SHORT COMMUNICATION

Computational Resources for Bioscience Education

Rajiv K. Kar [1]

Abstract

With the ongoing laboratory restrictions, it is often challenging for bioscience students to make satisfactory progress in their projects. A long-standing practice in multi-disciplinary research is to use computational and theoretical method to corroborate with experiment findings. In line with the lack of opportunity to access laboratory instruments, the pandemic situation is a win-win scenario for scholars to focus on computational methods. This communication outline some of the standalone tools and webservers that bioscience students can successfully learn and adopt to obtain in-depth insights into biochemistry, biophysics, biotechnology, and bioengineering research work.

Keywords Bioscience · Structural modeling · Visualizers · Classical simulations · Quantum mechanics

Introduction

The worldwide impact of COVID-19 has raised several challenges for research scholars [1]. The common problems include restricted access to lab workspace, delayed transportation of materials, and reduced technical assistance for experimental troubleshoot [2]. Though the online platforms have rescued to a certain level with one-on-one discussion, educational sessions, and virtual conferences, the problem still persists among the graduate students and forming the cloud of worries—"how to make progress in the research project?" (Fig. 1). While the laboratory experiments are not at full speed, it is still possible to use computational power. This communication aims to highlight a few useful computational tools that can help in making research progress, amid restricted laboratory access.

✉ Rajiv K. Kar
rajiv.kar@mail.huji.ac.il; rajivkarpharma@gmail.com

1 Faculty II-Mathematics and Natural Sciences, Technische Universität Berlin, Sekr. PC 14, Strasse des 17. Juni 135, D-10623 Berlin, Germany
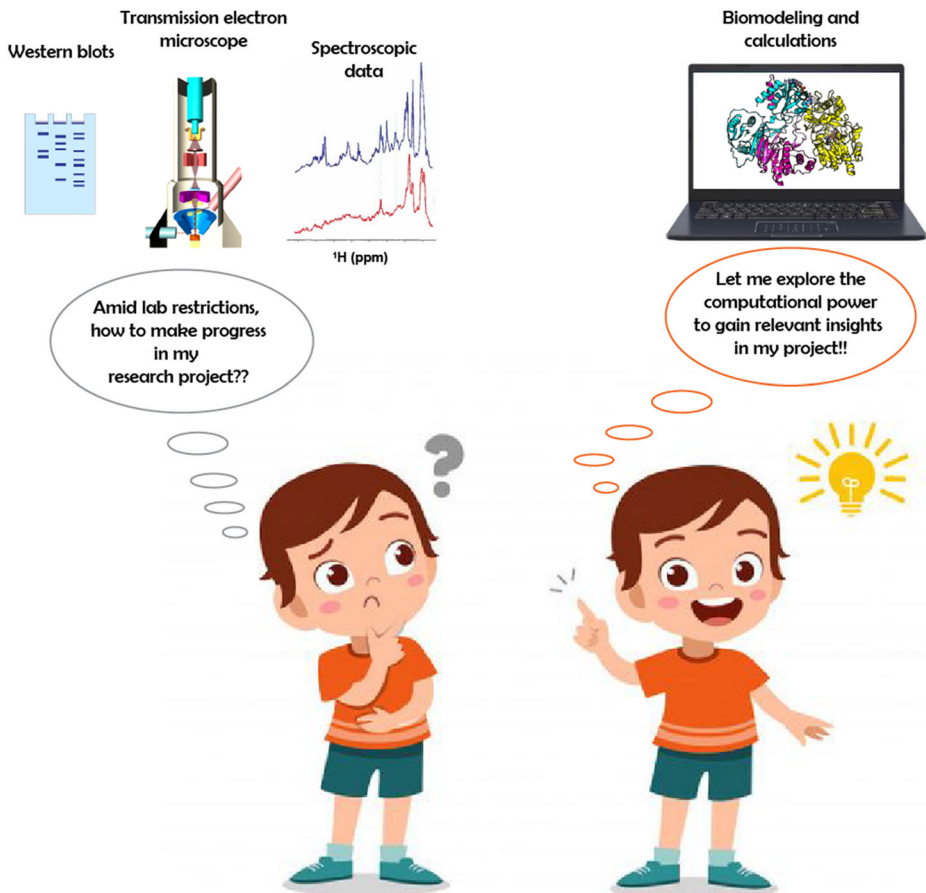
Fig. 1 Illustration suggesting that exploring computational powers could be valuable for scholars to make progress in bioscience projects

