HHS Public Access

Author manuscript

Methods. Author manuscript; available in PMC 2023 March 30.

Published in final edited form as:

Methods. 2021 January ; 185: 3–14. doi:10.1016/j.ymeth.2020.02.009.

The fine art of preparing membrane transport proteins for biomolecular simulations: concepts and practical considerations

Hana Shiref[1], **Shana Bergman**[2], **Sophie Clivio**[3], **Michelle A. Sahai**[1,*]

[1]Department of Life Sciences, University of Roehampton, London SW15 4JD, UK.

[2]Department of Physiology and Biophysics, Weill Cornell Medical College of Cornell University (WCMC), New York, NY, 10065, USA.

[3]Harvard University, Cambridge, Massachusetts 02138, USA.

Abstract

Molecular dynamics (MD) simulations have developed into an invaluable tool in bimolecular research, due to the capability of the method in capturing molecular events and structural transitions that describe the function as well as the physiochemical properties of biomolecular systems. Due to the progressive development of more efficient algorithms, expansion of the available computational resources, as well as the emergence of more advanced methodologies, the scope of computational studies has increased vastly over time. We now have access to a multitude of online databases, software packages, larger molecular systems and novel ligands due to the phenomenon of emerging novel psychoactive substances (NPS). With so many advances in the field, it is understandable that novices will no doubt find it challenging setting up a protein-ligand system even before they run their first MD simulation. These initial steps, such as homology modelling, ligand docking, parameterization, protein preparation and membrane setup have become a fundamental part of the drug discovery pipeline, and many areas of biomolecular sciences benefit from the applications provided by these technologies. However, there still remains no standard on their usage. Therefore, our aim within this review is to provide a clear overview of a variety of concepts and methodologies to consider, providing a workflow for a case study of a membrane transport protein, the full-length human dopamine transporter (hDAT) in complex with different stimulants, where MD simulations have recently been applied successfully.

Keywords

molecular dynamics; computational modelling; protein structure preparation; membrane protein simulations; small molecule parameterization; molecular docking

*corresponding author, michelle.sahai@roehampton.ac.uk.

Author Contributions

All authors contributed to writing the manuscript.

Supplementary data

The following are the Supplementary data to this article.

Homology Modelling files for the alignment, Modeller script and output are included.

1. Introduction

Molecular dynamics (MD) simulations have become a powerful and widely accepted technique for understanding the detailed structure and dynamics of biomolecular systems (Karplus and McCammon, 2002). In the last decade advances in computer hardware (Cheatham and Roe, 2015; Kutzner et al., 2015), algorithm and force field developments (Larsson et al., 2011; Aldeghi and Biggin, 2017; Nerenberg and Head-Gordon, 2018; Jing et al., 2019), and more recently in machine learning (Böhm, 1996; Dorn et al., 2014; Behler, 2016; Pérez et al., 2018; Chen et al., 2019; Hu et al., 2019; Plante et al., 2019; Rao et al., 2019; Romero et al., 2019; Wang et al., 2019) and applications of virtual reality (Glowacki et al., 2018; Amabilino et al., 2019; O'Connor et al., 2019) have enabled rapid progress in the field. Any seasoned expert would tell you that this is a rather exciting time, as we try to make sense of and then try to apply these new methods and tools to our own work. With that being said, we are now more so than ever, attracting a record number of junior colleagues who are being taught the practical concepts much earlier in their careers, such as in high school, leading them to pursue related degrees at University (Burkholder et al., 2009; Lundquist et al., 2016; Burgin et al., 2018; Taly et al., 2019). No doubt we will soon be looking at a new generation of molecular modellers feverishly entering the field, having been taught that the study of the macromolecular structure is a key point in the understanding of biology.

Putting aside the practical challenges of the applications of molecular modelling for biomolecular systems, there is a common problem amongst novices in the lead up and preparation for the actual MD simulation that causes much frustration and delays in projects. The simulation setup usually depends on following a series of steps. An expert modeller would normally carry out these steps skilfully with a set of tools (commercially available, open-source or in-house), because they have the necessary knowledge and resources to overcome any specific problems that would arise because of the countless hours of troubleshooting. However, without the adequate training a newcomer to MD simulation would be overcome even if one step fails. This usually happens because novices tend to blindly use default procedures sourced online or documented by a post-doc or fellow student in the lab setting. This not only leads to anguish from making spurious observations, that are hard to discern from the correct ones, but it also leads to the unpopularity of biomolecular simulations and the classification that "it is hard" or "makes no sense" by the biochemical community.

Hence, we only see MD simulations being restricted to research groups that have the necessary tools and expertise. As a community, this does not bode well in sharing best practices or making it easy for newcomers to learn how to perform the basic tasks. What is shared may not be documented well enough or is standardised for a particular molecular system. This leads to a lack of compatibility and interoperability when being applied to new systems. We can think of two ways to overcome this; a top-down approach, where we design one package that does everything and that becomes the standard or a bottoms-up approach, where we make it easier to incorporate the already existing packages. The former could be regarded as the wrong solution because it just introduces another set of protocols that would have to be learnt from scratch. Therefore, the latter would appear to point to a right solution,

a black box program for novices to use, with an easy to use Graphical User Interface (GUI) for the setup of the simulation system that can account for the many different codes, force fields, solvation representations, protein structures, ligands, and especially if simulating membrane proteins, the all so crucial lipids. If only it was that simple!

Several tools have attempted to address parts of this problem like InterMol, that allows the user to read and write different molecular formats, or LOOS, MDTraj and MDAnalysis that have tools for the end-point analysis of different outputs (Michaud-Agrawal et al., 2011; McGibbon et al., 2015; Gowers et al., 2016; Shirts et al., 2017). These packages have restricted applications and use different scripts for the different packages. The closet tool available for use today that addresses the interoperability and integration of different software packages is the joint EPSRC supported Collaborative Computational Project for Biomolecular Simulation (CCPBioSim) and High-End Computing Resources by the Biomolecular Simulation Community (HECBioSim) software called BioSimSpace biosimspace.org) (Hedges et al., 2019). BioSimSpace is intended primarily to be used by novices who may be unfamiliar with programming in general. Not only does it allow interoperability for common software packages to work together, but it ensures that outputs from one package can be easily used as inputs for another package. The flexibility of its workflow means that it can work on different hardware and can be run in a number of different ways, e.g. command line or JupyterLab. BioSimSpace offers some reprieve from the current situation but a novice still needs a working knowledge of the offered simulation packages and the different steps in setting up a simulation and most importantly how to obtain the raw input files, which will be discussed in the next section.

The most popular simulation packages (AMBER (Reese et al., 2018), CHARMM (Brooks et al., 2009), GROMACS (Pronk et al., 2013; Lundborg and Lindahl, 2015) and NAMD (Phillips et al., 2005)) do have accompanying tools, which perform most steps of the preparation. Acellera ACEMD (Harvey et al., 2009) and the HTMD platform also possess similar tools; although the latter is not freely available; some basic versions of the tools are available online at playmolecule.org (Doerr et al., 2017; Jiménez et al., 2017, 2018; Martínez-Rosell et al., 2017; Galvelis et al., 2019). There are also many other combined user-friendly interfaces that provide a solution to the simulation setup utilising GROMACS and CHARMM (Kota, 2007; Miller et al., 2008; Roopra et al., 2009; Sellis et al., 2009; Makarewicz and Kaźmierkiewicz, 2013; Paissoni et al., 2014; Lundborg and Lindahl, 2015), while VMD provides a number of plug-ins that facilitate simulations with NAMD (Phillips et al., 2005). These tools provide the automatic setup functionality to prepare systems for simulation without an in-depth knowledge of the inner workings of the software, thereby promoting accessibility to the field for novices. However, there still remains a lack of a standardised representation of the structural information needed for the input setup, most interfaces are restricted to a single MD package, and the data is not easily interchangeable, especially force fields. There is usually an embedded scripting language that complicates the automation process, and without experience in coding this would be a daunting task.

Furthermore, a novice would have to decide very early on, depending on the size of the system and the mechanism or interactions they would like to observe, the level of detail needed to represent their simulated system. There are a number of different levels of detail

using classical molecular mechanics (MM) and quantum mechanical (QM) representations. Generally, MM methods are the choice for protein simulations and can either be in an all-atom (atomistic) representation or a course-grained (CG) representation. An atomistic representation would give the best representation of the actual system; however, it can be unfeasible for very large systems and timescales because of the computational cost required. While a CG representation of a system that is reduced by a number of degrees of freedom, offers an alternative approach when studying large systems or when long simulations are required. The advantage of CG simulations, because of the reduced degrees of freedom and lack of detailed interactions, means that less resources are required, and the calculations are a lot faster than that of the same system in an atomistic representation. A third method, using a hybrid approach of QM/MM applies the accuracy of the QM and speed of MM approaches to systems where an important but small part can be described by QM and the rest modelled by MM.

In this review we aim to provide a concise overview of the various concepts and methodologies that are required for the initial steps of an *atomistic* membrane protein setup, present their strengths and limitations, and highlight the open challenges faced by novices. We will particularly underline how to overcome these challenges with recent developments in MD packages and tools. Furthermore, within the context of the methodologies discussed, we will provide a workflow for a case study of a membrane transport protein, the full-length human dopamine transporter (hDAT) in complex with different stimulants, where MD simulations have recently been applied successfully (Khelashvili et al., 2015a, 2015b; Razavi et al., 2017, 2018; Sahai et al., 2017, 2018; Loi et al., 2020). We will discuss the practical considerations of homology modelling, ligand docking and parameterization, protein preparation and membrane setup; which can be applied to most membrane protein systems before any MD simulations are performed.

2. Understanding and obtaining your protein structure

Before embarking on the various steps required for the simulation setup we feel it is important to ask a question recently seen on #SciTwitter (Morris, 2019) “If somebody gave you a protein structure and you had to open it and look around at it, would you know what you were looking at?” Surprisingly from nearly 300 people that participated in the poll (similar to the class size of the first year undergraduate Biochemistry course at the University of Roehampton), ~40% of them said no. This is a blunt reminder that a good place to start, for those unfamiliar and new to structural biology, would be to learn basic biochemistry to identify the 20 amino acids, different secondary structures (like alpha helices and beta sheets) and inter- and intra-molecular bonds (disulfide bond, hydrogen bond, Van der Waals and electrostatic interactions) as well as any features like conserved regions or domains that would allow you to categorise the various protein classes (GPCR, transporter, channel, kinase etc).

The US Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB; rcsb.org), and especially PDB-101 (Berman et al., 2000; Minor et al., 2016) is a good place to start accessing introductory material to help beginners learn about these 3-dimensional (3D) structures as well as advanced tutorials for extended learning.

Additionally, the RCSB PDB is a repository to obtain the 3D structures that have been solved by the various biophysical methods that include X-ray crystallography and Nuclear Magnetic Resonance spectroscopy (NMR) (Opella, 1997; Zhang and Cherezov, 2019).

The RSC PDB is a member of the Worldwide PDB consortium (wwPDB; wwpdb.org) (Burley et al., 2019), whose members also include the Protein Data Bank in Europe (PDBe; pdbe.org) (Mir et al., 2018), Protein Data Bank Japan (PDBj; pdbj.org) (Kinjo et al., 2017) and BioMagResBank (BMRB; bmrb.wisc.edu) for NMR data (Ulrich et al., 2008). Established in 1971, the PDB Core Archive now houses 3D atomic coordinates of >144 000 structural models of proteins, DNA/RNA, and their complexes with metals and small molecules and related experimental data and metadata. The wwPDB also works closely with the Electron Microscopy Data Bank (EMDB; emdb-empiar.org), a worldwide repository that houses related experimental data/metadata from 3D Electron Microscopy (3DEM) and Electron Tomography (ET) (Tagari et al., 2002; Milne et al., 2013). Publication of new macromolecular structures in most scientific journals is conditional on mandatory deposition of the 3D atomic coordinates comprising the structural model plus experimental data used to derive the structures and associated metadata to the PDB. As such these databases should be the first places to start looking for your 3D structural information.

