Computational and Structural Biotechnology Journal 19 (2021) 3708–3719

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Review

Structure-based *in silico* approaches for drug discovery against *Mycobacterium tuberculosis*

Alexander D.H. Kingdon, Luke J. Alderwick *

Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom

ARTICLE INFO

Article history:
Received 23 April 2021
Received in revised form 22 June 2021
Accepted 22 June 2021
Available online 24 June 2021

Keywords:
Drug discovery
Mycobacterium tuberculosis
In silico
Docking
Machine learning

ABSTRACT

Mycobacterium tuberculosis is the causative agent of TB and was estimated to cause 1.4 million death in 2019, alongside 10 million new infections. Drug resistance is a growing issue, with multi-drug resistant infections representing 3.3% of all new infections, hence novel antimycobacterial drugs are urgently required to combat this growing health emergency. Alongside this, increased knowledge of gene essentiality in the pathogenic organism and larger compound databases can aid in the discovery of new drug compounds. The number of protein structures, X-ray based and modelled, is increasing and now accounts for greater than > 80% of all predicted *M. tuberculosis* proteins; allowing novel targets to be investigated. This review will focus on structure-based *in silico* approaches for drug discovery, covering a range of complexities and computational demands, with associated antimycobacterial examples. This includes molecular docking, molecular dynamic simulations, ensemble docking and free energy calculations. Applications of machine learning onto each of these approaches will be discussed. The need for experimental validation of computational hits is an essential component, which is unfortunately missing from many current studies. The future outlooks of these approaches will also be discussed.

Contents

Abbreviations: cMD, Classical Molecular Dynamic; cryo-EM, cryogenic electron microscopy; CV, collective variable; LIE, Linear Interaction Energy; MD, Molecular Dynamic; MDR, multi-drug resistant; MMPB(GB)SA, Molecular Mechanics with Poisson Boltzmann (or generalised Born) and Surface Area solvation; Mt, *Mycobacterium tuberculosis*; ns, nanosecond; PTC, peptidyl transferase centre; RMSD, root-mean square-deviation; Tuberculosis, TB.

* Corresponding author.
E-mail address: l.alderwick@bham.ac.uk (L.J. Alderwick).

https://doi.org/10.1016/j.csbj.2021.06.034
2001-0370/

1. Introduction

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB), contributed to an estimated 1.4 million deaths in 2019, with approximately 10 million new infections in the same year [1]. Also, it is predicted that *M. tuberculosis* latently infects approximately one-third of the world's population [1,2] and resistance to the current drug-treatment regime is also on the rise, with 3.3% of new cases being multi-drug resistant (MDR); this number increases drastically to 17.7% for previously treated infections [1]. If this global epidemic is to be stopped, it requires the identification and exploitation of novel drug targets, alongside other preventative approaches and treatment options [1,3].

The development of new antimycobacterial drugs is particularly challenging, in part due to the unique adaptations that *M. tuberculosis* employs which are not present in other bacterial species. The unique mycobacterial cell envelope structure, composed of modified peptidoglycan, mycolic acids and arabinogalactan, provides a waxy hydrophobic barrier which prevents penetration of several antibiotics [4,5]. In addition, *M. tuberculosis* can enter a hypoxia-induced latent growth-state, characterised by reduced metabolic activity [2,3]. This has been coupled to lower efficacy of several antibiotics, including isoniazid and beta-lactams, as their killing activity relies on active growth or metabolism [6].

The four front-line antimycobacterial drugs in current use (ethambutol, isoniazid, pyrazinamide and rifampicin), were all discovered and developed through traditional compound screening experimental methodologies [7–9]. These studies resulted in the development of ethambutol from polyamines, isoniazid and pyrazinamide from nicotinamide and rifampicin from rifamycin [7–9]. In addition, drug repurposing studies have led to the identification of many second-line antimycobacterial drugs, including fluoroquinolones, linezolid and clofazimine [10]. Repurposed drugs also represent one-third of all the new TB drugs currently in clinical trials [11]. These phenotypic drug-to-target approaches have continued to be used to successfully identify new drugs, such as delamanid and pretomanid from nitroimidazooxazole [12,13]. However, the screening of large compound libraries is financially expensive and high re-discovery rates coupled with fewer novel hits per high-throughput screen, demonstrates that alternative approaches are required for the discovery and development of new anti-TB therapies. In this regard, the use of computational approaches for initial virtual screening, followed by concurrent experimental and computational analysis has the potential to reduce costs and increase the quality of compounds taken forward towards the developmental pipeline.

To date, two conventional computational approaches are utilised for drug discovery/repurposing projects which are either, ligand-based [14] or structure-based [15–17]. The former primarily focusses on data mining of chemical structures and associated biological activity, while the latter is concerned with the interactions of potential drugs with targets of biological interest. Both approaches aim to find chemical structures which are the most active against a particular target/organism, however, structure-based approaches have greater potential to find novel chemical structures [18].

This review focuses upon structure-based methods related to anti-TB drug discovery efforts. Several different *in silico* approaches will be covered, across a range of complexities and computational demands, and recent examples of their application to target *M. tuberculosis* highlighted. The application of machine-learning on several of these approaches will also be covered, alongside the increased need to perform experimental validation on computational predictions. However, before structure-based approaches can be undertaken, the selection of a target of interest and a chemical compound library to screen is essential [16,17], hence, these will be briefly covered.

2. Protein target selection and structures

Drug target selection is a major challenge in the field of drug discovery, as it usually requires a detailed understanding of the biological role and molecular genetics associated with genes that are required for bacterial survival or establishment of infection. Therefore, a common approach of target-based drug discovery research is to focus on only essential *M. tuberculosis* genes. In this regard, several highly useful studies detailing *M. tuberculosis* gene essentiality have provided guidance to the field [19,20].

Once a protein drug-target has been identified, protein structures required for downstream screening can be obtained in several ways, including crystallographic methods, cryogenic electron microscopy (cryo-EM) and homology modelling. Crystallographic methods are labour intensive and produce an average protein structure, normally utilising X-rays to solve experimentally obtained protein crystals. Cryo-EM is a more recent development, which rapidly freezes proteins in aqueous environments, trapping them in ice crystals, and then uses transmission electron microscopy to solve the structures. This allows structural determination of proteins which do not readily crystallise, including membrane proteins. Homology modelling is a computational approach which uses the primary protein sequence and known crystal structures of homologous proteins, to generate the most likely protein structure. In addition, newer *ab initio* methods are rapidly increasing in accuracy, such as AlphaFold [21], and these deep learning approaches may dominate computational methods in the future. However, crystallographic methods are currently still the preferred approach due to their accuracy and experimental validation.

Crystal structures are currently available for a large number of *M. tuberculosis* proteins within Protein Data Bank [22], with 2,630 structures based on X-ray diffraction and a further 41 derived from electron microscopy. The great quantity of *M. tuberculosis* protein structures is in part due to the effort of the TB-structural genomics consortium [23]. If the crystal structure is not available, then the SWISS-MODEL repository [24] provides an alternative, containing a collection of modelled structures, using ProMod3 to perform homology modelling [25]. *M. tuberculosis* is one of the core species for which new models are generated and updated on a weekly basis, to account for new crystal structures, and currently 3,366 protein-encoding sequences out of a predicted 3,993 have modelled structures [24]. Alternative services for structural modelling

include I-TASSER [26] and Phyre2 [27]; they utilise different methods to ProMod3, and hence may provide alternative structures. The use of both crystal structures and models can allow the majority of *M. tuberculosis* proteins to be utilised for drug screening and further structural studies.

3. Chemical libraries – virtual and tangible

Chemical libraries used for *in silico* screening can either be taken from databases containing modelled chemical structures or derived from tangible libraries normally utilised for high-throughput screening. The latter provides a pre-selected set of compounds, whereas the former approach generally requires selection of a sub-set of compounds. Hence, the compound libraries chosen can focus on either a broad-range of physiochemical properties, structural diversity, TB-specific compounds, or compounds for drug-repurposing. The type of compounds selected may also be influenced by the target of interest or the aim of the *in silico* screen, such as novel compound identification or drug-repurposing.

If the main goal of a virtual screening campaign is to identify a novel chemical compound with a strong predicted binding affinity, then large compound databases will contain the greatest structural diversity. These include the ZINC database, approximately 230 million compounds [28], the ChEMBL database, approximately 2 million compounds [29] and the Enamine REAL database, approximately 1.4 billion compounds [30], among many others [31]. These large general compound databases may be ideal for searching for unique compound structures; however, they have two main limitations; 1. their size might make them too computationally demanding for screening, 2. antimycobacterial drugs have been shown to possess different physiochemical properties compared to 'typical' drugs [32]. As these compound databases are normally generated around 'typical' drug physiochemical properties, they may not contain many compounds which are suitable for targeting *M. tuberculosis*. Hence, TB-specific compound libraries may provide a more focussed effort. These include the CDD-TB library of compounds, approximately 7,000 compounds [33] and the WuXi antituberculosis library, approximately 10,000 compounds [34]. The former is a virtual set of chemical compounds, whereas the latter is a physical compound library typically used in high-throughput screens which could be adapted for *in silico* use [34].

Drug repurposing provides an alternative strategy for drug-screening studies, utilising specific compound libraries containing only clinically approved drugs, which are also commercially available. One obvious inherent limitation of these focussed drug libraries is reduced compound structural diversity. However, since all drugs in clinical use come with an abundance of *in vivo* clinical data (absorption, distribution, metabolism, excretion, and toxicity etc), one clear potential benefit of screening repurposed libraries is the rapid expedition of hits towards the clinic. Examples of such libraries include the Prestwick library, 1,520 compounds [35] and the Broad Institute drug repurposing library, 6,798 compounds [36]. For the former, the molecules' 3D structures are available from the ZINC database [28], whereas for the later, the SMILES strings can be obtained from the Drug Repurposing Hub [36].

