2

Molecular Dynamics–Based Approaches Describing Protein Binding

Andrea Spitaleri and Walter Rocchia

CONCEPT Lab, Istituto Italiano di Tecnologia, via Morego, 30, 16163 Genoa, Italy

2.1 Introduction

Biomolecular recognition is at the heart of all biological processes that take place in living organisms. Understanding how a ligand binds to a biological receptor, how proteins interact with each other, how lipids and proteins aggregate in the cell membrane, and how these events trigger or block a wide range of biochemical reactions is of paramount importance, not only for the field of biophysics but also for other disciplines such as rational drug design.

The first breakthrough in the theory of biomolecular recognition was provided by Fischer, who as early as 1894, proposed his popular lock-and-key model for ligand binding [1]. According to this model, which was originally developed in the context of enzyme catalysis, ligands involved in biological reactions fit perfectly into their targets like a key into a lock. In this process, no changes in conformations are allowed. Later, however, the lock-and-key model based on rigid body interactions between ligands and proteins was severely challenged by the conformational plasticity of macromolecules, as revealed by X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and single-molecule fluorescence detection [2].

In recent years, more dynamic models of molecular recognition have superseded Emil Fischer's rigid lock-and-key binding paradigm. The idea of induced fit, introduced by Koshland in 1958, aimed at overcoming the limitations of the lock-and-key model, relies on the formation of an initial loose ligand–receptor complex that induces a conformational change in the protein, resulting in a series of rearrangements that lead to a complex with tighter binding.

In 1999, Nussinov and coworkers coined the term conformational selection, also known as population shift, selected fit, and population selection, which is based on the idea that all conformations are present when the ligand is not bound to the receptor and, then, the ligand acts to selectively stabilize specific receptor conformations, causing a shift in the populations observed in the unbound ensemble toward this specific conformational state [2, 3].

Although both theories, induced-fit and conformational selection, appear to be antagonistic, they are not necessarily mutually exclusive. Nowadays it is

Biomolecular Simulations in Structure-Based Drug Discovery,
First Edition. Edited by Francesco L. Gervasio and Vojtech Spiwok.

increasingly seen that these two major mechanisms contribute to molecular recognition, which requires a dynamic description over timescales spanning several orders of magnitude to be fully understood. Moreover, water molecules often play a crucial role in molecular association. In fact, experimental and theoretical studies have pointed out the capital importance of both entropic and enthalpic contributions of water networks to the free energy of binding [4]. All these findings not only served to advance the field toward a better understanding of receptor–ligand (e.g. protein, peptide, nuclei acid, and ligand) binding but also introduced an extra degree of complexity to the description of biomolecular recognition processes. To comprehend how biomolecular recognition occurs, we first need to understand the role of all different partners involved in this association process.

2.1.1 Protein Binding: Molecular Dynamics Versus Docking

Molecular docking programs are widely used to predict receptor–ligand complexes and are extensively and routinely used in computer-aided drug discovery, mostly in the framework of virtual screening (VS) campaigns [5]. Molecular docking, in practice, has two essential requirements: structural data, for candidate ligands and the protein target of interest, and a procedure to estimate protein–ligand interaction poses and strengths. Various molecular docking algorithms are available to predict receptor–ligand, e.g. protein, peptide, nucleic acid, and ligand, poses and to rank them based on scoring functions implemented in each specific docking approach.

In spite of the large usage in structure-based drug discovery (SBDD), docking retains crucial liabilities. The high computational speed in determining the putative poses comes at the cost of accuracy, especially when target rearrangements are required upon ligand binding. Sampling the conformational space remains challenging. All docking algorithms exhibit sensitivity to initial conditions; therefore, some variation in the poses and scores is expected when starting from different input conformations of a particular ligand. The vast majority of molecular docking applications considers the ligand conformational flexibility either during docking calculations or using libraries of conformers. Conversely, the flexibility of the receptor is usually neglected as the number of degrees of freedom, which should be considered in the calculations, is extremely computationally demanding. Moreover, docking algorithms lack explicit water treatments, which are crucial for reproducing specific drug–target complexes. All these approximations constitute a major drawback in docking calculations, limiting the success rate of the drug discovery campaign.