There are also general databases where you can find biological and structural information of the different protein classes such as The National Center for Biotechnology Information (NCBI; ncbi.nlm.nih.gov) and UniProt, the universal protein knowledgebase (UniProt; uniprot.org) (Bateman, 2019) as well as specialised databases like the Transporter Classification Database (TCDB) (Saier et al., 2016), GPCRdb: the G protein-coupled receptor Database (Vroling et al., 2011; Munk et al., 2016) and the Membrane Protein Data Bank (MPDB) (Raman et al., 2006).

These lists are not intended to be exhaustive, as there are many other resources available for the same purposes, but needless to say familiarising yourself with these databases and methods would help immensely when it comes time for understanding how the protein structure was obtained.

2.1 Visualising your protein structure

Going back to part of the #SciTwitter (Morris, 2019) question, would you know how to actually open and view the protein structure? A newcomer to structural biology needs to grasp many modern biological concepts as well as gain an appreciation of the interactions and relative sizes of molecular systems. While structure determines function is increasingly recognized as an important scientific concept for students, this relationship remains absent from many schematic depictions of biomolecules and processes found in biological textbooks. This is especially true when proteins appear as colourful blobs or structured boxes or even actual locks and keys when explaining enzyme activity or protein-ligand binding. The use of molecular visualization software provides a remedy to this however it is not without its own limitations; some of these tools have steep learning curves that limit the time that could be used to focus on the research project and they often lack the capability for viewing dynamic trajectories, when it comes time to analysing the molecular dynamics simulations. Despite its practicality, there still exists many hurdles for using molecular

visualization tools effectively. Choosing the right visualisation tool at the beginning of this journey and spending the necessary time to learn its features can be an invaluable advantage.

There are a number of common 3D molecular viewers, such as UCSF Chimera (Pettersen et al., 2004), OpenStructure (Biasini et al., 2010, 2013), Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.), Rasmol (Sayle and Milner-White, 1995), Swiss PDB Viewer (Guex et al., 1997) and Visual Molecular Dynamics (VMD) (Humphrey et al., 1996). The most commonly used software packages are free, at least for educational use and are available to run on a wide range of architectures and the common operating systems including Microsoft Windows, Apple macOS and LINUX. It is good to bear in mind that when choosing the software to use most likely this will be influenced by the preference of your research group, quality of graphical representations, and if there is a need to view MD trajectories. VMD is often used for this latter purpose as it can read the trajectory files created during simulations in formats produced by many different software packages (Humphrey et al., 1996). Pymol produces excellent images, but is less straightforward for viewing trajectories (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.). Other freely available software, such as UCSF Chimera (Pettersen et al., 2004), OpenStructure (Biasini et al., 2010, 2013), Rasmol (Sayle and Milner-White, 1995) and Swiss PDB Viewer (Guex et al., 1997) can be used to view individual conformations (snapshots) from a simulation, but they lack the ability to show trajectories as animations.

VMD is our preference for visualising protein systems. It has many added advantages with the foremost being that it supports the Tcl (Tool Command Language) scripting language (Dalke and Schulten, 1997). Users are supported by online documentation (tcl.tk) and by books targeting all levels of experience therefore even novice users are likely to gain experience with the language. Although this would be a new language for a newcomer to learn, the advantages far outweigh any initial disadvantage. Tcl can be used to parse the simulation configuration file, allowing variables and expressions to be used in initially defining options, and also change options during a running simulation, such as in running interactive MD like Steered Molecular Dynamics (SMD), implemented in the QwikMD plugin (Ribeiro et al., 2016). In addition, VMD can render publication quality animations and images from large trajectories stored on clusters and supercomputers with MPI implementations.

3. Stepping towards a molecular dynamics simulation: Our case study

Now that we have covered how to obtain your protein structure and how to visualise it in the above sections 2 and 2.1, we can now attempt to work through the computational system setup for the membrane transport protein, hDAT, in complex with various psychostimulant drugs. At this point we should have a working knowledge of a preferred visualisation software and background research on the protein system and ligand that will be needed in the atomistic MD simulations.

3.1 3D structure prediction when there is no protein structure in the databases

We previously introduced databases where you can obtain the 3D coordinates of protein structures in Section 2. Now what happens if your protein system does not appear in any of these databases?

Protein structures are uniquely determined by their primary amino acid sequences (Anfinsen, 1973), and in some cases, are unknown or unavailable (Anfinsen, 1973). However, identifying direct contacts between protein and ligand is contingent on having a 3D structure. This is especially important with hDAT, that is responsible for the reward and reinforcement properties of stimulants like cocaine and novel psychoactive substances (NPS) (Beuming et al., 2008; Bisgaard et al., 2011; Dawson et al., 2014; Sahai et al., 2017, 2018). Briefly, hDAT belongs to the family of neurotransmitter: sodium symporters, that also includes the serotonin transporter (SERT) and norepinephrine transporter (NET) and controls dopamine (DA) homeostasis by mediating Na+ and Cl- dependent reuptake of DA (Rudnick and Clark, 1993; Kristensen et al., 2011). hDAT, like the other members of this family of transporters, has both an intracellular amino- and carboxyl-termini and twelve transmembrane (TM) helical domains (Yamashita et al., 2005; Penmatsa et al., 2013b; Coleman and Gouaux, 2018) (Figure 1). Topological information like this can be easily obtained from UniProt (Bateman, 2019) to aid in understanding the structure of the protein and to confirm the location of TMs in the membrane environment. A high-resolution elucidation of hDAT structure, especially regarding its substrate and inhibitor recognition sites would be important for any studies linked to understanding the reinforcing properties of psychostimulants; none exist.

However, we do have access to X-ray crystal structures of proteins homologous to the NSS family and now more recently crystal structures of members of this family (Penmatsa et al., 2013a, 2013b, 2015; Wang et al., 2015; Coleman and Gouaux, 2018), however, still none for hDAT. Before 2013, the crystal structure of a bacterial leucine transporter (LeuTAa), a protein homologous with the NSS family, for which several crystal structures corresponding to various functional states (outward open, occluded, and inward open) had been determined (Singh et al., 2007, 2008; Krishnamurthy and Gouaux, 2012), was used as a ‘template’ to employ computational modelling techniques to produce a feasible 3D hDAT structure. Since then we have used the crystal structure of the *Drosophila melanogaster* dopamine transporter (dDAT) (Penmatsa et al., 2013a, 2015; Wang et al., 2015), which shows a remarkable similarity between the TM bundles of LeuTAa (Hansen et al., 2014). Because of computational modelling the substrate binding site (S1) is known to be deeply buried in the transporter structure (Beuming et al., 2008; Bisgaard et al., 2011; Sahai et al., 2017) and we can describe a site that overlaps with that of dopamine and many of the popular psychostimulants (Sahai et al., 2017, 2018). It is also clearly distinct from the site observed for antidepressant binding (S2 site) to the leucine transporter (LeuTAa) which is found facing the extracellular vestibule above the S1 site (Quick et al., 2009).

There are three main computational modelling approaches that may be employed in predicting a 3D protein structure: *ab initio* prediction, “fold” recognition, and comparative (homology) modelling (Petrey and Honig, 2005). These differ principally in the sequence and structural database information used. While true *ab initio* methods rely entirely on the

physical and chemical information contained in the primary amino acid sequence to predict the structure, fold recognition, or “threading,” relies heavily on the structural similarities between specific distantly related or unrelated proteins (Bonneau et al., 2001; Petrey and Honig, 2005; Das and Baker, 2008). Comparative (homology) modelling, enables us to construct a 3D structure of the unknown, ‘target’ protein based on the structure of a similar protein, principally considered the ‘template’ (Sali and Blundell, 1993). Homology modelling involves challenges in finding templates with relatively high sequence identities but if the proteins that share greater than 40% amino acid sequence identity, comparative modelling is straightforward (e.g. dDAT and hDAT). For proteins with less than 30% amino acid sequence identity (e.g., LeuTAa and hDAT), comparative modelling becomes more challenging. Still, the LeuTAa crystal structure has successfully guided the creation of useful comparative models for hDAT computational studies despite the absence of appreciable amino acid sequence identity (Beuming et al., 2008; Kniazeff et al., 2008; Shi et al., 2008; Quick et al., 2009; Shi and Weinstein, 2010; Bisgaard et al., 2011; Zhao et al., 2012; Shan et al., 2011; Zhao et al., 2011; Kantcheva et al., 2013; Stolzenberg et al., 2015). This has revealed the 3D structure of DAT as exhibiting a LeuT-like structure fold (Perez and Ziegler, 2013).

Conducting computational homology modelling, however, requires more than the first step; (1) simple ‘template’ recognition, in fact, it is a multi-step process with four more progressive stages; (2) sequence alignment, (3) model building for the intended ‘target’, which is based on the 3D structure of the ‘template’, (4) model refinement, analysis of alignments, gap deletions as well as additions and finally (5) model validation (Martí-Renom et al., 2000) (Figure 2). Often, alongside ‘template’ recognition, the subsequent ‘template’ alignment step is performed. The most popular server used to conduct sequence alignment using the ‘target’ to find the 3D coordinate for ‘template’ protein sequence is BLAST (Basic Local Alignment Search Tool), which is a database from the NCBI (NCBI Resource Coordinators, 2016). This usually relies on one obtaining the FASTA (Pearson and Lipman, 1988; Pearson, 2014) sequence of your ‘target’ protein (e.g. hDAT), from either NCBI or UniProt (Bateman, 2019) and then using the ‘Standard Protein BLAST’ feature, and selecting the ‘Protein Data Bank proteins (pdb)’ as the ‘Database’, to search for any sequence similarities from the deposited structures. Usually, the most significantly aligned sequences are presented first with the identifiable sequence identity, important in helping to choose the ‘template’.

There are alternative options and servers available in the case of BLAST being unable to find protein structures with an appropriate sequence identity. UniProt is one such option. After searching for the name of the protein, and selecting the correct organism and gene name, the results page displays a number of options. Under ‘Structure’, you would find a list of available PDB entries as well as a prediction from the 3D structure databases, SWISS-MODEL (SMR) (Bienert et al., 2017) and ModBase (Pieper et al., 2006), with prepared 3D predicted structures. The Protein Model Portal (PMP; proteinmodelportal.org) (Arnold et al., 2009), an online server, also consists of millions of model structures provided by different partner resources including ModBase and SMR. A careful warning at this stage; it is extremely important a user of these databases is aware of the ‘template’ being used as well as the alignment that is produced. Reproducibility of the ‘target’ produced

amongst the different tools should also be cross-linked with background literature i.e. (5) model validation, since steps (3) model building and (4) model refinement are performed automatically behind the scenes.

For step (2) sequence alignment in the hDAT modelling, a previously published sequence alignment of the NSS-family proteins (Beuming et al., 2006) is used to first construct homology models for the transmembrane (TM) part of the hDAT (contained in residues 57−590) based on the 'template', the outward-facing dDAT structure (PDB code: 4M48). See the Supplementary data for details about the alignment files, Modeller script and output. For the N- and C- termini we have employed *ab initio* methods to predict these segments [38]. All additional steps for the TM generation are then conducted using the MODELLER program (Eswar et al., 2006), available for most common operating systems. Incidentally, The ModWeb comparative protein structure modelling webserver is based on Modeller but has the added advantage of control of choosing the 'template' and the alignment. It is worth noting that there are some commercially available software packages that also include a homology modelling module that you can control, such as Schrödinger's Prime Homology Modelling workflow (Schrödinger, LLC: Portland, OR, 2007, Web address: www.schrodinger.com.) and MOE (Molecular Operating Environment) (C. C. G. I. Molecular Operating Environment (MOE), 1010 Sher- booke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013)

To use MODELLER, you only need the pdb file for the 'template', an alignment file for the 'target' and 'template' and the Python MODELLER script; salilab.org/modeller provides excellent tutorials on how to use MODELLER with examples of the syntax used in the alignment files (usually FASTA (Pearson and Lipman, 1988; Pearson, 2014)) as well as the scripts. Careful editing of the alignment file and script avoids any spurious errors with syntaxes that can cause initial errors. If there are internal ions they can also be added to the alignment file with specialised syntaxes, allowing ions to also be modeled into the 'target'.

