdisciplinary/Multidisciplinary, Textbooks/Reference Books, Distance Learning/Self Instruction, Inquiry-Based/Discovery Learning, Medicinal Chemistry, Molecular Modeling*

Understanding the basis of computer-aided drug design (CADD) is a vital part of the training of students specializing in medicinal chemistry, chemical biology, or pharmacy. The ability to select a target, explore the interactions between protein and ligand, design small molecule drug candidates and evaluate their potential *in silico* provides students with a crucial skill-set for a successful career as medicinal chemists.[1−4] In March 2020, at the beginning of the COVID-19 pandemic, there arose the need for simple access to training and software for this purpose that could be followed by students working remotely; universities around the world scrambled to translate their laboratory-based teaching and research projects to remote working. At University College London (UCL), we were faced with the daunting task of transferring 30 students on our MRes Organic Chemistry: Drug Discovery and MSc Drug Design and Development courses from laboratory research projects based around organic synthesis to meaningful and topical research projects in computational drug design. In the UK, students who have chosen to specialize in medicinal chemistry, chemical biology, or pharmacy generally do not have extensive backgrounds or

expertise in computer programming. The students were nervous about engaging with the "command line" and were not experienced in downloading specialist software, getting it operational, and transferring data from one stage of the CADD process to the next. Moreover, during lockdown the students had mostly left the university, and we were now without the means to deliver hands-on training or provide computational facilities on campus.

In pharmaceutical companies, CADD is carried out by well-resourced specialized groups, who have access to powerful software packages with high-end graphical user interfaces. The cost of providing such packages and training for the numbers of students involved was prohibitive. We wished to provide a
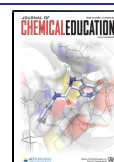
solution that would equip students to carry out such analyses in less well-resourced settings, and also give them more of an insight into the different stages of computational drug design. If students could use freely available (ideally open source) software, this would additionally allow them to work with other students (perhaps also those beyond UCL) using a common platform, which would improve their ability to share data and collaborate on the underlying science. However, open-source computational drug design software is usually written by experts, for experts, with little time or resource to invest in optimization. Thus, user interfaces are usually minimal and not written with the inexperienced user in mind. A limited number of groups have developed courses and protocols to teach CADD to undergraduate students. However, these have mainly focused on using open-source software to teach the basic principles of drug design,[5−10] or using specific tools and approaches to illustrate specific drug-target interactions.[11−13]

These factors motivated us to put together a "laboratory handbook" to enable students to carry out complete CADD-based research projects, using readily accessible, open-source software. More significantly it prompted us to reassess how our masters' courses were designed and delivered, specifically to integrate CADD with the synthesis of the resulting drug candidates. In this way we believed that we could deliver a flexible course in the most challenging of circumstances while still providing a relevant and rigorous academic training for the student. UCL has pioneered a framework for research-based education, the Connected Curriculum.[14−16] The central aim of this framework is to empower students at all levels to learn through participating in research, and to fully integrate research and teaching throughout all aspects of undergraduate and postgraduate courses. With this in mind, we sought to build an integrated line of research activity throughout the Masters' courses, connecting the students with current research projects in medicinal chemistry throughout UCL, and building research partnerships between the participating staff and students on the course. Hence, the "laboratory handbook" is not focused on one particular drug target or stage of the drug design process. Rather, it is intended to equip senior undergraduate and Masters' level students with the skills necessary to independently carry out their own research investigations, on targets chosen in discussion with their research supervisors.

We worked with students from our first such cohort to put together a handbook that would allow us to incorporate these tools into our course as a guide for future students who have little or no experience of this area. This resource is intended to help students to access, download, use, and interpret the outputs from a complete suite of open-source software and guide them through all stages of the process of computational drug design. This provides a complete workflow, from the initial downloading and evaluation of a target protein structure from the PDB, through searching databases for candidate structures, designing and inputting drug candidates, molecular docking experiments, evaluation of the results and subsequent rounds of molecular redesign. Importantly, all of the software used is open source and so can be downloaded onto the students' own laptops and run remotely: the laboratory handbook provides guidance to the students on how to accomplish this. Universities and colleges where students have access to central servers may wish to allow their students to use these for the molecular docking itself, as this is computationally intensive; however, we have also included instructions for alternatives that could be run locally or using GoogleCoLab.[17] Our intention was always that each student would work on their own project (target, hit, therapeutic area), and be supervised by one academic member of staff on a day-to-day basis. Yet staff and students have mixed levels of experience in CADD, so it was essential that the course supported all participants through the use of webinars, discussion fora, and email-based assistance where appropriate (outlined in the next section).