Selecting a suitable compound library for virtual screening largely depends on which chemical properties are the focus of the study and the parameters imposed to select for hits. Once a target protein and compound library have been selected, a variety of *in silico* approaches can be employed to identify compounds that bind the protein target.

4. Molecular docking

4.1. Molecular docking approaches

Molecular docking is a computational process which aims to study the interactions occurring between a protein and compound of interest. Various docking programmes exist based on this premise, but they all use different algorithms to try and fit a compound into the binding site of a protein [15]. They generate several potential conformations, or poses, which are then scored, and the top scoring poses, in a ranked order, are the main output [15]. These programmes focus on the compounds' flexibility while treating the protein as completely rigid [37,38]; this approach saves computational power, but also decreases the accuracy [39].

The range of docking programmes available includes: GOLD [40,41]; AutoDOCK Vina [42]; Glide [43,44] and PharmScreen [45], amongst several others [15,46,47]. These programmes display diversity in both their search and score methods. They also differ in the their availability due to licensing [48], with the majority being restricted to commercial use only.

As each of these programmes use different approaches for generating docking poses, their efficacy can be difficult to compare. Most programmes have been shown to generate binding site poses similar to crystallographic structures, but with varying abilities to reproduce binding affinity data [47]. A previous study found GOLD and Glide to be superior to other docking programmes [46], however many docking programmes, including AutoDOCK Vina and PharmScreen were developed more recently than this study was performed. The success of molecular docking can also be dependent on the protein–ligand pair being simulated and CNN_Dock-Bench, a deep-learning based programme, has been developed to predict which molecular docking programme would be most successful for correct pose predictions [49].

4.2. Molecular docking applied to M. tuberculosis

Molecular docking is the most widely used computational approach for virtual screening against *M. tuberculosis* proteins and has resulted in numerous published studies that are summarised in Table 1. The majority of proteins targeted by this approach represent proteins encoded by essential genes, based on the Himar1 transposon mutagenesis study of DeJesus (2017) [19]; with the exceptions being antigen 85c, BioA, EthR, NarL and LipU. However, non-essentiality assigned by this approach is based on *in vitro* growth and does not guarantee these genes are non-essential *in vivo* [50]. For example, the NarL protein is required for anaerobic survival during infection and BioA is essential for biotin synthesis during *M. tuberculosis* latency [51,52]. In addition, the EthR protein is involved in ethionamide resistance and hence survival during drug treatment [53].

Focussing on the types of proteins being targeted, the majority are involved in either intermediary metabolism or lipid metabolism within *M. tuberculosis*. Then regulatory proteins and cell wall regulator proteins make up the remainder of the targets. InhA and DprE1 have been the targets of the most virtual screening campaigns, with at least three each so far. InhA is the eventual target of both isoniazid and ethionamide, following activation of these prodrugs [54] and hence InhA represents a validated target, whose inhibition has an *in vivo* impact on *M. tuberculosis* survival. While DprE1 is targeted by several antimycobacterial drugs in the current anti-TB development pipeline, and hence it represents another validated target [1,10].

The majority of the studies outlined in Table 1 have taken compounds from general chemical databases containing millions of compounds. While the remainder are equally split between TB-

Table 1
Molecular Docking virtual screening studies against M. tuberculosis proteins.

Docking Programme	TB Protein	Protein function	Compounds Screened	Computational follow-up	Experimental Validation	Reference
AutoDOCK Vina	MurB	Peptidoglycan biosynthesis	FDA-approved compounds:Drug Bank (1932)eLEA3D (1852)	MD simulations and MMPBSA energy calculations	No	[68]
	MurE	Peptidoglycan biosynthesis				
	InhA	Mycolic acid biosynthesis	5.6 million compounds from:NCI; Enamine; Asinex; Chembridge & Vitas-M Labs	No	For InhA hits only – *in vivo* against *M. tuberculosis* H37Rv & *in vitro* against InhA protein + follow-up [69]	[70]
	DHFR	Nucleic acid biosynthesis				
	FabG	Mycolic acid biosynthesis				
	Cyclophilin A	Protein folding				
	DprE1	Arabinogalactan biosynthesis	ChemDiv dataset – 135,755 compounds	ADMET predictions	No	[71]
	PanK	Coenzyme A biosynthesis	78 phytochemicals	No	No	[72]
	DprE1	Arabinogalactan biosynthesis				
	PknB	Protein kinase				
	KasA	Mycolic acid biosynthesis				
AutoDOCK 3.05	Isocitrate lyase	Glyoxylate bypass	Malaysian Local Natural Compound Database – 3,000 compounds	MD simulations – then ensemble docking	*In vivo* testing against *M. smegmatis.*	[73]
AutoDOCK 4.0	RmlD	Carbohydrate biosynthesis	Super Natural-II database – 570 compounds	MD simulations – then ensemble docking	No	[74]
CDOCKER	BioA	Biotin biosynthesis	Enamine REAL database – 4.5 million compounds	ADMET predictions	*In vivo* confirmation against *M. tuberculosis* H37Rv	[51]
	LdtB	Peptidoglycan biosynthesis				[75]
FRIGATE	Antigen 85c	Lipid metabolism	ZINC database – 2 million compounds	No	NMR binding against Antigen 85c and MIC against M. smegmatis	[76]
Glide	LipU	Lipid hydrolysis	6,282 FDA-approved drugs	MD simulations and Prime MMGBSA calculations	No	[77]
	AroB	Shikimate pathway	1,082 compounds preselected from DrugBank database	MD simulations	No	[78]
	GlnA1	Glutamine biosynthesis	ChEMBL antimycobacterials – 56,400; FDA-approved drugs – 1596; natural products – 419 & phytochemicals – 918.	MD simulations and MMPBSA calculations	No	[79]
	DprE1	Arabinogalactan biosynthesis	30,789 ChEMBL antimycobacterial compounds	ADME predictions; MD simulations and MMPBSA & MMGBSA calculations	No	[80]
	PknA	Protein kinase	3,176 FDA-approved drugs	MD simulations and MMPBSA calculations	No	[81]
	NarL	Nitrate regulation	4,754 ChEMBL antimycobacterial compounds	MD simulations and MMPBSA calculations	No	[52]
	InhA	Mycolic acid biosynthesis	1,026 compounds pre-selected from Maybridge database	MD simulations	*In vivo* confirmation against *M. bovis BCG*	[82]
	MraY	Peptidoglycan biosynthesis	10,500 compounds from Asinex database	MD simulations and prime MMGBSA calculations	No	[83]
GOLD	EthR	Transcriptional regulator	*Drugs Now* subset of ZINC database – 409,201 compounds	Follow-up [84] performed MD simulations & binding energy calculations	*In vivo* confirmation against *M. tuberculosis* H37Rv & crystal structures of compound-EthR complexes	[53]
LibDock	KasA	Mycolic acid biosynthesis	Top 50 diverse compounds selected by machine learning[85]	No	*In vitro* binding to purified KasA protein	[86]
AutoDOCK, GOLD, FlexX, Surflex Dock	InhA	Mycolic acid biosynthesis	ZINC database – 999,853 compounds	Toxicity predictions	*In vitro* inhibition assay against InhA	[87]
GOLD & Plants	MbtI	Mycobactin synthesis	2,050 compounds pre-selected from Enamine database	MD simulations	*In vitro* inhibition assay against MtbI & *in vivo* MIC against *M. tuberculosis* H37Rv	[88]
GOLD & RF-Score	AroQ	Shikimate pathway	4379 diverse compounds, selected from 9 million	No	*In vitro* inhibition assay against AroQ	[89]
AutoDOCK 4.2 & Surflex Dock	FtsZ	Cell division	67 trisubstituted benzimidazoles analogues	MD simulations	No	[90]

specific, natural product and drug repurposing focussed libraries. Hence, completely novel chemical discovery remains the focus of these preliminary drug discovery campaigns. For the majority of these studies (Table 1) the apparent absence of follow-on experimental data (either *in vitro* or *in vivo*) confirming compound bioactivity is an obvious limitation that will prevent these predicted compounds from being taken forward.

4.3. Machine learning applied to molecular docking

A wider issue for the use of molecular docking is the propensity for false positives to be generated when screening compound libraires. This may be linked to many programmes' scoring functions containing inherent biases for large molecules [55]. False negatives can also occur if drug binding requires conformational changes to the protein [56], as molecular docking is focussed on compound flexibility. In contrast, machine learning approaches have been successfully used to predict known protein–ligand affinities with higher accuracy than conventional molecular docking [55,57]. Many comprehensive reviews have been undertaken on the application of machine learning on the scoring functions within molecular docking [58–60], and hence this information won't be repeated herein.

Several freely available machine-learning programmes exist as open-access resources, these target either the scoring-functions or pose generation of molecular docking programmes [57–59,61]. NNScore2.0 [62], SIEVE-Score [63] and RF-Score-VS [64] are some examples of machine-learning programmes that focus on scoring-functions, all of which are trained on ligand-receptor binding characteristics and associated K_d values. In turn, these programmes re-score pre-generated docked poses, thus providing an alternative ranking of all compounds being investigated. These programmes are particularly interesting as they can be applied retroactively to previous docking campaigns, to assess whether promising compounds have been missed. Currently, there is little wide-spread adoption of these new machine learning methods in the field of drug discovery. The lack of uptake is likely due to a lack of immediate accessibility of these programmes, compared to more established molecular docking programmes, and required computational knowledge represents a barrier to entry for programme implementation. These programmes have also been trained on specific sets of drug-protein binding data, SIEVE-Score and RF-Score-VS were trained on the Directory of Useful Decoys, Enhanced dataset [65], while NNScore was trained on 4,141 protein–ligand complexes selected from the protein data bank [66]. Training on specific drug-protein binding data can allow high quality predictions, if the downstream application's proteins and potential drugs align with the training set [59]. However, if the properties of the training and application proteins/drugs are different, then user-supplied input data to train the models could be utilised to increase the accuracy of predictions [58,59].

