Molecular dynamics (MD) is an important tool that can offer significant benefits to structure-based drug design. This review addresses the theoretical background and various applications of MD that can transform the current drug discovery efforts.

Molecular dynamics-driven drug discovery: leaping forward with confidence

Aravindhan Ganesan[1], Michelle L. Coote[2] and Khaled Barakat[1]

[1] Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, AB, Canada
[2] ARC Centre of Excellence for Electromaterials Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

Given the significant time and financial costs of developing a commercial drug, it remains important to constantly reform the drug discovery pipeline with novel technologies that can narrow the candidates down to the most promising lead compounds for clinical testing. The past decade has witnessed tremendous growth in computational capabilities that enable *in silico* approaches to expedite drug discovery processes. Molecular dynamics (MD) has become a particularly important tool in drug design and discovery. From classical MD methods to more sophisticated hybrid classical/quantum mechanical (QM) approaches, MD simulations are now able to offer extraordinary insights into ligand–receptor interactions. In this review, we discuss how the applications of MD approaches are significantly transforming current drug discovery and development efforts.

Introduction

The quest for new drugs has always remained crucial throughout human history. From the influenza epidemics of the 1800s and 1900s [1] to the recent Ebola virus outbreaks [2], the population of the world has constantly faced epidemics, in addition to life-threatening diseases, such as cancer. Thus, drug discovery continues to be the most significant challenge for the scientific community. The overall drug discovery process, from the identification of potential lead compounds to the US Food and Drug Administration (FDA) approval of a drug, is not only complex, but also expensive and time consuming. A recent report[1] published by the Tufts Center for the Study of Drug Development (CSDD) estimated the overall cost of developing an approved drug to be a staggering US$2.6 billion, with an average of approximately 14 years to complete the entire development cycle of a single drug (from the research lab to market) [3].

Drug design and development have matured over the past two decades by exploiting the advantages of new experimental techniques and complementary technologies. The early 1990s

Corresponding author: Barakat, K. (kbarakat@ualberta.ca)

[1] http://csdd.tufts.edu/news/complete_story/pr_tufts_csdd_2014_cost_study.

Aravindhan Ganesan has a Masters in Bioinformatics (from India) and a PhD awarded by Swinburne University of Technology, Australia, where he worked with Professor Feng Wang. Following the PhD, he worked with Professor Coote at the Australian National University (ANU), before joining Dr Barakat's team at the University of Alberta in 2015. In his current position, Dr Ganesan is applying his interdisciplinary skills towards discovering novel drugs for targeting immune-checkpoint receptors to fight human cancers. He has achieved prestigious awards, including the German DAAD research fellowship and the Australian Endeavour postdoctoral research fellowship.

Professor Michelle Coote, Australian National University (ANU), leads a multidisciplinary research team working in the broad area of computer-aided chemical design. She is a graduate of the University of New South Wales, where she completed a B.Sc. (Hons) in industrial chemistry (1995), followed by a PhD in polymer chemistry (2000). She joined the ANU in 2001, initially as a postdoctoral fellow with Professor Leo Radom. She has published extensively in the fields of computational chemistry, polymer chemistry and physical organic chemistry, and is a member of the ARC Centre of Excellence for Electromaterials Science. She has received many awards, including the Pople Medal of the Asia-Pacific Association for Theoretical and Computational Chemistry (2015). In 2014, she was elected to the Fellowship of the Australian Academy of Science.

Khaled Barakat is the leader of a multidisciplinary world-class research team developing novel immunotherapy drugs targeting immune checkpoint proteins. He received his PhD in biophysics from the University of Alberta in 2012, followed by a 2-year postdoctoral fellowship in Professor Michael Houghton's Lab. During his career, Dr Barakat has received numerous awards, including CIHR and AIHS postdoctoral fellowships, the prestigious UofA dissertation award, the ACRI Studentship and many distinction awards throughout his undergraduate and graduate studies. He is also the editor of several journals. Dr Barakat's lab is supported by various funding agencies, including the Alberta Cancer Foundation, Natural Sciences and Engineering Research Council (NSERC), Li Ka Shing Institute of Virology, and IC-IMPACTS Centres of Excellence.

http://dx.doi.org/10.1016/j.drudis.2016.11.001

GLOSSARY

Free energy of binding within the context of ligand–protein complexes in drug design, the free energy of binding is defined as the free energy difference between the ligand-bound state (complex) and the free unbound states (free protein and free ligand)
Induced-fit effect conformational changes in an enzyme triggered by the interactions with (or binding of) small molecules or other proteins
Molecular docking a method to predict the favoured binding orientations between two molecules to form a stable complex
Periodic boundary condition (PBC) a method used in MD simulations to eliminate the issues concerning boundary effects, arising from finite size, by treating the system as infinite with the help of a unit cell
Scoring function mathematical method to quantify the interactions between two molecules when they are docked together
Shape matching a sampling method that uses receptor complementarity as a criterion for identifying the ligand-binding conformations
Stochastic algorithms a sampling method that incorporates random changes to the ligand in transitional, rotational, and conformational space to identify the most suitable ligand-binding conformation
Systematic search a sampling method that utilises all degrees of freedom to sample the ligand-binding conformations
Virtual screening a computational approach used in structure-based drug design to screen a library (or libraries) of small molecules against the desired protein target to rank them based on their affinities to the concerned binding site of the target; also known as 'in silico screening'

saw rapid advancements in combinatorial chemistry and high-throughput gene-sequencing technology. These enabled the synthesis of huge compound libraries within a short span of time and their screening for various targets, thereby accelerating the discovery processes. This raised the hope of transforming the drug discovery field, making natural products obsolete. However, over time, the field of combinatorial chemistry began to face many technical challenges. In particular, the combinatorial libraries did not cover many structurally diverse compounds [4]. Furthermore, the compounds in these libraries were not as stereochemically rich as natural products. Therefore, the exploitation of these compound libraries did not result in expected fruitful outcomes; in fact, they escalated the costs of testing [5] and resulted in reduced success rates. For example, until recently, only two compounds generated *de novo* had reached the clinic [6]. One was sorafenib from Bayer, which was first approved by FDA in 2005 as a drug for cancer. The second was ataluren, which was approved in the European Union in 2014 as a drug for the treatment of genetic disorders [6]. Nevertheless, there have been some significant efforts towards improving the combinatorial chemistry field. For instance, schemes were developed to address the lack of diversity, and included diversity-oriented synthesis [7], which uses a 'build/couple/pair strategy' [8]. In addition, strategies, such as 'split and pool solid phase synthesis', were developed as a more powerful approach for synthesising huge combinatorial chemistry libraries. Despite many efforts, the field of combinatorial chemistry has still not reached full capacity. Kodadek [9] discussed various recent advances in combinatorial chemistry, which led to a focus on computational methods as low-cost tools for driving the early search process for compounds with desired biological activity and pharmacological profiles, before initiating experiments.

Structure-based drug design (SBDD) is one of various vital computational approaches that have been found to be effective in the identification of hits for *in vitro* testing. As its name indicates, in principle, knowledge of the 3D structures of proteins and ligands is required to perform SBDD. Recently, there has been a dramatic accumulation of biological data, from gene sequences to 3D structures of proteins and compound databases, which offers excellent support to SBDD research. As of June 2016, the Protein Data Bank (PDB) (www.pdb.org) contained more than 100 000 experimentally determined (e.g., via X-ray, NMR, and electron microscopy) protein structures, of which almost 26% correspond to human proteins. The UniProtKB/Swiss-Prot genome database (www.uniprot.org) contains ~540 000 amino acid sequences. These huge databases offer a gamut of potential targets for several human diseases. Moreover, when the experimentally determined 3D structures of any proteins (or enzymes) are not available in PDB, computational models of the unknown proteins for subsequent *in silico* studies can be constructed using SBDD-based methods, such as homology modelling, threading, and *de novo* designing [10]. The success of virtual screening (see Glossary) and SBDD is also dependent on the availability of different compound libraries that comprise chemical compounds from diverse structural classes, to increase the probability of obtaining novel hits. There are several freely available compound databases, such as ZINC15 [11,12] (~120 million compounds), Chemspider [13] (35 million compounds), ChEMBL [14] (~2 million compounds), DrugBank [15] (~14 000 compounds), PubChem [16] (64 million compounds), among others.

When a specific target and compound libraries are selected, molecular docking-based virtual high-throughput screening is used to identify only those compounds (from the libraries) with higher affinities to the active site of the protein [17]. The proteins are dynamic biological molecules and their flexibilities have vital roles in the process of ligand recognition and, thus, in SBDD. In addition, ligand binding also tends to induce measurable levels of conformational changes in the proteins to adapt a biophysical state that is suitable to form a strongly bound complex (known as 'induced-fit' effects). Nevertheless, accounting for receptor flexibilities remains a major challenge and regular molecular-docking methods are mostly unable to capture such conformational changes in proteins.

MD is a computational method that can take on this challenge and predict the time-dependent behaviour of a molecular system, thus becoming an invaluable tool in SBDD. It has been particularly valuable in exploring the energy landscapes of proteins and identifying their physiological conformations, which are, in many cases, not even accessible through high-resolution experimental techniques. MD is also useful in the structural refinements of postdocking complexes, such that the complementarity between the ligand and the receptor is enhanced in the complex state, allowing better rescoring of complexes.

In this review, we discuss in detail the various applications of MD approaches in modern drug discovery efforts. Although, there

have been several recent *Rev*. [18–26] focussing on the usefulness of MD in drug design, they mostly focussed on the theoretical background, applications of MD for accounting protein flexibility and a few binding free energy methods. However, here we complement these existing *Rev*. by addressing various aspects of SBDD, for which MD methods and QM/molecular mechanics (MM) approaches can offer some valuable solutions. The review begins by briefly introducing molecular docking and virtual screening in SBDD. We discuss the recent developments in docking methods and how they struggle to account protein flexibility in SBDD. Subsequently, we discuss in detail how MD is helping to fill this gap and various applications of MD in SBDD, including postdocking structural refinements and accurate binding free energy estimations. Various binding free energy methods and their recent developments are presented, along with several examples. In addition, we also discuss an emerging trend of using solvent information more explicitly from MD simulations, which provide significant information regarding the effects of water molecules in drug design. Furthermore, we also highlight the various limitations to MD methods and, subsequently, we discuss the applications of advanced hybrid QM/MM MD in drug design. Finally, we present a simple and practical workflow for integrating the various computational methods discussed in this review for SBDD.

Molecular docking and flexibility challenges

Molecular docking protocols predict the optimal placement of each compound within a predefined active site of a protein target. They generate a comprehensive set of conformations of the ligand–receptor complex (predominantly based on the ligand poses). These poses are subsequently ranked based on their stability using different scoring functions [27]. There are several programs for ligand–protein docking, including DOCK [28], AutoDock [29], Gold [30], and GLIDE [31]. These docking-based methods have been of great use in modern drug discovery campaigns, mainly because of their speed and simple set-ups.

Early docking methods assumed that the ligand–protein binding phenomenon could be modelled as a simple 'lock-and-key' scheme. That is, the aim was to identify a ligand (i.e., a key) with the exact shape complementarity to fit within a stiff active site cavity (as a keyhole) of the protein. In this way, most early docking algorithms treated the ligand and the receptor as two rigid counterparts. However, this assumption holds well only for rare cases, such as the trypsin–BPTI complex, in which the interfaces of the bound and unbound states are almost identical in their conformations [32]. However, it does not reflect the reality in most cases, where both ligands and receptors undergo mutual changes to accommodate each other in the complex state. Thus, the ligand–protein binding mechanism is now described as a 'hand-and-glove' scheme (Fig. 1), indicating that the best fit is still an essential factor but in a flexible environment [33]. Most of the current docking software programs have adopted ligand sampling as one of the basic elements in their docking protocols. Several sampling algorithms, such as shape matching, systematic search, and stochastic algorithms, are currently used in docking to generate several ligand conformers (often referred as 'poses') around the given receptor environment [34]. For example, software programs such as GLIDE [31] and LUDI [35] implement systematic search methods in docking, whereas AutoDock incorporates stochastic methods to account for ligand flexibility in docking. Thus, there have been significant advancements in the methods to allow exhaustive ligand flexibility in docking-based virtual screening [34].

By contrast, protein flexibility has been almost ignored in docking calculations. Few techniques, such as soft-docking and rotamer libraries [34,36], have been developed to tackle this problem. In soft-docking, the protein flexibility is implicitly included during the calculation, by softening the interatomic van der Waals terms in the scoring function such that it allows small levels of overlap between the receptor and ligand [34,36]. Software programs such as GOLD and AutoDock implement soft-docking. Some programs attempt to implement protein conformational changes into docking calculations by treating the side chains as flexible, while retaining the rigidity of backbone atoms [34,36]. These methods use rotamer libraries, which comprise a list of side-chain conformations determined by experiments and statistical analyses. GLIDE [31], for example, adapts an induced-fit docking method, where selected side chains are mutated into alanine residues to avert steric clashes during docking [31]. Later on, the conformations of these side-chains are adjusted to generate possible configurations that can adapt to the new environment, followed by energy minimisation of the binding site.

Nevertheless, such attempts only allow local movements of some selected residues in the active site, but are not able to capture the overall effects of ligand binding on the conformation of proteins. To overcome this issue, an ensemble of protein structures can be used to account for the full receptor flexibility during docking. This method has become one of the most widely accepted techniques in SBDD. In this approach, all protein structures are combined to form a single representation [18,36] that includes conformational changes that occur during the ligand-binding process. This is usually achieved by averaging the grids of the different protein conformations (in the ensemble) into a single global receptor grid that is then used in molecular docking. Knegtel *et al*. [37] were one of the first to use an averaged grid generated from different experimentally determined structures for ligand docking. The authors used this approach for different test cases, including HIV protease, ras p21 protein, uteroglobins, and retinol-binding protein. They found that the averaged grids approach exhibited better accuracies compared with those of a single structure. The issue of protein flexibility in docking was also addressed by using a united description scheme [38]. In this way, multiple experimentally derived protein structures are superimposed, where the similar segments in the ensemble structures are aligned and fused together, while the variable regions are used as an ensemble. The ensemble of varied segments of proteins is combinatorially explored to produce possible new conformations of proteins for docking calculations [18,34]. However, this approach relied heavily on the quality of the ligand conformational sampling. In addition, such approaches account only for the ligand–protein interaction energy where the internal energy of the protein is mostly neglected [18] (Fig. 2).