At this juncture it is important to stress that the development of this CADD project was not intended as a replacement for laboratory-based inquiry. Rather, we saw an opportunity to connect both CADD and "traditional" laboratory synthetic endeavor. Our vision has always been that the student would undertake a CADD project and then proceed to the laboratory where they can then prepare the molecules that they themselves have designed. Thus, we aim to deliver a rich, immersive learning experience connected throughout the year by a single theme and goal.

■ RATIONALE FOR THE SELECTION OF SOFTWARE PACKAGES

The software packages to be used by the students were selected with accessibility and usability in mind. While there are many excellent commercial software packages, these often require expensive licenses; in addition, many packages require access to dedicated software license management software. Since most of the students would be in lockdown in student halls, or at home in different countries, and perhaps using communal hardware, we felt the use of open-source software would offer greatest access. It was also recognized that multiple platforms would need to be supported: Mac OS, Windows, and Linux.

Much open-source software is written by experts to be used by other experts and not written for students with no prior knowledge of the use of the command line. We sought to minimize the learning curve by selecting packages that could be installed using conventional "drag and drop" installation. In the cases where there was no simple "drag and drop" installer, or where additional nonstandard software libraries needed to be installed, we chose to install these on a central server that provided students with access to a Jupyter notebook,[18] an open-source web application that allows anyone to create and share documents that contain live code, equations, visualizations, and narrative text. This meant the student would only need a web browser on their local machine to access the software. An additional advantage was that all students had access to effectively the same hardware for some of the more computationally intensive exercises.

■ EXAMPLE PROJECTS

The decision to promote the use of freely available software was taken in part to aid students' options for collaboration with others, and we envisaged adopting open science principles in future iterations if the students felt comfortable sharing their work as they went along.[19,20] In selecting targets, starting points, and therapeutic areas on which to work, the students were presented with two options: (i) work on a target of interest to their supervising academic; or (ii) work on an existing open science project. Both of these bring interesting opportunities for the students, though the second would allow

the student to interact with a potentially wider scientific team. Locally we have considerable experience in encouraging and mentoring students as part of open science projects such as "Open-Source Malaria"[21,22] and "Open-Source Antibiotics",[23] where validated small molecule hits currently require improvement against known but therapeutically novel targets. At the time of development, there were also emerging opportunities for students to make contributions to ongoing COVID-19 research.[24] The publication of the crystal structure of the SARS-CoV-2 main protease (M^{Pro}), with structural data on fragment screens and inhibitors,[25,26] along with continuously updated structural data[27] from the open source XChem fragment screen at the Diamond Light Source (UK national synchrotron science facility) made this a compelling target. In both cases, as for many of the less-open projects available locally, it was important to ensure that resources were available for biological evaluation of the proposed molecules, were they to be so promising that their synthesis was justified. In other words, it was motivating for the students to see that there was the possibility of a complete design–synthesis–test cycle within one project. Indeed, for the M^{Pro} project, there was the possibility of submitting promising structures to the COVID Moonshot, a fully open initiative to crowdsourcing, synthesizing, and assaying molecule designs from medicinal chemists worldwide.[28,29]

■ WORKFLOW

Software Employed

We wanted students to be able to open, and perhaps edit, files using a commonly available text editor (e.g., Notepad, BBEdit, or TextEdit) so we made the decision to use only generic chemical file types that use plain text (e.g., SMILES, sdf, PDB). All students had access to ChemDraw; however, alternative chemical drawing packages could also be used (Marvin, ChemDoodle, JME, Ketcher).

DataWarrior[30] is a desktop chemically intelligent spreadsheet and data analysis tool that combines spreadsheet filtering and interactive plotting. The website provides downloadable installers for Mac, Windows, and Linux. It comes with comprehensive help and there are several introductory videos available.[31]

PyMOL[32−34] is an open-source biomolecule viewer that can be used to view proteins and their ligands. It can be installed by downloading a compressed file or by using Anaconda. Again, there is a freely accessible introductory video.[33]

SMINA[35] is a fork of AutoDock Vina. There are many different docking software packages; most require familiarity with the command line and defining the binding site can be daunting for inexperienced users. We chose SMINA because the ligand in the crystal structure can be used to define the binding site. We also chose to install SMINA on the UCL cluster and provide access via a Jupyter notebook. While this notebook is specific for the UCL cluster, we have also provided two other versions. The first version runs locally and requires the user to install RDKit,[36] OpenBabel,[37] SMINA,[35] and py3Dmol.[38] The second version can be run using GoogleCoLab[17] and thus only a web browser is required. To enable facile communication, a Slack[39] channel was set up and all students given access. Individual channels were set up for the different software packages.