Machine learning has been applied to a docking-based approach targeting the ribosomal peptidyl transferase centre (PTC) of *M. tuberculosis* [67]. This study provided validation for machine learning to be applied for drug discovery, as the generated model predicted binding efficiency to the PTC which matched experimental results [67]. The model was trained using the pose outputs from molecular docking, focussing on the eleven-atom core structure of all the phenylthiazole compounds screened. One limitation of this study was the lack of testing for the model's ability to predict activity of novel untested structures.

To date, no scoring-function focussed machine learning study, for *M. tuberculosis* protein molecular docking, has been published in the literature, thereby highlighting a largely unexplored area of research. Providing suitable training sets become available, these machine learning approaches have the potential to greatly improve the efficiency of *M. tuberculosis* virtual screening studies.

5. Molecular dynamic simulations

5.1. Classical molecular dynamic simulations

Molecular dynamic (MD) simulations are a computational approach to model the motions of atoms over short nano or microsecond timeframes [91–93]. These simulations require an initial input structure such as a crystal structure or homology model, which specifies where all the atoms are and at what velocities they are currently travelling. Then the simulation uses predetermined force fields, CHARMM [94] and AMBER [95] force fields are commonly used [92,96], to calculate the forces acting on each atom. This information is used to solve the Newtonian equations of motion generating a trajectory and the atoms are moved to these new positions. This movement normally represents a timestep of two femtoseconds. This process is repeated for each new set of atom positions until a pre-set number of timesteps has elapsed [93]. For a short ten nanosecond simulation, five million timesteps need to be calculated. At specific timesteps, for example every 5,000 timesteps, the atom positions and trajectories can be saved as trajectory frames. This process allows a subset of the total number of timesteps, rather than the whole simulation output, to be processed during downstream analysis. Several programmes exist to undertake these MD simulations, including NAMD [92], GROMACS [97], AMBER [98] and CHARMM [99], with the first two being utilised the most. The necessary input files for these programmes can be generated using the CHARMM-GUI webserver [100] or manually using VMD [101], among several other resources [93].

Despite there being several programmes in currently use to perform MD simulations, no one programme is favoured in the published literature [96]. One main distinguishing feature of these MD simulation programmes is whether the license is free or commercial [48]. Generally, MD programmes are found to output similar results, despite them utilising different force fields and algorithms [96,102]. The efficacy of each programme will depend on the system being simulated, the parameters required, the type of MD simulation being undertaken and the hardware available. For example, GROMACS has been developed to allow its execution on any hardware from laptops to supercomputers [97], whereas, NAMD has been developed for optimum scalability across high performance clusters [92].

5.2. Enhanced sampling MD simulation methods

The above description of MD simulations has focussed upon classical MD simulations (cMD), using only the Newtonian equation of motion and force fields to solve them [92]. However, these approaches are limited in the amount of conformational space they can sample in the limited timescale of the simulation. This lack of sampling viable conformations and hence, the inability to simulate larger-scale structural changes, is one of the major challenges in MD simulations [103–105]. As a result, several modification to MD simulations have been introduced to include additional parameters which speed up this sampling process [48,92,103,104,106].

These alterations to the MD simulations can be split into two main groups: collective variable (CV) alterations and tempering [48]. The main difference is the former group of approaches requires existing knowledge of the system being modelled, whereas the latter set of approaches does not. However, these two groups of approaches have also been combined [103,107]. The CV MD approach, most commonly metadynamic MD simula-

tions, is based on simulation variables which can be measured, including: bond distances, dihedral angles, root-mean square-deviation (RMSD) from a known structure, radius of gyration or ligand–protein distances [105]. These variables are then used to determine when additional biasing parameters will be applied to the simulation [48,103,105,108] and two to three CVs are normally specified per simulation [48,103]. The two most common tempering approaches are accelerated MD [109,110] and replica-exchange MD [111]. Tempering MD simulations generally enhance the conformational sampling by uniformly boosting all degrees of freedom of the system. This is not an exhaustive list of enhanced sampling methods and several comprehensive review of these methods related to MD simulations have been undertaken elsewhere [103–105].

MD simulation outputs generally are not used alone for analysing drug-protein binding, rather data is extracted from the simulations and then used for further analysis within drug discovery. Two main pieces of information can be gained: 1, the conformational changes of the protein and 2, the energy of the protein and its associated interactions.

5.3. MD simulations applications to M. Tuberculosis

MD simulations have been applied to several *M. tuberculosis* proteins, as summarised in Table 1. The majority of simulations have been utilised either in ensemble docking or predictions of protein–ligand binding and these will be discussed later. Two studies are exceptions, using MD simulations to model drug-binding interactions in more detail. In the case of trisubstituted benzimidazoles binding to FtsZ, the MD simulations were used to validate the docking-predicted binding interactions and assess the protein–ligand complexes' stability [90]. The other paper, utilising MD simulations, focussed on simulations of the RND efflux pumps involved in antimycobacterial drug resistance, specifically MmpL5 [112]. This work focussed on the potential interactions of two antimycobacterial drugs with this potential efflux pump and explored the effects of binding on both the MmpL5 trimer and drugs. Both these papers highlight the greater level of information which can obtained on drug-protein interaction through MD simulations compared to molecular docking approaches. To date, there are no reported studies that have applied enhanced sampling MD simulation approaches to *M. tuberculosis* proteins for drug discovery purposes.

5.4. Machine learning applied to MD simulations

The main drawbacks of MD simulations are the computational resource demands, the large timescales required to sample many protein conformations and errors due to the underlying force fields. To combat these issues, machine learning has been applied to several different aspects of MD simulations, including improvements to the underlying force fields [113–115]; increasing the protein conformations which are sampled [116] and improving the analysis of MD simulations [116,117]. The application to the improvement of underlying forcefield, while able to achieve large timescale reductions [115], have so far only been applied to simple organic molecules [113,115,118] or large single-component systems [119], rather than biomolecules. Selected examples for machine learning-based enhanced sampling of protein conformations will be highlighted, as they have been applied most extensively to biomolecular MD simulations.

Several methods have been applied to enhance the sampling of protein conformations during MD simulations, either through biasing of ongoing MD simulations [120] or through selection of starting-point protein conformations for short MD simulations [121,122]. Anncolvar uses machine learning to approximate CVs for metadynamic MD simulations, which would otherwise be too computationally expensive to apply during a simulation, such as molecular surface area calculations [120]. Hence, allowing novel biases to be applied, to generate more diverse protein conformations. An alternative approach has been to use an MD trajectory as the machine learning input, and then generate predicted protein conformations, expanding the structures which are sampled without increased simulation times [123].

These machine learning approaches are still in the early stages of development, with many different methodologies being employed concurrently, and as yet, no one approach is favoured [113,116]. No applications of machine learning onto MD simulations for antimycobacterial drug discovery have been published. Once these approaches become more accessible, they may allow faster screening of compounds and may make MD simulations accessible to more resource-limited settings.

6. Ensemble docking

One of the major disadvantages of molecular docking of compounds against protein targets is the inability to account for the protein flexibility during these simulated drug interactions. Addition of protein flexibility to current docking programmes would impede their use for screening large compound libraries due to the computational demands. One method which has been adopted to help reduce this problem is use of ensemble docking [37,48,124]. This approach incorporates the flexibility of the protein by performing molecular docking on an ensemble of protein structures and then a weighted average score for each compound is calculated.

The structural diversity of a protein can be found in two ways. First, if several crystal structures of the protein exist, either the apo- or compound-bound forms, they can all be used as known protein conformations [37,125]. Alternatively, MD simulations can be performed on a protein structure, either a crystal structure or model, then the trajectory frames can be clustered. The most common or representative protein conformations found by clustering are used for ensemble docking [48,126]. Several methods are available for the clustering of MD trajectory frames, including: g_cluster within GROMACS, which clusters by a user-specified RMSD cut-off [97]; clustering within Chimera, which clusters by an algorithm-determined RMSD cut-off [127]; POVME which clusters by binding pocket shape [128] and principal component analysis which groups structures along trajectories representing the largest structural differences. A major issue of these clustering approaches is how to select for the best clustered structures to use for ensemble docking, as the most abundant conformations may not be the drug-binding conformations [106]. Hence, a balanced approach is required between selecting diverse conformations and selecting conformations that can discriminate between active and inactive ligands, to allow ensemble docking to be most effective.

Several different types of MD simulations can be used to generate the structural diversity. Replica exchange, accelerated and metadynamics MD simulations, described above, are some of the methods which have been used to generate increased structural diversity across the trajectory frames [48,104,106]. These enhanced sampling methods increase the likelihood that the ensemble of protein structures used for docking represent the physiological structures which occur within the cell and ideally the drug-binding conformations.