Structural Modeling

Most of the bioscience problems deal with biological macromolecules such as proteins, nucleic acids, and small molecules. Though the structural coordinates are available from the protein data bank (PDB) for some structures, there exists a large sequence-structure gap for many biomacromolecules [3]. Homology modeling tools are a valuable starting point to analyze the putative three-dimensional conformation, interacting residues, and active site arrangement in these structures. The available standalone tools include MODELLER [4], while webservers are Phyre2 [5], ROBETTA [6], and SWISS MODEL [7]. The structural validation of such models for the interface analysis, surface assemblies, and Ramachandran plot can be determined using PDBePISA server [8]. The solvent accessibilities can also be analyzed using Naccess standalone program [9]. The contact map details can be analyzed using ConPlot [10] and DISTEVAL [11]. Relevant details of normal mode for predicting collective protein domain motions is achievable using iMODS [12], ElNémo [13], and WEBnm@ [14]. To determine the protein stability, especially for projects with protein mutants, one can use servers like SDM [15], MAESTROweb [16], PoPMuSic [17], DUET [18], and pPerturb [19]. These algorithms

are accountable for predicting thermodynamics (free energy) properties based on input structural coordinates in a PDB format.

Bio-macromolecular Complex

The macromolecular interactions are an important area that helps in better understanding protein functioning, mechanism, and drug designing. The widely used server ClusPro [20] is one of such tools that predict protein-interface based on the energetic evaluation. Hex program offers interactive fast Fourier transform-based docking [21]. Several useful options are also available on HADDOCK [22] webserver that can help in blind or experimental constraint-guided protein docking. While protein-protein docking is more complicated due to the presence of several degrees of freedom, the algorithm for small molecule docking is relatively simpler. The widely used AutoDock [23] is suitable for such exercise, with AutoDock Vina [24] offering high-throughput screening. Notably, these tools must be benchmarked for the system under consideration, such as by docking a known complex and comparing the root-mean-squared deviation (docked vs original complex). Additionally, the docking outcomes can be benchmarked against a reliable set of decoy molecular sets, available from DUD-E [25] and DEKOIS 2.0 [26] database. Schematic diagrams for protein-ligand contacts can be plotted using Ligplot [27]. More advanced tools such as for pKa prediction include Protein-Sol server [28], DelPhiPKa [29], and PDB2PQR [30]. However, these tools come with limitation to process non-standard residues such as small organic and drug-like molecules.

Visualizers

The visualization tools add fun while depicting the spatial arrangement of amino acids and nucleobases, secondary structures, and water molecules in a three-dimensional network. Nevertheless, they can render publication-quality images (Fig. 2). One can also perform structural modeling with visualizers like PyMol [31], VMD [32], and Chimera [33].

Classical Simulations

The collaboration between classical simulation with biophysics is more common these days. Some of the relevant free tools to perform molecular simulation include
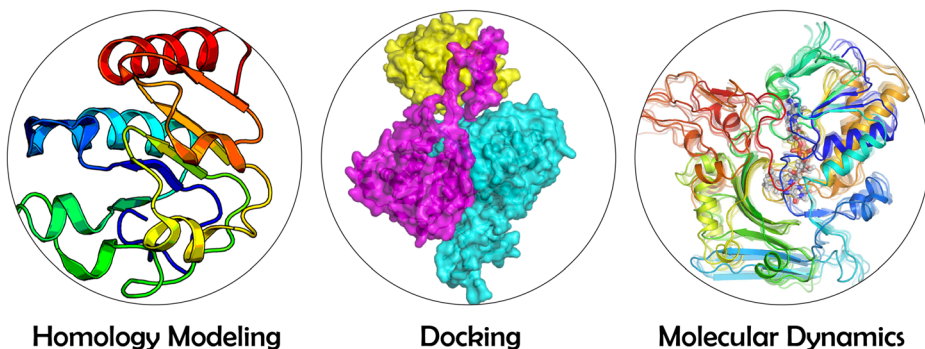