Step 1: Identifying a Target

Students select their protein of interest through discussion with their supervisor. Once they have settled upon a target, crystal structures of the protein must be obtained from either the Protein Data Bank or another online source. The students were directed to obtain their protein structures as pdb files, as discussed in chapter 3 of the laboratory handbook (see the Supporting Information).

Step 2: Visualizing the Target Protein

Students examine their protein of interest through molecular modeling in PyMOL.[17] Chapter 4 of the laboratory handbook details how to analyze the targets in PyMOL (through mouse-clicking and command line tutorials) and demonstrates how to derive meaningful binding information from ligand-bound structures.

Step 3: Developing a Compound Library

A collection of molecules for docking experiments are gathered through the use of online databases. There are many excellent databases available such as ZINC 15,[40,41] Enamine,[42] Pubchem,[43] and ChEMBL,[44,45] with millions of structures readily accessible. In addition to compiling data sets from online sources, students are encouraged to modify existing ligands and/or design their own novel compounds.

Step 4: Preparation for Docking Experiments

Preparing Compounds for Docking. The students review their compound data sets and filter out structures based on key properties such as LogP, TPSA, MW and other drug likeliness values in DataWarrior. After refining their lists of compounds, minimized 3D conformations of each structure are generated and written into sdf files for use in docking experiments as detailed in chapter 6 of the laboratory handbook.

Preparing PDB Files for Docking. The quality of the protein structure is critical for successful docking. Some of the potential issues and solutions are described as follows:

- Alternates. Residues with alternate locations and/or ambiguous sequence identities are examined, and those with the highest occupancy are chosen.
- Termini. Protein chain C- or N-termini that need to be charged or capped require attention. Similarly, for DNA structures the terminal phosphate may only have three oxygen atoms bonded to the phosphorus. Accordingly, an additional oxygen atom should be added.
- Sometimes loops are very disordered and appear as breaks in the chain. It may be possible to use a loop library to model a replacement.
- Hydrogen atoms are usually not visible and so need to be added and checked. In particular, hydrogen atoms on heteroatoms and water molecules should be checked, especially at the active site where the local environment may influence a residue's pKa value.
- Ligand. Novel ligands in particular need checking to confirm that atoms and bond orders are correct.
- Conformation. Check that torsions are reasonable and there are no clashes.
- Charge. The charge on all ionizable groups should be checked.