Step (3) model building requires you to run the MODELLER Python script. Despite any experimental errors in the 'template' such as poor electron density, backbone generation usually reveals a good model from a series of models irrespective of such errors. In some cases, the 'target' and 'template' alignment can comprise of gaps or missing very flexible regions that could not be resolved by crystallography. Therefore, to overcome this, helices are inserted or deleted from the alignment and instead annotated as loops; this is known as the loop modelling stage and forms part of step (4) model refinement. Figure 3 gives details of the extracellular loop 2 (EL2). It is much shorter in dDAT and is truncated in the crystal structure while the longer EL2 region in hDAT is then modeled in by the loop modelling. Side-chain modelling is performed to ensure the inclusion of all the atoms within the protein. This is a fundamental component as it enables us to understand further protein-ligand interactions.

MODELLER also includes step (5) model validation. Usually, this is included after the model building step to include a model optimisation step that outputs statistical potential, to help you assess the models built. Optimisation also can prevent incorrect backbone predictions on rotamers by restraining atom positions or applying energy minimisations. In

MODELLER, Discrete Optimized Protein Energy, abbreviated as DOPE, is the statistical potential that is produced as an output from MODELLER and can help you assess the protein structures (Verdonk et al., 2011). While increased energies are displayed with higher DOPE scores, these are usually disregarded as the best model due to potential unfavourable non-bonded interactions. A lower DOPE score also assesses the compatibility of the structure, both in regard to its sequence as well as the generated model. To aid in this assessment it is important to visualise a 3D alignment of the 'target' and the 'template' with your favourite visualisation tool (Figure 3) and consider the scoring potentials from the homology modelling software. This model validation stage is a common technique as it affirms the predicted 3D protein structure is free from any errors.

3.2 Ligand preparation and docking

Molecular docking has emerged as a powerful approach in the structure-based drug discovery pipeline (McConkey et al.; Bohacek et al., 1996; Chou, 2004; Hou and Xu, 2004; Kitchen et al., 2004). It can be used to model the interaction between a small molecule (ligand) and a protein at the atomic level, allowing us to characterize and understand fundamental biochemical processes when the ligand is in the binding site of its 'target' protein. The docking process involves two basic steps: (1) prediction of the ligand conformation as well as its position and orientation within the binding site (or 'pose') and (2) assessment of the binding affinity (Bohacek et al., 1996). These two steps are related to sampling methods and scoring algorithms based on different criteria including steric clashes, electrostatic interactions, force field interactions and hydrogen bonding to predict the most energetically favourable orientation of the ligand and thus to quantify the likelihood of the biochemical nature of docking in that position.

Knowing the location of the binding site before the docking process significantly increases the docking efficiency, which is the case when there are crystal structures of the protein in complex with ligands in the binding site. You can also obtain information about the sites by comparison of the 'target' protein with a family of proteins sharing a similar function (LeuTAa and hDAT) or with similar proteins co- crystallized with other ligands (dDAT and hDAT). If knowledge about the binding site is lacking, then there are various binding site prediction software or online servers that can be utilised to identify putative binding sites within proteins. Although not an exhaustive list, it includes software like CAVER (Jurcik et al., 2018), SURFNET (Laskowski, 1995), SiteMap module of Schrödinger suite of programs (Schrödinger, LLC: Portland, OR, 2007, Web address: www.schrodinger.com.), 3DLigandSite (Wass et al., 2010) and DeepSite, a protein-binding site predictor using 3D-convolutional neural networks (Wass et al., 2010). Docking without any prior knowledge about the binding site is called blind docking.

You can find other such tools or the database repositories for structure-based drug design such as Click2Drug, click2drug.org and BBCU, https://bip.weizmann.ac.il/toolbox/structure/binding.htm.

Docking programs, like the Autodock (Morris et al., 1998), AutoDock Vina (Trott and Olson, 2010), GOLD (Verdonk et al., 2003) and GLIDE module of Schrödinger suite of programs (Schrödinger, LLC: Portland, OR, 2007, Web address: www.schrodinger.com.) all

follow the same general steps of preparing the protein and the ligand, defining the binding site and adjusting the ligand pose according to thermodynamic principles to minimise free energy of binding. There are many choices for each user, including mode of docking (flexible ligand and rigid receptor docking or flexible ligand and flexible receptor docking), formatting of the input files, naming and indexing of residues, orientation of the protein, and identification of ions. The “lock- and-key” assumption was the basis for the earliest reported docking methods, which states the ligand and receptor are to be treated as rigid bodies, and their affinities for one another should be equivalent to a geometric fit amidst their shape (Mezei, 2003). The “induced-fit” theory was introduced years later, proposing that the ligand and receptor should be treated as flexible during the docking (Hammes, 2002). Under this assumption, the various backbone movements of protein and ligand residues are able to affect various side chains of each structure, allowing the residues of each part to better fit one another. Therefore, the sampling procedure is of a higher order of magnitude in terms of the number of degrees of freedom when it is within a fully flexible receptor/ligand docking, compared to flexible docking in a rigid receptor. Despite the larger computational cost of this methodology, it predicts the binding mode of the molecule with a higher accuracy than rigid body algorithms (Hammes, 2002).

Ultimately, the docking process should discover the most favourable pose between a receptor and a ligand. Prior to docking, a number of steps should be performed to prepare both the ligand and protein crystal structure or homology model for docking. This includes adding hydrogen atoms, optimising hydrogen bonds, removing atomic clashes in the protein structure. Similarly, the 3D coordinates of the ligand is either downloaded from databases like the PDB or the NCBI PubChem, pubchem.ncbi.nlm.nih.gov or manually prepared to manufacture 3D geometries, assign bond orders, and generate accessible ionisation and tautomer states (Madhavi Sastry et al., 2013).

Docking schemes comprise of two parts: (1) an algorithm that scans the translational, rotational and conformational space available to a chosen molecule in the binding region, alongside (2) an objective function that must be minimised in this process. The function calculates an approximate measure of binding affinity, normally referred to as a scoring function (Schulz-Gasch and Stahl, 2003). Verdonk et al. (2011) discovered six specific limitations for all molecular docking studies (Verdonk et al., 2011). These include 1) the quality of the docking programs used, although there may be variability with the same docking program; 2) the types of targets and ligands included in the studies; 3) the level of experience the user has with the docking software; 4) the quality of the X-ray structures, where structures may have poor electron density for the ligands, disorder, or not be fully refined; 5) the preparation of the binding sites and ligands (protonation states might be incorrect) and 6) protocol differences (site grid definitions can differ between user experience while some users pre-optimize complexes).

We perform our docking procedure using the Schrödinger software suite (Schrödinger, LLC: Portland, OR, 2007, Web address: www.schrodinger.com.), which requires a license for academic purposes. We have docked a number of ligands and psychostimulants into the binding site of full length hDAT (Khelashvili et al., 2015b; Sahai et al., 2017, 2018) using GLIDE implemented in the Induced fit docking (IFD) protocol, in the Schrödinger software

suite (Schrödinger, LLC: Portland, OR, 2007, Web address: www.schrodinger.com.). We are also limited in our assumption that the ligands being used will have been crystallised before. Without a crystallised structure, it is precarious to generate a structure and to trust the orientation, chirality and properties. However, if caution is given to the building, preparation, and docking of this ligand, and if the poses are considered in the context of an extensive literature review, it is still possible for this docking methodology to serve its purpose of revealing biochemical details at the binding site of a protein.

Here we will provide an example of how three very different ligands can be prepared for docking: cocaine, 5-IT and a more complex ligand diphenylprolinol (D2PM) (Figure 4). The 3D coordinates of cocaine can be retrieved from the crystal structure of dDAT in complex with cocaine, PDB ID: 4XP4; 5-IT is a phenethylamine derivative or substituted amphetamine and can be built by modifying an already existing crystal structure of amphetamine bound to dDAT (PDB ID: 4XP9). The 2D and 3D build function that is a part of Schrödinger's GUI Maestro can do this and is freely available. The 3D coordinates of D2PM can be retrieved from PubChem: Compound CID: 7045371 (NCBI Resource Coordinators, 2016) and then prepared with the LigPrep module in Schrödinger [44]. If you are inundated with search results from PubChem for your own compound, it is best to write out the name of the compound e.g. 'diphenylprolinol' for your search instead of its abbreviation. The prepared compounds should all carry a net positive charge and can be assigned by Epik, a module within Schrödinger that provides p*Ka* Prediction [44]. In preparation for molecular dynamics simulations, ligand parameterisation of novel compounds may need to be performed. Therefore, at this stage the prepared ligand should be save in .mol2 format. Ligand parameterisation will be discussed in Section 4.

The hDAT homology model was then prepared using the Protein Preparation Wizard module in Maestro, following which the Induced fit docking (IFD) protocol, in the Schrödinger software suite was implemented to dock the ligand. We used the binding site residues Phe76, Asp79, Ser149, Val152, Tyr156, Asn157, Phe326, Val328 and Ser422, previously identified as important for binding psychostimulants of comparable size to define the docking grid box (Beuming et al., 2008; Bisgaard et al., 2011). IFD Docking was then performed using standard precision (SP). Random initial positions and conformations of the ligand were screened for clashes with the protein and subsequently refined by allowing flexibility of the side-chains in the binding.

Depending on the number of poses generated, Schrödinger's IFD protocol, like other docking protocols will produce output structures ranked by a specific score. The IFDScore is calculated by 1.0 Glide_Gscore + 0.05 Prime_Energy, and accounts for the protein-ligand interaction energy alongside the overall energy of the system and utilised to rank the conformations. The lower the IFDScore, the more favourable the binding. A careful warning is to not take these scores at face value as it is important to visualise all the poses generated as many of the worst poses may meet the criteria of lowest IFDScore. As a result, once the docking procedure has completed, numerous potential poses will be provided, each with their associated docking scores and penalties associated with unfavourable biochemical interactions or steric clashes. To choose the most biologically realistic pose, it is suggested to compare each pose with known binding site interactions, specific electrostatic and

hydrogen bonds, crystallized orientations of similar ligands or similar proteins, and to consider the probability of each pose given what is known about the ligand and the protein into which it is docked.

Figures 4 and 5 explain the workflow for the preparation of the protein and ligands for docking as well as the final docked result (Figure 4).

3.3 Immersion of docked complex into a membrane environment

For most docked systems, it is necessary to consider the dynamics of the complex in the context of a membrane in which it would be natively found. Proteins are known to interact with the membrane in which they are embedded, and they can influence and be influenced by lipid-protein interactions (Mondal et al., 2013). Both experimental and computational studies have remarked that the environment surrounding NSS transporter proteins can play a critical role in their function and can influence reorientations of TM segments and their direct relation to the functional mechanism (Mondal et al., 2013; Khelashvili et al., 2015a). Careful evaluation of the membrane-protein interaction patterns is essential in order to understand the functional mechanisms of these important proteins.

Therefore, once the specific complex has been chosen (from Section 3.2), it should then be immersed in a realistic biological environment with lipids known to be found surrounding that transmembrane protein. In our hDAT system, a mixture of POPE/POPC/POPS/PIP2/cholesterol lipids, closely resembling the neuronal cell plasma membrane is used (Ariga et al., 1988). CHARMM-GUI membrane builder (Brooks et al., 2009) was used to create bilayer models of desired lipid compositions. The membrane preparation protocol in the software takes advantage of experimentally determined values for the area per molecule for different lipids and makes use of the extensive library of lipid conformations to generate a lipid bilayer.