These limitations can be addressed by molecular dynamics (MD) simulations. Thanks to the rapid development of faster architectures (graphics processing unit (GPU)-based clusters) and better algorithms for high-level computations, classical MD simulations nowadays allow the implementation of SBDD strategies that account for the structural flexibility of drug–target systems at a fully atomistic description. Therefore, MD simulations can be exploited to characterize the protein–ligand binding process at a fully dynamic level. As a result, classical MD is no longer considered prohibitive for effective drug design. Instead, it is pushing

the frontiers of computationally driven drug discovery in both academia and industry [6].

2.1.2 Molecular Dynamics – The Current State of the Art

MD is a physical method for studying the interaction and motion of atoms and molecules according to Newton's physics. A force field is used to estimate the forces between interacting atoms and calculate the overall energy of the system. MD simulations are used today to study nearly every type of macromolecule, including proteins, nucleic acids, and carbohydrates, of biological or medicinal interest. Simulations span wide spatial and temporal ranges and resolutions. To be explicit, all-atom MD, thousands to millions of individual atoms representing, for instance, all the atoms of a protein and surrounding water molecules, move in a series of short (e.g. 2 fs), discrete time steps. At each step, the forces on each atom, determined from the "force field," e.g. a collection of physics-based parameters that represent both bonded and nonbonded (e.g. van der Waals) interatomic forces, are computed and the atomic position and velocity are updated according to Newton's laws of motion [7]. Recently, computational speed has grown rapidly with the introduction of parallel computing based on GPU computing techniques and specialized computing machines, e.g. high-performance supercomputers [8], designing of specialized hardware [9], and development of intelligent MD protocols [10]. As a consequence, nowadays simulations lasting up to a few milliseconds are now possible. Assuming that the ergodic hypothesis holds, an infinitely long MD trajectory should be able to sample the entire conformational space. However, except for a few cases published recently [11, 12], long MD simulations are often relegated to some local minimum in the rugged and complex free energy surface (FES) that is accessible to the system. At room temperature, the probability of crossing high-energy barriers is often too small to be observed during a finite MD simulation. Most likely, even with several hundred nanoseconds of simulations, the system might be confined to limited regions of the conformational space. Therefore, besides the increasing computational power, various modeling techniques have been developed to accelerate simulations and/or to sample rare events during ligand binding processes. A solution usually applied to overcome the limited sampling efficiency of MD simulations at room temperature consists in raising the simulation temperature. The additional kinetic energy available at higher temperature allows the crossing of high-energy barriers, thus ensuring a wider sampling of the conformational space. This methodology is the basis of two computational approaches, simulated annealing (SA) and parallel tempering (also named replica-exchange molecular dynamics, REMD). SA consists in heating up the system in order to jump out from the initial local minimum to explore other minima [13]. The heating step is followed by a gradual cooling, which allows the system to slowly settle down to a lower energy minimum. REMD [14] is based on the run of multicopy MD simulations randomly initialized, at different temperatures. The conformations are then exchanged at different temperatures following the Metropolis criterion. The strength and robustness of this method allows sampling of both low- and high-energy configurations [15].

The well-recognized limitations of sampling in atomistic dynamics have led to many innovative alternatives to enhance the coverage of the thermally accessible conformational space and to capture *rare events* (events that might happen on a long timescale) [16]. One trick that is commonly used to address this problem is to add a potential bias in order to force the *rare event* to occur. In this context, several techniques, including the local elevation method [17], taboo search [18], the Wang–Landau method [19], adaptive force bias [20] conformational flooding [21], umbrella sampling [22], weighted histogram techniques [23], transition state theory and path sampling [24], steered MD [25], free energy–guided sampling [26], and metadynamics [27], have been developed to address the sampling problem. In this context, metadynamics has emerged as a powerful coarse-grained non-Markovian MD approach for the acceleration of rare events and the efficient and rapid computation of multidimensional FES as a function of a restricted number of degrees of freedom, named collective variables (CVs). Different from other sampling methods, in which the calculation of FES requires an additional step (such as the weighted histogram analysis method (WHAM) [19]), metadynamics directly provides a good estimate of the free energy of the system projected into the CVs. On one hand, the main advantage of metadynamics along with other methods based on CVs is that the sampling is enhanced toward the specific event of interest by biasing the MD simulations along chosen CVs, which are functions of the atomic coordinates. In this way, free energy barriers can be efficiently crossed. On the other hand, it can be difficult to choose the right set of CVs *a priori*, particularly when the system's reaction mechanism is not yet known. In the latter case, tempering methods, such as REMD, accelerated molecular dynamics (aMD) [28], and potential scaled [29] seem to be more appropriate since they act by heating all degrees of freedom of the system or by modifying the Hamiltonian [30].