Homology Modeling   **Docking**   **Molecular Dynamics**

Fig. 2 Representation of static structures from homology modeling and protein-protein docking, and dynamic structures from molecular dynamics simulation

GROMACS [34] and Desmond [35]. However, they require advanced hardware to compute physics-based time-dependent motions of biomolecules. A Perl-based toolset for structure preparation and analysis is available with MMTSB [36]. While CHARMM-GUI server [37] provides numerous options to prepare solvated structures, membrane bilayer, and coarse-grain systems, one can also use this server to prepare input complement to other simulation engines. Equilibrated trajectories for a short time-scale are possible to be calculated using MDWeb server [38] and ChemCompute server [39]. Furthermore, scholars having a basic knowledge of python can make use of MDAnalysis [40] to extract information. Both VMD and Chimera offer capabilities to visualize, and varieties of plugins to analyze the simulation trajectory.

Quantum Chemical Calculations

Calculations based on quantum mechanics (QM) are more accurate and reliable compared to the classical simulations; however, they come with their own set of limitations. Accurate modeling of protein-ligand interaction and elucidation of spectroscopic properties are some valuable insights that might be crucial to investigate biochemical and biophysical problems. Notably, rigorous quantum mechanics methods are mandated to achieve their transferability and robustness in connection to experimental corroboration. Orca [41] and GAMESS [42] are open-source quantum chemistry package that offer a wide range of capabilities, including geometry optimization, calculation of UV/Vis excitation energy, CD spectra, and vibrational frequencies. Avogadro [43] is user-friendly freeware that not only enables chemical editing but also to visualize frontier molecular orbitals and vibrational modes. Psi4 open-source package [44] is useful in analyzing the wavefunctions and Hartree-Fock energy decomposition. While these programs are accountable for high-computing time, one can run calculations with a smaller basis set on desktop and laptop with Windows or Linux operating system. Thankfully, it is also possible to perform QM calculations with a high basis set using the webserver ChemCompute [39], which offers computing time on cluster nodes for registered users with an academic electronic address.

Conclusion

In summary, this communication highlights various (not limited to) computational tools relevant to obtain atomistic-scale analysis of biomacromolecules. While having restricted access to laboratory space and equipment, learning computational methods could mark a substantial advancement. Most of the standalone tool and web servers listed are freely accessible, with academic registration. Nevertheless, before using these computational tools and corroborating with the experimental measurements, it is essential to understand their theoretical principles.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.s

Consent for Publication Not applicable.

Competing Interests The author declares no competing interests.

References

1. Schiermeier, B. Q., Else, H., Mega, E. R., Padma, T. V., & Gaind, N. (2020). What it's really like to do science. *Nature, 586*(7830), 486–487.
2. Myers, K. R., Tham, W. Y., Yin, Y., Cohodes, N., Thursby, J. G., Thursby, M. C., Schiffer, P., Walsh, J. T., Lakhani, K. R., & Wang, D. (2020). Unequal effects of the COVID-19 pandemic on. *Nature Human Behaviour, 4*(9), 880–883. https://doi.org/10.1038/s41562-020-0921-y.
3. Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., Christie, C. H., Dalenberg, K., di Costanzo, L., Duarte, J. M., Dutta, S., Feng, Z., Ganesan, S., Goodsell, D. S., Ghosh, S., Green, R. K., Guranović, V., Guzenko, D., Hudson, B. P., Lawson, C. L., Liang, Y., Lowe, R., Namkoong, H., Peisach, E., Persikova, I., Randle, C., Rose, A., Rose, Y., Sali, A., Segura, J., Sekharan, M., Shao, C., Tao, Y. P., Voigt, M., Westbrook, J. D., Young, J. Y., Zardecki, C., & Zhuravleva, M. (2021). RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research, 49*(November 2020), 437–451. https://doi.org/10.1093/nar/gkaa1038.
4. Webb, B., & Sali, A. (2014). Protein structure modeling with MODELLER. *Methods in Molecular Biology, 1137*(1137), 1–15. https://doi.org/10.1007/978-1-4939-0366-5.
5. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling , prediction and analysis. *Nature Protocols, 10*(6), 845–858. https://doi.org/10.1038/nprot.2015-053.
6. Yang, J., Anishchenko, I., Park, H., Peng, Z., & Ovchinnikov, S. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences of the United States of America, 117*(3), 1496–1503. https://doi.org/10.1073/pnas.1914677117.
7. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research, 46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427.
8. Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology, 372*(3), 774–797. https://doi.org/10.1016/j.jmb.2007.05.022.
9. Hubbard, S. J., & Thornton, J. M. (1993). "NACCESS", Computer Program.
10. Sa, F., Mesdaghi, S., Simpkin, A. J., Burgos-ma, J. J., Murphy, D. L., Uski, V., … Rigden, D. J. (2021). Structural bioinformatics ConPlot: Web-based application for the visualization of protein contact maps integrated with other data. *Bioinformatics*, (January), 1–3. https://doi.org/10.1093/bioinformatics/btab049