An alternate ensemble-based strategy to model protein flexibility in molecular docking is to explicitly consider multiple individual receptor conformations [39] and perform rigid docking of ligands against all those target structures. An ensemble of protein configurations is usually generated from an NMR structure of the

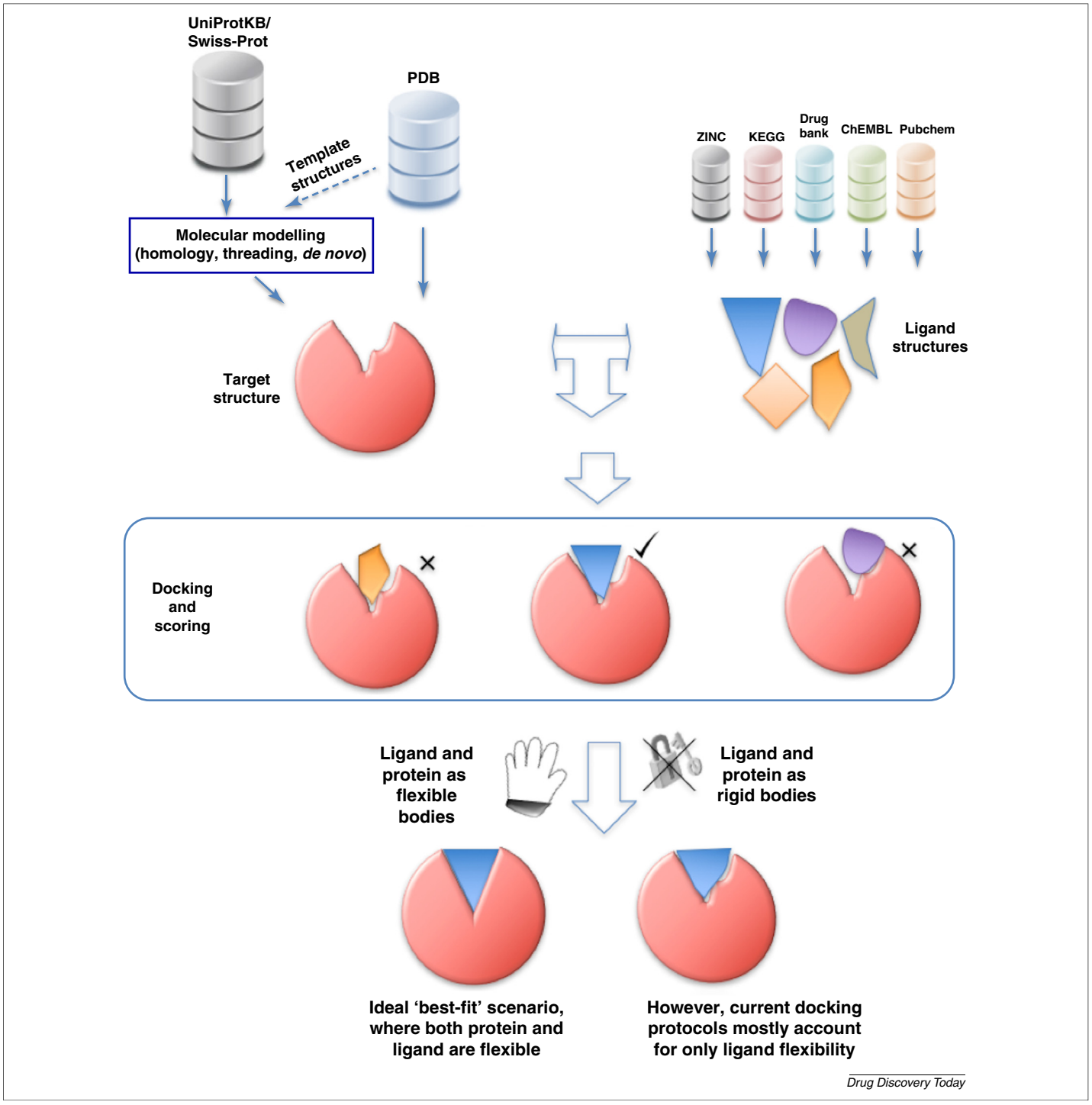

FIGURE 1

A schematic representation of induced effects of ligand binding to its receptor.

chosen receptor or a set of X-ray crystal structures for the same receptor but with different ligands. Nevertheless, the main pitfall with using an ensemble of X-ray crystal structures is that the subsequent docking (or virtual screening) could be biased towards the structures available. This could be even more troublesome if all the available structures are co-crystallised with analogous ligands. By contrast, in the absence of those experimental structures, modelling and MD simulations can be carried out to collect statistically significant protein conformations from the resulting (MD) trajectories. More discussions about this strategy are provided in the following sections. In fact, this combination-approach (i.e., mixing MD and molecular docking) is becoming more common [40–43], irrespective of the availability of experimental structures. For example, in a recent study, Campell *et al.* [44] presented an approach that uses a biased-MD simulation on the known X-ray crystal structure(s) of ligand–protein complex(es), followed by

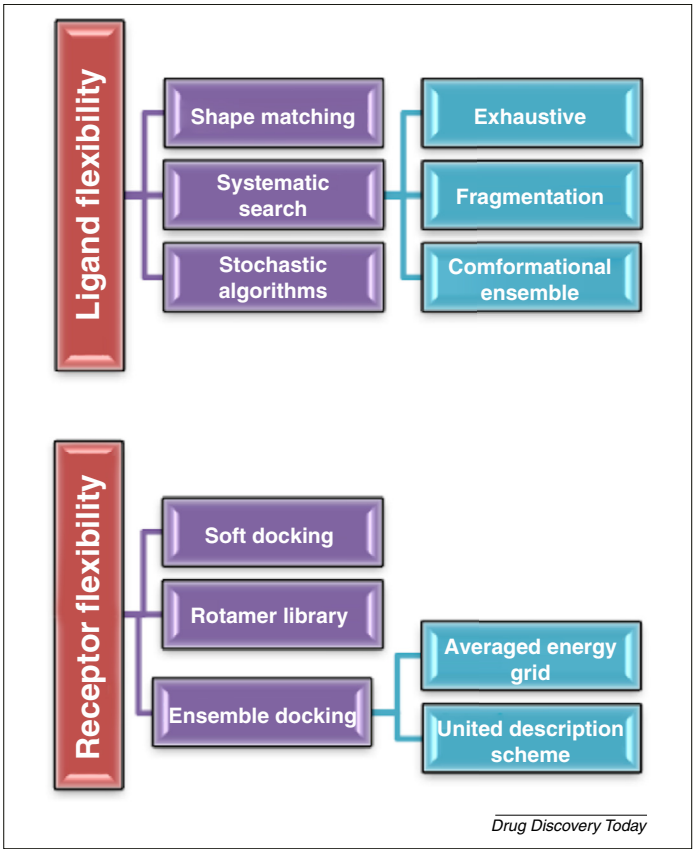

FIGURE 2

Different categories of method used for accounting ligand and receptor flexibilities in molecular docking.

rigid ligand docking to identify the best ranking pose for the complex(es). To demonstrate this scheme, the authors selected two test systems, cyclin-dependent kinase 2 (CDK2) and factor Xa (FXa) [44]. The authors collected the available crystal structures of these systems and performed MD simulations by introducing an external bias potential to retain the initial ligand conformation, thereby also maintaining the binding state that had been known to exist. Later, the authors collected a cluster of protein conformations from the MD trajectory and used them for the ensemble-based docking of a new set of ligands in the known pocket. This work [44] demonstrated that, despite the availability of crystal structures, MD simulations can be useful to account for protein flexibility in docking-based virtual screening. A recent study [45] showed that enrichment performances of virtual screening against three different targets (neuraminidase, HIV protease, and peroxisome proliferator-activated alpha receptors) displayed excellent improvements when using MD-based screening. Therefore, MD methods are now recognised as a valuable tool in SBDD.

Classical molecular dynamics

MD is the most widely used computational technique to study the equilibration structures and dynamic interactions of biological systems [20,24,40–43]. It provides additional insights into time-dependent variations and configurational changes in the structures of the biological systems, which can be related to their functionalities [46]. Classical MD regards atoms as solid spheres and the bonds connecting them as springs. This allows the atoms in the system to only oscillate within a specified distance. Classical MD is based on Newton's equations of motion (Eqn 1):

$$m_i \frac{\delta^r r_i}{\delta t^2} = F_i. \quad (1)$$

Here, F_i is the component of the net force acting on the ith atom with a mass, m_i. r_i denotes the position of the atom at time t. The force can then be computed as (Eqn 2):

$$F_i = -\frac{\delta U(r_1, r_2, \ldots, r_n)}{\delta r_i}, \quad (2)$$

where, $U(r_i, r_2, \ldots, r_n)$ is the potential energy function of the specific conformation and can be described by using the concept of a force field with predefined parameters [47]. A force field is a mathematical expression comprising the functional form of the potential energy, which includes the possible bonded (bonds, angles, and dihedrals) and nonbonded interaction (van der Waals potentials and Coulomb potentials) terms between the different atoms in the system. The bond stretching and angle terms are commonly modelled using a harmonic potential function, whereas the dihedrals are expressed as a cosine function. The nonbonded terms are modelled using Lennard-Jones potentials [48] and Coulomb's law. The particle-mesh Ewald (PME) method [49] under periodic boundary conditions is normally used in classical MD simulations to treat long-range electrostatic interactions in the system. Several force fields have been developed for MD simulations of biological systems, such as CHARMM [50], AMBER [51], and GROMOS [52]. Most of these methods have different functional forms to treat MD simulations, which makes it difficult to transfer parameters from one force field to another.

It is generally problematic to compare the performance of different force fields, because the outputs significantly depend on the type of system and properties studied [53]. However, there have been some efforts to compare different force fields and most of them find that the results concerning the structure and dynamics of systems vary depending on the force field. For example, Todorova *et al.* [54] compared five popular force fields (CHARMM27, OPLS, AMBER03, and the united-atom GROMOS 43A1 and GROMOS 53A6) for simulating insulin. The study addressed the effects of each force field on the conformational evolution and structural properties of the peptides and compared them against the established experimental data. The results found that different structural trends emerged depending on the force field used; however, CHARMM27 and GROMOS 43A1 delivered the best representation of the experimental behaviour [54]. Similar conclusions were drawn from several other studies, whereas other studies concluded that no major differences (in properties and performance) were detected when comparing different force fields. Therefore, it is important to make a careful selection of a force field before using it in MD simulations. 'Learning from experience' is one of the practical approaches for choosing a force field for MD simulations. Before choosing a force field, the users need to be clear about the system they are working on and what the key question (or property) is that they are trying to address through MD simulations. Subsequently, the users need to do a literature search to find out whether MD simulations of similar systems or properties have been previously reported and, if yes, what types of force field were applied to those simulations. If more than one force field

has been applied, then the one among them that was able to provide more accurate results needs to be identified. It is also important to benchmark the selected set of atomic force fields to test against reliable metrics. Sometimes, the choice of force field might also depend on the type of water models involved in the simulations, because force fields have been developed for certain water models (e.g., TIP3P, TIP4, and SPC) [54,55]. For instance, it has been suggested that the combinations of TIP3P-AMBER, TIP3P-CHARMM, TIP4P-OPLS, and SPC-GROMOS are more relevant to the experiments [54,55], although there are some exceptions (e.g., [55]). Becker *et al.* [56] listed several considerations for choosing an appropriate force field in material science and engineering, and these suggestions also hold for biomolecular simulations.

Solving Newton's equations of motion analytically is unpractical for the thousands of degrees of freedom typically involved in many MD problems. As a result, numerical integration algorithms, such as Verlet integrator [57], velocity Verlet integrator [58], and leapfrog integrator [59], are usually used to solve these equations and predict the next move for all atoms during MD simulations. Given that the dynamics of the covalent bonds involving hydrogen atoms are not crucial in biological problems, they are usually constrained using integration algorithms, such as SHAKE [60], RATTLE [61], and LINCS [62]. Hence, a time-step value in the range of 1.5–2 fs is possible and has been shown to be suitable for MD simulations of many biological systems [46].

The main advantage of the MD approach is its ability to mimic the experimental conditions in which a typical biological question is addressed. For instance, experiments are carried out by controlling different factors, such as temperature, pressure, number of atoms, ionic concentration, and the type of solvent used to solvate the interacting molecules. All these factors can be readily adjusted and controlled in MD simulations within the context of statistical mechanics ensembles [63]. These ensembles include the microcanonical ensemble (constant total energy), canonical ensemble (constant temperature), and isothermal-isobaric ensemble (constant temperature and pressure). The microcanonical ensemble is the most basic approach and involves a constant number of particles (N), a constant volume (V), and constant energy (E). However, because the condition of maintaining a constant total energy is not realistic [64], the canonical ensemble (NVT) [65] and isothermal-isobaric ensemble (NPT) [66] are commonly used. Several thermostats and barostats, such as Langevin [67], Berendsen [68], and Nose-Hoover [69,70], have been developed to fix the temperature and pressure in MD simulations. In fact, the isothermal-isobaric (NPT) ensemble is the most widely used ensemble in MD simulations, because it reflects the actual experimental conditions. There are several classical MD programs, including but not limited to, AMBER (www.ambermd.org), CHARMM (www.charmm.org), NAMD (www.ks.uiuc.edu/Research/namd/), GROMACS (www.gromacs.org), Desmond (www.deshawresearch.com), and Hyperchem (www.hyper.com). Some important quantities that are frequently used when analysing MD trajectories are provided in Box 1.

BOX 1

Important quantities in MD analyses

Root mean square deviation

RMSD is a measure of the average deviation or distance between the atoms when 3D structures are superimposed on each other. When analysing an MD trajectory, this value (or RMSD) could be an important quantity that is useful to trace how much the structure that underwent MD simulations has deviated from its starting structure.

Interaction energy

The interaction energy is the amount of energy that is caused by the interaction(s) between two residues (or objects) and its contribution towards the total energy of the system. Interaction energies between different amino acid residues from the target and the bound ligand could have a significant impact on the binding affinity of the complex. Thus, identifying the key residues that have high interaction energies against the ligand is important in binding mode analyses.

Interaction distance

The interaction distance is a minimum distance between two nonbonded residues of proteins or between residues and ligand that could affect each other, thereby impacting the total energy of the system.

Correlation functions

Correlation functions are mathematical descriptors that connect the properties of protein structures with that of their significance. Thus, correlation function remains an important tool for protein structure analyses from the MD trajectories.

Radial distribution function

The radial distribution function is a quantity that describes the average radial packaging of atoms in a system. It can be calculated by constructing normalised histograms of atom pair distances with respect to an ideal gas (Eqn I).

$$g(r) = \frac{n(r)}{4\pi r^2 \rho \Delta r} \quad \text{(I)}$$

where, $n(r)$ is the number of atoms in a shell of width Δr at distance r and ρ is the mean atom density. This quantity can be useful, for instance, to identify how many waters are coordinating with a metal ion in the active site of the protein during the course of MD simulation.

Hydrogen bond

A hydrogen bond (H-bond) reflects the electrostatic force that attracts the H attached to one electronegative atom to another electronegative atom holding a lone pair of electrons. Thus, identifying the number of H-bonds between the bound ligand and its surrounding amino acid residues of the protein is one of the key steps while analysing MD trajectories.

MD simulations and protein flexibility

The dynamic nature of proteins is a well-established phenomenon [71]. Proteins are flexible biological molecules that can adopt multiple conformational states in solution [18]. Few of these conformations are able to bind efficiently to the ligands and/or other systems in the environment. For example, certain configurations of proteins can adapt an open state that keeps the channels accessible for water molecules and ligands to bind/unbind freely [72,73]. By contrast, in some other conformations of the same proteins, the highly malleable loops can block the channel partially or completely, thereby restricting ligand access. In addition, binding of the ligand can also lead to conformational changes in proteins, from local reorganisation of side chains to hinge dynamics of domains [40–42]. As a result, proteins often shift between different conformational states separated by low- and high-energy barriers in the free-energy landscapes during chemical reactions. Histone deacetylase 8 (HDAC8) is one of the best examples of dynamic mobility in proteins. These unusual

dynamics of HDAC8 have been captured by at least 21 different experimentally determined structures [Protein Data Bank (PDB) IDs: 1T64, 1T67, 1T69, 1VKG, 2W22, 2V5W, 2V5X, 3EW8, 3EWF, 3EZF, 3EZT, 3F06, 3F07, 3F0R, 3SFF, 3SFH, 3MZ3, 3MZ4, 3MZ6, 3MZ7, and 3RQD] [46,71,74]. By comparison of all the reported experimental structures, it was found that an 11 Å-deep active-site pocket of the enzyme changes between a broadly open conformation to a partially open state and a fully closed structure [71,74]. Some experimental structures of HDAC8 also display an extra pocket that lies parallel to the main pocket (Fig. 3). All these structures are proposed to exist in equilibrium and are involved in ligand binding/unbinding, product release, or water transfers [46,71]. Furthermore, some proteins could have additional druggable binding sites, which are cryptic in nature and have the potency to modulate the functionalities of the concerned receptors allosterically. Such cryptic or allosteric binding sites are usually not easily detectable in the ligand-free structures, as in TEM1 β-lactamase [75] and p38 MAP kinase [76] for instance, and require significant conformational changes in the receptors to become visible. Hence, these sites are usually not detectable from a single representative structure and require large conformational sampling to reveal them. A well-known success story of MD in such applications is with regards to the discovery of a novel-ligand binding trench in the HIV-integrase enzyme. In 2004, Schames *et al.* [77] performed MD simulations of the HIV-integrase enzyme along with the docked ligand and discovered a novel ligand-binding region, the trench. The existence of this cryptic site was later also confirmed by X-crystallography. Subsequently, scientists from Merck along with their collaborators performed intense experimental research [78] on this novel binding site, which eventually led to the development of novel anti-HIV inhibitors, such as raltegravir [19].