Finally, MD simulations are capable of providing quantitative binding estimates with high accuracy, i.e. free energy of binding, but it has been typically regarded as too computationally expensive. A variety of free energy simulation methods, such as free energy perturbation (FEP), thermodynamic integration (TI), and λ dynamics, employ an analysis of atomistic MD simulations to determine the free energy difference between two related ligands via either a chemical or alchemical path. In recent years, FEP calculations have benefited from improved force fields, new sampling algorithms, and the emergence of low-cost parallel computing, which have resulted in the level of accuracy and turnaround time needed to impact lead optimization efforts.

The examples in the following sections indicate some of the current capabilities of MD simulations in three different fields: protein–protein, protein–peptide, and protein–ligand binding and the insights they can provide.

2.2 Protein–Protein Binding

Protein–protein interactions (PPIs) are ubiquitous in biological systems as they are the building blocks for the complex molecular machinery that drives critical

functions within the cell. Protein–protein association signals transduction and inhibition, immune response, and protein recruitment to membrane scaffolds. Mutations and perturbations within the proteins that result in aberrant interactions often lead to a number of diseases, including cancer. It has been estimated that the human interactome involves between 130 000 and 600 000 PPIs [31–33]. Although the Protein Data Bank (PDB) contains more than 100 000 entries, structural data on protein–protein complexes remain scarce [34]. In this context, MD simulations along with docking aim at offering opportunities to fill this gap and provide valuable information on the structure of PPIs as well as their specific interactions, and furthermore support the design of novel PPI modulators or inhibitors. PPIs are often characterized by the cooperative formation of many weak interatomic interactions over an extended binding surface rather than by few strong interactions, as in the case of protein–ligand binding. In comparison to classical small-molecule targets, the binding site of protein–protein interfaces is relatively large, with an area of about 1500–3000 $Å^2$ [35]. Nevertheless, slight conformational changes can be temporarily induced to enable PPIs, which mostly involve side-chain motions or small fluctuations of loops [36].

Despite the availability of several experimental data on the PPIs, such as NMR, X-ray structure, and isothermal titration calorimetry (ITC) thermodynamic data, the description of the protein–protein system is limited to static snapshots, which are not sufficient to describe coherently the molecular processes occurring upon binding. In this scenario, atomistic simulations could greatly contribute to linking binding mechanism and function. MD simulation could adequately represent the complex physics and thermodynamics behind the protein–ligand complex formation in order to investigate the underlying mechanisms of complex formation and modulation or inhibition. The mechanism of regulated recognition involves conformational changes in the binding partners, which are prevented or induced by external signals. This implies that protein–protein binding motifs can be labile and intrinsically flexible with binding sites that are shallow and rather featureless compared to "druggable" sites [36]. In comparison to classical ligand–protein approaches, this step is far from trivial as high-resolution structural information is needed as a prerequisite. These data can be obtained from experiments (NMR, X-ray crystallography, etc.), PPI databases, or modeling approaches such as homology modeling or protein–protein docking. As an alternative to simulating already existing PPIs, it has been shown that the initial formation of PPIs during MD simulations starting from two separate proteins is possible [37]. Nevertheless, this approach still is computationally demanding and requires experimental validation. This section highlights the benefits of applying MD to PPI investigations.

With MD simulations, protein interactions, flexibility, and sampling of the conformational space can be characterized, which is of great value for the investigation of the interacting surfaces of protein–protein complexes. MD simulations can also be used to identify binding hot spots and transient pockets that play a role in protein–protein binding but could not be captured by crystallography [38, 39]. Due to the dynamic nature of protein–protein interfaces, transient pockets and buried binding hot spots can emerge on the protein surfaces; and binding of small molecules to these transient areas have been reported [40].

Combining MD and principal component analysis (PCA), Dixit and Verkhivker identified conserved functional regions of Hsp90, an important chaperone protein [41]. These findings help in the design of inhibitors targeting these protein–protein interfaces and furthermore provide mechanistic insights into the allosteric regulation of the complex dynamics and motions of the chaperone machinery.

Bastianelli et al. exploited MD simulations to refine the protein–protein docking results and to perform free energy calculations [42]. In this work, the authors were able to predict the interactions between two proteins, i.e. PcFK1, a small protein extracted from spider venom, and PfSUB1, a subtilisin-like serine protease. The results were then confirmed by experimental validation.