Ensemble docking has also been utilised against *M. tuberculosis* proteins specifically, using cMD trajectories for ensemble generation, for GlfT2 [129], RmlD [74] and isocitrate lyase [73]. The trajectory lengths of the MD simulations used were 100, 50 or 18 ns, respectively; with 10 ns MD simulations previously being suffi-

cient to sample druggable binding pockets in multiple targets [106]. Based on these trajectories, 100, 13 and 22 different structures were obtained by clustering for downstream ensemble docking, using cpptraj ([129,130], no further information provided), g_cluster using a 2.0 Å cut-off [74] or active site clustering by backbone RMSD [73], respectively. Validation of these conditions/methods for drug-binding predictions, through *in vitro* protein-binding analysis, was not undertaken during these studies [73,74,129]. However, *in vitro* antimycobacterial testing against *M. smegmatis*, undertaken in one study, indicated potential antimycobacterial inhibitors from the crude plant extracts under investigation [73].

An example of the application of machine learning onto ensemble docking is ENRI, a programme that has been developed to enrich protein conformations which can discriminate between active and inactive compounds [131]. The ENRI programme initially focussed on nuclear receptors, which are not known to be present within *M. tuberculosis*; hence, if suitable training sets for *M. tuberculosis* protein families were developed, this could then be used to adapt ENRI's use. Ideally, this approach will allow fewer protein conformers to be utilised for ensemble docking, representing more accurate drug-binding conformations, leading to more successful hits from these virtual screens. In addition, as ensemble docking utilises both MD simulations and molecular docking, any machine learning programmes which have previously been mentioned, could be combined and applied to further increase ensemble docking success.

7. Protein-Ligand binding energies

Ensemble docking allows the protein flexibility to be accounted for during molecular docking compound screening, improving upon one limitation. However, the other main limitation of molecular docking programmes are the scoring functions, which are used to predict the binding energies of compounds. The scoring functions generally only consider bonding interactions, such as hydrogen bonding, electrostatic and hydrophobic interactions [15]. This excludes the contributions of solvation, both the displacement of solvent from the active site and the stability of the molecule in solvent, and entropic contributions [39,132,133]. Several more advanced approaches exist to take account of these discrepancies, including Linear Interaction Energy (LIE), Molecular Mechanics with Poisson Boltzmann (or generalised Born) and Surface Area solvation (MM-PB(GB)SA) and alchemical methods [132–135]. The first two methods generally utilise the end-point outputs from MD simulations for their calculations, whereas the last set of approaches require new MD simulations. These methods can generate accurate predictions of the ligand–protein binding energies, but the current computational demands make this inaccessible for more than a few compounds.

MM-PB(GB)SA is the most widely adopted approach for estimating ligand binding energy, after molecular docking [132,136,137]. This approach performs calculations on the free ligand, free protein and ligand–protein complex, then the differences are used to estimate the free energy of ligand binding [132,138]. These calculations are performed in three sections, Molecular Mechanics (MM), Poisson Boltzmann (PB) (or generalised Born (GB)) and Surface Area solvation (SA), before the summation is used to estimate the binding energy [137]. Using one MD simulation of the ligand–protein complex, rather than three separate MD simulations for each component, appears to be the favoured approach, giving similar accuracy with lower computational requirements [132,139,140].

The calculation of drug-protein binding energies using MMPB(GB)SA has been applied to several drug discovery attempts against *M. tuberculosis* proteins. These include: LipU [77], GlnA1 [79], DprE1[80], PknA [81], NarL [52] , PanC [141], MurB [68] and MurE [68]. In the majority of cases, Glide was used to perform the initial virtual screen. For the MMPB(GB)SA calculations, either the Prime MMGBSA method was used, or MD simulations were performed, and the trajectory outputs used for the calculations. The former approach performs energy minimisation of the ligand, protein and docked ligand–protein complex in place of an MD simulation, then performs MMGBSA calculations [142], this was used for LipU and MraY drug-binding calculations [77,83]. The latter approach used MD simulations which varied between 15 and 100 ns per ligand–protein complex. Depending on the MD programme used, either GROMACS or AMBER, the associated script to calculate MMPBSA energies, or MMPB(GB)SA energies for DprE1 & PanC [80,141], was used. Unfortunately, as no *in vitro* binding studies were performed following these computational studies, a comparison of the relative accuracies of Prime MMGBSA vs MD simulation-MMPBSA is not possible.

The other methods to calculate drug-protein energies have not been applied to *M. tuberculosis* proteins as often. However, LIE has been used for drug binding energies against EthR [84] and an alchemical method to estimate the absolute ligand binding energy has been applied to RmlC [143].

Machine learning has rarely been applied to these current binding energy calculation methods, likely due to large computational costs of performing them. The only application found was the use of machine learning to filter the MD trajectory frames used to calculate MMPB(GB)SA binding energies [144]. This led to an increased correlation between experimental and predicted binding energies for MMPBSA calculations, but no difference for MMGBSA calculations [144]. More generally, any machine learning approaches which improve the efficiency of MD simulations, will also increase the efficiency of these follow-up binding energies calculations, as they also required MD simulations.

In contrast to the lower number of applications of machine learning onto current methods, many machine learning based programmes have been developed to predict protein–ligand binding affinity [57]. The focus on novel approaches is likely due to the high computational costs of the current methods. However, these approaches have not been widely tested, with only one machine-learning based on approach used in the latest SAMPL host–guest challenge [145].

8. Hit analysis

Computational approaches for drug discovery have the obvious drawback of generating many potential hits, which often include many false positives due to inherent software biases. This is especially true of molecular docking and hence further down-stream analysis of the hits should be undertaken in confirmatory studies. Potential hits can also be mis-classified as false negatives, especially during molecular docking due to the forced protein rigidity [37,38,56]. The true scale of the false negatives problem may be underestimated, as they are rarely evaluated following virtual screening campaigns, despite one study finding two of five inhibitor classes were missed by molecular docking compared to high-throughput screening [56]. However, ensemble docking may reduce this issue by accounting for protein flexibility [56,126,146] and allowing multiple drug-binding modes to be assessed.

When selection of follow-up hits is being undertaken on the basis of physiochemical properties, it should be noted that the current TB-drug space has significantly different physiochemical properties in comparison to those specified by the Lipinski's rule of five [32,147]. The difference in compound properties is in part due to the hydrophobic *M. tuberculosis* cell envelope, which pre-

vents penetration of many drug compounds. To aid in the evaluation of hits from an *in silico* drug discovery screen, several computational approaches exist for predicting penetration through the *M. tuberculosis* membrane [148,149] and predicting *in vivo M. tuberculosis* activity [150]. This can allow downstream development of hits which are predicted to both strongly bind a protein of interest and penetrate the *M. tuberculosis* cell envelope.

An initial study modelled the *M. tuberculosis* cell envelope, using MD simulations to evaluate compound penetration, however, this could only test a small number of compounds at once [151]. More recently machine learning has been used to predict compound permeability and the MycPermCheck programme has been developed to allow for routine screening of compounds [149]. Alongside predictions of *M. tuberculosis* membrane penetration, further machine learning models have been developed to predict compound activity against *M. tuberculosis in vivo* [85,150]. This has been undertaken using either whole cell *M.* tuberculosis activity data [85] or *in vivo* data from mouse infection models [150]. The former machine learning model has been successfully applied to InhA docked hits, to prioritise molecules with potential whole-cell *M. tuberculosis* activity [69]. While the activity data used to train these *in vivo* activity models is publicly available, the trained models themselves are not, and hence are less accessible for use within hit analysis, compared to MycPermCheck.

The predictive power of machine learning based models is heavily tied to the training dataset used, both the breath of compound properties and the quality of the data [152]. Hence, these programmes will effectively predict membrane penetration/compound activity for some novel compounds but will generate inaccurate predictions for many others. In the short-term, experimental validation of any predictions would need to be undertaken rather than complete reliance on these computational analysis methods.

9. Experimental validation

Whilst many of the published studies summarised in this review (Table 1) have provided many novel "potential" hits, and are useful for testing computational drug discovery approaches, the lack of follow-up *in vitro* or *in vivo* confirmatory studies means that many of these projects have stalled. To allow the hit compounds of these studies to be turned into antimycobacterial lead compounds, at least some experimental validation is required. As a minimum, either the hits should be tested for binding against the protein of interest or tested for bacteriostatic and/or bactericidal activity against *Mycobacteria,* ideally *M. tuberculosis*. Both of these approaches provide evidence that development of these new hits may be successful. The latter validation method is more informative for further development, as it confirms a drug can penetrate the *M. tuberculosis* cell envelope and is not removed by efflux or deactivated by metabolism before it exerts its bacteriostatic/bactericidal activity. In addition, due to the increasing levels of *M. tuberculosis* drug resistance, testing of novel drug compounds against clinical MDR-TB strains would provide strong evidence for their future development [10], due to limited number of treatment options [1].

An exemplary approach for *in silico* structure-based drug discovery against *M. tuberculosis* was undertaken by Tatum *et al.* [53,84] (Table 1). An initial virtual screening approach, using GOLD, docked 409,201 compounds, selected from the *Drugs Now* subset of the ZINC database, against EthR. Then, hits were shortlisted for chemical diversity and by visual inspection of binding poses. This led to 85 compounds which were tested for *in vitro* binding against EthR, and 20 compounds showed binding. Fifteen of these hits were then tested against *M. tuberculosis* H37Rv and co-crystallisation trials of EthR-drug were undertaken, leading to four crystal structures. This study provides compelling evidence for the development of the four potential EthR inhibitors [53] and for adopting similar approaches for future drug discovery efforts. A follow-up study then focussed on MD simulations and LIE calculations for the four potential inhibitors, providing computational comparisons of the experimental binding modes [53]. Further studies similar to this work are required to move from promising drug hits into lead compounds that can be tested within preclinical trials.

10. Summary and outlook

The current antimycobacterial drug discovery approach has focussed on *in vitro* high—throughput screening and drug repurposing. These are likely to remain important areas of research and the *in-silico* approaches discussed herein can be complementary methods and aid in the discovery of new compounds. These approaches should allow a larger number of compounds to be virtually screened, before testing *in vitro* and allow the diversity and ideally quality of hits to increase. This will hopefully increasing the likelihood of this ligand discovery being translated into novel drug discovery [152].