Besides the CHARMM-GUI membrane builder, VMD also has a Membrane plugin. Additionally, the Acellera suite of tools (Doerr et al., 2017); has introduced MembraneBuilder on playmolecule.org, an application to build complex membranes, necessary to build and run atomistic simulations of membrane proteins. However, at the moment the lipid components are limited to POPC and POPE for the VMD plugin and POPC, POPE and cholesterol for MembraneBuilder.

Following the generation of the lipid bilayer, the docked complex can then be inserted. We refer back to the common simulation codes (AMBER (D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGra, 2018; Reese et al., 2018), CHARMM (Brooks et al., 2009), GROMACS (Pronk et al., 2013; Lundborg and Lindahl, 2015) and NAMD (Phillips et al., 2005)) that have accompanying tools to help in this process. Considerations here would be orientation in the membrane and removing any membrane lipids overlapping into the protein environment. No doubt the automation of this crucial step can be invaluable to a novice’s toolkit.

4. Ligand Parameterization and final steps before MD simulations

Molecular dynamics (MD) simulations must overcome many barriers to be considered a viable method for probing biological systems, of which includes the limitations of time scale, system size and accuracy in the representation of the underlying molecular system. With the first two relying on advances within the formations in hardware and algorithms, the latter requires diligent development of better force fields that are to sufficiently describe key interactions in the simulation system. An important limitation is the complexity of developing missing force field parameters for novel chemicals, such as small molecule ligands. These chemical entities are often vital components in the biological system of interest, but can obstruct the utility of the molecular dynamics technologies for fields that include drug discovery when difficulties arise in accurately parameterising small molecules (Durrant and McCammon, 2011; Borhani and Shaw, 2012).

Various empirical force fields including OPLS, AMBER, CHARMM as well as GROMOS have been developed for MD simulations (Guvench and MacKerell, 2008). In order to apply these general force fields to an arbitrary drug-like molecule, functionality for assignment of atom types, parameters and charges is required. Historically, quantum mechanics (QM) can parameterize them with better accuracy but they are computationally expensive and slow, which limits applicability to a small number of molecules. While accurate force fields are not generally available for all molecules, like novel drug-like molecules. Therefore, it is biased to suggest a single parameter set can sufficiently characterise a large number of compounds. Therefore, to address this “small molecule problem”, one approach was to develop a limited set of building blocks covering a particular class or family of molecules. This has been a principle of CHARMM General Force Field (CGenFF) (Vanommeslaeghe et al., 2010) as well as General Amber Force Field (GAFF) (Wang et al., 2004), of which only target drug-like molecules within a biological environment. Commonly used AMBER and CHARMM force fields contain parameters for biomolecules (proteins, nucleotides, saccharides, lipids, etc.), but lack parameters for other biologically relevant molecules (co-factors, drugs, etc.) and are not guaranteed to be transferable to all possible chemical environments.

There have been a variety of tools that have been developed in assigning missing parameters directly from analogy to already pre-existing ones, which depend on databases of molecules of already parameterized compounds for a particular force field. Some examples of these tools include ParamChem (Vanommeslaeghe et al., 2012a, 2012b), and MATCH (Yesselman et al., 2012) web servers for CGenFF. VMD also offers a parameterization tool known as the force field toolkit (ffTK) (Mayne et al., 2013 which is also designed specifically for the parameterization of small molecules. ffTK provides many advantages including 1) the optimization of the energetic contribution of each component contributing to intramolecular interactions including charges, bonds, angles and dihedrals; 2) user-selected level of QM theory; 3) visualisation at each step; and 4) a manual or automatic option in parameter guessing and refining. With so many options, however, there are molecule-dependent hurdles at each step of ffTK that require user-manipulation and troubleshooting. Newer tools like Parameterize found in the Acellera HTMD platform (Galvelis et al., 2019) or online on playmolecule.org, can also improve the quality of the parameters by QM data, by refitting

electrostatic potentials (ESP) charges and rotatable dihedral angle parameters. As such, tools like Parameterize and ffTK are poised to fundamentally solve the problem of transferability for atom types and parameters. One only needs to prepare a .mol2 file of your ligand, with added hydrogens and knowledge of the charge to use tools like Parametrize, ffTK, PubChem and MATCH. However, careful inspection of the prepared topologies and parameters, in addition to a simple minimisation in water can reveal if the parameters are correct.

Following the steps in Section 3 and this section, we have all the parts to put together the complex for an MD simulation. The standard MD packages will allow the protein-membrane complex to be solvated and ionized in order to mimic a near-physiological environment. The system dimensions and approximate number of atoms can vary depending on the hydration number of water molecules per lipid and concentration of ions used in the system. Following this the system can then be evaluated by MD techniques in your favourite MD package, such as AMBER (Reese et al., 2018), CHARMM (Brooks et al., 2009), GROMACS (Pronk et al., 2013; Lundborg and Lindahl, 2015) and NAMD (Phillips et al., 2005) and Acellera ACEMD (Harvey et al., 2009).

5. MD Analysis

MD simulations produce very complex data where essentially the cartesian coordinates of each atom of the system are recorded at every timestep of the trajectory. Depending on the size of the system this could be thousands or millions of steps with huge MD generated trajectories taking up gigabytes or more of space. Therefore, data analysis has to be specialised to extract useful information in addition to data management and storage, which pose a major challenge for accessibility. We will concentrate our discussion on the types of analyses that one can perform on the data from MD simulations.

In general, there are many tools that can analyse the trajectories. It all depends on what data you want extracted and importantly how it is extracted. If you are familiar with programming languages, you can either write the code yourself in your favourite language or you can find already built code on open-source forums like GitHub (github.com) which hosts version-controlled software.

MDTraj and MDAnalysis, which have been introduced previously, can analyse the simulation trajectory for many different packages (Michaud-Agrawal et al., 2011; McGibbon et al., 2015; Gowers et al., 2016), while CHARMM (Brooks et al., 2009) and GROMACS (Pronk et al., 2013; Lundborg and Lindahl, 2015) packages contain a large number of programs that perform particular analysis tasks. Various VMD plug-ins can analyse NAMD trajectories (Phillips et al., 2005) and HTMD can analyse ACEMD trajectories (Harvey et al., 2009). This is of course not an exhaustive list but they all perform standard tasks including calculating a root mean square deviation (RMSD) i.e. the structural distance between coordinates, root mean square fluctuation (RMSF) i.e. the average deviation over time of a protein residue from a reference position or for performing principal component analysis (PCA) on the trajectory i.e. the conversion of the movement of all atoms in the protein into a set of principal components which are

linearly independent. Other measurements can include hydrogen bonding, dihedral angle and distance measurements or solvent accessibility.

All these tools and more solve the data generation and analysis problem, but it is up to the user to understand what information needs to be extracted and which tool matches their skills and expertise to be utilised efficiently.

6. Concluding remarks

The field of molecular mechanics has developed enormously since its inception in the 1970s. Due to the rapid improvements of hardware, algorithms, and force fields we can now simulate complex systems at even better and longer time scales and resolutions. By applying molecular modelling methods from ligand docking to MD simulations, scientists are able to discover details of biochemical events such as oligomerization and neurotransmitter transport mechanisms of novel compounds. Nevertheless, we must equip newcomers to the field with the basics in a simple and straightforward manner.

Although the procedures detailed above provide advantages to existing methods, list solutions and provides limitations, gaps will still remain in consistency as new software and force fields are created. To overcome this weakness, we recommend taking robust measures in examining the software and outputs generated. We cannot account for all methodologies, but in establishing means to overcome these challenges at each step, it is our hope that we begin to bridge the disparities between existing methods and to pave the way for new methods that researchers may find workable for them.

In this way the great hurdle that appears to be the field of molecular modelling can be approached willingly to achieve higher accuracy when studying systems with increased chemical complexity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge the computational resources from the Department of Life Sciences, University of Roehampton and the computational resources of the David A. Cofrin Center for Biomedical Information in the HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine at Weill Cornell Medical College, USA.

References

Aldeghi M, and Biggin PC (2017). Advances in Molecular Simulation Third Edit. Elsevier doi:10.1016/B978-0-12-409547-2.12343-1.

Amabilino S, Bratholm LA, Bennie SJ, Vaucher AC, Reiher M, and Glowacki DR (2019). Training Neural Nets to Learn Reactive Potential Energy Surfaces Using Interactive Quantum Chemistry in Virtual Reality. J. Phys. Chem. A 123, 4486–4499. doi:10.1021/acs.jpca.9b01006. [PubMed: 30892040]

Anfinsen CB (1973). Principles that govern the folding of protein chains. Science 181, 223–30. doi:10.1126/science.181.4096.223. [PubMed: 4124164]

Ariga T, Macala LJ, Saito M, Margolis RK, Greene LA, Margolis RU, et al. (1988). Lipid composition of PC12 pheochromocytoma cells: characterization of globoside as a major neutral glycolipid. Biochemistry 27, 52–58. doi:10.1021/bi00401a010. [PubMed: 3349045]

Arnold K, Kiefer F, Kopp J, Battey JND, Podvinec M, Westbrook JD, et al. (2009). The Protein Model Portal. J. Struct. Funct. Genomics 10, 1–8. doi:10.1007/s10969-008-9048-5. [PubMed: 19037750]

Bateman A (2019). UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res 47, D506–D515. doi:10.1093/nar/gky1049. [PubMed: 30395287]

Behler J (2016). Perspective: Machine learning potentials for atomistic simulations. J. Chem. Phys 145, 170901. doi:10.1063/1.4966192. [PubMed: 27825224]

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. (2000). The Protein Data Bank. Nucleic Acids Res 28, 235–242. doi:10.1093/nar/28.1.235. [PubMed: 10592235]

Beuming T, Kniazeff J, Bergmann ML, Shi L, Gracia L, Raniszewska K, et al. (2008). The binding sites for cocaine and dopamine in the dopamine transporter overlap. Nat. Neurosci 11, 780–9. doi:10.1038/nn.2146. [PubMed: 18568020]

Beuming T, Shi L, Javitch JA, and Weinstein H (2006). A comprehensive structure-based alignment of prokaryotic and eukaryotic neurotransmitter/Na+ symporters (NSS) aids in the use of the LeuT structure to probe NSS structure and function. Mol. Pharmacol 70, 1630–42. doi:10.1124/mol.106.026120. [PubMed: 16880288]

Biasini M, Mariani V, Haas J, Scheuber S, Schenk AD, Schwede T, et al. (2010). OpenStructure: a flexible software framework for computational structural biology. Bioinformatics 26, 2626–8. doi:10.1093/bioinformatics/btq481. [PubMed: 20733063]

Biasini M, Schmidt T, Bienert S, Mariani V, Studer G, Haas J, et al. (2013). OpenStructure: An integrated software framework for computational structural biology. Acta Crystallogr. Sect. D Biol. Crystallogr 69, 701–709. doi:10.1107/S0907444913007051. [PubMed: 23633579]

Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, et al. (2017). The SWISS-MODEL Repository—new features and functionality. Nucleic Acids Res 45, D313–D319. doi:10.1093/nar/gkw1132. [PubMed: 27899672]

Bisgaard H, Larsen MAB, Mazier S, Beuming T, Newman AH, Weinstein H, et al. (2011). The binding sites for benztropines and dopamine in the dopamine transporter overlap. Neuropharmacology 60, 182–90. doi:10.1016/j.neuropharm.2010.08.021. [PubMed: 20816875]

Bohacek RS, McMartin C, and Guida WC (1996). The art and practice of structure-based drug design: A molecular modeling perspective. Med. Res. Rev 16, 3–50. doi:10.1002/(SICI)1098-1128(199601)16:1<3::AID-MED1>3.0.CO;2-6. [PubMed: 8788213]

Böhm G (1996). New approaches in molecular structure prediction. Biophys. Chem 59, 1–32. doi:10.1016/0301-4622(95)00120-4. [PubMed: 8867324]

Bonneau R, Tsai J, Ruczinski I, Chivian D, Rohl C, Strauss CEM, et al. (2001). Rosetta in CASP4: Progress in ab initio protein structure prediction. Proteins Struct. Funct. Genet 45, 119–126. doi:10.1002/prot.1170.