Therefore, it is logical to use an ensemble of protein conformations in SBDD instead of a single representation. Nevertheless, because of high costs and technical complexities, experimentally determined structures for different conformations are only available for few proteins. As discussed in earlier sections, MD simulations are now being used to collect ensembles of protein structures for SBDD to close this gap. Under this MD scheme, the target structure (obtained from PDB or computational modelling) is initially subjected to large-scale MD simulations followed by root mean squared deviation (RMSD) conformational clustering to accumulate all possible conformations of a typical protein structure. Subsequently, statistical analysis methods, such as principal component analysis (PCA), are then used to transform the original space of correlated variables into a reduced set of independent

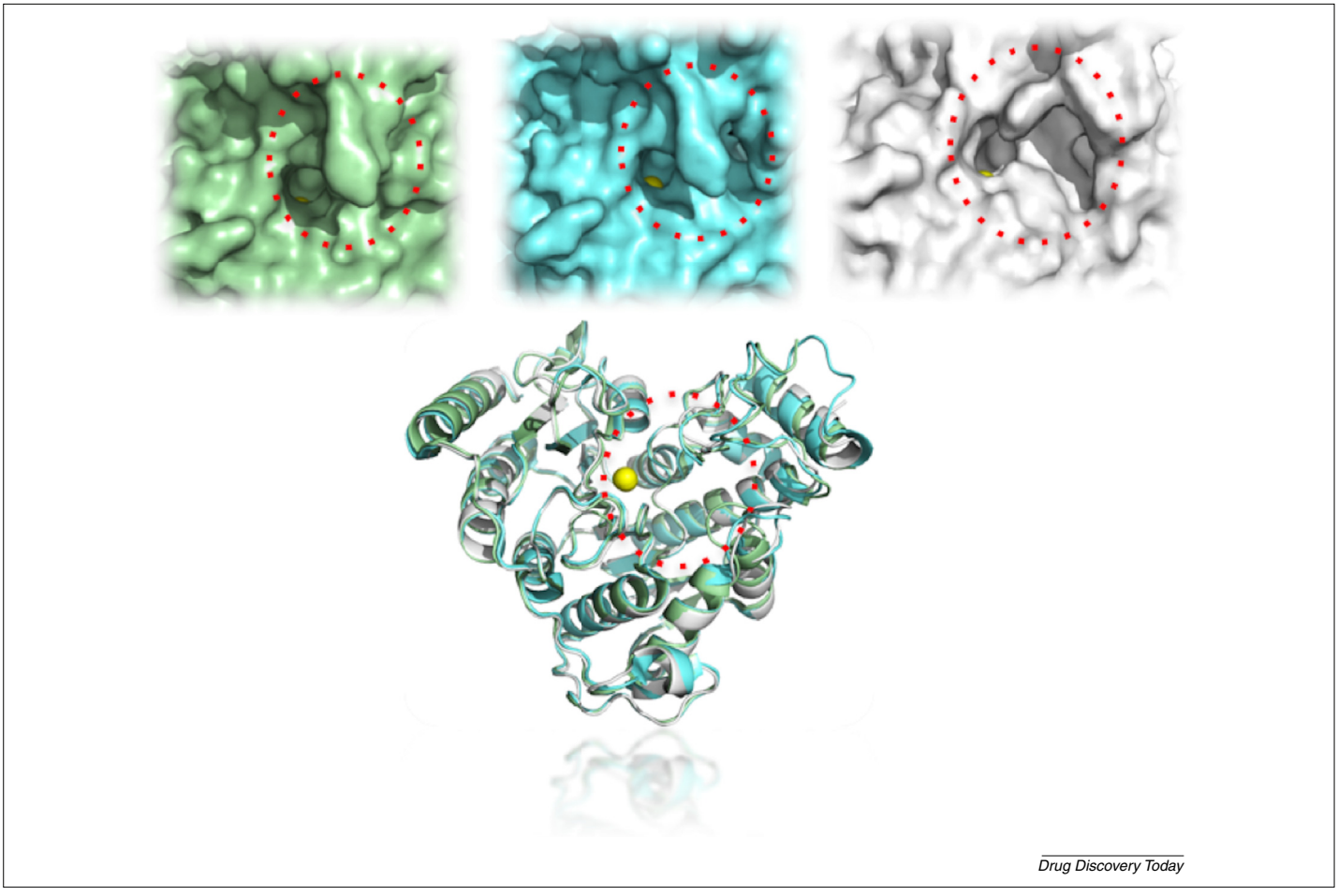

FIGURE 3

Different X-ray crystal structures of histone deacetylase 8 (HDAC 8) showing the different conformations of the binding-site pockets. An overlap of the structures is also shown as a ribbon structure.

variables comprising the most vital dynamics of the system [40–42,79]. This results in an ensemble of protein structures that can be used in docking-based virtual screening. This MD scheme to account for receptor flexibility is popularly known as the 'relaxed complex scheme' (RCS) [42]. RCS has been successfully used in several studies [40–43], for instance, to conduct an ensemble-based virtual screening against the MDM2 protein [41], a main regulator for p53. Over 50-ns MD simulations of the structure of MDM2 were performed using the AMBER99SB force field and NAMD program and 28 distinct conformations of MDM2 were sampled for further virtual screening of several ligand databases [41]. The 28 structures included 22 structures that comprised ~75% of the apo trajectory, five structures representing ~80% of the bound trajectory, and a single MDM2 conformation from the MDM2-p53 crystal structure [41]. The study revealed that MDM2 is a highly flexible protein that adopts distinct conformational changes [41], which were effectively captured using MD simulations, as shown in Fig. 4a. In another study, Bowman *et al.* [80] performed MD simulations of the p53-MDM2 complex and generated multiple structures of the system, so as to account for protein flexibility in their subsequent docking-based virtual screening. This led to the discovery of five small-molecule inhibitors of the human MDM2–p53 interaction, with one compound exhibiting a Ki of 110 ± 30 nM [80]. These small molecules have scaffolds that distinct from nutlin, a known inhibitor of the MDM2–p53 interaction [80]. Thus, by incorporating the RSC approach, it is possible to discover novel therapeutically attractive small molecules. In another study, a MD-based RSC approach was used to develop a computational atomistic model of a human ether-á-go-go-related (hERG) ion channel [40]. Conformational sampling of the MD trajectory of hERG resulted in 45 different clusters that made a comprehensive description of backbone (Fig. 4b) and side-chain dynamics (Fig. 4c) of the inner cavity of the ion channel [40]. This model serves as a powerful tool to predict hERG blocking and can be useful in developing safer and more efficient drugs [40].

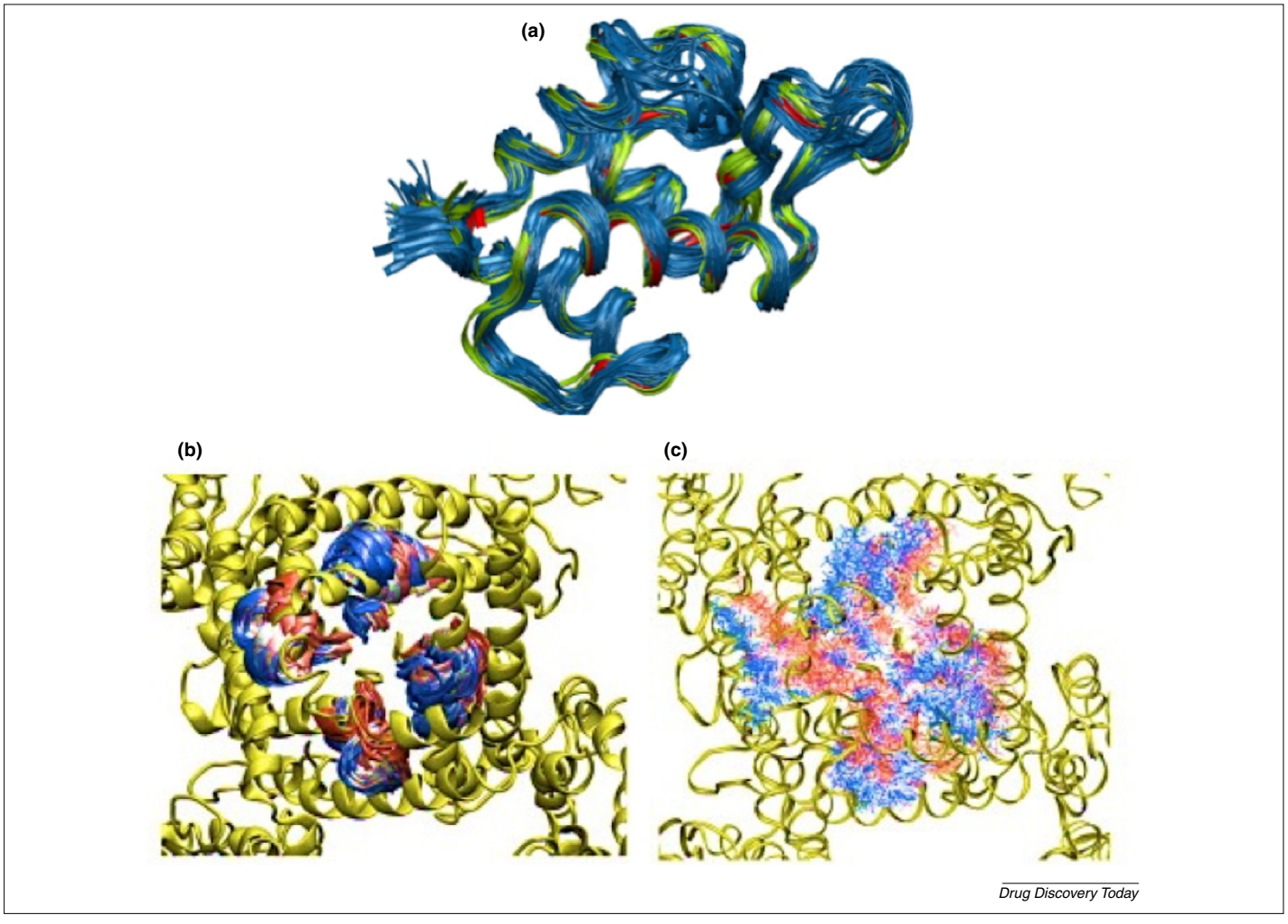

FIGURE 4

Ensembles of structures sampled from long molecular dynamics (MD) trajectories. Twenty-eight structures of the MDM2 protein sampled from a 50 ns-long MD trajectory **(a)** and 45 structures of the hERG ion channel captured from a 500 ns-long MD simulation showing the flexibilities of the backbone region **(b)** and side-chain dynamics **(c)**. In (a), the structures of holo-, and apo-trajectories of MDM2 protein are shown in green and blue, respectively. In (b) and (c), the colours indicate the flexibility of the relevant segments in the dominant conformations. The representative conformation of the target is shown in red and the other conformations in the clusters are provided in colours ranging from red to blue.
Reproduced, with permission, from [41] (a) and [204] (b,c).

In combination with other computational approaches, MD simulations can help in characterising protein–protein interactions. These types of interactions have important roles in several biological processes, including signal transduction, cell metabolism, and/or transport. Understanding these interactions can access a new era of drug discovery, expanding the target space for new and more effective drugs [81]. Although the protein–protein interfaces are generally large, only selected subsets of residues are responsible for the strong binding at these sites. Such residues, along with the surrounding domains, are known as 'hotspots'. Protein–protein interactions are also known to have transient binding pockets that are not captured in experimentally determined structures. MD simulation has become routine in approaches for identifying these hotspots and predicting binding sites for their regulation. For instance, MD simulations have provided a detailed understanding of the dimer interface in the HIV 1 protease enzyme, which is characterised by solvent accessible surface areas and interdimeric hydrogen bonds [82]. In a recent study [83], MD simulations were used to model and characterise the human programmed death-1 (PD-1) bound to its two human ligands, PDL-1 and PDL-2. Table 1 lists some of the studies that have used MD simulations for various applications (such as accounting protein flexibility and dynamics, postdocking structural refinements, and free energy of binding calculations) on different target enzymes (or proteins) over the past 5 years.

MD simulations and postdocking structural refinements

Although docking can predict the optimal placement of a ligand within the active site of a receptor, not all of the key interactions between the ligand and receptor are usually depicted accurately. Hence, it is generally recommended to perform MD simulations on the complexes obtained from docking because this can help in optimising their interactions. For instance, in a previous study [84], molecular docking predicted that sulfonamide derivatives bind effectively into the active site of aldose reductase, which was contrary to the lower activity found for these compounds in experiments. *In silico* refinements of these structures using MD revealed that a water molecule from the exterior migrated to the binding site and interrupted the key interactions between sulphonamide ligands and the receptor. This was identified to be a reason for the reduced activity of the tested compounds in experiments [84]. In another study, MD simulations were used to discern the different docked complexes of propidium and human acetylcholinesterase based on their stability [85]. The most stable structures identified with the help of MD simulations were in excellent correlation with the binding modes proposed by experiments [85]. Similarly, a combination of ensemble-based molecular docking and MD refinements of postdocking complexes helped reveal for the first time a unique symmetrical binding mode of daclatasvir (a drug in Phase 3 clinical trials) with the hepatitis C virus (HCV) NS5A protein and for different HCV genotypes [43] (Fig. 5).