Although the interaction area between proteins is relatively large, it has been shown that only a small subset of key residues contribute the most to high-affinity binding in PPIs [43]. These protein domains with key interacting residues are termed hot spots, and hot spot detection is an efficient strategy to identify potential druggable areas of PPIs [43]. This can be done by computational alanine scanning (CAS), by which a single residue is mutated in alanine in order to determine its energetic contribution to identifying hot spots. Using the CAS method, Sousa et al. provided an evaluation of the energetic contribution of the amino acid residues at the subunit interface and also potential starting points for HIV inhibitors that block the dimerization of the protein [44].

2.3 Protein–Peptide Binding

In many PPIs, it has been shown that the interactions are dominated by short linear recognition motifs that are part of the larger protein [45]. Protein–peptide recognition has a crucial role in various fundamental aspects of cellular homeostasis, such as signal transduction, protein trafficking, and immune response. Peptides often serve as signaling vehicles as, for example, in the case of hormones and neurotransmitters. Peptides in organisms often occur as part of the breakdown of proteins. One important example is the amyloid beta (Abeta) peptide which results from the hydrolysis of the APP protein on the surface of neurons. Plaques formed by the aggregation of Abeta peptides are recognized as one of the hallmarks of Alzheimer's disease [46]. Antimicrobial peptides are an important component of the innate immune system, and peptides presented and recognized on the surface of antigen-presenting cells by the fragmentation of foreign proteins are recognized as one of the main processes of the adaptive immune response [47].

Protein–peptide complexes often display interaction patterns different from the ones often encountered in protein–protein complexes. For example, protein–peptide interactions tend to be more hydrogen bond–dependent than PPIs, usually involving the peptide backbone. Finally, the protein–peptide binding process frequently involves significant conformational rearrangements of peptide chains (e.g. from unfolded to folded structure).

The efficient treatment of the large-scale changes remains one of the major challenges for molecular docking and dynamics simulations [48]. The inefficiency

in classic MD sampling is a result of the highly rugged nature of the folding energy landscape of the peptide chains. The prediction of the structures of protein–peptide complexes and the quantitative prediction of their binding affinities are far from a solved problem. In the following, we report the few examples in literature exploiting MD simulations of protein–peptide systems. In 2009, Pietrucci et al. used the bias-exchange metadynamics technique to successfully describe the binding mechanism of a small peptide to the HIV-1 protease [49]. Even though the authors accurately computed the free energy associated with ligand binding and unbinding as a function of 7 CVs, almost 2 μs of simulation were required to converge the free energy. Despite this, they managed to characterize the kinetics of the binding/unbinding process using a discrete-states kinetic model, including the relevant metastable states along the recognition pathway.

Tortorici et al. performed MD simulations in explicit solvent as well as in temperature replica exchange MD simulations to explore the peptide conformations derived from the "SNAG" domain of SNAIL1 in solution [50]. True to the experimental data, MD simulations revealed that the peptides exist as random structures, showing transient secondary structures with an α-helical propensity of only about 7%. The authors used this evidence to suggest that the peptides do indeed exist as disordered conformers in solution and only form a secondary structure upon binding, in agreement with experimental data (CD spectra) and with the analysis of seven crystal structures of SNAIL-derived peptides bound to LSD1/CoREST in which the peptides were bound in a helical structure.

Efficiently predicting binding affinities of protein–peptide binding affinities from structural models would be an extremely important step in our overall understanding of the regulation of biological systems and critical for drug design and development. MD simulations together with binding free energy calculation under the molecular mechanics-generalized born surface area (MM-GBSA) approximation was used to study the binding affinities of a series of short peptides against the catalytic site of protein kinase A [51]. To reduce computational cost, the entropy term was not computed. This approach was justified by the observation that other studies on similar systems had concluded that the entropy term did not contribute significantly to the binding affinity and/or would not vary appreciably across the peptide library. The authors reported high correlation between experimental free energy values and computational values ($r^2 = 0.98$). The authors also concluded that the most influential term in the binding free energy is the electrostatic term, again showing that the success of this model is intimately tied to the electrostatic interactions present.