Borhani DW, and Shaw DE (2012). The future of molecular dynamics simulations in drug discovery. J. Comput. Aided. Mol. Des 26, 15–26. doi:10.1007/s10822-011-9517-y. [PubMed: 22183577]

Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al. (2009). CHARMM: The biomolecular simulation program. J. Comput. Chem 30, 1545–1614. doi:10.1002/jcc.21287. [PubMed: 19444816]

Burgin SR, Oramous J, Kaminski M, Stocker L, and Moradi M (2018). High school biology students use of visual molecular dynamics as an authentic tool for learning about modeling as a professional scientific practice. Biochem. Mol. Biol. Educ 46, 230–236. doi:10.1002/bmb.21113. [PubMed: 29457336]

Burkholder PR, Purser GH, and Cole RS (2009). Using Molecular Dynamics Simulation To Reinforce Student Understanding of Intermolecular Forces. J. Chem. Educ 85, 1071. doi:10.1021/ed085p1071.

Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Costanzo L Di, et al. (2019). Protein Data Bank: The single global archive for 3D macromolecular structure data. Nucleic Acids Res 47, D520–D528. doi:10.1093/nar/gky949. [PubMed: 30357364]

C. C. G. I. Molecular Operating Environment (MOE), 1010 Sher- booke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013.

Cheatham TE, and Roe DR (2015). The impact of heterogeneous computing on workflows for biomolecular simulation and analysis. Comput. Sci. Eng 17, 30–39. doi:10.1109/MCSE.2015.7.

Chen S, Wiewiora RP, Meng F, Babault N, Ma A, Yu W, et al. (2019). The dynamic conformational landscape of the protein methyltransferase SETD8. Elife 8. doi:10.7554/elife.45403.

Chou K-C (2004). Structural Bioinformatics and its Impact to Biomedical Science. Curr. Med. Chem 11, 2105–2134. doi:10.2174/0929867043364667. [PubMed: 15279552]

Coleman JA, and Gouaux E (2018). Structural basis for recognition of diverse antidepressants by the human serotonin transporter. Nat. Struct. Mol. Biol 25, 170–175. doi:10.1038/s41594-018-0026-8. [PubMed: 29379174]

Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE III, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi D, Gilson MK, Gohlke H, Goetz AW, Greene D, Harris R, Homeyer N, Izadi S, Kovalenko A, Kurtzman T, Lee TS, LeGra S, M. Y. D and A. K. P (2018). No Title.

Dalke A, and Schulten K (1997). Using Tcl for molecular visualization and analysis. Pac. Symp. Biocomput, 85–96. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9390282 [Accessed July 30, 2019].

Das R, and Baker D (2008). Macromolecular Modeling with Rosetta. Annu. Rev. Biochem 77, 363–382. doi:10.1146/annurev.biochem.77.062906.171838. [PubMed: 18410248]

Dawson P, Opacka-Juffry J, Moffatt JD, Daniju Y, Dutta N, Ramsey J, et al. (2014). The effects of benzofury (5-APB) on the dopamine transporter and 5-HT2-dependent vasoconstriction in the rat. Prog. Neuropsychopharmacol. Biol. Psychiatry 48, 57–63. doi:10.1016/j.pnpbp.2013.08.013. [PubMed: 24012617]

Doerr S, Giorgino T, Martínez-Rosell G, Damas JM, and De Fabritiis G (2017). High-Throughput Automated Preparation and Simulation of Membrane Proteins with HTMD. J. Chem. Theory Comput 13, 4003–4011. doi:10.1021/acs.jctc.7b00480. [PubMed: 28723224]

Dorn M, E Silva MB, Buriol LS, and Lamb LC (2014). Three-dimensional protein structure prediction: Methods and computational strategies. Comput. Biol. Chem 53, 251–276. doi:10.1016/j.compbiolchem.2014.10.001.

Durrant JD, and McCammon JA (2011). Molecular dynamics simulations and drug discovery. BMC Biol 9, 71. doi:10.1186/1741-7007-9-71. [PubMed: 22035460]

Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M, et al. (2006). “Comparative Protein Structure Modeling Using Modeller,” in Current Protocols in Bioinformatics (Hoboken, NJ, USA: John Wiley & Sons, Inc.), 5.6.1–5.6.30. doi:10.1002/0471250953.bi0506s15.

Galvelis R, Doerr S, Damas JM, Harvey MJ, and De Fabritiis G (2019). PlayMolecule Parameterize: a Scalable Molecular Force Field Parameterization Method Based on Quantum-Level Machine Learning. J. Chem. Inf. Model, acs.jcim.9b00439. doi:10.1021/acs.jcim.9b00439.

Glowacki BR, Bates P, Tew P, Thomas LM, Glowacki DR, Metatla O, et al. (2018). Sampling molecular conformations and dynamics in a multiuser virtual reality framework. Sci. Adv 4, eaat2731. doi:10.1126/sciadv.aat2731.

Gowers R, Linke M, Barnoud J, Reddy T, Melo M, Seyler S, et al. (2016). MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In Proceeds of the 15th Python in Science Conference (SciPy), 98–105. doi:10.25080/majora-629e541a-00e.

Guex N, Peitsch MC, and Peitsch MC (1997). SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis 18, 2714–2723. doi:10.1002/elps.1150181505. [PubMed: 9504803]

Guvench O, and MacKerell AD (2008). “Comparison of Protein Force Fields for Molecular Dynamics Simulations,” in (Humana Press), 63–88. doi:10.1007/978-1-59745-177-2_4.

Hammes GG (2002). Multiple Conformational Changes in Enzyme Catalysis† doi:10.1021/BI0260839.

Hansen FH, Skjørringe T, Yasmeen S, Arends NV, Sahai MA, Erreger K, et al. (2014). Missense dopamine transporter mutations associate with adult parkinsonism and ADHD. J. Clin. Invest 124, 3107–20. doi:10.1172/JCI73778. [PubMed: 24911152]

Harvey MJ, Giupponi G, and De Fabritiis G (2009). ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. J. Chem. Theory Comput 5, 1632–1639. doi:10.1021/ct9000685. [PubMed: 26609855]

Hedges LO, Mey ASJS, Laughton CA, Gervasio FL, Mulholland AJ, Woods CJ, et al. (2019). BioSimSpace: An interoperable Python framework for biomolecular simulation doi:10.21105/joss.01831.

Hou T, and Xu X (2004). Recent Development and Application of Virtual Screening in Drug Discovery: An Overview. Curr. Pharm. Des 10, 1011–1033. doi:10.2174/1381612043452721. [PubMed: 15078130]

Hu X, Wang Y, Hunkele A, Provasi D, Pasternak GW, and Filizola M (2019). Kinetic and thermodynamic insights into sodium ion translocation through the μ-opioid receptor from molecular dynamics and machine learning analysis. PLoS Comput. Biol 15, e1006689. doi:10.1371/journal.pcbi.1006689. [PubMed: 30677023]

Humphrey W, Dalke A, and Schulten K (1996). VMD-Visual Molecular Dynamics. J MolGraphics 14, 33–38.

Jiménez J, Doerr S, Martínez-Rosell G, Rose AS, and De Fabritiis G (2017). DeepSite: protein-binding site predictor using 3D-convolutional neural networks. Bioinformatics 33, 3036–3042. doi:10.1093/bioinformatics/btx350. [PubMed: 28575181]

Jiménez J, Škalič M, Martínez-Rosell G, and De Fabritiis G (2018). K DEEP : Protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks doi:10.1021/acs.jcim.7b00650.

Jing Z, Liu C, Cheng SY, Qi R, Walker BD, Piquemal J-P, et al. (2019). Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. Annu. Rev. Biophys 48, 371–394. doi:10.1146/annurev-biophys-070317-033349. [PubMed: 30916997]

Jurcik A, Bednar D, Byska J, Marques SM, Furmanova K, Daniel L, et al. (2018). CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. Bioinformatics 34, 3586–3588. doi:10.1093/bioinformatics/bty386. [PubMed: 29741570]

Kantcheva AK, Quick M, Shi L, Winther A-ML, Stolzenberg S, Weinstein H, et al. (2013). Chloride binding site of neurotransmitter sodium symporters. Proc. Natl. Acad. Sci. U. S. A 110, 8489–94. doi:10.1073/pnas.1221279110. [PubMed: 23641004]

Karplus M, and McCammon JA (2002). Molecular dynamics simulations of biomolecules. Nat. Struct. Biol 9, 646–652. doi:10.1038/nsb0902-646. [PubMed: 12198485]

Khelashvili G, Doktorova M, Sahai MA, Johner N, Shi L, and Weinstein H (2015a). Computational modeling of the N-terminus of the human dopamine transporter and its interaction with PIP2 -containing membranes. Proteins 83, 952–69. doi:10.1002/prot.24792. [PubMed: 25739722]

Khelashvili G, Stanley N, Sahai MA, Medina J, LeVine MV, Shi L, et al. (2015b). Spontaneous Inward Opening of the Dopamine Transporter Is Triggered by PIP2-Regulated Dynamics of the N-Terminus. ACS Chem. Neurosci 6, 1825–37. doi:10.1021/acschemneuro.5b00179. [PubMed: 26255829]

Kinjo AR, Bekker GJ, Suzuki H, Tsuchiya Y, Kawabata T, Ikegawa Y, et al. (2017). Protein Data Bank Japan (PDBj): Updated user interfaces, resource description framework, analysis tools for large structures. Nucleic Acids Res 45, D282–D288. doi:10.1093/nar/gkw962. [PubMed: 27789697]

Kitchen DB, Decornez H, Furr JR, and Bajorath J (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. Nat. Rev. Drug Discov 3, 935–949. doi:10.1038/nrd1549. [PubMed: 15520816]

Kniazeff J, Shi L, Loland CJ, Javitch JA, Weinstein H, and Gether U (2008). An intracellular interaction network regulates conformational transitions in the dopamine transporter. J. Biol. Chem 283, 17691–701. doi:10.1074/jbc.M800475200. [PubMed: 18426798]

Kota P (2007). GUIMACS - a Java based front end for GROMACS. In Silico Biol 7, 95–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17688433 [Accessed July 30, 2019]. [PubMed: 17688433]

Krishnamurthy H, and Gouaux E (2012). X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. Nature 481, 469–74. doi:10.1038/nature10737. [PubMed: 22230955]

Kristensen AS, Andersen J, Jørgensen TN, Sørensen L, Eriksen J, Loland CJ, et al. (2011). SLC6 neurotransmitter transporters: structure, function, and regulation. Pharmacol. Rev 63, 585–640. doi:10.1124/pr.108.000869. [PubMed: 21752877]

Kutzner C, Páll S, Fechner M, Esztermann A, De Groot BL, and Grubmüller H (2015). Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. J. Comput. Chem 36, 1990–2008. doi:10.1002/jcc.24030. [PubMed: 26238484]

Larsson P, Hess B, and Lindahl E (2011). Algorithm improvements for molecular dynamics simulations. Wiley Interdiscip. Rev. Comput. Mol. Sci 1, 93–108. doi:10.1002/wcms.3.

Laskowski RA (1995). SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. J Mol Graph 13, 307–308,323–330. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8603061 [Accessed July 31, 2019].