Step 5: Performing Molecular Docking Experiments

In chapter 9 of the laboratory handbook, students are taken through the process of running their own docking experiments.
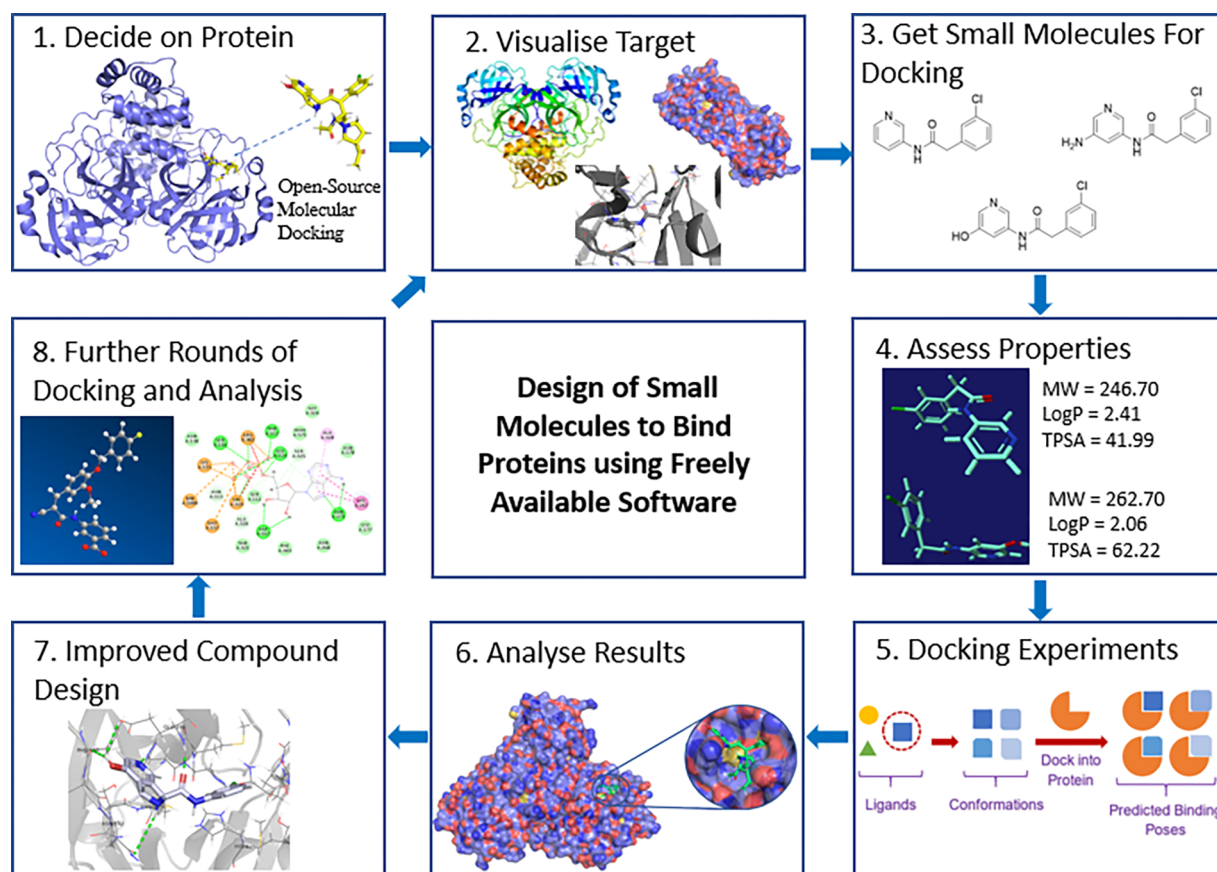
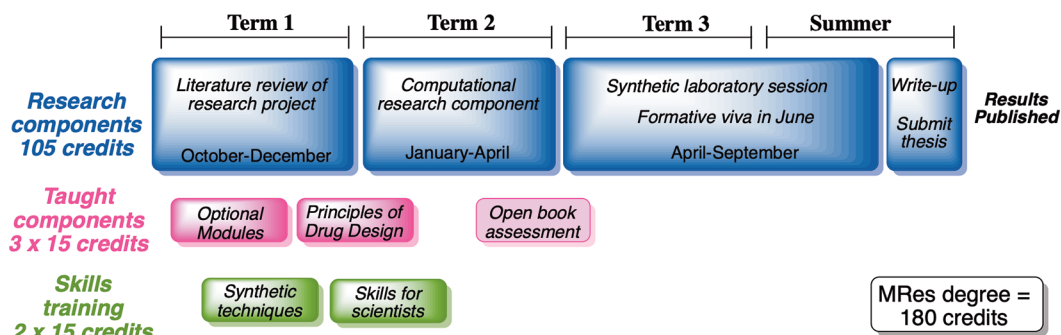**Figure 1.** Workflow for open source computer-aided drug design.

Figure 2. Redesigned course following the Covid-19 pandemic.

Experiments are designed by the students to dock specific lists of compounds into their target protein. The docking is facilitated by using a Jupyter Notebook that can be accessed on the students' own machines through use of a VPN to the UCL Teaching Cluster or via alternative methods (such as GoogleCoLab). The notebook provides an environment suitable for performing docking simulations, including RDKit to generate conformations for docking and SMINA to perform the docking experiments.

Steps 6, 7, and 8: Analysis of Docking Results

Finally, the students are shown how they might go about analyzing the results from their docking experiments. The poses generated from SMINA are assessed through visualization in PyMOL and the docking scores (RFScore) and binding affinities (minimized affinity) calculated by SMINA for each pose are examined in DataWarrior. Students then go on

to perform repeated cycles of design, docking, and analysis to further develop their compounds.

The workflow process is illustrated diagrammatically in Figure 1, which gives an overview of the process related to the sections included in the laboratory handbook.