MD has made significant contributions to the understanding the structure properties of G-protein-coupled receptors (GPCRs). For instance, a previous study used postdocking MD simulations to reveal significant dynamic changes in the human CC chemokine receptor 3 (CCR3) and the human muscarinic acetylcholine receptor 3 (hM3R) that influence their ligand-binding modes [24]. In particular, MD simulations found a strong H-bond between the docked ligand and N508 residue of hM3R that is key to holding the complex. This was again confirmed by performing MD simulations of N508A mutant hM3R and the ligand complex, in which the

TABLE 1

Examples of recent studies that used MD simulations for various applications, such as accounting protein flexibility, postdocking structure refinement, and binding free energy calculations

Protein flexibility and conformational analysis		Structure and post-docking refinement		Binding free energy calculations	
Enzyme/target protein	**Refs**	**Enzyme/target protein**	**Refs**	**Enzyme/target protein**	**Refs**
Nucleoprotein of Influenza A virus	[171]	CASP8 and 9 targets	[172]	Cytidine deaminase	[103]
AcrB transporter	[173]	CASP10 targets	[174]	hERG	[40,104]
M3 muscarinic acetylcholine receptor 3	[175]	CASP11 targets	[174]	Mur ligase	[86,168]
Isomerase	[176]	CASP8 and 9 targets	[177]	Enoyl-ACP reductase	[107]
GPCR	[178]	M3 muscarinic acetylcholine receptor 3	[24]	HIV protease	[109]
MDM2-p53	[41,80]	MurD Ligase	[170]	Multiple targets	[110,179,180]
α-Spectrin SH3 protein	[181]	NS5A	[43]	NS5B	[182]
Nalp domain	[183]	Acetylcholinesterase	[184]	Cytochrome P450	[185]
hERG	[40]	Galectin 8C domain	[186]	MDM2-p53	[187]
Melanocortin 4 receptor	[188]	Multiple targets	[189]	NS2B/NS3 Dengue virus	[190]
Histone deacetylases	[46,73,191]	β-Lactamase	[192]	HIV-1 RT RNase	[193]
Glycoprotein	[194]	Neurotoxin serotype A	[195]	ERCC1-XPF	[114]
Lysozyme	[196]	Phosphorylase kinase	[197]	STAT3 and STAT5	[198]
Mad2	[199]	Tubulin	[200]	Phosphorylase kinase	[201]
MurD ligase	[167]	Monoamine oxidase B	[87]	Human biliverdin-IXα reductase	[88]
***Giardia duodenalis* 14-3-3**	[202]	Aldolase reductase	[89]	AF9 protein	[203]

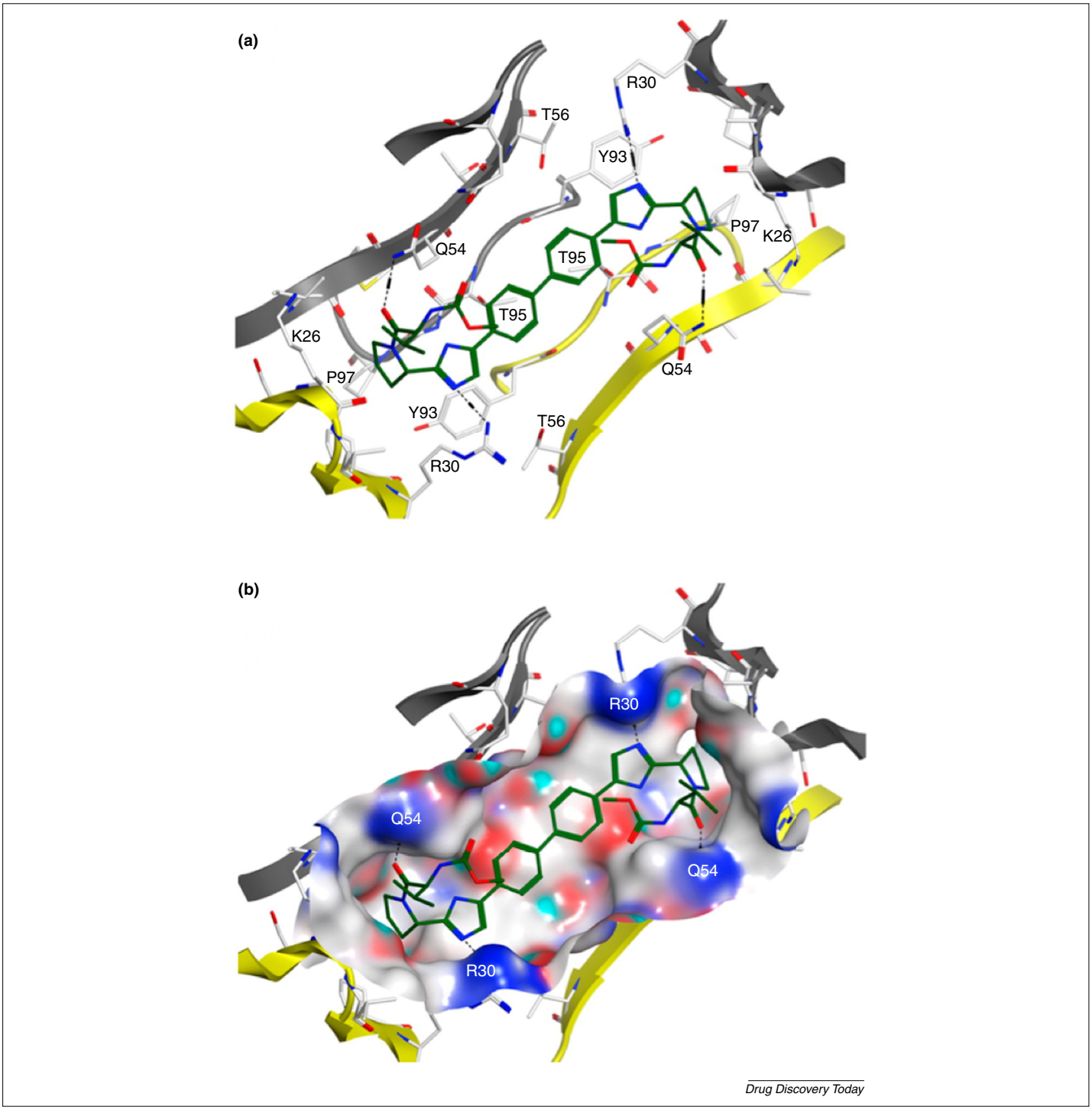

FIGURE 5

Binding mode of daclatasvir with the NS5A protein [43]. The bound drug is shown as a green-coloured stick representation and the protein residues are displayed as white sticks **(a)**. The binding sites within the NS5A receptor are also provided as a surface representation **(b)**.
Reprinted, with permission, from [205].

ligand was found to be pushed to the exit [24]. In another study by Perdih *et al.* [86], the authors used molecular docking and MD simulations, along with a range of experiments, to identify furan-based benzene mono- and dicarboxylic acid derivatives as potential inhibitors of all four bacterial Mur ligases. The authors initially performed *in vitro* testing of seven furan-based benzene-1,3-dicarboxylate derivatives, based on their previous studies, and found out that one of the compounds was able to inhibit all Mur ligases in the micromolar range [86]. Subsequently, this compound was docked into the active site of the MurD enzyme and two different ligand-binding modes were identified. Subsequently, the authors performed ~20 ns-long MD simulations and interaction energy calculations to further refine the postdocking complex and also identify the best binding mode of the ligand. Finally, based on the

results obtained, four novel furan-based benzene-monocarboxylic acids were discovered to inhibit multiple Mur ligases in the low micromolar range [86]. Moreover, one of the newly discovery compounds also exhibited promising antibacterial activity against *Staphylococcus aureus* [86].

Previous studies [87–89] have also shown that integrating an induced-fit docking (IFD) method along with MD and/or QM/MM simulations can be useful for the efficient description of induced molecular flexibilities within the protein–ligand complexes and also for accurate binding-mode analysis of ligands. For example, in a recent study, Distinto *et al.* [87] used IFD and MD simulations to unravel the putative binding modes and activities of 1-aryliden-2-[4-(4-chlorophenyl)thiazol-2-yl]hydrazines against the monoamine oxidase B (MAO-B) enzyme, an attractive target for treating neurodegenerative diseases. By structural alignment of several X-ray structures of MAO-B co-crystallised with different inhibitors, it was understood that the enzyme adopted induced-fit changes with respect to the bound ligands. Hence, the authors initially performed IFD using the Schrodinger drug discovery suite, during which the side chains near the inhibitor were kept flexible. The results from the IFD explained how ligand binding tended to induce structural changed in the protein. However, many of the compounds showing two binding modes were ranked high in IFD. To determine the best binding mode of the inhibitors, the authors performed 3–5 ns-long MD simulations for both the binding modes of two of the top-ranking compounds from IFD. The MD results followed by the free energy calculations highlighted the significance of the fluorine atom interacting with water near the cofactor and the influence of the steric bulkiness of substituents in the arylidene moiety. The authors proposed that the pharmacophore features of these experimentally synthesised compounds, developed using combined IFD, MD and free energy calculations, should be useful for achieving novel high-affinity MAO-B inhibitors for the treatment of neurodegenerative disorders [87].

In another study, Fu *et al.* [88] combined IFD with classical MD simulations, free energy of binding calculations, and QM/MM calculations to study substrate binding to human biliverdin-IXα reductase (hBVR-A) of biliverdin-IXα and four analogues. hBVR-A is a key enzyme regulating a range of cellular processes and is involved in the conversion of biliverdin-IXα to bilirubin-IXα. In this work [88], the authors initially used the structure of the hBVR-A/NADPH/substrate I complex for the docking of analogues into the binding pocket using the IFD procedure implemented in the Schrodinger program. During the IFD, a tyrosine residue in the active site was treated with flexibility. Subsequently, the best-ranking ternary complex structures from IFD were subjected to classical MD simulations [88]. Multiple snapshots obtained from the MD simulations were used for performing free energy of binding calculations. The predicted free energies of binding for five analogues agreed well with the experimental binding affinities and also helped to identify the best binding pose for the complexes [88]. Finally, the authors investigated the catalytic mechanisms of the ternary complex structure (in this study) by calculating the reaction energy profiles using advanced QM/MM calculations. These advanced calculations were useful to understand the reaction mechanisms of the system studied, which in the long term, should assist in the design of potent hBVR-A inhibitors [88]. Thus, MD serves as an important tool for not only refining the postdocking complexes, but also revealing more appropriate binding modes of the ligands within the receptor structures.

MD simulations and predicting the free energy of binding

Molecular recognition is critical in several biochemical and biological processes [90]. Many biological processes are initiated by specific binding between two interacting entities in the cell. Although docking combined with MD simulations can provide a clear image of the shape complementarity between these entities at their binding interface, whether these interactions are significant or realistic requires an additional and essential piece of information, namely the free energy of binding, which is the driving force toward forming this complex. Calculation of the binding free energy [ΔG_{bind}; i.e., the free energy difference between the ligand-bound state (complex) and the corresponding unbound states of proteins and ligands) is used to quantify the affinity of a ligand to its target. Assessing the ΔG_{bind} of a series of ligands against a particular target can discern those ligands with higher binding affinities with the target. Thus, the ΔG_{bind} calculations are important in drug design, and normally follow the docking-based virtual screening processes. Several computational methods, from computationally rigorous thermodynamics pathways approaches to less complex end-point methods, have been developed for ΔG_{bind} calculations. The former methods include thermodynamic integration (TI) and free energy perturbation (FEP) methods, whereas liner interaction energy (LIE), MM-generalised Born surface area (MM-GBSA), and MM-Poisson–Boltzmann surface area (MM-PBSA) are end-point methods. Each of these methods has its own strengths and limitations, and their computational requirements and speed are inversely correlated with their accuracy.

TI and FEP methods are thermodynamic pathways approaches that are commonly used for the calculation of relative binding free energies [91,92]. These methods are mainly based on the application of thermodynamic cycles and, thus, require the transformation of the system from the initial state to the final state through alchemical changes of the system energy function [91]. These methods involve the change of a ligand A into ligand B in two states, such as a solvent-only unbound state (of the ligands) and a bound state (i.e., ligand–protein complexes). This provides free energy changes for the unbound states ($\Delta G_{unbound}$) and bound states (ΔG_{bound}) of the ligands [91]. It is also possible to mutate ligand A to 'nothing', which in principle can provide absolute free energies of binding. Understandably, these methods (TI and FEP) demand multiple MD simulations and rigorous sampling of ligand, protein, and solvent degrees of freedom. As a result, the thermodynamic pathways methods are in general able to provide accurate estimates of the free energy of binding at a cost of high computational time [93,94]. For instance, the TI method coupled with MD simulations has been used to identify potential huperzine derivatives with higher binding affinity towards acetylcholinesterases [95]. Similarly, the FEP approach has also shown to predict more accurate binding free energies for several inhibitor–enzyme complexes [93,96]. However, estimating the ΔG_{bind} values using these methods requires large numbers of conformational samples, which in turn inflates the computational costs heavily [93,94]. Given the need for enormous computational resources, these methods have mostly been applied to only small sets of

ligand–protein complexes. Nevertheless, with increasing supercomputing capabilities and more improved methods, TI and FEP are gradually being included in the SBDD pipeline, especially in guiding lead optimisation [92,97–100]. For instance, a recent study [100] by a large team of authors, from Schrödinger, Nimbus, Columbia, Yale, and UC-Irvine, showed that FEP calculations are able to make highly accurate affinity predictions across a range of ligands and targets. This work included diverse sets of targets, such as BACE, CDK2, JNK1, MCL1, p38, PTP1b, Tyk2, and thrombin. The estimated binding free energies reported in this study were in good agreement with the experiments. Indeed, most of the values were within 1 kcal mol^{-1}, with only nine out of 199 ligands studied deviated from their experimental values by over 2 kcal mol^{-1} [100].

Less rigorous alternatives to thermodynamic pathways are the end-point approaches, which include methods such as LIE, MM-PBSA, and MM-GBSA. Unlike thermodynamic pathways approaches, these end-point methods sample only structures involved at either ends of the reaction pathways; that is, the free receptors (proteins) and ligands and the final ligand–protein complexes. The ΔG_{bind} in this approach can be written as Eqn 3:

$$\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand}) \quad (3)$$

The LIE method, developed by Aqvist *et al.* [101], considers the process of a ligand (*L*) binding to a protein as a partition problem, in which the free ligand (*F*) from the bulk solvent medium is transferred to a solvated protein environment (*P*). Therefore, two independent MD simulations, one for the complex and the other for solvated ligand, needs to be performed to calculate ΔG_{bind} using the LIE method. Nonetheless, the reliance of the LIE on the end-points of binding makes it an attractive (and affordable) approach for lead optimisation in drug discovery. Several studies have used the LIE method for the computational analyses of inhibitors against a variety of targets. This includes benzamide-based thrombin inhibitors [102], inhibitors of *Mycobacterium tuberculosis* H37Rv cytidine deaminase [103], sertindole analogues to block the hERG potassium channel [104], and allophenylnorstatine molecules to inhibit Plm4 enzyme, a target for *Plasmodium malariae* [105], for instance. The LIE method has been shown to predict binding free energies with a root mean square error (RMSE) <1 kcal mol^{-1} compared with the experimental values [106]. As indicated above, the thermodynamic pathways methods are also able to make predictions with a similar, if not better, accuracy range for different targets.

MM-GBSA and MM-PBSA are two other well-established end-point methods that are most popular in SBDD. The two methods use an implicit solvent model to account for the solvent molecules and use dielectric continuum models to obtain the electrostatic components for the solvation energy. The MM-PB(GB)SA ΔG_{bind} can be estimated using Eqn 4:

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{Solv} - T\Delta S \quad (4)$$

Here, ΔE_{MM} refers to the molecular mechanical energy and it is the sum of all energies from the bonded and nonbonded interactions. The solvation energy, ΔG_{solv}, is the sum of the polar and nonpolar contributions of solvation. The polar solvation terms ($\Delta G_{PB/GB}$) are estimated using a generalised-Born model or a Poisson–Boltzmann solver. The nonpolar contributions are computed based on the size of the solvent-accessible surface area (ΔG_{SASA}) in the ligand and protein. The final component of the ΔG_{bind} equation (Eqn 4) is $T\Delta S$, which corresponds to the conformational entropy changes in the reaction product (i.e., protein–ligand complex) upon ligand binding.

The inclusion of conformational entropy ($T\Delta S$) in the MMPB(GB)SA calculations to obtain absolute ΔG_{bind} remains challenging. The accurate calculation of ($T\Delta S$) is computationally expensive and, in many cases, its inclusion does not guarantee better accuracies in the final energies [46,107,108]. Rather, previous studies have shown that accounting for conformational entropy obtained from insufficient MD sampling has adversely affected the calculations [107,109]. For instance, Su *et al.* [107] showed that the accuracies of their MM-PBSA and MM-GBSA calculations for 16 known benimidazole inhibitors against *Francisella tularensis* enoyl-ACP reductase were significantly affected because of using different number of frames for their enthalpy and entropy calculations. The authors sampled 2400 frames from the MD trajectory used for their enthalpy calculations; however, because of limited computational resources, they only used 48 frames (evenly selected from the trajectory) for the entropy calculations. Therefore, it is important to have large numbers of MD snapshots to derive a reliable estimate of absolute ΔG_{bind}, which can significantly increase the computational costs. As a result, many studies tend to neglect $T\Delta S$ and use the 'relative' ΔG_{bind} instead. Relative ΔG_{bind} energies can be predicted with a reasonable accuracy, and are generally sufficient to rank a group of compounds against the same target in SBDD [110].