Loi B, Sahai M, De Luca MA, Shiref H, and Opacka-Juffry J (2020). The role of dopamine in the stimulant characteristics of novel psychoactive substances (NPS) – the case of desoxypipradol (2-DPMP). Front. Pharmacol (submitted)

Lundborg M, and Lindahl E (2015). Automatic GROMACS Topology Generation and Comparisons of Force Fields for Solvation Free Energy Calculations. J. Phys. Chem. B 119, 810–823. doi:10.1021/jp505332p. [PubMed: 25343332]

Lundquist K, Herndon C, Harty TH, and Gumbart JC (2016). Accelerating the use of molecular modeling in the high school classroom with VMD Lite. Biochem. Mol. Biol. Educ 44, 124–129. doi:10.1002/bmb.20940. [PubMed: 26751137]

Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, and Sherman W (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J. Comput. Aided. Mol. Des 27, 221–234. doi:10.1007/s10822-013-9644-8. [PubMed: 23579614]

Makarewicz T, and Kaźmierkiewicz R (2013). Molecular Dynamics Simulation by GROMACS Using GUI Plugin for PyMOL. J. Chem. Inf. Model 53, 1229–1234. doi:10.1021/ci400071x. [PubMed: 23611462]

Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, and Šali A (2000). Comparative Protein Structure Modeling of Genes and Genomes. Annu. Rev. Biophys. Biomol. Struct 29, 291–325. doi:10.1146/annurev.biophys.29.1.291. [PubMed: 10940251]

Martínez-Rosell G, Giorgino T, and De Fabritiis G (2017). PlayMolecule ProteinPrepare: A Web Application for Protein Preparation for Molecular Dynamics Simulations. J. Chem. Inf. Model 57, 1511–1516. doi:10.1021/acs.jcim.7b00190. [PubMed: 28594549]

Mayne CG, Saam J, Schulten K, Tajkhorshid E, and Gumbart JC (2013). Rapid parameterization of small molecules using the force field toolkit. J. Comput. Chem 34, 2757–2770. doi:10.1002/jcc.23422. [PubMed: 24000174]

McConkey BJ, Sobolev V, and Edelman M The performance of current methods in ligand–protein docking. Curr. Sci 83, 845–856. doi:10.2307/24107087.

McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernandez CX, et al. (2015). MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys. J 109, 1528–1532. doi:10.1016/j.bpj.2015.08.015. [PubMed: 26488642]

Mezei M (2003). A new method for mapping macromolecular topography. J. Mol. Graph. Model 21, 463–472. doi:10.1016/S1093-3263(02)00203-6. [PubMed: 12543141]

Michaud-Agrawal N, Denning EJ, Woolf TB, and Beckstein O (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem 32, 2319–2327. doi:10.1002/jcc.21787. [PubMed: 21500218]

Miller BT, Singh RP, Klauda JB, Hodoscek M, Brooks BR, Woodcock HL, et al. (2008). CHARMMing: a new, flexible web portal for CHARMM. J. Chem. Inf. Model 48, 1920–9. doi:10.1021/ci800133b. [PubMed: 18698840]

Milne JLS, Borgnia MJ, Bartesaghi A, Tran EEH, Earl LA, Schauder DM, et al. (2013). Cryo-electron microscopy - A primer for the non-microscopist. FEBS J 280, 28–45. doi:10.1111/febs.12078. [PubMed: 23181775]

Minor W, Dauter Z, and Jaskolski M (2016). The young person's guide to the PDB. Postepy Biochem 62, 242–249. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28132477 [Accessed July 30, 2019]. [PubMed: 28132477]

Mir S, Alhroub Y, Anyango S, Armstrong DR, Berrisford JM, Clark AR, et al. (2018). PDBe: Towards reusable data delivery infrastructure at protein data bank in Europe. Nucleic Acids Res 46, D486–D492. doi:10.1093/nar/gkx1070. [PubMed: 29126160]

Mondal S, Khelashvili G, Shi L, and Weinstein H (2013). The cost of living in the membrane: a case study of hydrophobic mismatch for the multi-segment protein LeuT. Chem. Phys. Lipids 169, 27–38. doi:10.1016/j.chemphyslip.2013.01.006. [PubMed: 23376428]

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem 19, 1639–1662. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.

Morris R (2019). No Title.

Munk C, Isberg V, Mordalski S, Harpsøe K, Rataj K, Hauser AS, et al. (2016). GPCRdb: the G protein-coupled receptor database - an introduction. Br. J. Pharmacol 173, 2195–2207. doi:10.1111/bph.13509. [PubMed: 27155948]

NCBI Resource Coordinators NR (2016). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 44, D7–19. doi:10.1093/nar/gkv1290. [PubMed: 26615191]

Nerenberg PS, and Head-Gordon T (2018). New developments in force fields for biomolecular simulations. Curr. Opin. Struct. Biol 49, 129–138. doi:10.1016/j.sbi.2018.02.002. [PubMed: 29477047]

O'Connor MB, Bennie SJ, Deeks HM, Jamieson-Binnie A, Jones AJ, Shannon RJ, et al. (2019). Interactive molecular dynamics in virtual reality from quantum chemistry to drug binding: An open-source multi-person framework. J. Chem. Phys 150. doi:10.1063/1.5092590.

Opella SJ (1997). NMR and membrane proteins. Nat. Struct. Biol 4 Suppl, 845–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9377156 [Accessed July 30, 2019]. [PubMed: 9377156]

Paissoni C, Spiliotopoulos D, Musco G, and Spitaleri A (2014). GMXPBSA 2.0: A GROMACS tool to perform MM/PBSA and computational alanine scanning. Comput. Phys. Commun 185, 2920–2929. doi:10.1016/j.cpc.2014.06.019.

Pearson WR (2014). "BLAST and FASTA Similarity Searching for Multiple Sequence Alignment," in (Humana Press, Totowa, NJ), 75–101. doi:10.1007/978-1-62703-646-7_5.

Pearson WR, and Lipman DJ (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. U. S. A 85, 2444–8. doi:10.1073/pnas.85.8.2444. [PubMed: 3162770]

Penmatsa A, Wang KH, and Gouaux E (2013a). X-ray structure of dopamine transporter elucidates antidepressant mechanism. Nature 503, 85–90. doi:10.1038/nature12533. [PubMed: 24037379]

Penmatsa A, Wang KH, and Gouaux E (2013b). X-ray structure of the dopamine transporter in complex with tricyclic antidepressant. Nature 503, 85. doi:10.1038/NATURE12533. [PubMed: 24037379]

Penmatsa A, Wang KH, and Gouaux E (2015). X-ray structures of Drosophila dopamine transporter in complex with nisoxetine and reboxetine. Nat. Struct. Mol. Biol 22, 506–508. doi:10.1038/nsmb.3029. [PubMed: 25961798]

Pérez A, Martínez-Rosell G, and De Fabritiis G (2018). Simulations meet machine learning in structural biology. Curr. Opin. Struct. Biol 49, 139–144. doi:10.1016/j.sbi.2018.02.004. [PubMed: 29477048]

Perez C, and Ziegler C (2013). Mechanistic aspects of sodium-binding sites in LeuT-like fold symporters. Biol. Chem doi:10.1515/hsz-2012-0336.

Petrey D, and Honig B (2005). Protein Structure Prediction: Inroads to Biology. Mol. Cell 20, 811–819. doi:10.1016/J.MOLCEL.2005.12.005. [PubMed: 16364908]

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. (2004). UCSF Chimera?A visualization system for exploratory research and analysis. J. Comput. Chem 25, 1605–1612. doi:10.1002/jcc.20084. [PubMed: 15264254]

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem 26, 1781–1802. doi:10.1002/jcc.20289. [PubMed: 16222654]

Pieper U, Eswar N, Davis FP, Braberg H, Madhusudhan MS, Rossi A, et al. (2006). MODBASE: a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res 34, D291–D295. doi:10.1093/nar/gkj059. [PubMed: 16381869]

Plante A, Shore DM, Morra G, Khelashvili G, and Weinstein H (2019). A machine learning approach for the discovery of ligand-specific functional mechanisms of GPCRs. Molecules 24. doi:10.3390/molecules24112097.

Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29, 845–854. doi:10.1093/bioinformatics/btt055. [PubMed: 23407358]

Quick M, Winther A-ML, Shi L, Nissen P, Weinstein H, and Javitch JA (2009). Binding of an octylglucoside detergent molecule in the second substrate (S2) site of LeuT establishes an inhibitor-bound conformation. Proc. Natl. Acad. Sci. U. S. A 106, 5563–8. doi:10.1073/pnas.0811322106. [PubMed: 19307590]

Raman P, Cherezov V, and Caffrey M (2006). The Membrane Protein Data Bank. Cell. Mol. Life Sci 63, 36–51. doi:10.1007/s00018-005-5350-6. [PubMed: 16314922]

Rao S, Klesse G, Stansfeld PJ, Tucker SJ, and Sansom MSP (2019). A heuristic derived from analysis of the ion channel structural proteome permits the rapid identification of hydrophobic gates. Proc. Natl. Acad. Sci 116, 201902702. doi:10.1073/pnas.1902702116.

Razavi AM, Khelashvili G, and Weinstein H (2017). A Markov State-based Quantitative Kinetic Model of Sodium Release from the Dopamine Transporter. Sci. Rep 7, 1–14. doi:10.1038/srep40076. [PubMed: 28127051]

Razavi AM, Khelashvili G, and Weinstein H (2018). How structural elements evolving from bacterial to human SLC6 transporters enabled new functional properties. BMC Biol 16, 1–14. doi:10.1186/s12915-018-0495-6. [PubMed: 29325545]

Reese JB, Bober SL, Daly MB, Program C, Chase F, Sciences B, et al. (2018). HHS Public Access 123, 4757–4763. doi:10.1002/cncr.31084.Talking.

Ribeiro JV, Bernardi RC, Rudack T, Stone JE, Phillips JC, Freddolino PL, et al. (2016). QwikMD - Integrative Molecular Dynamics Toolkit for Novices and Experts. Sci. Rep 6, 26536. doi:10.1038/srep26536. [PubMed: 27216779]

Romero R, Ramanathan A, Yuen T, Bhowmik D, Mathew M, Munshi LB, et al. (2019). Mechanism of glucocerebrosidase activation and dysfunction in Gaucher disease unraveled by molecular dynamics and deep learning. Proc. Natl. Acad. Sci 116, 5086–5095. doi:10.1073/pnas.1818411116. [PubMed: 30808805]

Roopra S, Knapp B, Omasits U, and Schreiner W (2009). jSimMacs for GROMACS: A Java Application for Advanced Molecular Dynamics Simulations with Remote Access Capability. J. Chem. Inf. Model 49, 2412–2417. doi:10.1021/ci900248f. [PubMed: 19852516]

Rudnick G, and Clark J (1993). From synapse to vesicle: The reuptake and storage of biogenic amine neurotransmitters. BBA - Bioenerg 1144, 249–263. doi:10.1016/0005-2728(93)90109-S.

Sahai MA, Davidson C, Khelashvili G, Barrese V, Dutta N, Weinstein H, et al. (2017). Combined in vitro and in silico approaches to the assessment of stimulant properties of novel psychoactive substances – The case of the benzofuran 5-MAPB. Prog. Neuro-Psychopharmacology Biol. Psychiatry 75, 1–9. doi:10.1016/j.pnpbp.2016.11.004.

Sahai MA, Davidson C, Dutta N, and Opacka-Juffry J (2018). Mechanistic Insights into the Stimulant Properties of Novel Psychoactive Substances (NPS) and Their Discrimination by the Dopamine Transporter—In Silico and In Vitro Exploration of Dissociative Diarylethylamines. Brain Sci 8, 63. doi:10.3390/brainsci8040063. [PubMed: 29642450]

Saier MH, Reddy VS, Tsu BV, Ahmed MS, Li C, and Moreno-Hagelsieb G (2016). The Transporter Classification Database (TCDB): Recent advances. Nucleic Acids Res 44, D372–D379. doi:10.1093/nar/gkv1103. [PubMed: 26546518]

Sali A, and Blundell TL (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol 234, 779–815. Available at: 10.1006/jmbi.1993.1626 [Accessed July 31, 2012]. [PubMed: 8254673]

Sayle RA, and Milner-White EJ (1995). RASMOL: biomolecular graphics for all. Trends Biochem. Sci 20, 374–376. doi:10.1016/S0968-0004(00)89080-5. [PubMed: 7482707]

Schrödinger, LLC: Portland, OR, 2007, >Web address>: www.schrodinger.com. Available at: www.schrodinger.com.