■ NEW DEGREE STRUCTURE: PEDAGOGICAL RATIONALE

The development of the "laboratory handbook" and the restrictions imposed by the COVID-19 pandemic presented a pedagogical opportunity for us to refine and improve the original design of the course.[46] In particular, the migration of taught material and examinations to online platforms gave us the freedom to organize the course in a way that was relevant to the students while still giving them maximum exposure to a
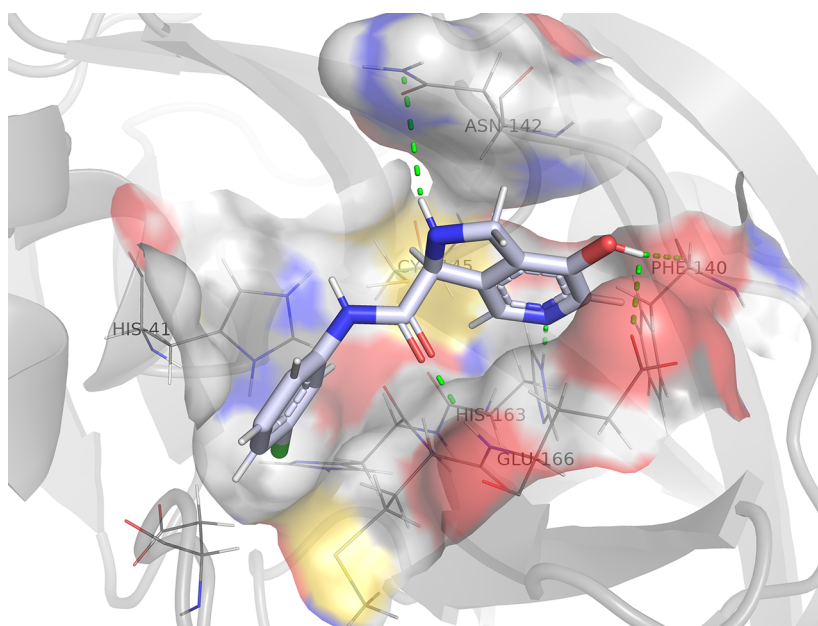
Figure 3. An example of one of the final compounds selected for synthesis from the CADD projects of the 2019−2020 cohort. The compound is the student's own design for a small molecule targeting the SARS-CoV-2 M^{pro}, visualized in PyMOL.

diverse skillset and a holistic overview of the drug discovery process.

The new structure (Figure 2) allows a more streamlined approach. Academic modules are taken in the first term, with the choice of subject (particularly Principles of Drug Design) supporting the ethos of the course and preparing students for later activities. Significantly, students choose a project at the beginning of their course and then use the first term to research the literature background to the field of their interest. This allows students to immerse themselves in the theoretical aspects of their project before beginning any "hands-on" computational techniques. The early choice of project also gives students the incentive to learn the techniques that they would later use.

The incorporation of all (or most) of the academic and theoretical aspects of the course into term 1 allowed us to be creative in the training and research experience that we could offer in terms 2 and 3. We were keen to offer both the vital synthetic chemistry experience that is required in modern pharmaceutical and biotechnology industries as well as an industry-leading training in computational molecular docking. With our new degree structure, we realized that we could offer a rigorous and valuable training in molecular docking techniques in term 2, using the "laboratory handbook" and supported by webinars, discussion fora, and email-based help, as described above. This is then followed by an immersive synthetic laboratory experience in term 3 and through the summer. Crucially, the students would work on the same target throughout their course; writing a literature review on the target in term 1, undertaking molecular docking studies on this target in term 2, and then, for the remainder of the course, preparing molecules in the laboratory that they have designed themselves. We have found that this structure works well because it gives students a sense of value and ownership of their project, it trains them in all of the modern drug discovery techniques, and is an original research area to justify the award of an advanced degree.

OUTCOMES

Following the first iteration of this redeveloped course (May−September 2020) we were delighted with the achievements of the students. Several students generated novel compound structures that had profiles worthy to be taken forward to be synthesized and tested as inhibitors of their chosen target (Figure 3). In terms of the pedagogical outcomes, from student feedback, despite some initial reticence, most students reported having a good experience and to have learned skills that they felt would be valuable for their future careers. Although this is only anecdotal and meaningful data are difficult to collect in such a disrupted year, this was reflected in the degree outcomes, with more than three-quarters of the cohort obtaining the highest (distinction) classification for their degree.