Two strategies are commonly used in MM-GBSA and MM-PBSA calculations: (i) the three-trajectory scheme; and (ii) the single trajectory scheme [20,46,107,111]. In the former, three different MD trajectories that pertain to the 'apo' protein, free ligand, and the ligand–protein complex (i.e., the end-point structures) are sampled for snapshots. In principle, this three-trajectory scheme provides more accurate results than the single-trajectory approach; however, it has high computational costs [110,111]. By contrast, the single trajectory scheme requires only a single MD simulation for the ligand–protein complex, which significantly reduces the required computational time [20,110–112]. Apart from the choice of strategy, there are several factors that can affect MMPB(GB)SA calculations, which includes length of simulations, choice of the force field, solute dielectric constants, solvent model, and the net charge of the systems. For instance, it has been argued that using multiple short and independent MD simulations, instead of one long MD trajectory, can provide better ΔG_{bind} predictions [107,108,111,113].

There have been several studies that compared and tested the efficiencies of MM-GBSA and MM-PBSA in predicting accurate ΔG_{bind} energies for different ligand–protein complexes. Their general conclusion is that the accuracy of these methods tends to be system dependent. For example, Hou *et al.* [110] found MM-GBSA to predict accurate relative ΔG_{bind} for 59 ligands against six different protein targets when compared with MM-PBSA that outperformed the former in making the absolute ΔG_{bind} predictions. The MM-PBSA method has been extensively applied for a range of studies, including screening and ranking of ligands against the ERCC–XPA complex [41], understanding the binding mode of daclatasvir onto the NS5A viral protein [43], and binding of

human programmed death-1 of T cells with its ligands [83]. By contrast, Oeheme *et al.* [109] concluded that MM-GBSA performed better than MM-PBSA in computing ΔG_{bind} of their ligand–HIV protease systems. Thus, it is clear that neither of these two methods is universally superior and the choice of the method should be made on a case-by-case basis. For example, Jordheim *et al.* [114] combined MD simulations, virtual screening, and MM-PBSA-based binding free energy calculations, along with different experimental techniques, to identify potential inhibitors of ERCC1-XPF protein–protein interactions. The authors performed 20 independent virtual screening runs against the 20 XPF structures present in an NMR ensemble after their MD equilibration. Top hits from each screening were extracted and then ranked based on their binding free energies. From these results, 73 compounds were subjected to a range of experiments, including cytotoxicity assays, steady-state fluorescence and synchronous fluorescence experiments, and immunocytochemistry. The hits were evaluated on A549 and HCT116 cancer cells. Finally, one compound was found to exhibit promising activity in all the experiments and was also able to interact with the domain of XPF that is responsible for interacting with ERCC1, thus disrupting the protein–protein interactions. Thus, MD-based binding free energy calculations are helpful in guiding the hit identification stage. However, one of the significant drawbacks of both methods is their inability to make accurate predictions for ligands with formal charges [109,111,115]. Hence, it is important to improve the existing methods or develop new methods that can account for charged ligands (including tautomers), which form a significant area of drug research.

MD simulation and solvent dynamics analyses

Computational analyses of the structure and thermodynamic properties of water have recently become a useful tool in SBDD [116–119]. The properties of surface water molecules have been proposed to have important roles in molecular recognition and ligand–protein (and/or protein–protein) interactions in solution [116,120]. Although small in size, water molecules are involved in a range of interactions, including H-bonds and van der Waals contacts [120]. As a result of such interactions, it is often difficult to displace water molecules to facilitate the binding of a drug. Therefore, the energy involved in relocating water molecules between surface layer and bulk water upon binding of a macromolecule (protein for instance) with another macromolecule and/or ligands can significantly impact the overall free energy of binding [116–120]. Hence, the hydration patterns of a binding pocket can offer important insights into the properties of the pocket and also quantify the hydrophobic forces involved in the binding of small-molecule drugs with proteins. There are several *in silico* tools that can help in extensive molecular descriptor analyses of solvation from the MD trajectories. These algorithms include WaterMap (from Schrodinger) [121], WaterFLAP (from Molecular Discovery) [122], SZMAP (from OpenEye) [123], GIST (in Amber) [120,124], WatMD (in-house tool of Novartis) [116,125], SPAM (from GlaxoSmithKline) [126], STOW [127], and WatClust [127]. Some of these methods, WaterMap [121], STOW [127] and WatClust for instance, are based on inhomogeneous fluid solvation theory (IST), where enthalpy is accounted directly from nonbonded interactions and entropy is calculated from a local expansion in terms of correlation functions [121].

For instance, the WaterMap program [121] initially clusters the water molecules (in the MD trajectory) based on their spatial distribution, such that they form individual hydration sites. Subsequently, the hydration sites are analysed using IST to determine the enthalpy and entropy properties of water molecules within each site. This method has been successfully used to gain insights into binding sites for various systems, including peptides binding to PDZ domains [128], the FKBP12 protein [129], protease and kinase binding affinity [130,131], and the A2A GPCR [132]. For example, Beuming *et al.* [129] used the WaterMap tool to analyse the hydration sites for a panel of 27 different protein targets across a range of families. Initially, the authors [129] performed ~2 ns-long MD simulations for each of the targets and the resultant trajectories were subjected to analyses with the WaterMap program. The results [129] revealed ~31 500 hydration sites in the targets, for which the authors calculated their thermodynamic information (such as free energy, entropy, and enthalpy). The authors further demonstrated that such thermodynamic properties of the hydration sites could be used to identify potential binding sites and evaluate their druggability [129]. It was found that clusters of high-energy solvation sites were mostly related with binding sites. However, Ramsey *et al.* [120] noted that the IST-based methods are limited to the analyses of high-occupancy hydration sites and do not describe significantly the hydration structures in low-density regions. As an alternative to these methods, the authors developed a tool named grid IST (or GIST) [120] and added it to the CPPTRAJ toolset of AmberTools. GIST discretises the integrals of IST onto a 3D grid, which fills the binding pocket region, thus covering both high-density and low-density regions [120]. As a result, unlike IST methods, GIST is able to offer a smoothed map of water structure and the corresponding thermodynamic information for the complete region of interest. For instance, GIST analyses of the molecular host cucurbit[7]uril revealed significant information about the hydration structure and thermodynamic properties of this receptor [124]. The results revealed a toroidal region of high-density hydration sites at the centre of the nonpolar cavity of the host. The results [124] also showed that this specific hydration site, despite having a high density of water molecules, is energetically and entropically unfavourable. The authors related this to the known ability of this receptor to bind external molecules with unusually high affinities [124]. Henceforth, a combination of MD simulations and explicit analyses of solvent dynamics are helpful to advance knowledge of the effects of water molecules in structural biology and drug design [124].

Constant pH molecular dynamics

Ligand–protein complex formation not only leads to conformational changes in the structures of the proteins and/or ligands, but can also impact the pK_a values of their charged side chains. The most common practice in molecular docking and standard molecular dynamics is to assign fixed protonation states for the protein residues, substrates, and ligands, based on prior chemical knowledge. However, it is a known fact that the protonation states of a typical ionisable group involve dynamic processes that can alter the chemical environment during binding. Previous studies [133,134] noted that the p*K*a values of titratable residues can change as a result of several factors. This includes the solvation

of the group upon ligand binding, electrostatic interactions between the ligand and protein, and structural reorganisations within ligand–protein complexes after binding. Thus, the protonation states of ionisable amino acid residues and nonprotein molecules (ligands and substrates) can be subjected to a change during the course of MD simulations. By preserving the protonation states, the MD simulations ignore any binding-induced pK_a changes within the systems. This missing information can limit our complete understanding of the underlying biological processes.

Constant pH molecular dynamics (CpHMD) has been developed for the computational prediction of p*K*a values [135] for ionisable residues in the biological systems under study. The early CpHMD approach used GB solvent as the continuum aqueous environment and Langevin dynamics for the propagation through the nonsolvent (or solute) trajectories [136]. However, this approach has been found to be less accurate for many systems, particularly when water molecules have an active role. Alternately, Donnini *et al.* [137] developed a fully atomistic CpHMD method with a λ-dynamics approach, which can be carried out in explicit solvents. This method allows for the dynamic change of protonation states of titratable groups, thus being able to predict the possible average protonation states at a given pH. This method samples the relevant configurations of the end states of titration groups by considering the protonated states as $\lambda = 0$ and deprotonated states as $\lambda = 1$ [137]. Given the importance of the protonation states of titratable groups in SBDD, it is suggested that a constant pH MD simulation be performed for the ligand–protein complexes before any production MD simulations are initiated. This way, the protonation states of the ionisable groups in the system can be accurately described.

More recently, there have been significant developments in improving the CpHMD [138–141]. For instance, attempts have been made to improve CpHMD using the replica exchange concept (*vide supra*). The basic idea is to perform simulations of different replicas at different pH values. After a set number of steps, the pHs are exchanged between the replicas so as to sample a wider range of protonation states [139]. This approach has been shown to greatly improve the convergence rate and accuracy of CpHMD simulations [140].

Limitations of MD

Classical MD simulations remain a valuable tool in drug design. They are helpful in understanding key molecular motions, energetics, ligand–protein interactions, receptor flexibilities, and conformational changes in the molecular systems, which facilitate the identification of potential candidates with higher affinities to targets. However, it is also important to acknowledge that MD also has some potential limitations and pitfalls, most particularly those concerning time limitations, force-field issues, and quantum effects [53,142].

Time limits and the sampling problem

Currently, typical MD simulations are carried out on systems containing hundreds to millions of atoms, and for several nanoseconds to microseconds. Although these are impressive developments in the field (of MD), it is possible that such time limits might not be sufficient to relax the systems to study certain quantities. For instance, several physical properties of biological systems, such as protein folding, ligand binding, and unbinding processes, mostly occur on long timescales that are normally inaccessible using classical mechanics MD simulations. Furthermore, it is known that biological systems can get trapped in deep energy wells of their potential energy surfaces [143], which can result in sampling insufficient and/or nonrelevant conformations even from long MD trajectories [144]. Improper preparation of the initial structure or insufficient equilibration of the initial structure(s) can impact the quality of the MD results. Therefore, sampling (or) equilibration of an ensemble of structures remains a key issue in MD simulations. Such challenges can be tackled by using alternative strategies. One of the solutions is to apply an enhanced sampling MD approach [46,145], in which an additional bias, such as an external force, is applied to the system to explore the different potential energy surfaces. Although this strategy introduces some artefacts from external bias, it is useful to allow large-scale conformational changes in the systems within the affordable computational cost. Several enhanced sampling approaches have been developed, including metadynamics, replica exchange molecular dynamics (REMD), random acceleration molecular dynamics (RAMD), steered molecular dynamics (SMD), and adaptive bias force steering (ABFS). There are several reviews (e.g., [145–147]) discussing the applications of these methods in SBDD. Alternatively, coarse-grained MD (CG-MD) [148], which reduces the degrees of freedom in large systems by clustering groups of atoms into CG beads, has been developed to deal with large dynamic changes in more complex macromolecules.

Force field issues and quantum effects

The MM force field used in the simulation has vital roles in defining the structural model of the studied system. Force fields are usually developed by combining available experimental data and the results from high-level *ab initio* calculations on small models that form larger systems and, hence, they are fundamentally approximations [53,142]. Furthermore, force fields are parameterised such that they include several atom types describing varied situations of the same atoms (or functional groups). As a result, the transferability of force fields is restricted. Thus, results of MD simulations are reliable only as long as the potential energy functions (or force fields) mimic the forces experienced by the atoms in the real system under study [142].

Classical MD, because of its capabilities to handle large-size systems using affordable computational resources, has gained extraordinary popularity in SBDD. Classical approximations are mostly well suited for nonreactive molecular interactions in biological systems [149,150]. However, they are not able to effectively describe the chemical reactions occurring in biological systems. For example, classical MD might not be able to offer a solution for understanding the reaction mechanisms of drug/substrate–protein complexes, chemical processes of proton transfer within active site, and binding/cleaving processes of certain covalently bonded ligands. In such cases, the use of QM, which explicitly models the electrons in the system, becomes essential at the expense of computational time. To overcome this challenge, reactive force fields have been developed recently that allow chemical reactivity to be treated to some extent [53,149,150]. In reactive force fields, the interatomic potential defines chemical reactions by implementing a bond-order formulation. Within this

scheme, the bond orders in the system are empirically calculated using interatomic distances between atoms during MD simulation, whereas the electronic interactions driving chemical bonding are treated implicitly facilitating the modelling of changes in atom connectivity [149,150]. Recent reviews (e.g., [149,150]) discuss various applications and challenges of such reactive force fields.

Another important challenge faced by classical MD is accounting for electronic polarisation, a significant quantum effect [142]. Within the classical MD framework, each atom in the system is assigned with a pre-set partial charge and is maintained throughout the simulation. Nevertheless, this is not always true, because the biomolecules are in general polarisable, which means that the electron clouds encircling the atoms constantly shift in response to their chemical environment. Thus, it would be effective if the partial changes could be represented as a dynamic parameter, which is not the case with most of the current classical force fields. In response to the importance of this challenge, there have been significant ongoing efforts to develop robust polarisable force fields for MD simulations [151]. Some of the current-generation polarisable force fields include AMOEBA [152,153], CHARMM Drude, and AMBER ff02 [151]. Indeed, polarisable force fields also have their own challenges and should be used with caution. For example, these polarisable force fields are in general slower than nonpolarisable force fields and, as a result, they are more vulnerable to sampling issues. Therefore, polarisable force fields might not be suitable systems where large conformational sampling has important roles. Although having some weaknesses, current polarisable force fields are promising. Given the importance of electrostatic interactions in biological systems, and with more developmental efforts underway, polarisable force fields will soon become an inevitable choice for future classical MD simulations. There are some recent articles that discuss the current status and future directions for polarisable forces and MD simulations [151–153].

Advanced hybrid QM/MM MD

Although there have been significant efforts to fix the issues (concerning chemical reactivity and electronic polarisation) within the classical MD framework, using QM MD, which explicitly models the electrons in the system, has been an alternative practical strategy in biomolecular simulations and SBDD. QM-MD generates dynamical trajectories by using the forces obtained from the electronic structure calculations that are performed at every time step of simulation. Therefore, it able to accurately describe any reactions involving significant electronic effects, such as electron correlation and electron polarisation effects [154,155]. Nevertheless, QM-MD simulations are computationally intensive, which limits the practicality of applying this approach only to smaller sized systems ($\sim 10^2$ atoms) and for limited time scales ($\sim 10^{-12}$ s) [156]. Hence, it was important to find a mid-point that offers both 'the chemical accuracy' of QM-MD and 'feasibility' of MM-MD. To address this problem, Warshel and Levitt [157] introduced a state-of-the-art hybrid MD scheme popularly known as QM/MM. In this approach, a chemically reactive region in the ligand–protein complex (mostly binding site residues and bound ligand) are treated with more accurate QM methods, and the rest of the system is described using MM force fields (Fig. 6). To date, several QM/MM implementations have been developed [158–160] and applied in many studies that focussed on large drug–protein and/or protein–protein systems. For example, in their recent study, Chen *et al.* [161] used QM/MM MD and QM/MM GBSA methods for studying the interactions of benzamide inhibitors with trypsin. In this study, the authors treated the active site residues of the receptor and the inhibitors with QM methods (B3LYP/6-31G(d), PM3, PM6, and RM1) and the rest of the system with the classical ff99SB force field and AMBER program. The authors found that binding free energies calculated with the snapshots obtained from QM/MM MD trajectories displayed excellent agreements with experimental values [161]. In another study [162], QM/MM MD simulations revealed that the fourth ligand coordinating with the active site zinc ion in the Acutolysin A enzyme is a water molecule, rather than a hydroxide anion, correcting a misconception from the low-resolution X-ray crystal structure. It was also revealed by a study that the QM/MM FEP approach outperformed the conventional FEP scheme in