Schulz-Gasch T, and Stahl M (2003). Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. J. Mol. Model 9, 47–57. doi:10.1007/s00894-002-0112-y. [PubMed: 12638011]

Sellis D, Vlachakis D, and Vlassi M (2009). Gromita: a fully integrated graphical user interface to gromacs 4. Bioinform. Biol. Insights 3, 99–102. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20140074 [Accessed July 30, 2019]. [PubMed: 20140074]

Shan J, Javitch JA, Shi L, and Weinstein H (2011). The substrate-driven transition to an inward-facing conformation in the functional mechanism of the dopamine transporter. PLoS One 6, e16350. doi:10.1371/journal.pone.0016350. [PubMed: 21298009]

Shi L, Quick M, Zhao Y, Weinstein H, and Javitch JA (2008). The mechanism of a neurotransmitter:sodium symporter--inward release of Na+ and substrate is triggered by substrate in a second binding site. Mol. Cell 30, 667–77. doi:10.1016/j.molcel.2008.05.008. [PubMed: 18570870]

Shi L, and Weinstein H (2010). Conformational rearrangements to the intracellular open states of the LeuT and ApcT transporters are modulated by common mechanisms. Biophys. J 99, L103–5. doi:10.1016/j.bpj.2010.10.003. [PubMed: 21156121]

Shirts MR, Klein C, Swails JM, Yin J, Gilson MK, Mobley DL, et al. (2017). Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. J. Comput. Aided. Mol. Des 31, 147–161. doi:10.1007/s10822016-9977-1. [PubMed: 27787702]

Singh SK, Piscitelli CL, Yamashita A, and Gouaux E (2008). A competitive inhibitor traps LeuT in an open-to-out conformation. Science 322, 1655–61. doi:10.1126/science.1166777. [PubMed: 19074341]

Singh SK, Yamashita A, and Gouaux E (2007). Antidepressant binding site in a bacterial homologue of neurotransmitter transporters. Nature 448, 952–6. doi:10.1038/nature06038. [PubMed: 17687333]

Stolzenberg S, Quick M, Zhao C, Gotfryd K, Khelashvili G, Gether U, et al. (2015). Mechanism of the association between Na+ binding and conformations at the intracellular gate in neurotransmitter:sodium symporters. J. Biol. Chem 290, 13992–14003. doi:10.1074/jbc.M114.625343. [PubMed: 25869126]

Tagari M, Newman R, Chagoyen M, Carazo JM, and Henrick K (2002). New electron microscopy database and deposition system. Trends Biochem. Sci 27, 589. doi:10.1016/S0968-0004(02)02176-X. [PubMed: 12417136]

Taly A, Nitti F, Baaden M, and Pasquali S (2019). Molecular modelling as the spark for active learning approaches for interdisciplinary biology teaching. Interface Focus 9, 20180065. doi:10.1098/rsfs.2018.0065. [PubMed: 31065338]

The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC

Trott O, and Olson AJ (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem 31, 455–61. doi:10.1002/jcc.21334. [PubMed: 19499576]

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, et al. (2008). BioMagResBank. Nucleic Acids Res 36, D402–8. doi:10.1093/nar/gkm957. [PubMed: 17984079]

Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem 31, 671–90. doi:10.1002/jcc.21367. [PubMed: 19575467]

Vanommeslaeghe K, MacKerell AD, and Jr (2012a). Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. J. Chem. Inf. Model 52, 3144–54. doi:10.1021/ci300363c. [PubMed: 23146088]

Vanommeslaeghe K, Raman EP, MacKerell AD, and Jr (2012b). Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. J. Chem. Inf. Model 52, 3155–68. doi:10.1021/ci3003649. [PubMed: 23145473]

Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, and Taylor RD (2003). Improved protein-ligand docking using GOLD. Proteins Struct. Funct. Bioinforma 52, 609–623. doi:10.1002/prot.10465.

Verdonk ML, Giangreco I, Hall RJ, Korb O, Mortenson PN, and Murray CW (2011). Docking performance of fragments and druglike compounds. J. Med. Chem 54, 5422–5431. doi:10.1021/jm200558u. [PubMed: 21692478]

Vroling B, Sanders M, Baakman C, Borrmann A, Verhoeven S, Klomp J, et al. (2011). GPCRDB: an information system for G protein-coupled receptors. Nucleic Acids Res 39, D309–D319. doi:10.1093/nar/gkq1009. [PubMed: 21045054]

Wang J, Olsson S, Wehmeyer C, Pérez A, Charron NE, De Fabritiis G, et al. (2019). Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. ACS Cent. Sci 5, 755–767. doi:10.1021/acscentsci.8b00913. [PubMed: 31139712]

Wang J, Wolf RM, Caldwell JW, Kollman PA, and Case DA (2004). Development and testing of a general amber force field. J. Comput. Chem 25, 1157–1174. doi:10.1002/jcc.20035. [PubMed: 15116359]

Wang KH, Penmatsa A, and Gouaux E (2015). Neurotransmitter and psychostimulant recognition by the dopamine transporter. Nature 521, 322–7. doi:10.1038/nature14431. [PubMed: 25970245]

Wass MN, Kelley LA, and Sternberg MJE (2010). 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Res 38, W469–73. doi:10.1093/nar/gkq406. [PubMed: 20513649]

Yamashita A, Singh SK, Kawate T, Jin Y, and Gouaux E (2005). Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. Nature 437, 215–23. doi:10.1038/nature03978. [PubMed: 16041361]

Yesselman JD, Price DJ, Knight JL, Brooks CL, and III (2012). MATCH: An Atom- Typing Toolset for Molecular Mechanics Force Fields. J. Comput. Chem 33, 189. doi:10.1002/JCC.21963. [PubMed: 22042689]

Zhang Q, and Cherezov V (2019). Chemical tools for membrane protein structural biology. Curr. Opin. Struct. Biol doi:10.1016/J.SBI.2019.06.002.

Zhao C, Stolzenberg S, Gracia L, Weinstein H, Noskov S, and Shi L (2012). Ion-controlled conformational dynamics in the outward-open transition from an occluded state of LeuT. Biophys. J 103, 878–88. doi:10.1016/j.bpj.2012.07.044. [PubMed: 23009837]

Zhao Y, Terry DS, Shi L, Quick M, Weinstein H, Blanchard SC, et al. (2011). Substrate-modulated gating dynamics in a Na + -coupled neurotransmitter transporter homologue. Nature 474, 109–113. doi:10.1038/nature09971. [PubMed: 21516104]

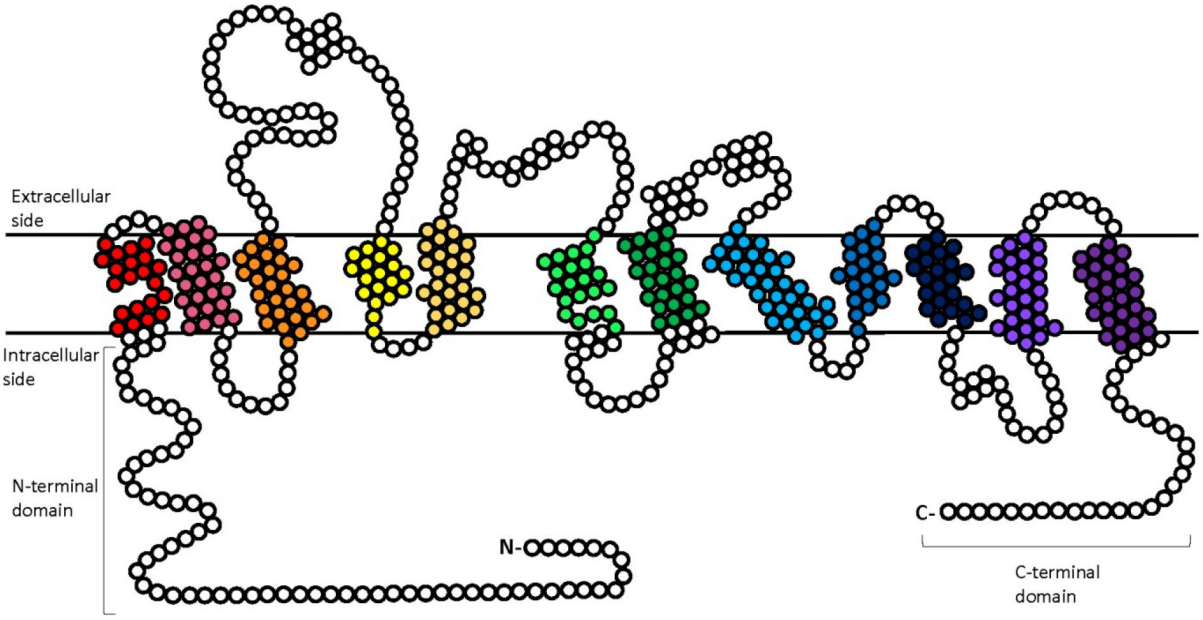

Figure 1. Two-dimensional schematic representation of the topology of the human dopamine transporter (hDAT).

Colored regions indicate the transmembrane (TM) domains that are embedded in the lipid bilayer. Areas outside of this region appear as not coloured and are either extracellular or intracellular loops or the N- and C- termini that both reside on the intracellular side.

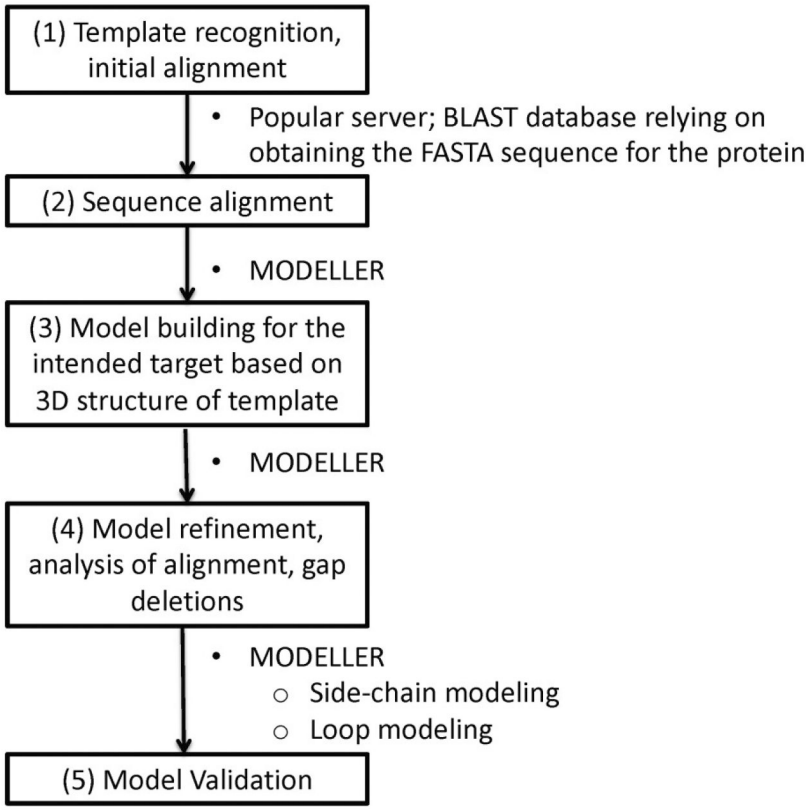

Figure 2. A schematic of the multi-step process for molecular docking

(1) simple ‘template’ recognition; (2) sequence alignment; (3) model building for the intended ‘target’, which is based on the 3D structure of the ‘template’; (4) model refinement, analysis of alignments, gap deletions as well as additions and finally (5) model validation.