SUMMARY

We envisage that this will be an invaluable resource for students undertaking postgraduate MSc, MRes, or MPharm programs in which there is a substantial element of drug design, and potentially for final-year students on MSci, MChem, or MPharm four-year undergraduate courses. The development of this course has been undertaken specifically to give the students a research experience in which laboratory work was not viable but still has wider relevance to organic chemistry and the development of skills within that discipline. Pedagogically, this has both engaged and enthused the students, and this has been reflected in their degree outcomes. For the former students who are authors on this work (B.A.C., Y.W.) and who worked to prepare the handbook for future cohorts, their efforts have been received well by the current cohort who appreciate the fact that the handbook has been written by students for students. Accordingly, future iterations of the handbook will evolve with subsequent cohorts' feedback. To facilitate this, and to encourage contributions from the broader community, we have made the manual available online and are now converting it to a wiki-based living document for ease of future editing.[47]

More widely, although disruptive at the time, the development of this course was a timely one for the University College London cohorts. For some time focus at UCL has been moving from "teach-and-exam"-based assessment, (particularly at the postgraduate level) toward a more immersive approach informed by research (the Connected Curriculum). The necessity for reinvention in mid-2020 allowed us to effect such changes with a degree of freedom that rarely presents itself. A particular focus of this redevelopment was that all aspects (projects, software packages) should be open-source and therefore the research undertaken should not only be valuable to the students themselves but freely available to the scientific community. Because access to powerful computer servers or expensive software packages is not required, we believe that this laboratory protocol will help to democratize the scientific process in this field, making computational drug design readily accessible to students and researchers in a variety of settings, including less well-funded institutions, and even to final-year students in secondary schools.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.1c00289.

> Degree structure contrasting the previous versions with the new one (PDF, DOCX)
> Docking files in a folder for performing molecular docking experiments (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

Matthew H. Todd − *School of Pharmacy, University College London, London WC1N 1AX, United Kingdom;* ⓞ orcid.org/0000-0001-7096-4751; Email: matthew.todd@ucl.ac.uk

Jonathan D. Wilden − *Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom;* ⓞ orcid.org/0000-0003-4269-2714; Email: j.wilden@ucl.ac.uk

Alethea B. Tabor − *Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom;* ⓞ orcid.org/0000-0001-8216-0347; Email: a.b.tabor@ucl.ac.uk

Authors

Bethanie A. Clent − *Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom*

Yuhang Wang − *School of Pharmacy, University College London, London WC1N 1AX, United Kingdom*

Hugh C. Britton − *Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom*

Frank Otto − *Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom;* ⓞ orcid.org/0000-0003-4009-1974

Christopher J. Swain − *Cambridge MedChem Consulting, Cambridge CB22 4RN, United Kingdom*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jchemed.1c00289

Notes

The authors declare no competing financial interest.

A copy of the first version of the laboratory handbook is available to edit and download via github. https://github.com/UCL/Open_Docking_Lab_Handbook

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Lolli, M.; Narramore, S.; Fishwick, C. W. G.; Pors, K. Refining the Chemical Toolbox to Be Fit for Educational and Practical Purpose for Drug Discovery in the 21st Century. *Drug Discovery Today* **2015**, *20* (8), 1018−1026.

(2) Rafferty, M. F. No Denying It: Medicinal Chemistry Training Is In Big Trouble. *J. Med. Chem.* **2016**, *59* (24), 10859−10864.

(3) McInally, T.; Macdonald, S. J. F. Unusual Undergraduate Training in Medicinal Chemistry in Collaboration between Academia and Industry. *J. Med. Chem.* **2017**, *60* (19), 7958−7964.

(4) Miller, S. M.; Moos, W. H.; Munk, B. H.; Munk, S. A. Graduate and Postgraduate Education at a Crossroads. In *Managing the Drug Discovery Process: How to Make It More Efficient and Cost-Effective*; Woodhead Publishing:: Cambridge, UK, 2017; Chapter 6, pp 103−128.

(5) Price, G. W.; Gould, P. S.; Marsh, A. Use of Freely Available and Open-Source Tools for In Silico Screening in Chemical Biology. *J. Chem. Educ.* **2014**, *91* (4), 602−604.