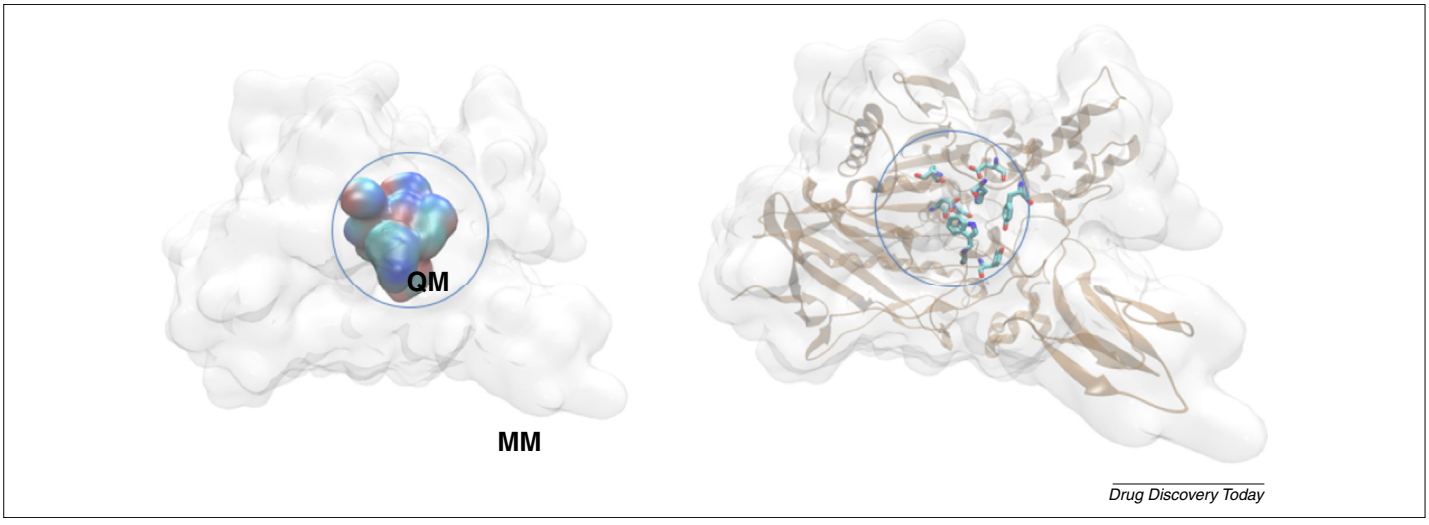

FIGURE 6

A quantum mechanics/molecular mechanics (QM/MM) model of human acetylcholinesterase, where residues that can be treated under QM are shown as ball-and-sticks; the remaining systems shown as surface representations and cartoons can be treated with a MM force field.

predicting accurate binding free energies for a set of fructose 1,6-bisphosphatase inhibitors [93]. Cui and coworkers [163] showed that a hybrid QM/MM-FEP approach could be used to predict accurate p*Ka* values of biological systems. Thus, QM/MM MD simulations have the ability to offer accurate dynamic information that is significant in understanding the structure–function relations of proteins and their interactions with different classes of ligand, the key in drug discovery research.

Nevertheless, it is also important to acknowledge the fact that QM/MM MD simulations also have some clear pitfalls. One of the most important problems in QM/MM simulations is the treatment of the interface region that connects the QM part with that for MM, particularly if they are covalently bonded, as in the case of ligand–protein systems. When a complete system is explicitly cut into QM and MM parts, then it will leave the former region with incomplete valences, which can lead to failed QM treatment [164]. The most common strategy to overcome this issue is to cap the bordering QM residues, which undergo partition, with hydrogen atoms. However, such hydrogen capping introduces atoms into the QM region that differ from those that were originally present in the real system, which can lead to artefacts [164]. Furthermore, QM/MM MD simulations of large protein–ligand systems are still computationally expensive. Hence, they can only be applied to select systems in drug design, such as for those top-ranking hits filtered from thorough virtual screening and classical MD simulations, where follow-up details about key ligand–protein interactions for pharmacophore modelling are computationally justified.

Perspectives on integrating the computational approaches

The past decade has seen tremendous developments in the field of molecular modelling and drug design methods. As discussed above, several modelling and MD-based approaches are available to help modern drug design and discovery efforts. Nevertheless, how we integrate these methods, along with other *in silico* approaches and experiments, is important for increasing our chances of identifying more promising hits from the chemical pool of compounds. Although there are no specific set of rules on how these methods should be combined, extensive knowledge and experience gained over years have provided some logical strategy of implementing them. In Fig. 7, we present a more simplified and practical workflow that assembles classical MD, binding free energy calculations, and QM/MM methods at various stages leading from hit identification to lead optimisation. For instance, the need for classical MD simulations could first arise upon having one (or) more initial 3D structures either from the PDB or through molecular modelling methods. Given that most of these methods are single snapshots of the target, long classical MD simulations (usually a timescale of a few hundred nanoseconds) are required so that large conformational changes in the target can be captured during the simulation. At this stage, the user needs to make several cautious choices, such as the software program, the empirical force field and simulation parameters that are suitable to perform stable MD simulations.

Following this stage, different clustering algorithms, such as RMSD-based clustering or PCA analyses, can be performed to sample the dominant conformations of the target from the MD trajectory. The target conformations obtained from the MD simulations will serve the purpose of addressing the protein flexibility concerns during the subsequent virtual screening procedures. Indeed, there are some computational methods to identify possible cryptic binding sites from these ensembles of target structures and target them during screening. Following the docking and scoring, MD simulations can again be performed on the ligand–protein complexes to refine the complexes and calculate their binding affinities. At this point, the user must make several careful selections, including the length of MD simulations, force field for simulation, and methods for binding-free estimations. Usually, it is suggested that short MD trajectories (~1–2 ns long) are collected for each ligand–target complexes and use them for free energy of binding calculations. The end-point methods, MM-PBSA or MM-GBSA, are mostly popular for these calculations, although other methods, such as FEP, are gaining popularity in the field. Once the high-ranking compounds are identified, they can be experimentally tested using different kinds of assay. Currently, it is hoped that a 5–10% of hit rate (during experimental testing) can be achieved by incorporating rigorous computational modelling and prescreening protocols, although this might not be the case always.

Regardless of whether the results from the experiments are positive or negative, they can be back-fed to the computational protocol to improve it for subsequent phases of screening. For example, if the results are negative (meaning no significant hits were identified), then (i) the lengths of initial MD simulations can be increased to enlarge the sample size for target conformations; (ii) the chemical search space can be increased by expanding the numbers of compounds in the libraries; (iii) the parameters in the docking protocols can be refined; (iv) the length of MD simulations for binding free energy calculations can be increased; and (v) change the method used for free energy estimations. In the event of obtaining good hits from the experiments, then the user might wish to perform extended MD simulations (now for hundreds of ns) to understand the key dynamic interactions between the hits and the targets. Binding free energy methods (or) other enhanced sampling MD methods can also be applied at this stage to gain in-depth knowledge about binding mode(s) of the hits. Based on this information, an effective pharmacophore model or quantitative structure–activity relation (QSAR) model (and/or experimental SAR) can be developed and implemented in subsequent screening protocols. When one or more promising hits, those showing attractive inhibition potentials, promising immunological activity and also nontoxic profiles, are identified, then complexes of such hits can be taken forward for more advanced and computationally expensive QM/MM simulations. At this stage, the user must be cautious in defining the QM and MM segments in the system and also choose a cost-effective (but also accurate) QM model and a suitable MM force field for treating classical segment. The choice of software program is also a key, because using the one that scales well could be helpful to run the QM/MM simulations for large timescales. Such rigorous hybrid simulations can offer extraordinary insights into the reaction mechanisms involved between the selected hit(s) and the target(s). Understanding the reaction mechanisms can be useful towards achieving a better lead compound(s). Those leads showing promising *in vitro* and *in vivo* activities can be taken to further lead optimisation and lengthy and expensive clinical trial stages. Indeed, off-target interactions of drug are

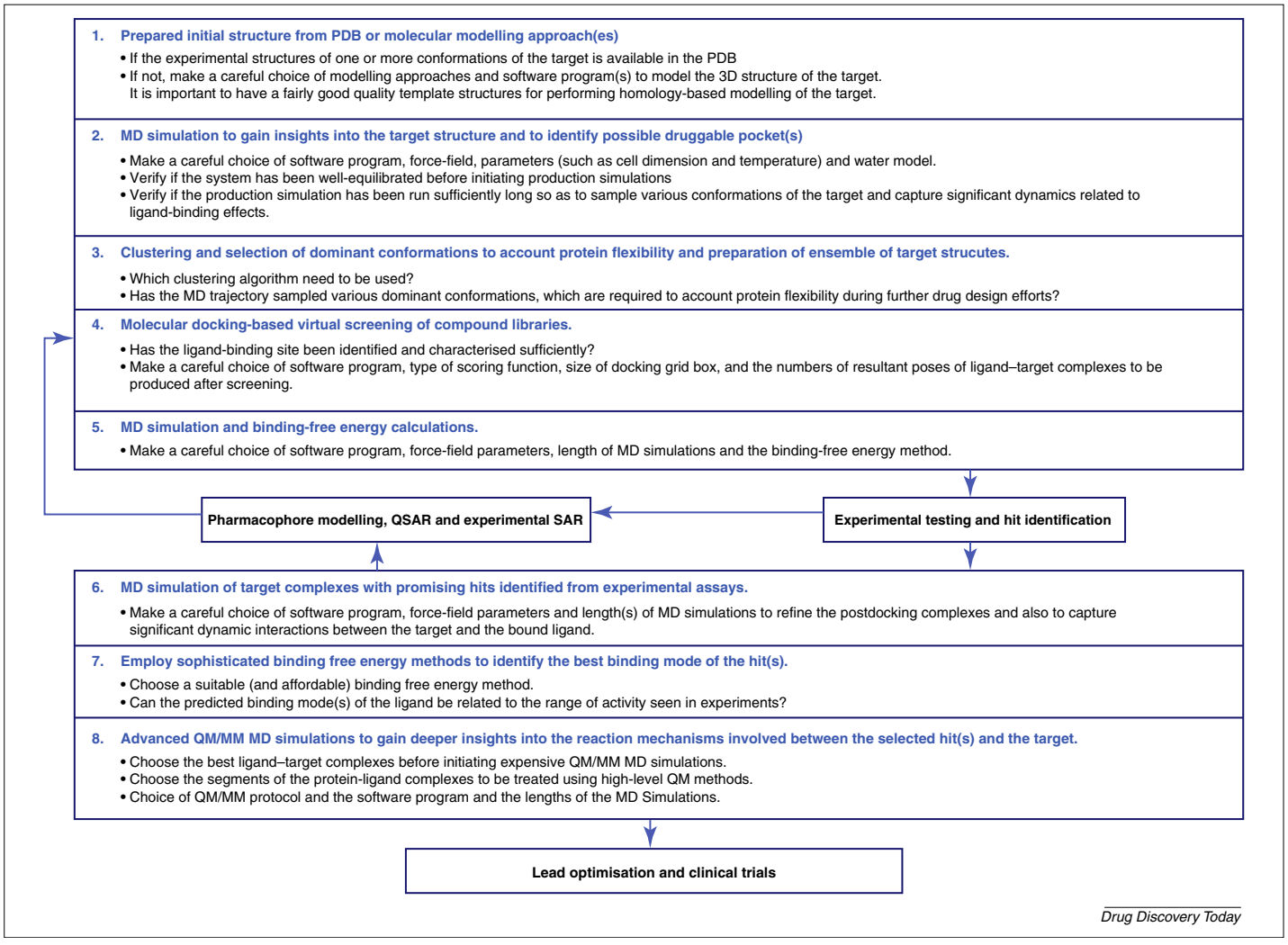

FIGURE 7

A simplified and practical workflow for molecular modelling and drug design. This workflow lists a sequence of steps that provides an overview of how molecular dynamics (MD) approaches can be stacked along with other modelling and experimental procedures during drug design and discovery efforts. In addition, several key decisions that need to be taken during each of the modelling and MD stages are listed. This workflow does not underestimate the role of experiments in drug discovery; rather it highlights the roles of computational approaches, given that an in-depth discussion of the experimental techniques is not within the scope of present review.

yet another important challenge facing the community; and computational methods are also helpful to address this challenge, although this is not discussed here.

The potentials of combining all the computational methods discussed in this review can be best demonstrated, for instance, by a series of studies [86,165–169] carried out on a bacterial enzyme, namely bacterial MurD ligase. A team of scientists from the National Institute of Chemistry, Slovenia, along with their collaborators, carried out several studies on this enzyme, including studying the domain flexibility using MD simulations followed by drug design efforts [165–167], postdocking refinements of the complexes using MD approaches [167,170], understanding the reaction mechanism(s) of the identified hit–enzyme complexes using QM/MM methods [169], and free energy calculations to understand the binding of inhibitors to the MurD ligase and further drive the design processes [86,168]. In one of the preliminary studies [165], the authors performed extensive targeted MD (TMD) simulations to gain some insights into the substrate-binding process and also the structural changes in the enzyme during the transition(s) between the experimentally determined closed and open states [165]. In another study [166], the authors used this information to perform off-path simulation to obtain a relative energy comparison pathway of the two TMD-generated closing pathways. This study also discerned the pathway that had high-energy demands when performing the biochemical processes [166]. The authors claimed that the results from their studies agreed well with experimental findings [166]. Subsequently, the authors selected three MurD ligase conformations from their MD simulations and used them for a two-stage docking-based virtual screening study [167]. The screening identified a panel of promising hits, out of which one (an aminothiazole class inhibitor) was confirmed experimentally to act against dual targets, MurD and MurC. The authors redocked this inhibitor against all the target structures and performed extended classical MD simulations to gain atomistic insights into the ligand–target interactions [167]. The authors also identified another inhibitor class of benzene-1,3-dicarboxylic acid 2,5-dimethylpyrrole

derivatives that showed dual MurD/MurE inhibition properties [170]. In the follow-up study, the authors performed extended MD simulations of this inhibitor–MurD complex to explore their geometrical behaviours. Later, they also performed binding free energy calculations using a linear interaction energy (LIE) method that described the energetic behaviour and binding affinity of the compound [170]. Using the information gathered from these studies, the authors again developed new pharmacophore models and performed a new phase of virtual screening to discover a novel set of compounds that showed promising effects in the experiments [170]. A similar combination of MD and LIE-based binding free energy calculations were also carried out for Furan-based benzene mono- and dicarboxylic acid derivatives against the bacterial Mur ligases [86]. In their ongoing computational and experimental efforts to design drugs for Mur ligases, the authors also performed advanced QM/MM simulations [169], using a B3LYP level of QM theory and CHARMM MM force fields, of the experimental structure of MurD in the PDB (code: 2UAG). This QM/MM study [169] was useful to understand the intermediate tetrahedral formation in the enzyme complex, which was not previously known [169]. Hence, the set of studies by these authors demonstrates how a series of computational studies (along with experiments) can be set up to advance our knowledge about the structural properties of a specific target and make progress towards achieving the goal(s) of drug discovery.

Concluding remarks

It has been 38 years since the first MD simulations of bovine pancreatic trypsin inhibitor were carried out for 9.2 ps. Since then, there has been tremendous growth in supercomputing power and significant developments in the accuracy and efficiency of MD-based computational methods. In addition, MD is now well established as an important contributor to drug design and development. With current capacities, MD simulations can be used for larger biological systems and for microsecond timescales. Such longer classical MD simulations help in effective treatments of the induced-fit effects of the drug binding onto receptors, and can be used to realise optimal drug–receptor binding modes and collect larger conformational samples of the complexes that allow more accurate binding free energy estimations. Alternate versions of classical MD methods, such as CpHMD and enhanced sampling MD approaches, allow chemical changes and other intricate biological events to be traced, which normally occur within ligand–protein complexes but are not observed within the practical limits of classical MD simulations. By contrast, advanced hybrid QM/MM MD methods are useful in revealing the actual reaction mechanisms occurring at the ligand-binding site of the receptor, which are important when designing potent ligands that could trigger effective inhibition of the disease targets. Thus, MD approaches offer range of opportunities and capabilities. Assembling them appropriately with other *in silico* approaches and experiments can enhance the possibilities of identifying more credible hits that can eventually become effective next-generation drugs to serve the human population.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.drudis.2016.11.001.

