Teaching Virtual Protein-Centric CUREs and UREs Using Computational Tools

Anthony Bell[1], Laura Christian[2], David Hecht[3], Kathryn Huisinga[4], John Rakus[5] & Ellis Bell[1]

1, University of San Diego, 2, RPI, 3, SWCC, 4, Malone University, 5, Marshall University

Responding to the need to teach remotely due to COVID-19, we used readily available computational approaches (and developed associated tutorials (https://mdh-cures-community.squarespace.com/virtual-cures-and-ures)) to teach virtual Course-Based Undergraduate Research Experience (CURE) laboratories that fulfil generally accepted main components of CUREs or Undergraduate Research Experiences (UREs): Scientific Background, Hypothesis Development, Proposal, Experiments, Teamwork, Data Analysis, Conclusions, and Presentation[1]. We then developed and taught remotely, in three phases, protein-centric CURE activities that are adaptable to virtually any protein, emphasizing contributions of noncovalent interactions to structure, binding and catalysis (an ASBMB learning framework[2] foundational concept).

The courses had five learning goals (unchanged in the virtual format),focused on i) use of primary literature and bioinformatics, ii) the roles of non-covalent interactions, iii) keeping accurate laboratory notebooks, iv) hypothesis development and research proposal writing, and, v) presenting the project and drawing evidence based conclusions

The first phase, Developing a Research Proposal, contains three modules, and develops hallmarks of a good student-developed hypothesis using available literature (PubMed[3]) and preliminary observations obtained using bioinformatics, Module 1**:** Using Primary Literature and Data Bases (Protein Data Base[4], Blast[5] and Clustal Omega[6]), Module 2: Molecular Visualization (PyMol[7] and Chimera[8]), culminating in a research proposal (Module 3). Provided rubrics guide student expectations. In the second phase, Preparing the Proteins, students prepared necessary proteins and mutants using Module 4: Creating and Validating Models, which leads users through creating mutants with PyMol, homology modeling with Phyre2[9] or Missense[10], energy minimization using RefineD[11] or ModRefiner[12], and structure validation using MolProbity[13].

In the third phase, Computational Experimental Approaches to Explore the Questions developed from the Hypothesis, students selected appropriate tools to perform their experiments, chosen from computational techniques suitable for a CURE laboratory class taught remotely. Questions, paired with computational approaches were selected from Modules 5: Exploring Titratable Groups in a Protein using H++[14], 6: Exploring Small Molecule Ligand Binding (with SwissDock[15]), 7: Exploring Protein-Protein Interaction (with HawkDock[16]), 8: Detecting and Exploring Potential Binding Sites on a Protein (with POCASA[17] and SwissDock), and 9: Structure-Activity Relationships of Ligand Binding & Drug Design (with SwissDock, Open Eye[18] or the Molecular Operating Environment (MOE)[19]).

All involve freely available computational approaches on publicly accessible web-based servers around the world (with the exception of MOE). Original literature/Journal club activities on approaches helped students suggest tie-ins to wet lab experiments they could conduct in the future to complement their computational approaches.

This approach allowed us to continue using high impact CURE teaching, without changing our course learning goals. Quantitative data (including replicates) was collected and analyzed during regular class periods. Students developed evidence-based conclusions and related them to their research questions and hypotheses. Projects culminated in a presentation where faculty feedback was facilitated with the Virtual Presentation platform from QUBES[20]

These computational approaches are readily adaptable for topics accessible for first to senior year classes and individual research projects (UREs). We used them in both partial and full semester CUREs in various institutional settings. We believe this format can benefit faculty and students from a wide variety of teaching institutions under conditions where remote teaching is necessary.

This work is supported by NSF-1726932 EHR-IUSE, Principal Investigator: Ellis Bell
The authors declare no conflicts of interest

References.