Molecular docking represents the most widely applied method for discovery of drug compounds targeting *M. tuberculosis* proteins, likely due to its wider accessibility. There is a higher availability of user-friendly programmes for this approach, whereas MD simulations and its associated analysis have a greater barrier to entry, hence explaining their lower frequency of utilisation. However, the other *in silico* methods outlined here, especially MMPBSA calculations and machine learning-based hit analysis, provide other approaches for filtering molecular docking hits before *in vitro* testing needs to be undertaken. Ideally, they will allow fewer, but higher quality, hit compounds to be tested, reducing the overall costs of early drug screening, and providing a stronger basis for lead development and testing.

Out of all the applications of machine learning described herein, its application to molecular docking is likely to have the greatest impact in the short-term, as this initial step is becoming more universal in the early stages of target-based drug discovery. Hence, as machine-learning applications have been shown to generate more accurate protein binding predictions, this will likely increase the number of true positive hits that are tested *in vitro*. In the longer-term, the use of machine-learning methods to accurately calculate ligand–protein binding energies may revolutionise how protein specific inhibitors could be found. However, this would be dependent on the calculations being at a computational efficiency accessible for large-scale screening of compounds and accurate enough to fully model *in vitro* binding.

Alongside the application of these current and novel methods to *M. tuberculosis* drug discovery, they need to be tied to experimental validation of any hit compounds. This helps to validate both the methodology and generates more activity data that can be used to train machine learning models. Only through validation of computational hits can novel drug compounds be found to combat the growing drug resistance present within *M. tuberculosis* worldwide.

Funding information

AK was funded via a Wellcome Trust Doctoral Training Program (Antimicrobials and Antimicrobial Resistance); grant reference: 108876/B/15/Z.

CRediT authorship contribution statement

Alexander D.H. Kingdon: Writing - original draft, Investigation. **Luke J. Alderwick:** Writing - review & editing, Resources, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] WHO. Global Tuberculosis Report 2020; 2020.
[2] Barry III CE, Boshoff H, Dartois V, Dick T, Ehrt S, Flynn J, et al. The spectrum of latent tuberculosis: rethinking the goals of prophylaxis. Nat Rev Microbiol 2009;7:845–55. https://doi.org/10.1038/nrmicro2236.
[3] Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, et al. Tuberculosis Nat Rev Dis Prim 2016;2(1). https://doi.org/10.1038/nrdp.2016.76.
[4] Maitra A, Munshi T, Healy J, Martin LT, Vollmer W, Keep NH, et al. Cell wall peptidoglycan in Mycobacterium tuberculosis: An Achilles' heel for the TB-causing pathogen. FEMS Microbiol Rev 2019;43:548–75. https://doi.org/10.1093/femsre/fuz016.
[5] Dulberger CL, Rubin EJ, Boutte CC. The mycobacterial cell envelope — a moving target. Nat Rev Microbiol 2020;18(1):47–59. https://doi.org/10.1038/s41579-019-0273-7.
[6] Dutta NK, Karakousis PC. Latent Tuberculosis Infection: Myths, Models, and Molecular Mechanisms. Microbiol Mol Biol Rev 2014;78(3):343–71. https://doi.org/10.1128/MMBR.00010-14.
[7] Barry E, Lessons C. from Seven Decades of Antituberculosis Drug Discovery. Curr Top Med Chem 2011;11:1216–25. https://doi.org/10.2174/156802611795429158.
[8] Zhang Y. The Magic Bullets and Tuberculosis Drug Targets. Annu Rev Pharmacol Toxicol 2005;45(1):529–64. https://doi.org/10.1146/annurev.pharmtox.45.120403.100120.
[9] Sensi P. History of the Development of Rifampin. Rev Infect Dis 1983;5: S402–6. https://doi.org/10.1093/clinids/5.supplement_3.s402.
[10] Zumla A, Nahid P, Cole ST. Advances in the development of new tuberculosis drugs and treatment regimens. Nat Rev Drug Discov 2013;12(5):388–404. https://doi.org/10.1038/nrd4001.
[11] WHO. Global Tuberculosis Report 2019. Geneva; 2019.
[12] Stover CK, Warrener P, VanDevanter DR, Sherman DR, Arain TM, Langhorne MH, et al. A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis. Nature 2000;405(6789):962–6. https://doi.org/10.1038/35016103.
[13] Matsumoto M, Hashizume H, Tomishige T, Kawasaki M, Tsubouchi H, Sasaki H, et al. OPC-67683, a Nitro-Dihydro-Imidazooxazole Derivative with Promising Action against Tuberculosis In Vitro and In Mice. PLoS Med 2006;3(11):e466. https://doi.org/10.1371/journal.pmed.0030466.
[14] Geppert H, Vogt M, Bajorath J. Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. J Chem Inf Model 2010;50 (2):205–16. https://doi.org/10.1021/ci900419k.
[15] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: Methods and applications. Nat Rev Drug Discov 2004;3(11):935–49. https://doi.org/10.1038/nrd1549.
[16] Lohning AE, Levonis SM, Williams-Noonan B, Schweiker SS. A practical guide to molecular docking and homology modelling for medicinal chemists. Curr Top Med Chem 2017;17:2023–40. https://doi.org/10.2174/1568026617666170130110827.
[17] Shaker B, Yu M-S, Lee J, Lee Y, Jung C, Na D. User guide for the discovery of potential drugs via protein structure prediction and ligand docking simulation. J Microbiol 2020;58(3):235–44. https://doi.org/10.1007/s12275-020-9563-z.
[18] Swift RV, Jusoh SA, Offutt TL, Li ES, Amaro RE. Knowledge-Based Methods To Train and Optimize Virtual Screening Ensembles. J Chem Inf Model 2016;56 (5):830–42. https://doi.org/10.1021/acs.jcim.5b0068410.1021/acs.jcim.5b00684.s001.
[19] DeJesus MA, Gerrick ER, Xu W, Park SW, Long JE, Boutte CC, et al. Comprehensive Essentiality Analysis of the *Mycobacterium tuberculosis* Genome via Saturating Transposon Mutagenesis. MBio 2017;8(1). https://doi.org/10.1128/mBio.02133-16.
[20] Kolly GS, Boldrin F, Sala C, Dhar N, Hartkoorn RC, Ventura M, et al. Assessing the essentiality of the decaprenyl-phospho-d-arabinofuranose pathway in *Mycobacterium tuberculosis* using conditional mutants. Mol Microbiol 2014;92:194–211. https://doi.org/10.1111/mmi.12546.
[21] Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature 2020;577(7792):706–10. https://doi.org/10.1038/s41586-019-1923-7.
[22] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res 2000;28:235–42. https://doi.org/10.1093/nar/28.1.235.
[23] Terwilliger TC, Park MS, Waldo GS, Berendzen J, Hung L-W, Kim C-Y, et al. The TB structural genomics consortium: a resource for *Mycobacterium tuberculosis* biology. Tuberculosis 2003;83(4):223–49. https://doi.org/10.1016/S1472-9792(03)00051-9.
[24] Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, et al. The SWISS-MODEL Repository–new features and functionality. Nucleic Acids Res 2017;45:313–9. https://doi.org/10.1093/nar/gkw1132.
[25] Studer G, Tauriello G, Bienert S, Biasini M, Johner N, Schwede T, et al. ProMod3—A versatile homology modelling toolbox. PLoS Comput Biol 2021;17(1):e1008667. https://doi.org/10.1371/journal.pcbi.1008667.
[26] Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinf 2008;9:1–8. https://doi.org/10.1186/1471-2105-9-40.
[27] Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 2015;10 (6):845–58. https://doi.org/10.1038/nprot.2015.053.
[28] Sterling T, Irwin JJ. ZINC 15 – Ligand Discovery for Everyone. J Chem Inf Model 2015;55(11):2324–37. https://doi.org/10.1021/acs.jcim.5b00559.
[29] Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Paula Magariños M, et al. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res 2019;47:D930–40. https://doi.org/10.1093/nar/gky1075.
[30] Gorgulla C, Boeszoermenyi A, Wang Z-F, Fischer PD, Coote PW, Padmanabha Das KM, et al. An open-source drug discovery platform enables ultra-large virtual screens. Nature 2020;580(7805):663–8. https://doi.org/10.1038/s41586-020-2117-z.
[31] Hoffmann T, Gastreich M. The next level in chemical space navigation: going far beyond enumerable compound libraries. Drug Discov Today 2019;24 (5):1148–56. https://doi.org/10.1016/j.drudis.2019.02.013.
[32] Fullam E, Young RJ. Physiochemical properties and *Mycobacterium tuberculosis* transporters: key to efficacious antitubercular drugs?. RSC Med Chem 2021;12:43–56. https://doi.org/10.1039/d0md00265h.
[33] Ekins S, Bradford J, Dole K, Spektor A, Gregory K, Blondeau D, et al. A collaborative database and computational models for tuberculosis drug discovery. Mol BioSyst 2010;6(5):840. https://doi.org/10.1039/b917766c.
[34] Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A Deep Learning Approach to Antibiotic Discovery. Cell 2020;180(4):688–702. e13. https://doi.org/10.1016/j.cell.2020.01.021.
[35] Kanvatirth P, Jeeves RE, Bacon J, Besra GS, Alderwick LJ, Shin SJ. Utilisation of the Prestwick Chemical Library to identify drugs that inhibit the growth of mycobacteria. PLoS ONE 2019;14(3):e0213713. https://doi.org/10.1371/journal.pone.0213713.
[36] Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, et al. The Drug Repurposing Hub: a next-generation drug library and information resource. Nat Med 2017;23(4):405–8. https://doi.org/10.1038/nm.4306.
[37] Carlson HA, McCammon JA. Accommodating Protein Flexibility in Computational Drug Design. Mol Pharmacol 2000;57:213–8.
[38] Klebe G. Recent developments in structure-based drug design. J Med Mol 2000;78(5):269–81. https://doi.org/10.1007/s001090000084.
[39] Mobley DL, Dill KA. Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get". Structure 2009;17(4):489–98. https://doi.org/10.1016/j.str.2009.02.010.
[40] Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–48. https://doi.org/10.1006/JMBI.1996.0897.
[41] Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins 2003;52(4):609–23. https://doi.org/10.1002/prot.10465.
[42] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem 2010;31:455–61. https://doi.org/10.1002/jcc.21334.
[43] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 2004;47:1739–49. https://doi.org/10.1021/jm0306430.
[44] Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. J Med Chem 2004;47:1750–9. https://doi.org/10.1021/jm030644s.
[45] Vázquez J, Deplano A, Herrero A, Ginex T, Gibert E, Rabal O, et al. Development and Validation of Molecular Overlays Derived from 3D Hydrophobic Similarity with PharmScreen. J Chem Inf Model 2018;58:1596–609. https://doi.org/10.1021/acs.jcim.8b00216.
[46] Kellenberger E, Rodrigo J, Muller P, Rognan D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. Proteins Struct Funct Genet 2004;57(2):225–42. https://doi.org/10.1002/prot.20149.
[47] Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular Docking and Structure-Based Drug Design Strategies. Molecules 2015;20:13384–421. https://doi.org/10.3390/molecules200713384.
[48] Gioia D, Bertazzo M, Recanatini M, Masetti M, Cavalli A. Dynamic docking: A paradigm shift in computational drug discovery. Molecules 2017;22:1–21. https://doi.org/10.3390/molecules22112029.
[49] Jiménez-Luna J, Cuzzolin A, Bolcato G, Sturlese M, Moro S. A Deep-Learning Approach toward Rational Molecular Docking Protocol Selection. Molecules 2020;25:1–12. https://doi.org/10.3390/molecules25112487.
[50] Abrahams KA, Besra GS. Mycobacterial drug discovery. RSC. Med Chem 2020;11(12):1354–65. https://doi.org/10.1039/D0MD00261E.
[51] Billones JB, Carrillo MCO, Organo VG, Sy JBA, Macalino SJY, Emnacen IA, et al. In silico discovery and in vitro activity of inhibitors against *Mycobacterium tuberculosis* 7,8-diaminopelargonic acid synthase (Mtb BioA). Drug Des Devel Ther 2017;11:563–74. https://doi.org/10.2147/DDDT.S119930.