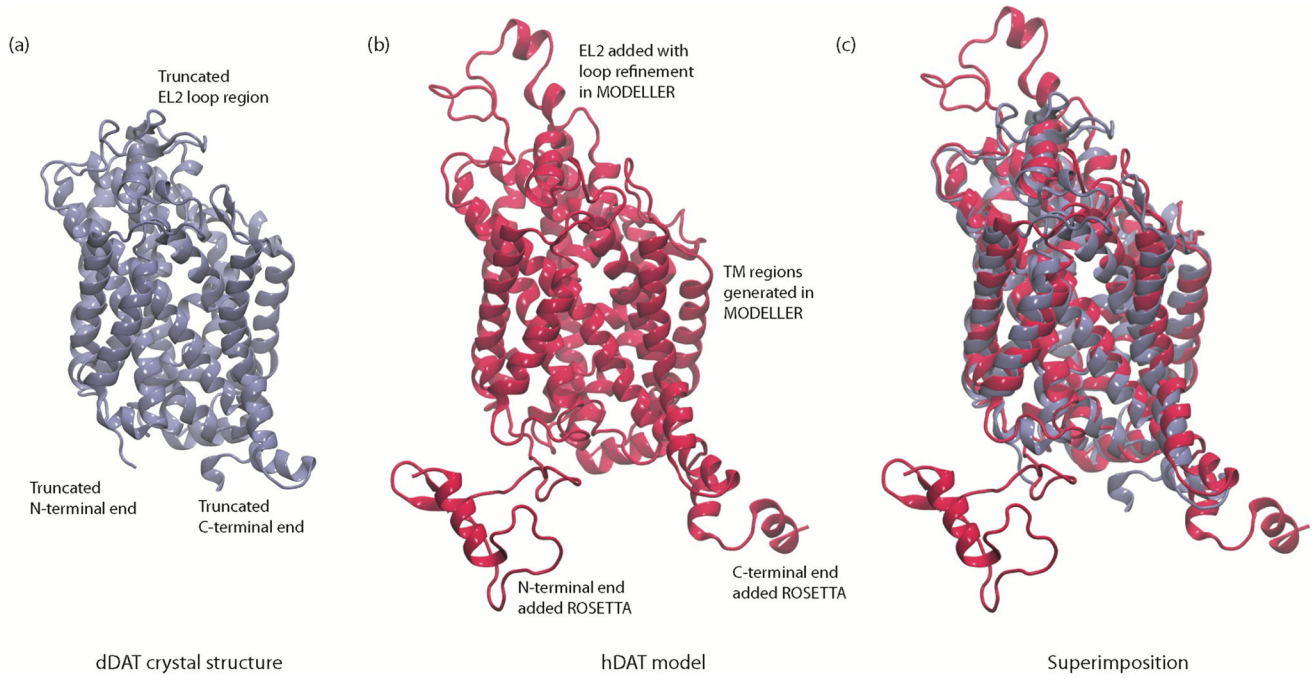

Figure 3. hDAT homology model predicted from the dDAT crystal structure

(a) The dDAT crystal structure (Protein Database (PDB) ID: 4M48) exhibiting the LeuT-like structure fold. (b) A model of the hDAT tertiary structure based on the alignment in [110] with predicted extracellular loop 2 (EL2) and N- and C- termini regions indicated. (c) Superimposition of (a) and (b).

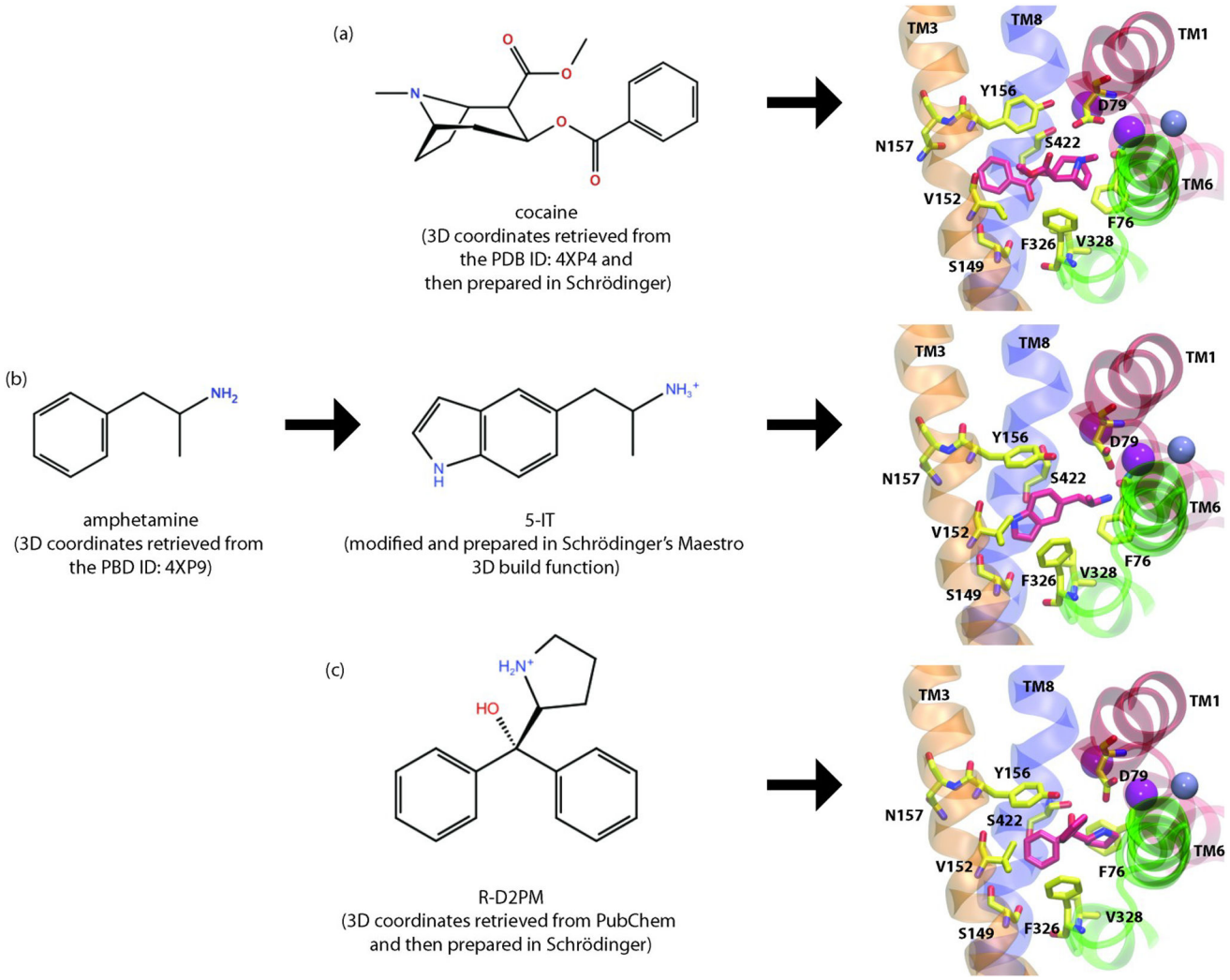

Figure 4. Workflow for preparation for IFD docking of three compounds.

(a) The 3D coordinates of cocaine were retrieved from the PBD (PDB ID: 4XP4) and then prepared in Schrödinger to add hydrogens and a positive charge to the nitrogen on the tropane group, (b) 5-IT was built in Schrödinger by modifying the 3D coordinates of amphetamine (PDB ID: 4XP9); hydrogens were added and a positive charge was added to the amine nitrogen (c) the 3D coordinates for D2PM were retrieved from PubChem: Compound CID: 704537; hydrogens were added and a positive charge was added to the nitrogen in the pyrrole ring. In the last panel for each compound you can see that each of these distinct ligands (shown in red) occupies a binding pocket that is deeply buried in the transporter structure and overlaps with the binding site of the substrate dopamine. Selected central binding site residues from each ligand are shown in yellow and labelled respectively. The internal sodium and chloride ions are shown in magenta and purple, respectively.

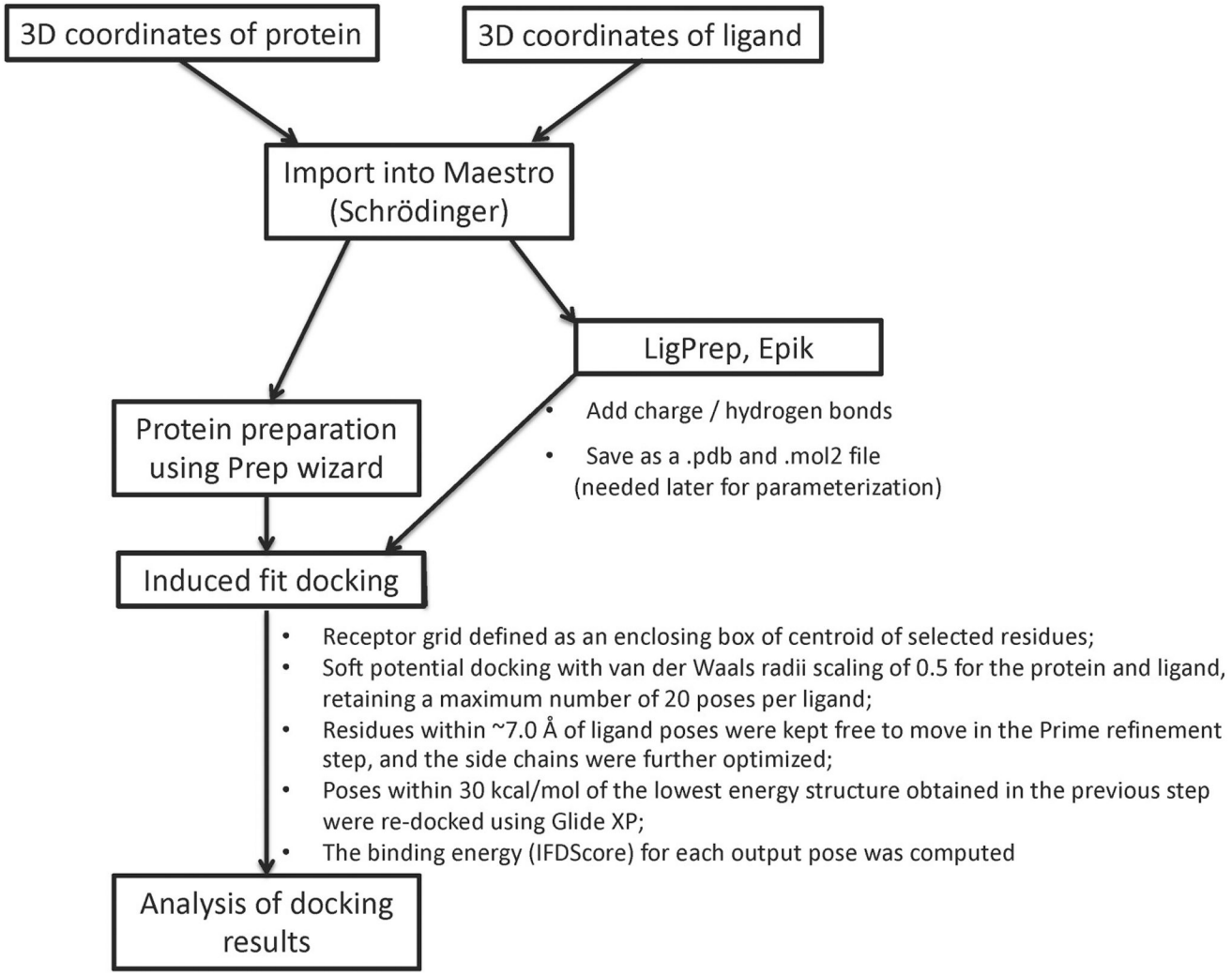

Figure 5. A schematic workflow of the docking procedure in Schrödinger.

Individual descriptions for the ligand and protein preparation as well as details for the IFD protocol are given. The 3D coordinates for the protein are either obtained from the PDB or homology modeling, while the 3D coordinates for the ligand are either obtained from PubChem or built manually (Figure 4).