11. Adhikari, B., Shrestha, B., Bernardini, M., Hou, J., & Lea, J. (2021). DISTEVAL: A web server for evaluating predicted protein distances. *BMC Bioinformatics, 22*(8), 1–9. https://doi.org/10.1186/s12859-020-03938-z.

12. Aliaga, I., Quintana-ort, E. S., & Chac, P. (2014). iMODS: Internal coordinates normal mode analysis server. *Nucleic Acids Research, 42*(April), 271–276. https://doi.org/10.1093/nar/gku339.

13. Suhre, K., & Sanejouand, Y. (2004). ElNemo: A normal mode web server for protein ElNe movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Research, 32*, 610–614. https://doi.org/10.1093/nar/gkh368.

14. Tiwari, S. P., Fuglebakk, E., Hollup, S. M., Skjærven, L., Cragnolini, T., Grindhaug, S. H., Tekle, K. M., & Reuter, N. (2015). WEBnm @ v2.0: Web server and services for comparing protein flexibility. *BMC, 15*(427), 1–12. https://doi.org/10.1186/s12859-014-0427-6.

15. Ascher, D. B., Pandurangan, A. P., Ochoa-monta, B., & Blundell, L. (2017). SDM: A server for predicting effects of mutations on protein stability. *Nucleic Acids Research, 45*(May), 229–235. https://doi.org/10.1093/nar/gkx439.

16. Laimer, J., Hiebl-flach, J., Lengauer, D., & Lackner, P. (2016). MAESTROweb: A web server for structure-based protein stability prediction. *Bioinformatics, 32*(9), 11414–11416.

17. Dehouck, Y., Kwasigroch, J. M., Gilis, D., & Rooman, M. (2011). PoPMuSiC 2.1: A web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics, 12*(151), 1–12.

18. Pires, D. E. V., Ascher, D. B., & Blundell, T. L. (2014). DUET: A server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Research, 42*(May), 314–319. https://doi.org/10.1093/nar/gku411.

19. Gopi, S., Devanshu, D., Rajasekaran, N., Anantakrishnan, S., & Naganathan, A. N. (2020). pPerturb: A server for predicting long-distance energetic couplings and mutation-induced stability changes in proteins via perturbations. *ACS Omega, 5*(2), 1142–1146. https://doi.org/10.1021/acsomega.9b03371.

20. Desta, I. T., Porter, K. A., Xia, B., Kozakov, D., Vajda, S., Desta, I. T., et al. (2020). Resource performance and its limits in rigid body protein- protein docking ll. *Structure, 28*(9), 1071–1081.e3. https://doi.org/10.1016/j.str.2020.06.006.

21. Ritchie, D. W. (2013). Protein docking using case-based reasoning. *Proteins: Structure, Function, and Bioinformatics, 81*(12), 2150–2158. https://doi.org/10.1002/prot.24433.

22. Van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., & Schmitz, C. (2016). The HADDOCK2.2 Web Server: User-friendly integrative modeling of biomolecular complexes. *Journal of Molecular Biology, 428*(4), 720–725. https://doi.org/10.1016/j.jmb.2015.09.014.

23. Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., & Olson, A. J. (2016). Computational protein – ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols, 11*(5), 905–919. https://doi.org/10.1038/pj.2016.37.

24. Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry, 31*(2), 455–461. https://doi.org/10.1002/jcc.