[52] Kumar N, Srivastava R, Prakash A, Lynn AM. Structure-based virtual screening, molecular dynamics simulation and MM-PBSA toward identifying the inhibitors for two-component regulatory system protein NarL of *Mycobacterium tuberculosis*. J Biomol Struct Dyn 2020;38(11):3396–410. https://doi.org/10.1080/07391102.2019.1657499.
[53] Tatum NJ, Liebeschuetz JW, Cole JC, Frita R, Herledan A, Baulard AR, et al. New active leads for tuberculosis booster drugs by structure-based drug discovery. Org Biomol Chem 2017;15(48):10245–55. https://doi.org/10.1039/C7OB00910K.
[54] Chakraborty S, Rhee KY. Tuberculosis Drug Development: History and Evolution of the Mechanism-Based Paradigm. Cold Spring Harb Perspect Med 2015;5:1–11. https://doi.org/10.1101/cshperspect.a021147.
[55] Sieg J, Flachsenberg F, Rarey M. In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening. J Chem Inf Model 2019;59(3):947–61. https://doi.org/10.1021/acs.jcim.8b0071210.1021/acs.jcim.8b00712.s001.
[56] Ferreira RS, Simeonov A, Jadhav A, Eidam O, Mott BT, Keiser MJ, et al. Complementarity Between a Docking and a High-Throughput Screen in Discovering New Cruzain Inhibitors. J Med Chem 2010;53(13):4891–905. https://doi.org/10.1021/jm100488w.
[57] Li H, Sze K, Lu G, Ballester PJ. Machine-learning scoring functions for structure-based drug lead optimization. WIREs Comput Mol Sci 2020;10:1–20. https://doi.org/10.1002/wcms.1465.
[58] Shen C, Ding J, Wang Z, Cao D, Ding X, Hou T. From machine learning to deep learning: Advances in scoring functions for protein–ligand docking. Wiley Interdiscip Rev Comput Mol Sci 2020;10:1–23. https://doi.org/10.1002/wcms.1429.
[59] Ballester PJ. Selecting machine-learning scoring functions for structure-based virtual screening. Drug Discov Today Technol 2019;32:81–7. https://doi.org/10.1016/J.DDTEC.2020.09.001.
[60] Li H, Peng J, Leung Y, Leung K-S, Wong M-H, Lu G, et al. The Impact of Protein Structure and Sequence Similarity on the Accuracy of Machine-Learning Scoring Functions for Binding Affinity Prediction. Biomolecules 2018;8(1):12. https://doi.org/10.3390/biom8010012.
[61] Wójcikowski M, Zielenkiewicz P, Siedlecki P. Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field. J Cheminform 2015;7:1–6. https://doi.org/10.1186/s13321-015-0078-2.
[62] Durrant JD, McCammon JA. NNScore 2.0: A Neural-Network Receptor-Ligand Scoring Function. J Chem Inf Model 2011;51(11):2897–903. https://doi.org/10.1021/ci2003889.
[63] Yasuo N, Sekijima M. Improved Method of Structure-Based Virtual Screening via Interaction-Energy-Based Learning. J Chem Inf Model 2019;59(3):1050–61. https://doi.org/10.1021/acs.jcim.8b0067310.1021/acs.jcim.8b00673.s001.
[64] Wójcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. Sci Rep 2017;7:1–10. https://doi.org/10.1038/srep46710.
[65] Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. J Med Chem 2012;55(14):6582–94. https://doi.org/10.1021/jm300687e.
[66] Durrant JD, McCammon JA. NNScore: A Neural-Network-Based Scoring Function for the Characterization of Protein-Ligand Complexes. J Chem Inf Model 2010;50(10):1865–71. https://doi.org/10.1021/ci100244v.
[67] Tam B, Sherf D, Cohen S, Eisdorfer SA, Perez M, Soffer A, et al. Discovery of small-molecular inhibitors targeting the ribosomal peptidyl transferase center (PTC) of *M. tuberculosis*. Chem Sci 2019;10:8764–7. https://doi.org/10.1039/c9sc02520k.
[68] Rani J, Silla Y, Borah K, Ramachandran S, Bajpai U. Repurposing of FDA-approved drugs to target MurB and MurE enzymes in *Mycobacterium tuberculosis*. J Biomol Struct Dyn 2020;38(9):2521–32. https://doi.org/10.1080/07391102.2019.1637280.
[69] Wang X, Perryman AL, Li S-G, Paget SD, Stratton TP, Lemenze A, et al. Intrabacterial Metabolism Obscures the Successful Prediction of an InhA Inhibitor of *Mycobacterium tuberculosis*. ACS Infect Dis 2019;5(12):2148–63. https://doi.org/10.1021/acsinfecdis.9b0029510.1021/acsinfecdis.9b00295.s00110.1021/acsinfecdis.9b00295.s00210.1021/acsinfecdis.9b00295.s003.
[70] Perryman AL, Yu W, Wang X, Ekins S, Forli S, Li S-G, et al. A Virtual Screen Discovers Novel, Fragment-Sized Inhibitors of *Mycobacterium tuberculosis* InhA. J Chem Inf Model 2015;55(3):645–59. https://doi.org/10.1021/ci500672v.
[71] Zhang G, Guo S, Cui H, Qi J. Virtual Screening of Small Molecular Inhibitors against DprE1. Molecules 2018;23:524–33. https://doi.org/10.3390/molecules23030524.
[72] Tuhin Ali M, Blicharska N, Shilpi JA, Seidel V. Investigation of the anti-TB potential of selected propolis constituents using a molecular docking approach. Sci Rep 2018;8:1–8. https://doi.org/10.1038/s41598-018-30209-y.
[73] Lee Y-V, Choi SB, Wahab HA, Lim TS, Choong YS. Applications of Ensemble Docking in Potential Inhibitor Screening for *Mycobacterium tuberculosis* Isocitrate Lyase Using a Local Plant Database. J Chem Inf Model 2019;59(5):2487–95. https://doi.org/10.1021/acs.jcim.8b0096310.1021/acs.jcim.8b00963.s001.
[74] Ravichandran R, Farrah Wahidah Ridzwan N, Bin MS. Ensemble-based high-throughput virtual screening of natural ligands using the Super Natural-II database against cell-wall protein dTDP-4-dehydrorhamnose reductase (RmlD) in *Mycobacterium tuberculosis*. J Biomol Struct Dyn Dyn 2020;1–10. https://doi.org/10.1080/07391102.2020.1867641.
[75] Billones JB, Carrillo MCO, Organo VG, Macalino SJY, Sy JBA, Clavio NAB, et al. Toward antituberculosis drugs: in silico screening of synthetic compounds against *Mycobacterium tuberculosis* l, d-transpeptidase 2. Drug Des Devel Ther 2016;10:1147–57. https://doi.org/10.2147/DDDT.S97043.
[76] Scheich C, Szabadka Z, Vértessy B, Pütter V, Grolmusz V, Schade M, et al. Discovery of Novel MDR-*Mycobacterium tuberculosis* Inhibitor by New FRIGATE Computational Screen. PLoS ONE 2011;6(12):e28428. https://doi.org/10.1371/journal.pone.0028428.
[77] Kaur G, Pandey B, Kumar A, Garewal N, Grover A, Kaur J. Drug targeted virtual screening and molecular dynamics of LipU protein of *Mycobacterium tuberculosis* and *Mycobacterium leprae*. J Biomol Struct Dyn 2019;37(5):1254–69. https://doi.org/10.1080/07391102.2018.1454852.
[78] Sivaranjani P, Naik VU, Madhulitha NR, Kumar KS, Chiranjeevi P, Alexander SP, et al. Design of Novel Antimycobacterial Molecule Targeting Shikimate Pathway of *Mycobacterium tuberculosis*. Indian J Pharm Sci 2019;81:438–47. , https://doi.org/10.36468/pharmaceutical-sciences.528.
[79] Kumari M, Subbarao N. Virtual screening to identify novel potential inhibitors for Glutamine synthetase of *Mycobacterium tuberculosis*. J Biomol Struct Dyn 2020;38(17):5062–80. https://doi.org/10.1080/07391102.2019.1695670.
[80] Niranjan Kumar, Srivastava R, Prakash A, Lynn AM. Virtual screening and free energy estimation for identifying *Mycobacterium tuberculosis* flavoenzyme DprE1 inhibitors. J Mol Graph Model 2021;102:107770. https://doi.org/10.1016/j.jmgm.2020.107770.
[81] Sundar S, Thangamani L, Manivel G, Kumar P, Piramanayagam S. Molecular docking, molecular dynamics and MM/PBSA studies of FDA approved drugs for protein kinase a of *Mycobacterium tuberculosis*; application insights of drug repurposing. Informatics Med Unlocked 2019;16:100210. https://doi.org/10.1016/j.imu.2019.100210.
[82] Kuldeep J, Sharma SK, Sharma T, Singh BN, Siddiqi MI. Targeting *Mycobacterium Tuberculosis* Enoyl-acyl Carrier Protein Reductase using Computational Tools for Identification of Potential Inhibitor and their Biological Activity. Mol Inform 2021;40(5):2000211. https://doi.org/10.1002/minf.v40.510.1002/minf.202000211.
[83] Mallavarapu BD, Abdullah M, Saxena S, Guruprasad L. Inhibitor binding studies of *Mycobacterium tuberculosis* MraY (Rv2156c): Insights from molecular modeling, docking, and simulation studies. J Biomol Struct Dyn 2019;37(14):3751–63. https://doi.org/10.1080/07391102.2018.1526715.
[84] Tatum NJ, Duarte F, Kamerlin SCL, Pohl E. Relative Binding Energies Predict Crystallographic Binding Modes of Ethionamide Booster Lead Compounds. J Phys Chem Lett 2019;10(9):2244–9. https://doi.org/10.1021/acs.jpclett.9b0074110.1021/acs.jpclett.9b00741.s001.
[85] Lane T, Russo DP, Zorn KM, Clark AM, Korotcov A, Tkachenko V, et al. Comparing and Validating Machine Learning Models for *Mycobacterium tuberculosis* Drug Discovery. Mol Pharm 2018;15(10):4346–60. https://doi.org/10.1021/acs.molpharmaceut.8b0008310.1021/acs.molpharmaceut.8b00083.s001.
[86] Puhl AC, Lane TR, Vignaux PA, Zorn KM, Capodagli GC, Neiditch MB, et al. Computational Approaches to Identify Molecules Binding to *Mycobacterium tuberculosis* KasA. ACS Omega 2020;5(46):29935–42. https://doi.org/10.1021/acsomega.0c0427110.1021/acsomega.0c04271.s001.
[87] Pauli I, dos Santos RN, Rostirolla DC, Martinelli LK, Ducati RG, Timmers LFSM, et al. Discovery of New Inhibitors of *Mycobacterium tuberculosis* InhA Enzyme Using Virtual Screening and a 3D-Pharmacophore-Based Approach. J Chem Inf Model 2013;53(9):2390–401. https://doi.org/10.1021/ci400202t.
[88] Chiarelli LR, Mori M, Barlocco D, Beretta G, Gelain A, Pini E, et al. Discovery and development of novel salicylate synthase (MbtI) furanic inhibitors as antitubercular agents. Eur J Med Chem 2018;155:754–63. https://doi.org/10.1016/j.ejmech.2018.06.033.
[89] Ballester PJ, Mangold M, Howard NI, Robinson RLM, Abell C, Blumberger J, et al. Hierarchical virtual screening for the discovery of new molecular scaffolds in antibacterial hit identification. J R Soc Interface 2012;9(77):3196–207. https://doi.org/10.1098/rsif.2012.0569.
[90] Li D, Chi Bo, Wang W-W, Gao J-M, Wan J. Exploring the possible binding mode of trisubstituted benzimidazoles analogues *in silico* for novel drug design targeting Mtb FtsZ. Med Chem Res 2017;26(1):153–69. https://doi.org/10.1007/s00044-016-1734-4.
[91] McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature 1977;267(5612):585–90. https://doi.org/10.1038/267585a0.
[92] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable Molecular Dynamics with NAMD. J Comput Chem 2005;26(16):1781–802. https://doi.org/10.1002/(ISSN)1096-987X10.1002/jcc.v26:1610.1002/jcc.20289.
[93] Hospital A, Goñi JR, Orozco M, Gelpí JL. Molecular dynamics simulations: advances and applications. Adv Appl Bioinforma Chem 2015;8:37–47. https://doi.org/10.2147/AABC.S70333.
[94] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 2010;31:671–90. https://doi.org/10.1002/jcc.21367.
[95] Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. PNAS 2018;115(21):E4758–66. https://doi.org/10.1073/pnas.1800690115.
[96] Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. Neuron Rev 2018;99(6):1129–43. https://doi.org/10.1016/j.neuron.2018.08.011.
[97] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism

from laptops to supercomputers. SoftwareX 2015;1-2:19–25. https://doi.org/10.1016/j.softx.2015.06.001.
[98] Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. J Comput Chem 2005;26(16):1668–88. https://doi.org/10.1002/(ISSN)1096-987X10.1002/jcc.v26:1610.1002/jcc.20290.
[99] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4(2):187–217. https://doi.org/10.1002/(ISSN)1096-987X10.1002/jcc.v4:210.1002/jcc.540040211.
[100] Lee J, Cheng Xi, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput 2016;12(1):405–13. https://doi.org/10.1021/acs.jctc.5b0093510.1021/acs.jctc.5b00935.s001.
[101] Humphrey W, Dalke A, Schulten K. Visual Molecular Dynamics. J Mol Graph 1996;14(1):33–8. https://doi.org/10.1016/0263-7855(96)00018-5.
[102] Shirts MR, Klein C, Swails JM, Yin J, Gilson MK, Mobley DL, et al. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. J Comput Aided Mol Des 2017;31(1):147–61. https://doi.org/10.1007/s10822-016-9977-1.
[103] Maximova T, Moffatt R, Ma B, Nussinov R, Shehu A, de Groot BL. Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics. PLoS Comput Biol 2016;12(4):e1004619. https://doi.org/10.1371/journal.pcbi.1004619.
[104] Kalyaanamoorthy S, Chen Y-P. Modelling and enhanced molecular dynamics to steer structure-based drug discovery. Prog Biophys Mol Biol 2014;114(3):123–36. https://doi.org/10.1016/j.pbiomolbio.2013.06.004.
[105] Spiwok V, Sucur Z, Hosek P. Enhanced sampling techniques in biomolecular simulations. Biotechnol Adv 2015;33(6):1130–40. https://doi.org/10.1016/j.biotechadv.2014.11.011.
[106] Amaro RE, Baudry J, Chodera J, Demir Ö, McCammon JA, Miao Y, et al. Ensemble Docking in Drug Discovery. Biophys J 2018;114(10):2271–8. https://doi.org/10.1016/j.bpj.2018.02.038.
[107] Tribello GA, Ceriotti M, Parrinello M. A self-learning algorithm for biased molecular dynamics. Proc Natl Acad Sci 2010;107:17509–14. https://doi.org/10.1073/pnas.1011511107/-/DCSupplemental.
[108] Basciu A, Malloci G, Pietrucci F, Bonvin AMJJ, Vargiu AV. Holo-like and Druggable Protein Conformations from Enhanced Sampling of Binding Pocket Volume and Shape. J Chem Inf Model 2019;59(4):1515–28. https://doi.org/10.1021/acs.jcim.8b0073010.1021/acs.jcim.8b00730.s001.
[109] Wang Y, Harrison CB, Schulten K, Mccammon JA. Implementation of Accelerated Molecular Dynamics in NAMD. Comput Sci Discov 2012;4:1–14. https://doi.org/10.1088/1749-4699/4/1/015002.
[110] Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. J Chem Phys 2004;120(24):11919–29. https://doi.org/10.1063/1.1755656.
[111] Sugita Y, Okamoto Y. Replica exchange molecular dynamics method for protein folding simulation. Chem Phys Lett 1999;314:141–51. https://doi.org/10.1385/1-59745-189-4:205.
[112] Sandhu P, Akhter Y. The drug binding sites and transport mechanism of the RND pumps from *Mycobacterium tuberculosis*: Insights from molecular dynamics simulations. Arch Biochem Biophys 2016;592:38–49. https://doi.org/10.1016/J.ABB.2016.01.007.
[113] Noé F, Tkatchenko A, Müller K-R, Clementi C. Machine Learning for Molecular Simulation. Annu Rev Phys Chem 2020;71:361–90. https://doi.org/10.1146/annurev-physchem-042018.
[114] Behler J. Perspective: Machine Learning potentials for atomistic simulations. J Chem Phys 2016;145:170901–9. https://doi.org/10.1063/1.4966192.
[115] Morawietz T, Artrith N. Machine learning-accelerated quantum mechanics-based atomistic simulations for industrial applications. J Comput Aided Mol Des 2021;35(4):557–86. https://doi.org/10.1007/s10822-020-00346-6.
[116] Wang Y, Lamim Ribeiro JM, Tiwary P. Machine learning approaches for analyzing and enhancing molecular dynamics simulations. Curr Opin Struct Biol 2020;61:139–45. https://doi.org/10.1016/J.SBI.2019.12.016.
[117] Taufer M, Estrada T, Johnston T. A survey of algorithms for transforming molecular dynamics data into metadata for in situ analytics based on machine learning methods. Philos Trans R Soc London A 2020;378:1–11. https://doi.org/10.1098/rsta.2019.0063.
[118] Schütt KT, Kessel P, Gastegger M, Nicoli KA, Tkatchenko A, Müller K-R. SchNetPack: A Deep Learning Toolbox For Atomistic Systems. J Chem Theory Comput 2019;15(1):448–55. https://doi.org/10.1021/acs.jctc.8b0090810.1021/acs.jctc.8b00908.s001.
[119] Lu D, Wang H, Chen M, Lin L, Car R, E W, et al. 86 PFLOPS Deep Potential Molecular Dynamics simulation of 100 million atoms with ab initio accuracy. Comput Phys Commun 2021;259:107624. https://doi.org/10.1016/j.cpc.2020.107624.
[120] Trapl D, Horvacanin I, Mareska V, Ozcelik F, Unal G, Spiwok V. Anncolvar: Approximation of complex collective variables by artificial neural networks for analysis and biasing of molecular simulations. Front Mol Biosci 2019;6:1–9. https://doi.org/10.3389/fmolb.2019.00025.
[121] Shin K, Tran DP, Takemura K, Kitao A, Terayama K, Tsuda K. Enhancing Biomolecular Sampling with Reinforcement Learning: A Tree Search Molecular Dynamics Simulation Method. ACS Omega 2019;4(9):13853–62. https://doi.org/10.1021/acsomega.9b0148010.1021/acsomega.9b01480.s001.
[122] Terayama K, Iwata H, Araki M, Okuno Y, Tsuda K. Machine learning accelerates MD-based binding pose prediction between ligands and proteins. Bioinformatics 2018;34:770–8. https://doi.org/10.1093/bioinformatics/btx638.
[123] Degiacomi MT. Coupling Molecular Dynamics and Deep Learning to Mine Protein Conformational Space. Structure 2019;27(6):1034–1040.e3. https://doi.org/10.1016/j.str.2019.03.018.
[124] Lin J-H, Perryman AL, Schames JR, McCammon JA. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. Biopolymers 2003;68(1):47–62. https://doi.org/10.1002/(ISSN)1097-028210.1002/bip.v68:110.1002/bip.10218.
[125] Österberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in autodock. Proteins Struct Funct Genet 2002;46(1):34–40. https://doi.org/10.1002/prot.10028.
[126] Amaro RE, Baron R, McCammon JA. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. J Comput Aided Mol Des 2008;22(9):693–705. https://doi.org/10.1007/s10822-007-9159-2.
[127] Kelley LA, Gardner SP, Sutcliffe MJ. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Eng 1996;9(11):1063–5. https://doi.org/10.1093/protein/9.11.1063.
[128] Wagner JR, Sørensen J, Hensley N, Wong C, Zhu C, Perison T, et al. POVME 3.0: Software for Mapping Binding Pocket Flexibility. J Chem Theory Comput 2017;13(9):4584–92. https://doi.org/10.1021/acs.jctc.7b0050010.1021/acs.jctc.7b00500.s002.
[129] Ortiz CLD, Completo GC, Nacario RC, Nellas RB. Potential Inhibitors of Galactofuranosyltransferase 2 (GlfT2): Molecular Docking, 3D-QSAR, and *In Silico* ADMETox Studies. Sci Rep 2019;9:1–28. https://doi.org/10.1038/s41598-019-52764-8.
[130] Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput 2013;9(7):3084–95. https://doi.org/10.1021/ct400341p.
[131] Akbar R, Jusoh SA, Amaro RE, Helms V. ENRI: A tool for selecting structure-based virtual screening target conformations. Chem Biol Drug Des 2017;89(5):762–71. https://doi.org/10.1111/cbdd.2017.89.issue-510.1111/cbdd.12900.
[132] Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov 2015;10(5):449–61. https://doi.org/10.1517/17460441.2015.1032936.
[133] Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA. Predicting absolute ligand binding free energies to a simple model site. J Mol Biol 2007;371(4):1118–34. https://doi.org/10.1016/j.jmb.2007.06.002.
[134] Shirts MR, Mobley DL, Chodera JD. Alchemical Free Energy Calculations: Ready for Prime Time? Annu. Rep. Comput. Chem., vol. 3, Elsevier B.V.; 2007, p. 41–59. https://doi.org/10.1016/S1574-1400(07)03004-6.
[135] Deng Y, Roux B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. J Phys Chem B 2009;113(8):2234–46. https://doi.org/10.1021/jp807701h.
[136] Wang C, Nguyen PH, Pham K, Huynh D, Le T-B, Wang H, et al. Calculating Protein-Ligand Binding Affinities with MMPBSA: Method and Error Analysis. J Comput Chem 2016;37(27):2436–46. https://doi.org/10.1002/jcc.v37.2710.1002/jcc.24467.
[137] Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, et al. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chem Rev 2019;119(16):9478–508. https://doi.org/10.1021/acs.chemrev.9b00055.
[138] Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–97. https://doi.org/10.1021/ar000033j.
[139] Swanson JMJ, Henchman RH, McCammon JA. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. Biophys J 2004;86(1):67–74. https://doi.org/10.1016/S0006-3495(04)74084-9.
[140] Genheden S, Ryde U. Comparison of end-point continuum-solvation methods for the calculation of protein-ligand binding free energies. Proteins Struct Funct Bioinforma 2012;80(5):1326–42. https://doi.org/10.1002/prot.24029.
[141] Ntie-Kang F, Kannan S, Wichapong K, Owono Owono LC, Sippl W, Megnassan E. Binding of pyrazole-based inhibitors to *Mycobacterium tuberculosis* pantothenate synthetase: docking and MM-GB(PB)SA analysis. Mol BioSyst 2014;10(2):223–39. https://doi.org/10.1039/C3MB70449A.
[142] Li J, Abel R, Zhu K, Cao Y, Zhao S, Friesner RA. The VSGB 2.0 Model: A Next Generation Energy Model for High Resolution Protein Structure Modeling. Proteins 2011;79(10):2794–812. https://doi.org/10.1002/prot.23106.
[143] Lawrenz M, Baron R, Wang Yi, McCammon JA. Effects of Biomolecular Flexibility on Alchemical Calculations of Absolute Binding Free Energies. J Chem Theory Comput 2011;7(7):2224–32. https://doi.org/10.1021/ct200230v.
[144] Wang Bo, Li L, Hurley TD, Meroueh SO. Molecular Recognition in a Diverse Set of Protein-Ligand Interactions Studied with Molecular Dynamics Simulations and End-Point Free Energy Calculations. J Chem Inf Model 2013;53(10):2659–70. https://doi.org/10.1021/ci400312v.
[145] Amezcua M, El Khoury L, Mobley DL. SAMPL7 Host-Guest Challenge Overview: assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations. J Comput Aided Mol Des 2021;35(1):1–35. https://doi.org/10.1007/s10822-020-00363-5.

