For reprint orders, please contact: reprints@future-science.com

Navigating the intricacies of molecular docking

Stephanie S Schweiker^{*,1} & Stephan M Levonis¹

¹Medicinal Chemistry Group, Faculty of Health Sciences & Medicine, Bond University, Robina, QLD 4229, Australia

*Author for correspondence: sschweik@bond.edu.au

“we outline some available software and how it might be used practically as a solution for computational drug design.”

First draft submitted: 9 December 2019; Accepted for publication: 11 December 2019; Published online: 6 February 2020

Keywords: AutoDock • computational chemistry • drug design • molecular docking

Computational molecular modeling refers to a collection of techniques used to explore and establish *in silico* models for the understanding of molecular interactions present in systems. These systems include those within the biological environment, for which *in silico* techniques have been in use since the 1980s [1,2]. Presently, computational chemistry techniques are at a level of unprecedented availability, with computational power and software availability making computational studies more feasible than ever before. In this way, computational chemistry has become a mainstay in the academic process of molecular design, and particularly in drug design as used in medicinal chemistry. In this editorial, we will focus on common software packages and techniques that can be performed on typical desktop computer systems available to any academic researcher for use.

Molecular docking

Molecular docking describes the process of producing a 3D computer model of a potential ligand and attempting to dock it to a protein macromolecule of interest using calculated interactions. Currently, this process is possible using the software freely available for academic use and can be performed via graphic user interfaces without requiring typed commands or experience with computer programming languages. The process uses algorithms to search the binding site and ‘fit’ the ligand into the defined 3D space. Depending on the algorithm, scoring functions are used to compare the different ligands and in combination with their docked pose can give valuable insight into prospective compound design. Practically, before molecular docking can be performed, a computer file of the protein of interest is required, as well as prepared files of the ligand/s to be evaluated.

Target site & ligand preparation

X-ray crystallographic structures are currently available on the protein data bank (PDB; www.rcsb.org) [3], which is a global repository of biological entities. If no x-ray structure is available for the target, then one can be produced using homology modeling. A homology model will create the structure based on the closest sequence available (see how to guide in [4]). X-ray structures are a subjective interpretation of the electron density within the static molecule. Most molecules are dynamic in nature which represents one of the key limitations and criticisms of molecular docking. Even so, some programs are demonstrated to give accurate predictions of binding preference most of the time with the receptor rigid and only the ligand made flexible [5]. Some structures in the PDB are determined from the solution state, while others are from the crystalline form. X-ray structures with high resolution are available, and the correct chemical structure can be confirmed by examining known flexible regions and cross-checking with other x-ray structures. Amino acid sidechains, such as glutamine, asparagine and histidine often require modification. This process is possible via a graphical user interface within software free for academic use, including Chimera and AutoDockTools [6,7]. Once the structure and identity of sidechains are correct, the protonation of the sidechains may be considered.

newlands
press

Acidic and basic sidechains of amino acids may need to be modified, as their protonation will vary depending on the pH. Typically physiological pH is considered, and this will correspond to aspartic acid (Asp, D, pKa 3.90, deprotonated [negative charge]), glutamic acid (Glu, E, pKa 4.07, deprotonated [negative charge]), histidine (His, H, pKa 6.04, deprotonated [neutral charge]), lysine (Lys, K, pKa 10.79, protonated [positive charge]) and arginine (Arg, R, pKa 12.48, protonated [positive charge]). Before molecular docking, the target site needs to be prepared, which includes removing the ligand from the crystal structure (if present), removing the structural water molecules, adding hydrogen atoms to the sidechains, assigning atoms types and adding charges. Many of these final preparation steps can be performed automatically by the software. Once the target is prepared this file can then be used as the working target site for molecular docking. The ligand library that is to be used for docking must be prepared, with some different considerations. This process involves identifying atom types, structural minimization, adding hydrogens and assigning charges to offer reasonable starting conformation for the molecular docking experiment.

Rigid & flexible docking

The two main types of molecular docking experiments we consider are flexible and rigid target-site docking. In molecular docking experiments, the ligands are almost always made flexible and are minimized within the binding site to produce the best fit based on computed interactions. Rigid docking searching algorithms treat the target site as being rigid and immobile and fit the ligand into the predefined pocket. The calculations for rigid docking are faster to complete and may not represent an accurate pose or global minimum conformation. Flexible docking algorithms vary in their complexity and may consider the target to have only flexible hydrogen atoms or may have only preselected flexible residues. It is possible to consider the entire protein target of interest to be flexible, but this may require a long computation time and the accuracy of the results may vary. In recent years with advances in computational power, more advanced molecular dynamics simulations have become possible, taking into account the interactions of water molecules present in a completely flexible environment, and even considering the polarization of water [8,9].