(6) Rodrigues, R. P.; Andrade, S. F.; Mantoani, S. P.; Eifler-Lima, V. L.; Silva, V. B.; Kawano, D. F. Using Free Computational Resources To Illustrate the Drug Design Process in an Undergraduate Medicinal Chemistry Course. *J. Chem. Educ.* **2015**, *92* (5), 827−835.

(7) Sydow, D.; Morger, A.; Driller, M.; Volkamer, A. TeachOpenCADD: A Teaching Platform for Computer-Aided Drug Design Using Open-Source Packages and Data. *J. Cheminf.* **2019**, *11*, 29.

(8) Sydow, D.; Wichmann, M.; Rodríguez-Guerra, J.; Goldmann, D.; Landrum, G.; Volkamer, A. TeachOpenCADD-KNIME: A Teaching Platform for Computer-Aided Drug Design Using KNIME Workflows. *J. Chem. Inf. Model.* **2019**, *59* (10), 4083−4086.

(9) Shi, X.-X.; Li, J.-Y.; Chen, Q.; Zhu, X.-L.; Hao, G.-F.; Yang, G.-F. Development of a Web-Based Laboratory Class to Reduce the Challenges in Teaching Fragment-Based Drug Design *J. J. Chem. Educ.* **2020**, *97* (2), 427−436.

(10) Acuna, V. V.; Hopper, R. M.; Yoder, R. J. Computer-Aided Drug Design for the Organic Chemistry Laboratory Using Accessible Molecular Modelling Tools. *J. Chem. Educ.* **2020**, *97* (3), 760−763.

(11) Helgren, T. R.; Hagen, T. J. Demonstration of AutoDock as an Educational Tool for Drug Discovery. *J. Chem. Educ.* **2017**, *94* (3), 345−349.

(12) Kholod, Y.; Hoag, E.; Muratore, K.; Kosenkov, D. Computer-Aided Drug Discovery: Molecular Docking of Diminazene Ligands to DNA Minor Groove. *J. Chem. Educ.* **2018**, *95* (5), 882−227.

(13) Tantillo, D. J.; Siegel, J. B.; Saunders, C. M.; Palazzo, T. A.; Painter, P. P.; O'Brien, T. E.; Nuñez, N. N.; Nouri, D. H.; Lodewyk, M. W.; Hudson, B. M.; Hare, S. R.; Davis, R. L. Computer-Aided Drug Design for Undergraduates. *J. Chem. Educ.* **2019**, *96* (5), 920−925.

(14) Fung, D. *A Connected Curriculum for Higher Education*; UCL Press: London, UK, 2017.

(15) *Shaping Higher Education with Students: Ways to Connect Research and Teaching,* Tong, V. C. H.; Standen, A.; Sotiriou, M., Eds.; UCL Press:: London, UK, 2018.

(16) *Teaching and Learning in Higher Education: Perspectives from UCL*; Davies, J. P., Pachler, N., Eds.; UCL Institute of Education Press:: London, UK, 2018.

(17) Welcome to CoLaboratory. https://colab.research.google.com/notebooks/intro.ipynb (accessed 2021-05-20).

(18) Project Jupyter home. https://jupyter.org (accessed 2021-05-20).

(19) Woelfle, M.; Olliaro, O.; Todd, M. H. Open Science is a Research Accelerator. *Nat. Chem.* **2011**, *3*, 745−748.

(20) Todd, M. H. Six Laws of Open Source Drug Discovery. *ChemMedChem* **2019**, *14*, 1804−1809.

(21) Williamson, A. E.; Ylioja, P. M.; Robertson, M. N.; Antonova-Koch, Y.; Avery, V.; Baell, J. B.; Batchu, H.; Batra, S.; Burrows, J. N.; Bhattacharyya, S.; Calderon, F.; Charman, S. A.; Clark, J.; Crespo, B.; Dean, M.; Debbert, S. L.; Delves, M.; Dennis, A. S. M.; Deroose, F.; Duffy, S.; Fletcher, S.; Giaever, G.; Hallyburton, I.; Gamo, F.-J.; Gebbia, M.; Guy, R. K.; Hungerford, Z.; Kirk, K.; Lafuente-Monasterio, M. J.; Lee, A.; Meister, S.; Nislow, C.; Overington, J. P.; Papadatos, G.; Patiny, L.; Pham, J.; Ralph, S. A.; Ruecker, A.; Ryan, E.; Southan, C.; Srivastava, K.; Swain, C.; Tarnowski, M. J.; Thomson, P.; Turner, P.; Wallace, I. M.; Wells, T. N. C.; White, K.; White, L.; Willis, P.; Winzeler, E. A.; Wittlin, S.; Todd, M. H. Open Source Drug Discovery: Highly Potent Antimalarial Compounds Derived from the Tres Cantos Arylpyrroles. *ACS Cent. Sci.* **2016**, *2*, 687−701.