[146] Weiss DR, Karpiak J, Huang X-P, Sassano MF, Lyu J, Roth BL, et al. Selectivity Challenges in Docking Screens for GPCR Targets and Antitargets. J Med Chem 2018;61(15):6830–45. https://doi.org/10.1021/acs.jmedchem.8b00718.
[147] Motamen S, Quinn RJ. Analysis of Approaches to Anti-tuberculosis Compounds. ACS Omega 2020;5(44):28529–40. https://doi.org/10.1021/acsomega.0c0317710.1021/acsomega.0c03177.s00110.1021/acsomega.0c03177.s002.
[148] Janardhan S, Ram Vivek M, Narahari Sastry G. Modeling the permeability of drug-like molecules through the cell wall of *Mycobacterium tuberculosis*: an analogue based approach. Mol BioSyst 2016;12(11):3377–84. https://doi.org/10.1039/C6MB00457A.
[149] Merget B, Zilian D, Müller T, Sotriffer CA. Structural bioinformatics MycPermCheck: the *Mycobacterium tuberculosis* permeability prediction tool for small molecules. Bioinformatics 2013;29:62–8. https://doi.org/10.1093/bioinformatics/bts641.
[150] Ekins S, Pottorf R, Reynolds RC, Williams AJ, Clark AM, Freundlich JS. Looking Back to the Future: Predicting in Vivo Efficacy of Small Molecules versus *Mycobacterium tuberculosis*. J Chem Inf Model 2014;54(4):1070–82. https://doi.org/10.1021/ci500077v.
[151] Hong X, Hopfinger AJ. Molecular Modeling and Simulation of *Mycobacterium tuberculosis* Cell Wall Permeability. Biomacromolecules 2004;5(3):1066–77. https://doi.org/10.1021/bm0345155.
[152] Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. Drug Discov Today 2021;26(2):511–24. https://doi.org/10.1016/j.drudis.2020.12.009.