Validation of model

The docking algorithm can be validated and assessed by comparing the extracted x-ray ligand (native ligand that was extracted from the original crystal structure) and the docked x-ray ligand to determine the root mean square deviation. A root mean square deviation of less than 2 Å is generally considered as an acceptable validation of the docking run [10,11]. Another process to assess the docking is to include known inhibitors and known decoy inhibitors in the docking run. Typically, for each docking run a library of ligands are produced, and within this library known inhibitors and known decoys are included. Once a docking run is complete, the ligands are ranked in terms of their binding affinity to the target site, and this allows the comparison of the ligands to one another. For example, AutoDock Vina used with PyRx ligands display a binding affinity in kcal/mol, with the more negative the binding affinity representing a better the fit into the binding pocket [12].

Interpretation of results

Results can be interpreted in terms of binding affinity and also key residue interactions. The key binding residues that are found within the site can be identified, and hydrogen-bonding interactions can be identified. This is in addition to determining where possible hydrophobic interactions may occur, performed by measuring the distance between the ligand and the residues functional group. Examining how the ligand folds or is positioned within the binding site and comparing the docked pose with known inhibitors or the decoys that were in the library can also give valuable insight.

Example protocol

In this section, we outline some available software and how it might be used practically as a solution for computational drug design. First, the PDB file of interest may be downloaded from the protein data bank, selecting a high-resolution PDB that is available for use [3]. The PDB file can then be opened with Chimera for preparation. Typically, a researcher might remove the ligand from the crystal structure, remove the structural water molecules, add hydrogen atoms to sidechains, assign atoms types and finally add charges before saving this as a new file. AutoDockTools may then be used to further manipulate and examine the PDB file if desired, before docking with AutoDock Vina via the PyRx graphical user interface. Binding data for the examined ligands are given in the form of kcal/mol, and can be used to guide further molecular design.

Conclusion

In recent times, the hurdle for entry into computational chemistry studies has become lowered to the point where routine experiments are possible in any academic environment where computers are present. This is particularly true for medicinal chemistry and drug design, with many programs available to run experiments locally on relatively low cost devices. In the interest of brevity, this editorial did not mention all of the multitudes of freely available databases and computer server systems that are available. Used for the collection of data and the computation of experiments on remote systems, this abundance of resources further lowers the device requirements for analysis. In the current era, these processes and technologies have largely become established. Computational tools are highly available and are primed for use by the end user, the academic researcher, to make discoveries and advance their field of knowledge.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1. van Gunsteren WF, Karplus M. Protein dynamics in solution and in a crystalline environment: a molecular dynamics study. *Biochemistry* 21(10), 2259–2274 (1982).
2. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28(7), 849–857 (1985).
3. Berman HM, Westbrook J, Feng Z *et al.* The protein data bank. *Nucleic Acids Res.* 28(1), 235–242 (2000).
4. Anna EL, Stephan ML, Billy W-N, Stephanie SS. A practical guide to molecular docking and homology modeling for medicinal chemists. *Curr. Top. Med. Chem.* 17(18), 2023–2040 (2017).
5. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Computation. Chem.* 31(2), 455–461 (2010).
6. Pettersen EF, Goddard TD, Huang CC *et al.* UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25(13), 1605–1612 (2004).
7. Morris GM, Huey R, Lindstrom W *et al.* AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Computation. Chem.* 30(16), 2785–2791 (2009).
8. Riniker S, van Gunsteren WF. A simple, efficient polarizable coarse-grained water model for molecular dynamics simulations. *J. Chem. Phys.* 134(8), 084110 (2011).
9. Lamoureux G, Harder E, Vorobyov IV, Roux B, MacKerell AD. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* 418(1), 245–249 (2006).
10. Warren GL, Andrews CW, Capelli AM *et al.* A critical assessment of docking programs and scoring functions. *J. Med. Chem.* 49(20), 5912–5931 (2006).
11. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL. Assessing scoring functions for protein–ligand interactions. *J. Med. Chem.* 47(12), 3032–3047 (2004).
12. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Computation. Chem.* 31(2), 455–461 (2010).