References

1 Jeffery, T. and David, M.M. (2006) 1918 influenza: the mother of all pandemics. *Emerg. Infect. Dis. J.* 12, 15
2 Clinton, T.R. *et al.* (2015) Design and characterization of ebolavirus GP prehairpin intermediate mimics as drug targets. *Protein Sci.* 24, 446–463
3 Myers, S. and Baker, A. (2001) Drug discovery: an operating model for a new era. *Nat. Biotechnol.* 19, 727–730
4 Feher, M. and Schmidt, J.M. (2003) Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* 43, 218–227
5 Lahana, R. (1999) How many leads from HTS? *Drug Discov. Today* 4, 447–448
6 Newman, D.J. and Cragg, G.M. (2016) Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661
7 Schreiber, S.L. (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* 287, 1964
8 Luo, T. and Schreiber, S.L. (2009) Gold(I)-catalysed coupling reactions for the synthesis of diverse small molecules using the build/couple/pair strategy. *J. Am. Chem. Soc.* 131, 5667–5674
9 Kodadek, T. (2011) The rise, fall and reinvention of combinatorial chemistry. *Chem. Commun.* 47, 9757–9763
10 Dorn, M. *et al.* (2014) Three-dimensional protein structure prediction: methods and computational strategies. *Comput. Biol. Chem. B* 53, 251–276
11 Irwin, J.J. and Shoichet, B.K. (2005) ZINC – a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45, 177–182
12 Sterling, T. and Irwin, J.J. (2015) ZINC 15 – ligand discovery for everyone. *J. Chem. Inf. Model.* 55, 2324–2337
13 Pence, H.E. and Williams, A. (2010) ChemSpider: an online chemical information resource. *J. Chem. Educ.* 87, 1123–1124
14 Gaulton, A. *et al.* (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1100–D1107
15 Wishart, D.S. *et al.* (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 34 (Suppl. 1), D668–D672
16 Bolton, E.E. *et al.* (2008) PubChem: integrated platform of small molecules and biological activities. *Annu. Rep. Comput. Chem.* 4, 217–241
17 Glaab, E. (2016) Building a virtual ligand screening pipeline using free software: a survey. *Brief. Bioinform.* 17, 352–366
18 Alonso, H. *et al.* (2006) Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* 26, 531–568
19 Durrant, J.D. and McCammon, J.A. (2011) Molecular dynamics simulations and drug discovery. *BMC Biol.* 9, 1–9
20 Kerrigan, J. (2013) Molecular dynamics simulations in drug design. *In Silico Models Drug Discov.* 993, 95–113
21 Zhang, J.-L. *et al.* (2013) Drug design benefits from molecular dynamics: some examples. *Curr. Comput. Aided Drug Des.* 9, 532–546
22 Nair, P. and Miners, J. (2014) Molecular dynamics simulations: from structure function relationships to drug discovery. *In Silico Pharmacol.* 2, 1–4
23 Mortier, J. *et al.* (2015) The impact of molecular dynamics on drug design: applications for the characterization of ligand–macromolecule complexes. *Drug Discov. Today* 20, 686–702
24 Tautermann, C.S. *et al.* (2015) What can we learn from molecular dynamics simulations for GPCR drug design? *Comput. Struct. Biotechnol. J.* 13, 111–121
25 Zhao, H. and Caflisch, A. (2015) Molecular dynamics in drug design. *Eur. J. Med. Chem.* 91, 4–14
26 Borhani, D.W. and Shaw, D.E. (2011) The future of molecular dynamics simulations in drug discovery. *J. Comput. Aided Mol. Des.* 26, 15–26
27 Huang, S.-Y. *et al.* (2010) Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* 12, 12899–12908
28 Lang, P.T. *et al.* (2009) DOCK 6: combining techniques to model RNA–small molecule complexes. *RNA* 15, 1219–1230
29 Morris, G.M. *et al.* (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791
30 Jones, G. *et al.* (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267, 727–748
31 Friesner, R.A. *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749

32 Kastritis, P.L. and Bonvin, A.M.J.J. (2012) On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *J. R. Soc. Interface* 10, 20120835
33 Koshland, D.E. (1995) The key-lock theory and the induced fit theory. *Angew. Chem. Int. Ed. Engl.* 33, 2375–2378
34 Huang, S.-Y. and Zou, X. (2010) Advances and challenges in protein–ligand docking. *Int. J. Mol. Sci.* 11, 3016
35 Böhm, H.-J. (1992) The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* 6, 61–78
36 Ferrari, A.M. *et al.* (2004) Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* 47, 5076–5084
37 Knegtel, R.M.A. *et al.* (1997) Molecular docking to ensembles of protein structures. *J. Mol. Biol.* 266, 424–440
38 Clauben, H. *et al.* (2001) FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* 308, 377–395
39 Fischer, M. *et al.* (2014) Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat. Chem.* 6, 575–583
40 Anwar-Mohamed, A. *et al.* (2014) A human ether-á-go-go-related (hERG) ion channel atomistic model generated by long supercomputer molecular dynamics simulations and its use in predicting drug cardiotoxicity. *Toxicol. Lett.* 230, 382–392
41 Barakat, K. *et al.* (2010) Ensemble-based virtual screening reveals dual-inhibitors for the p53–MDM2/MDMX interactions. *J. Mol. Graphics Model.* 28, 555–568
42 Barakat, K. and Tuszynski, J. (2011) Relaxed complex scheme suggests novel inhibitors for the lyase activity of DNA polymerase beta. *J. Mol. Graphics Model.* 29, 702–716
43 Barakat, K.H. *et al.* (2015) A refined model of the HCV NS5A protein bound to daclatasvir explains drug-resistant mutations and activity against divergent genotypes. *J. Chem. Inf. Model.* 55, 362–373
44 Campbell, A.J. *et al.* (2014) Ensemble-based docking using biased molecular dynamics. *J. Chem. Inf. Model.* 54, 2127–2138
45 Ge, H. *et al.* (2013) Molecular dynamics-based virtual screening: accelerating the drug discovery process by high-performance computing. *J. Chem. Inf. Model.* 53, 2757–2764
46 Kalyaanamoorthy, S. and Chen, Y.-P.P. (2014) Modelling and enhanced molecular dynamics to steer structure-based drug discovery. *Prog. Biophys. Mol. Biol.* 114, 123–136
47 Mackerell, A.D. (2004) Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* 25, 1584–1604
48 Jones, J.E. (1924) On the determination of molecular fields. II. From the equation of state of a gas. *Proc. R. Soc. Lond. A* 106, 463–477
49 Darden, T. *et al.* (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092
50 MacKerell, A.D., Jr *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616
51 Cornell, W.D. *et al.* (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117, 5179–5197
52 Daura, X. *et al.* (1998) Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J. Comput. Chem.* 19, 535–547
53 González, M.A. (2011) Force fields and molecular dynamics simulations. *JDN* 12, 169–200
54 Todorova, N. *et al.* (2008) Systematic comparison of empirical forcefields for molecular dynamic simulation of insulin. *J. Phys. Chem. B* 112, 11137–11146
55 Nguyen, T.T. *et al.* (2014) Effects of water models on binding affinity: evidence from all-atom simulation of binding of tamiflu to A/H5N1 neuraminidase. *Sci. World J.* 2014, 536084
56 Becker, C.A. *et al.* (2013) Considerations for choosing and using force fields and interatomic potentials in materials science and engineering. *Curr. Opin. Solid State Mater. Sci.* 17, 277–283
57 Verlet, L. (1967) Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* 159, 98–103
58 Swope, W.C. *et al.* (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: application to small water clusters. *J. Chem. Phys.* 76, 637–649
59 Cuendet, M.A. and van Gunsteren, W.F. (2007) On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm. *J. Chem. Phys.* 127, 184102
60 Ryckaert, J.-P. *et al.* (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341
61 Andersen, H.C. (1983) Rattle: a 'velocity' version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* 52, 24–34
62 Hess, B. *et al.* (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472
63 McQuarrie, D.A. (2000) *Statistical Mechanics*. University Science Books
64 Beard, D.A. and Qian, H. (2008) *Chemical Biophysics: Quantitative Analysis of Cellular Systems*. Cambridge University Press
65 Richardson, J.M. and Brinkley, S.R., Jr (1960) Cell method in grand canonical ensemble. *J. Chem. Phys.* 33, 1467–1478
66 Corti, D.S. and Soto-Campos, G. (1998) Deriving the isothermal-isobaric ensemble: the requirement of a 'shell' molecule and applicability to small systems. *J. Chem. Phys.* 108, 7959–7966
67 Adelman, S.A. and Doll, J.D. (1976) Generalized Langevin equation approach for atom/solid-surface scattering: general formulation for classical scattering off harmonic solids. *J. Chem. Phys.* 64, 2375–2388
68 Berendsen, H.J.C. *et al.* (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690
69 Nosé, S. (1984) A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 81, 511–519
70 Hoover, W.G. (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* 31, 1695–1697
71 Deschamps, N. *et al.* (2015) How the flexibility of human histone deacetylases influences ligand binding: an overview. *Drug Discov. Today* 20, 736–742
72 Ma, B. *et al.* (1999) Folding funnels and binding mechanisms. *Protein Eng.* 12, 713–720
73 Kalyaanamoorthy, S. and Chen, Y.-P.P. (2014) A steered molecular dynamics mediated hit discovery for histone deacetylases. *Phys. Chem. Chem. Phys.* 16, 3777–3791
74 Estiu, G. *et al.* (2010) On the inhibition of histone deacetylase 8. *Bioorg. Med. Chem.* 18, 4103–4110
75 Horn, J.R. and Shoichet, B.K. (2004) Allosteric inhibition through core disruption. *J. Mol. Biol.* 336, 1283–1291
76 Diskin, R. *et al.* (2008) A novel lipid binding site formed by the MAP kinase insert in p38α. *J. Mol. Biol.* 375, 70–79
77 Schames, J.R. *et al.* (2004) Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* 47, 1879–1881
78 Hazuda, D.J. *et al.* (2004) A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11233–11238
79 David, C. and Jacobs, D. (2014) Principal component analysis: a method for determining the essential dynamics of proteins. *Protein Dyn.* 1084, 193–226
80 Bowman, A.L. *et al.* (2007) Small molecule inhibitors of the MDM2–p53 interaction discovered by ensemble-based receptor models. *J. Am. Chem. Soc.* 129, 12809–12814
81 Rakers, C. *et al.* (2015) Computational close up on protein–protein interactions: how to unravel the invisible using molecular dynamics simulations? *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 5, 345–359
82 Sousa, S.F. *et al.* (2011) Detailed atomistic analysis of the HIV-1 protease interface. *J. Phys. Chem. B* 115, 7045–7057
83 Viricel, C. *et al.* (2015) Human PD-1 binds differently to its human ligands: a comprehensive modeling study. *J. Mol. Graphics Model.* 57, 131–142
84 Rastelli, G. *et al.* (2002) Discovery of new inhibitors of aldose reductase from molecular docking and database screening. *Bioorg. Med. Chem.* 10, 1437–1450
85 Cavalli, A. *et al.* (2004) A computational study of the binding of propidium to the peripheral anionic site of human acetylcholinesterase. *J. Med. Chem.* 47, 3991–3999
86 Perdih, A. *et al.* (2015) Furan-based benzene mono- and dicarboxylic acid derivatives as multiple inhibitors of the bacterial Mur ligases (MurC–MurF): experimental and computational characterization. *J. Comput. Aided Mol. Des.* 29, 541–560
87 Distinto, S. *et al.* (2012) Synthesis and biological assessment of novel 2-thiazolylhydrazones and computational analysis of their recognition by monoamine oxidase B. *Eur. J. Med. Chem.* 48, 284–295
88 Fu, G. *et al.* (2012) Molecular modeling to provide insight into the substrate binding and catalytic mechanism of human biliverdin-IXα reductase. *J. Phys. Chem. B* 116, 9580–9594
89 Vijjulatha, M. *et al.* (2014) Induced fit docking, pharmacophore modeling, and molecular dynamic simulations on thiazolidinedione derivatives to explore key interactions with Tyr48 in polyol pathway. *J. Mol. Model.* 20, 1–14
90 Huang, Y.M. *et al.* (2012) Insights from free-energy calculations: protein conformational equilibrium, driving forces, and ligand-binding modes. *Biophys. J.* 103, 342–351
91 Michel, J. *et al.* (2010) Rigorous free energy calculations in structure-based drug design. *Mol. Inform.* 29, 570–578