25. Mysinger, M. M., Carchia, M., Irwin, J. J., & Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *Journal of Medicinal Chemistry, 55*(14), 6582–6594.

26. Boeckler, F. M., Bauer, M. R., Ibrahim, T. M., & Vogel, S. M. (2014). Use of DEKOIS 2.0 to gain insights for virtual screening. *Journal of Cheminformatics, 6*(Suppl 1), 2014. https://doi.org/10.1186/1758-2946-6-S1-O24.

27. Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions Clean up structure. *Protein Engineering, 8*(2), 127–134.

28. Hebditch, M., Carballo-amador, M. A., Charonis, S., Curtis, R., & Warwicker, J. (2017). Sequence analysis Protein – Sol: A web tool for predicting protein solubility from sequence. *Bioinformatics, 33*(19), 3098–3100. https://doi.org/10.1093/bioinformatics/btx345.

29. Pahari, S., Sun, L., Basu, S., & Alexov, E. (2018). DelPhiPKa: Including salt in the calculations and enabling polar residues to titrate. *Proteins: Structure, Function, and Bioinformatics, 86*(12), 1277–1283. https://doi.org/10.1002/prot.25608.

30. Unni, S., Huang, Y., Hanson, R. M., Tobias, M., Krishnan, S., Li, W. W., et al. (2011). Web servers and services for electrostatics calculations with APBS and PDB2PQR. *Journal of Computational Chemistry, 32*, 5–8. https://doi.org/10.1002/jcc.

31. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. (n.d.).

32. Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics, 14*(1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

33. Huang, C. C., Meng, E. C., Morris, J. H., Pettersen, E. F., & Ferrin, T. E. (2014). Enhancing UCSF Chimera through web services. *Nucleic Acids Research, 42*(May), 478–484. https://doi.org/10.1093/nar/gku377.

34. James, M., Murtola, T., Schulz, R., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS : High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX, 2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

35. Gregersen, B. A., Klepeis, J. L., Kolossvary, I., & Mark, A. (2006). Scalable algorithms for molecular dynamics simulations on commodity clusters. In *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida , 2006,*.

36. Feig, M., Karanicolas, J., & Iii, C. L. B. (2004). MMTSB Tool Set: Enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics and Modelling, 22*(5), 377–395. https://doi.org/10.1016/j.jmgm.2003.12.005.

37. Qi, Y., Lee, J., Singharoy, A., Mcgreevy, R., Schulten, K., & Im, W. (2017). CHARMM-GUI MDFF/ xMDFF utilizer for molecular dynamics flexible fitting simulations in various environments. *Journal of Physical Chemistry B, 121*(15), 3718–3723. https://doi.org/10.1021/acs.jpcb.6b10568.

38. Hospital, A., Andrio, P., Fenollosa, C., Cicin-sain, D., Orozco, M., & Gelpí, J. L. (2012). MDWeb and MDMoby : An integrated web-based platform for molecular dynamics simulations. *Bioinformatics, 28*(9), 1278–1279. https://doi.org/10.1093/bioinformatics/bts139.

39. Perri, M. J., & Weber, S. H. (2014). Web-based job submission interface for the GAMESS computational chemistry program. *Journal of Chemical Education, 91*(12), 2206–2208.

40. Michaud-agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis : A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry, 32*, 2319–2327. https:// doi.org/10.1002/jcc.

41. Neese, F. (2018). Software update: The ORCA program system, version 4.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 8*(1), 4–9. https://doi.org/10.1002/wcms.1327.

42. Phys, J. C., Barca, G. M. J., Bertoni, C., Carrington, L., Fedorov, D. G., Gour, J. R., et al. (2020). Recent developments in the general atomic and molecular electronic structure system Recent developments in the general atomic and molecular electronic structure system. *Journal of Chemical Physics, 154102*(February), 154102. https://doi.org/10.1063/5.0005188.

43. 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), 1–17.

44. Phys, J. C., Smith, D. G. A., Galvelis, R., James, M., Schriber, B., Burns, L. A., et al. (2020). Throughput quantum chemistry P SI4 1.4: Open-source software for high-throughput quantum chemistry. *Journal of Chemical Physics, 184108*(February), 184108. https://doi.org/10.1063/5.0006002.