1. Provost, J., Bell, JK., and Bell, E, "Development and Use of CUREs in Biochemistry" Biochemistry Education: From Theory to Practice, (2019) Chapter 7, 143-171. DOI: 10.1021/bk-2019-1337.ch007
2. https://www.coursesource.org/courses/biochemistry-and-molecular-biology
3. PubMed: https://www.ncbi.nlm.nih.gov/pubmed/
4. Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, Philip E. Bourne, The Protein Data Bank, *Nucleic Acids Research*, Volume 28, Issue 1, 1 January 2000, Pages 235–242, https://doi.org/10.1093/nar/28.1.235
5. *Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ "*Gapped BLAST and PSI-BLAST: a new generation of protein database search programs".*Nucleic Acids Res. 1997 Sep 1; 25(17):3389-402.*
6. Madeira, F., Lee, J., Buso, N., Gur, T., Madhusoodnan, N., Basutkar, P., Tivey, ARN., Potter, SC., Finn, RD., Lopez, R "The EMBL-EBI search and sequence analysis tools APIs in 2019". Nucleic Acids Research, 30 Jun 2019, 47(W1):W636-W641
7. The PyMOL Molecular Graphics System, Version 2.0 Schrodinger LLC, https://pymol.org/edu/?q=educational/
8. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE."UCSF Chimera--a visualization system for exploratory research and analysis." *J Comput Chem.* 2004 Oct;25(13):1605-12.

7. Kelley,L.A., Mezulis,S., Yates,, C.M., Wass M. & Sternberg, M.J.E. “The Phyre2 web portal for protein modeling, prediction and analysis” *Nature Protocols* volume 10, pages845–858(2015) (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index)
8. Ittisoponpisan, S., Islam, S.A., Khanna, T., Alhuzimi, E., David, A., & Sternberg, M.J.E (2109) “Can predicted protein 3D structures provide reliable insights into whether Missense variants are disease associated?” J.Mol Biol 431, 2197-2212 (http://www.sbg.bio.ic.ac.uk/~missense3d/)
9. Bhattacharya D “refineD: improved protein structure refinement using machine learning based restrained relaxation”.Bioinformatics. 2019 Sep 15;35(18):3320-3328
10. Dong Xu and Yang Zhang. Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization. Biophysical Journal, vol 101, 2525-2534 (2011) (https://zhanglab.ccmb.med.umich.edu/ModRefiner/)
11. Williams et al. (2018) MolProbity: More and better reference data for improved all-atom structure validation. Protein Science 27: 293-315 (http://molprobity.biochem.duke.edu/)
12. Ramu Anandakrishnan, Boris Aguilar and Alexey V. Onufriev, "H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulation", Nucleic Acids Res., 40(W1):W537-541. (2012) (http://biophysics.cs.vt.edu/)
13. Aurélien Grosdidier,[1] Vincent Zoete,[1,*] and Olivier Michielin “SwissDock, a protein-small molecule docking web service based on EADock DSS” Nucleic Acids Res. 2011 Jul 1; 39(Web Server issue): W270–W277 (http://www.swissdock.ch/)
14. Gaoqi Weng, Ercheng Wang, Zhe Wang, Hui Liu, Feng Zhu, Dan Li, Tingjun Hou “HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA” *Nucleic Acids Research*, Volume 47, Issue W1, 02 July 2019, Pages W322–W330 (http://cadd.zju.edu.cn/hawkdock/)
15. Yu J, Zhou Y, Tanaka I, Yao M “Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere”. Bioinformatics. 2010 Jan 1;26(1):46-52 (http://g6altair.sci.hokudai.ac.jp/g6/service/pocasa/)
16. Open Eye, https://www.eyesopen.com/orion
17. MOE: Molecular Operating Environment: https://www.chemcomp.com/
18. Donovan, S., Diaz-Eaton, C., Gower, S., Jenkins, K, LaMar, M., Poli, D., Sheehy, R., and Wojdak, J. (2015). “QUBES: a community focused on supporting teaching and learning in quantitative biology”. *Letters in Biomathematics*. 2(1), 46-55. doi.org/10.1080/23737867.2015.1049969. QUBES: https://qubeshub.org/