Another important field in protein–peptide binding is the epigenetic recognition by histone tails. Epigenetic regulation of gene transcription relies on an array of recurring structural domains that have evolved to recognize posttranslational modifications on histones. Misinterpretation of epigenetic mechanisms is often associated with disease conditions. Epigenetic readers are therefore rapidly becoming a privileged target for drug discovery study. Plant homeodomain (PHD) fingers represent one of the largest families of epigenetic readers capable of decoding posttranslationally modified or unmodified histone H3 tails. Spiliotopoulos et al. presented a computational case study describing

the dynamics and the thermodynamics at the basis of unmodified histone H3 (H3K4me0) recognition by the first PHD finger of the transcriptional activator Autoimmune Regulator (AIRE-PHD1) [52]. MD simulations together with PCA showed that histone peptide binding strongly affects AIRE-PHD1 internal correlation motions. In particular, analysis of the essential domain motions highlighted the presence of a "flapping" movement in free AIRE-PHD1, which might be relevant for domain function. Importantly, this intrinsic domain "breathing" was blocked upon H3K4me0 binding. Moreover, calculations of the free energy of binding obtained through molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) methodology allowed us to dissect the energetic terms associated with native and alanine mutants of AIRE-PHD1/H3K4me0 complexes. The calculated binding free energies were in good qualitative agreement with experiments. Prompted by these encouraging results, the authors extended MM/PBSA calculations to other PHD fingers recognizing the unmodified status of histone H3K4 in order to investigate whether this subclass of histone readers shared similar thermodynamic behavior. Indeed, comparison of calculations with experimental binding data confirmed that MM/PBSA can be considered as a valuable tool to rapidly analyze the energetic determinants dictating histone decoding by this class of epigenetic effectors.

Unlike endpoint methods, such as MM/PBSA/GBSA, pathway free energy methods compute the difference in binding free energy between the unbound and bound states of the ligand–protein complex by linking them through a suitable thermodynamic path. Oftentimes the path is parameterized by a progress parameter λ ranging from 0 to 1. A λ-dependent potential energy function $U(x_A, x_B, \zeta_B, r_s; \lambda)$ is designed in such a way that at $\lambda = 0$ it represents the unbound state of the complex (when protein and ligand are not interacting), and at $\lambda = 1$ it represents the bound state of the complex. The reversible work along the path yields the binding free energy of the complex [53]. As the decoupling transformations are too extensive and the corresponding free energies are too large to routinely obtain converged binding free energy estimates, there has been only one report of the application of the double decoupling method to the calculation of absolute binding free energies of peptides [54].

2.4 Protein–Ligand Binding

MD simulations have been successfully and widely utilized in drug design and development [30]. Usually, protein–ligand binding is studied by docking simulations. If we are interested in determining a plausible pose, the receptor flexibility during the docking is modeled using an ensemble of structures, obtained through either experimental techniques (X-ray crystallography or NMR spectroscopy) or by exploiting sampling engines such as Monte Carlo (MC) or MD. The latter one has the advantage of taking into account the global flexibility of the protein, and the only drawback is the narrowness of sampling. First, a long MD simulation of the protein in the unbound state is performed to exhaustively sample the conformational space. Then, MD snapshots are selected either at regular time intervals or after cluster analysis to perform docking simulations.

Moving beyond protocols that use MD to incorporate target flexibility into standard docking calculations, it is now possible to run MD simulations for long enough to explore the free energy landscape and kinetic profile associated with the overall drug-binding process (i.e. from the drug fully solvated in water to the drug–target bound state) [55]. The major advantage of MD simulation with respect to docking simulation in studying protein–ligand is twofold: (i) investigating possible allosteric binding mode; (ii) structural characterization of the binding path and *a posteriori* calculations of the binding free energy minimum path. Even though the MD framework is certainly appealing and will most likely be employed more in the future, at present, some efficiency limitations prevent its widespread use in SBDD. Long *plain* MD simulations have already described the spontaneous binding of small organic molecules to biological targets of pharmacological interest, demonstrating both the adequacy of the present force fields and the feasibility, at least in some cases, of observing the rare event.

A seminal example of the useful insights offered by MD simulations in structure-based drug design was reported in 2011, when Buch and Giorgino performed on GPU-based infrastructures 495 MD simulations of 100 ns each to simulate the complete binding process of the inhibitor benzamidine to β-trypsin [56]. The approach allowed the identification of the lowest energy binding mode of a ligand to a receptor. Monitoring of the binding process at an atomic resolution can potentially assist the development of drugs able to control and modulate the ligand–receptor recognition process. In this context, Shaw and collaborators formulated a mechanism for the flipping of a conserved motif of Abl tyrosine kinases combining microsecond MD simulations with crystallographic and kinetic experiments. Importantly, the conformation of this motif was crucial to discriminate between active/inactive kinase conformations. Their results led to the identification of a class of potent inhibitors of both Src and Abl that recognize the inactive kinase conformations [12].