92 Christ, C.D. and Fox, T. (2014) Accuracy assessment and automation of free energy calculations for drug design. *J. Chem. Inf. Model.* 54, 108–120
93 Rathore, R.S. *et al.* (2013) Advances in binding free energies calculations: QM/MM-based free energy perturbation method for drug design. *Curr. Pharm. Des.* 19, 4674–4686
94 Wereszczynski, J. and McCammon, J.A. (2012) Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. *Q. Rev. Biophys.* 45, 1–25
95 Camps, P. *et al.* (2000) New tacrine–huperzine A hybrids (Huprines): highly potent tight-binding acetylcholinesterase inhibitors of interest for the treatment of Alzheimer's disease. *J. Med. Chem.* 43, 4657–4666
96 Rami Reddy, M. *et al.* (2000) Free energy calculations: use and limitations in predicting ligand binding affinities. *Rev. Comput. Chem.* 16, 217–304
97 Jorgensen, W.L. *et al.* (2006) Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* 16, 663–667
98 Kim, J.T. *et al.* (2006) FEP-guided selection of bicyclic heterocycles in lead optimization for non-nucleoside inhibitors of HIV-1 reverse transcriptase. *J. Am. Chem. Soc.* 128, 15372–15373
99 Michel, J. *et al.* (2009) In silico improvement of β3-peptide inhibitors of p53·hDM2 and p53·hDMX. *J. Am. Chem. Soc.* 131, 6356–6357
100 Wang, L. *et al.* (2015) Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* 137, 2695–2703
101 Åqvist, J. *et al.* (1994) A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* 7, 385–391
102 Nicolotti, O. *et al.* (2010) Screening of benzamidine-based thrombin inhibitors via a linear interaction energy in continuum electrostatics model. *J. Comput. Aided Mol. Des.* 24, 117–129
103 Timmers, L. *et al.* (2012) Combining molecular dynamics and docking simulations of the cytidine deaminase from *Mycobacterium tuberculosis* H37Rv. *J. Mol. Model.* 18, 467–479
104 Boukharta, L. *et al.* (2011) Computer simulations of structure–activity relationships for hERG channel blockers. *Biochemistry* 50, 6146–6156
105 Gutiérrez-de-Terán, H. *et al.* (2006) Computational analysis of plasmepsin IV bound to an allophenylnorstatine inhibitor. *FEBS Lett.* 580, 5910–5916
106 Gutiérrez-de-Terán, H. and Åqvist, J. (2012) Linear interaction energy: method and applications in drug design. *Comput. Drug Discov. Des.* 819, 305–323
107 Su, P.-C. *et al.* (2015) Comparison of radii sets, entropy, QM methods, and sampling on MM-PBSA, MM-GBSA, and QM/MM-GBSA ligand binding energies of *F. tularensis* enoyl-ACP reductase (FabI). *J. Comput. Chem.* 36, 1859–1873
108 Genheden, S. and Ryde, U. (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* 10, 449–461
109 Oehme, D.P. *et al.* (2012) Effect of atomic charge, solvation, entropy, and ligand protonation state on MM-PB(GB)SA binding energies of HIV protease. *J. Comput. Chem.* 33, 2566–2580
110 Hou, T. *et al.* (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* 51, 69–82
111 Sun, H. *et al.* (2014) Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys. Chem. Chem. Phys.* 16, 22035–22045
112 Hou, X. *et al.* (2013) How to improve docking accuracy of AutoDock4.2: a case study using different electrostatic potentials. *J. Chem. Inf. Model.* 53, 188–200
113 Genheden, S. and Ryde, U. (2010) How to obtain statistically converged MM/GBSA results. *J. Comput. Chem.* 31, 837–846
114 Jordheim, L.P. *et al.* (2013) Small molecule inhibitors of ERCC1-XPF protein–protein interaction synergize alkylating agents in cancer cells. *Mol. Pharmacol.* 84, 12–24
115 Kalyaanamoorthy, S. and Chen, Y.-P.P. (2013) Quantum polarized ligand docking investigation to understand the significance of protonation states in histone deacetylase inhibitors. *J. Mol. Graphics Model.* 44, 44–53
116 Velez-Vega, C. *et al.* (2015) Estimation of solvation entropy and enthalpy via analysis of water oxygen–hydrogen correlations. *J. Chem. Theory Comput.* 11, 5090–5102
117 Wong, S.E. and Lightstone, F.C. (2011) Accounting for water molecules in drug design. *Expert Opin. Drug Discov.* 6, 65–74
118 Riniker, S. *et al.* (2012) Free enthalpies of replacing water molecules in protein binding pockets. *J. Comput. Aided Mol. Des.* 26, 1293–1309
119 Li, Z. and Lazaridis, T. (2006) Thermodynamics of buried water clusters at a protein–ligand binding interface. *J. Phys. Chem. B* 110, 1464–1475
120 Ramsey, S. *et al.* (2016) Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *J. Comput. Chem.* 37, 2029–2037
121 Young, T. *et al.* (2007) Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proc. Natl. Acad. Sci. U. S. A.* 104, 808–813
122 Baroni, M. *et al.* (2007) A common reference framework for analysing/comparing proteins and ligands. Fingerprints for ligands and proteins (FLAP): theory and application. *J. Chem. Inf. Model.* 47, 279–294
123 Bayden, A.S. *et al.* (2015) Evaluating free energies of binding and conservation of crystallographic waters using SZMAP. *J. Chem. Inf. Model.* 55, 1552–1565
124 Nguyen, C.N. *et al.* (2012) Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit [7]uril. *J. Chem. Phys.* 137, 044101
125 Velez-Vega, C. *et al.* (2014) Time-averaged distributions of solute and solvent motions: exploring proton wires of GFP and PfM2DH. *J. Chem. Inf. Model.* 54, 3344–3361
126 Cui, G. *et al.* (2013) SPAM: a simple approach for profiling bound water molecules. *J. Chem. Theory Comput.* 9, 5539–5549
127 Li, Z. and Lazaridis, T. (2012) Computing the thermodynamic contributions of interfacial water. In *Computational Drug Discovery and Design* (Baron, R., ed.), pp. 393–404, Springer
128 Beuming, T. *et al.* (2009) High-energy water sites determine peptide binding affinity and specificity of PDZ domains. *Protein Sci.* 18, 1609–1619
129 Beuming, T. *et al.* (2012) Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. *Proteins* 80, 871–883
130 Abel, R. *et al.* (2011) Contribution of explicit solvent effects to the binding affinity of small-molecule inhibitors in blood coagulation factor serine proteases. *ChemMedChem* 6, 1049–1066
131 Guimarães, C.R.W. and Mathiowetz, A.M. (2010) Addressing limitations with the MM-GB/SA scoring procedure using the WaterMap method and free energy perturbation calculations. *J. Chem. Inf. Model.* 50, 547–559
132 Higgs, C. *et al.* (2010) Hydration site thermodynamics explain SARs for triazolylpurines analogues binding to the A2A receptor. *ACS Med. Chem. Lett.* 1, 160–164
133 Kim, M.O. and McCammon, J.A. (2016) Computation of pH-dependent binding free energies. *Biopolymers* 105, 43–49
134 Onufriev, A.V. and Alexov, E. (2013) Protonation and pK changes in protein–ligand binding. *Q. Rev. Biophys.* 46, 181–209
135 Wallace, J.A. and Shen, J.K. (2012) Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH. *J. Chem. Phys.* 137, 184105
136 Adcock, S.A. and McCammon, J.A. (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* 106, 1589–1615
137 Donnini, S. *et al.* (2011) Constant pH molecular dynamics in explicit solvent with λ-dynamics. *J. Chem. Theory Comput.* 7, 1962–1978
138 Lee, J. *et al.* (2016) Computational scheme for pH-dependent binding free energy calculation with explicit solvent. *Protein Sci.* 25, 231–243
139 Wallace, J.A. and Shen, J.K. (2011) Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange. *J. Chem. Theory Comput.* 7, 2617–2629
140 Swails, J.M. and Roitberg, A.E. (2012) Enhancing conformation and protonation state sampling of hen egg white lysozyme using pH replica exchange molecular dynamics. *J. Chem. Theory Comput.* 8, 4393–4404
141 Lee, J. *et al.* (2014) Constant pH molecular dynamics in explicit solvent with enveloping distribution sampling and Hamiltonian exchange. *J. Chem. Theory Comput.* 10, 2738–2750
142 Meller, J.A. (2001) *Molecular Dynamics*. John Wiley & Sons
143 Onuchic, J.N. *et al.* (1997) Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48, 545–600
144 Marsili, S. *et al.* (2010) ORAC: a molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level. *J. Comput. Chem.* 31, 1106–1116
145 Bernardi, R.C. *et al.* (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta* 1850, 872–877
146 De Vivo, M. *et al.* (2016) Role of molecular dynamics and related methods in drug discovery. *J. Med. Chem.* 59, 4035–4061
147 Rodriguez-Bussey, I.G. *et al.* (2016) Enhanced molecular dynamics sampling of drug target conformations. *Biopolymers* 105, 35–42
148 Saunders, M.G. and Voth, G.A. (2013) Coarse-graining methods for computational biology. *Annu. Rev. Biophys.* 42, 73–93
149 Senftle, T.P. *et al.* (2016) The ReaxFF reactive force-field: development, applications and future directions. *NPJ Comput. Mater.* 2, 15011
150 Han, Y. *et al.* (2016) Development, applications and challenges of ReaxFF reactive force field in molecular simulations. *Front. Chem. Sci. Eng.* 10, 16–38

151 Baker, C.M. (2015) Polarizable force fields for molecular dynamics simulations of biomolecules. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 5, 241–254
152 Ren, P. and Ponder, J.W. (2003) Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* 107, 5933–5947
153 Ponder, J.W. *et al.* (2010) Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* 114, 2549–2564
154 Tuckerman, M.E. (2002) Ab initio molecular dynamics: basic concepts, current trends and novel applications. *J. Phys. Condens. Matter* 14, R1297–R1355
155 Car, R. and Parrinello, M. (1985) Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* 55, 2471–2474
156 Lorenz, C. and Doltsinis, N.L. (2012) Molecular dynamics simulation: from 'ab initio' to 'coarse grained'. In *Handbook of Computational Chemistry* (Leszczynski, J., ed.), pp. 195–238, Springer
157 Warshel, A. and Levitt, M. (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103, 227–249
158 Laio, A. *et al.* (2002) A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. *J. Chem. Phys.* 116, 6941–6947
159 Svensson, M. *et al.* (1996) ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels–Alder reactions and $Pt(P(t\text{-}Bu)3)_2 + H_2$ oxidative addition. *J. Phys. Chem.* 100, 19357–19363
160 Vreven, T. *et al.* (2006) Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J. Chem. Theory Comput.* 2, 815–826
161 Chen, J. *et al.* (2015) A comparative study of trypsin specificity based on QM/MM molecular dynamics simulation and QM/MM GBSA calculation. *J. Biomol. Struct. Dyn.* 1–13
162 Wu, E.L. *et al.* (2009) Determination of the structure form of the fourth ligand of zinc in acutolysin A using combined quantum mechanical and molecular mechanical simulation. *J. Phys. Chem. B* 113, 2477–2485
163 Riccardi, D. *et al.* (2005) p*K*a calculations in solution and proteins with QM/MM free energy perturbation simulations: a quantitative test of QM/MM protocols. *J. Phys. Chem. B* 109, 17715–17733
164 Grotendorst, J. and Simulation, I.F.A. (2012) *Hierarchical Methods for Dynamics in Complex Molecular Systems*. Forschungszentrum
165 Perdih, A. *et al.* (2007) Targeted molecular dynamics simulation studies of binding and conformational changes in *E. coli* MurD. *Proteins* 68, 243–254
166 Perdih, A. and Solmajer, T. (2012) MurD ligase from *Escherichia coli*: C-terminal domain closing motion. *Comput. Theor. Chem.* 979, 73–81
167 Perdih, A. *et al.* (2014) Inhibitor design strategy based on an enzyme structural flexibility: a case of bacterial MurD ligase. *J. Chem. Inf. Model.* 54, 1451–1466
168 Perdih, A. *et al.* (2013) Molecular dynamics simulation and linear interaction energy study of d-Glu-based inhibitors of the MurD ligase. *J. Comput. Aided Mol. Des.* 27, 723–738
169 Perdih, A. *et al.* (2009) MurD ligase from *E. coli*: tetrahedral intermediate formation study by hybrid quantum mechanical/molecular mechanical replica path method. *Proteins* 74, 744–759
170 Perdih, A. *et al.* (2014) Benzene-1,3-dicarboxylic acid 2,5-dimethylpyrrole derivatives as multiple inhibitors of bacterial Mur ligases (MurC–MurF). *Bioorg. Med. Chem.* 22, 4124–4134
171 Tarus, B. *et al.* (2012) Molecular dynamics studies of the nucleoprotein of influenza A virus: role of the protein flexibility in RNA binding. *PLoS ONE* 7, e30038
172 Mirjalili, V. *et al.* (2014) Physics-based protein structure refinement through multiple molecular dynamics trajectories and structure averaging. *Proteins* 82, 196–207
173 Vargiu, A.V. and Nikaido, H. (2012) Multidrug binding properties of the AcrB efflux pump characterized by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20637–20642
174 Feig, M. (2016) Local protein structure refinement via molecular dynamics simulations with locPREFMD. *J. Chem. Inf. Model.* 56, 1304–1312
175 Kruse, A.C. *et al.* (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* 482, 552–556
176 Römer, R.A. *et al.* (2016) The flexibility and dynamics of protein disulphide-isomerase. *Proteins* 84, 1776–1785
177 Mirjalili, V. and Feig, M. (2013) Protein structure refinement through structure selection and averaging from molecular dynamics ensembles. *J. Chem. Theory Comput.* 9, 1294–1303
178 Preininger, A.M. *et al.* (2013) Conformational flexibility and structural dynamics in GPCR-mediated G protein activation: a perspective. *J. Mol. Biol.* 425, 2288–2298
179 Sun, H. *et al.* (2014) Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys. Chem. Chem. Phys.* 16, 16719–16729
180 Sgobba, M. *et al.* (2012) Application of a post-docking procedure based on MM-PBSA and MM-GBSA on single and multiple protein conformations. *Eur. J. Med. Chem.* 58, 431–440
181 Fraccalvieri, D. *et al.* (2011) Conformational and functional analysis of molecular dynamics trajectories by Self-Organising Maps. *BMC Bioinform.* 12, 1–19
182 Pan, D. *et al.* (2016) Computational study on the drug resistance mechanism of hepatitis C virus NS5B RNA-dependent RNA polymerase mutants to BMS-791325 by molecular dynamics simulation and binding free energy calculations. *Chemom. Intell. Lab. Syst.* 154, 185–193
183 Sahoo, B.R. *et al.* (2014) A conformational analysis of mouse Nalp3 domain structures by molecular dynamics simulations, and binding site analysis. *Mol. BioSyst.* 10, 1104–1116
184 Fang, J. *et al.* (2014) Inhibition of acetylcholinesterase by two genistein derivatives: kinetic analysis, molecular docking and molecular dynamics simulation. *Acta Pharm. Sin. B* 4, 430–437
185 Lu, H. *et al.* (2014) Binding free energies for nicotine analogs inhibiting cytochrome P450 2A6 by a combined use of molecular dynamics simulations and QM/MM-PBSA calculations. *Bioorg. Med. Chem.* 22, 2149–2156
186 Kumar, S. *et al.* (2013) Understanding the specificity of human galectin-8C domain interactions with its glycan ligands based on molecular dynamics simulations. *PLoS ONE* 8, e59761
187 Niu, R.-J. *et al.* (2013) Molecular dynamics simulations studies and free energy analysis on inhibitors of MDM2-p53 interaction. *J. Mol. Graphics Model.* 46, 132–139
188 Shahlaei, M. and Mousavi, A. (2015) A conformational analysis study on the melanocortin 4 receptor using multiple molecular dynamics simulations. *Chem. Biol. Drug Des.* 86, 309–321
189 Proctor, E.A. *et al.* (2012) Discrete molecular dynamics distinguishes nativelike binding poses from decoys in difficult targets. *Biophys. J.* 102, 144–151
190 Mukhametov, A. *et al.* (2014) Allosteric pocket of the dengue virus (serotype 2) NS2B/NS3 protease: in silico ligand screening and molecular dynamics studies of inhibition. *J. Mol. Graphics Model.* 52, 103–113
191 Kalyaanamoorthy, S. and Chen, Y.-P.P. (2013) Ligand release mechanisms and channels in histone deacetylases. *J. Comput. Chem.* 34, 2270–2283
192 Kumar, K.M. *et al.* (2014) Molecular docking and molecular dynamics studies on [small beta]-lactamases and penicillin binding proteins. *Mol. Biosyst.* 10, 891–900
193 Poongavanam, V. *et al.* (2014) Binding free energy based structural dynamics analysis of HIV-1 RT RNase H-inhibitor complexes. *Integr. Biol.* 6, 1010–1022
194 Wen, P.-C. *et al.* (2013) On the origin of large flexibility of P-glycoprotein in the inward-facing state. *J. Biol. Chem.* 288, 19211–19220
195 Hu, X. *et al.* (2014) Structural insight into exosite binding and discovery of novel exosite inhibitors of botulinum neurotoxin serotype A through in silico screening. *J. Comput. Aided Mol. Des.* 28, 765–778
196 Wang, K. *et al.* (2015) Hierarchical conformational analysis of native lysozyme based on sub-millisecond molecular dynamics simulations. *PLOS ONE* 10, e0129846
197 Hayes, J.M. *et al.* (2011) Kinetics, in silico docking, molecular dynamics, and MM-GBSA binding studies on prototype indirubins, KT5720, and staurosporine as phosphorylase kinase ATP-binding site inhibitors: the role of water molecules examined. *Proteins* 79, 703–719
198 Shao, S. *et al.* (2014) Dual-inhibitors of STAT5 and STAT3: studies from molecular docking and molecular dynamics simulations. *J. Mol. Model.* 20, 1–17
199 Li, C. *et al.* (2014) Molecular dynamics simulation on the conformational transition of the Mad2 protein from the open to the closed state. *Int. J. Mol. Sci.* 15, 5553–5569
200 Yadava, U. *et al.* (2015) Stabilization of microtubules by taxane diterpenoids: insight from docking and MD simulations. *J. Biol. Phys.* 41, 117–133
201 Begum, J. *et al.* (2015) An evaluation of indirubin analogues as phosphorylase kinase inhibitors. *J. Mol. Graphics Model.* 61, 231–242
202 Cau, Y. *et al.* (2015) Molecular dynamics simulations and structural analysis of *Giardia duodenalis* 14-3-3 protein–protein interactions. *J. Chem. Inf. Model.* 55, 2611–2622
203 Wang, Q. *et al.* (2016) Exploring the mechanism how AF9 recognizes and binds H3K9ac by molecular dynamics simulations and free energy calculations. *Biopolymers* 105, 779–786
204 Anwar Anwar-Mohamed, A. *et al.* (2014) A human ether-á-go-go-related (hERG) ion channel atomistic model generated by long supercomputer molecular dynamics simulations and its use in predicting drug cardiotoxicity. *Toxicol. Lett.* 230, 382–392
205 Barakat, K.H. *et al.* (2014) A refined model of the HCV NS5A protein bound to daclatasvir explains drug-resistant mutations and activity against divergent genotypes. *J. Chem. Inf. Model.* 55, 362–373