(22) OpenSourceMalaria. https://github.com/OpenSourceMalaria (accessed 2021-05-20).

(23) OpenSourceAntibiotics. https://github.com/opensourceantibiotics (accessed 2021-05-20).

(24) Tse, E. G.; Klug, D. M.; Todd, M. H. Open Science Approaches to COVID-19. *F1000Research* **2020**, *9*, 1043.

(25) Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.; Owen, C. D.; Resnick, E.; Strain-Damerell, C.; Aimon, A.; Ábrányi-Balogh, P.; Brandão-Neto, J.; Carbery, A.; Davison, G.; Dias, A.; Downes, T. D.; Dunnett, L.; Fairhead, M.; Firth, J. D.; Jones, S. P.; Keeley, A.; Keserü, G. M.; Klein, H. F.; Martin, M. P.; Noble, M. E. M.; O'Brien, P.; Powell, A.; Reddi, R. N.; Skyner, R.; Snee, M.; Waring, M. J.; Wild, C.; London, N.; von Delft, F.; Walsh, M. A. Crystallographic and Electrophilic Fragment Screening of the SARS-CoV-2 Main Protease. *Nat. Commun.* **2020**, *11*, 5047.

(26) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of M^{pro} from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, *582*, 289−293.

(27) Main Protease Structure and XChem Fragment Screen. https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem.html (accessed 2021-05-20).

(28) Covid Moonshot—Taking Fragments to Impact. https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem/COVID-MoonShot.html (accessed 2021-05-20).

(29) PostEra: COVID Moonshot. https://covid.postera.ai/covid (accessed 2021-05-20).

(30) DataWarrior home page. http://www.openmolecules.org/datawarrior/index.html (accessed 2021-05-20).

(31) DataWarrior workshop by Isabelle Giraud. https://youtu.be/Is2hLqqSFvM (accessed 2021-05-20).

(32) PyMol by Schrödinger. https://pymol.org/2/ (accessed 2021-05-20).

(33) PyMOL workshop by Garrett Morris. https://youtu.be/qOxS2wqajdg (accessed 2021-05-20).

(34) PyMol has two versions: Incentive (requires a license, both academic and commercial versions are suitable) and open-source (freely available on GitHub).

(35) Smina. https://sourceforge.net/projects/smina/ (accessed 2021-05-20).

(36) RDKit: Open-Source Cheminformatics Software. https://www.rdkit.org (accessed 2021-05-20).

(37) Open Babel: The Open Source Chemistry Toolbox. http://openbabel.org/wiki/Main_Page (accessed 2021-05-20).

(38) py3Dmol 0.9.1. https://pypi.org/project/py3Dmol/ (accessed 2021-05-20).

(39) Slack homepage. https://slack.com (accessed 2021-05-20).

(40) ZINC15. http://zinc.docking.org (accessed 2021-05-20).

(41) Sterling, T.; Irwin, J. J. ZINC 15-Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324−2337.

(42) Enamine homepage. https://enamine.net/ (accessed 2021-05-20).

(43) Explore Chemistry. https://pubchem.ncbi.nlm.nih.gov/ (accessed 2021-05-20).

(44) ChEMBL. https://www.ebi.ac.uk/chembl/ (accessed 2021-05-20).

(45) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* **2014**, *42*, D1083−D1090.

(46) There has been a long tradition of medicinal chemistry research and teaching at UCL with its pioneering undergraduate degree in medicinal chemistry being established in 1974. In 2012, UCL Chemistry established a new Master's level (level 7) course in "Organic Chemistry-Drug Discovery" to meet the growing needs of the pharmaceutical and biotechnology industry for well-qualified graduates and to respond to student desire to have a one-year research experience prior to committing to a three- or four-year Ph.D. program. The original structure of the degree and other relevant information is detailed in the Supporting Information.

(47) UCL/Open_Docking_Lab_Handbook. https://github.com/UCL/Open_Docking_Lab_Handbook (accessed 2021-05-20).