The G protein–coupled receptor β_2-adrenergic (β_2-AR) is an important target for hypertension and several heart diseases. Shaw and coworkers have used extensive plain MD to study how this receptor interacts and binds with the inverse agonist alprenolol [11]. The binding paths followed a two-step molecular mechanism. First, alprenolol reorients so that its hydrophobic group binds to the extracellular vestibule surface. Then, it penetrates into the binding pocket through the gate closing–opening mechanism between Tyr308 and Phe193 to form the salt bridges between the ammonium moiety and the carboxylate of Asp113.

More recently, Decherchi et al. ran extensive unbiased MD simulations (about 1 μs each) to investigate the tight binding event of a transition state analog (DADMe-immucilin H) into the pharmaceutical target, purine nucleoside phosphorylase [57]. Despite observing spontaneous binding through different routes, the authors used the pathways obtained by MD simulations to determine the free energy profiles associated with the diverse binding mechanisms.

The aforementioned examples require a huge amount of computational resources, far more than the average time allocated to a real-world drug discovery project. A valid alternative is to accelerate the sampling while preserving a statistical mechanics distribution of states through so-called enhanced sampling

methods. Researchers have recently demonstrated the power of these methods for studying protein–ligand binding and estimating the associated free energy and kinetics [38, 58].

Among all these, the metadynamics-based method, which allows fast exploration of the underlying free energy landscape of rare events, has been applied to a number of ligand–target complexes, demonstrating its ability to characterize binding and unbinding paths, to treat conformation flexibility, and to compute free energy profiles. Since Gervasio et al. first applied metadynamics to ligand–target complexes, several other informative studies have been reported [58, 59].

2.5 Future Directions

Nowadays, computational power has increased with respect to the past to such an extent that most MD simulations are run on personal powerful workstations. The increase in computational power implies a different future scenario. Simulations will become much larger and will reach longer timescales. This reach for ever larger and longer simulations will also be increasingly aided by improved algorithmic methods to increase the sampling of conformational space. The unbiased assessment of protein–protein and protein–nucleic acid interactions, in particular, will present new drug discovery opportunities. MD simulations can address two aspects of this problem. First, how and where do the partners interact? And as they interact, are novel ligand binding sites created, at the interface or at allosteric sites? MD simulations seem to be good at identifying low-energy protein conformations that harbor cryptic drug-binding sites. Second, how can small-molecule ligands modulate those interactions? How can we optimize their binding and druglike properties? Simulations of macromolecular assembly will also extend to include very large complexes, such as the nuclear pore complex or even entire organelles.

Although we have been able to simulate on a millisecond timescale, there are still some parameters including torsions in amino acids, carbohydrates, and single-stranded nucleic acids for which the force fields need further improvement. The existing force fields are not efficient for studying the crowded environment inside the cells, since these interactions involve multiple factors in real time. Therefore, the improved force fields may provide the opportunities for their wider applications on the complex biosystems in diverse cellular conditions.

2.5.1 Modeling of Cation-p Interactions

These interactions, driven by electric quadrupole moments and polarization effects, are now recognized to be quite important to both protein structure (e.g. arginine–tryptophan ladders) and protein function, for instance, in a wide variety of protein–ligand interactions. Polarizable models should enable us to more accurately describe (inter)molecular interactions, and, indeed, several polarizable force fields have begun to demonstrate their value. The AMOEBA

force field, for instance, includes some cross-terms and, significantly, polarizable atomic multipoles (up to quadrupoles) replace fixed partial charges [6].

2.6 Grand Challenges

There are several grand challenges that will help make MD simulations a standard "tool" to study and describe life sciences. First of all, free energy calculations should be more reliable and rapid. The calculation of free energy should answer to different questions for both macromolecule–ligand and macromolecule–macromolecule interactions. The importance of this computationally demanding task is crucial to answer whether either a ligand or a macromolecule will interact or not. Moreover, the magnitude of the interactions needs to be accurate (less than 1 kcal mol^{-1}). Finally, the free energy is of vital importance in drug discovery. Improvement of free energy calculations has to go through both the exhaustive sampling of configurational space and better force field. The abovementioned force field improvements should help significantly. Converged calculations will enable rigorous determination of both force field accuracy and the need for specific force field improvements.

References

1 Changeux, J.-P. and Edelstein, S. (2011). Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol. Rep.* 3: 19.
2 Boehr, D.D., Nussinov, R., and Wright, P.E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5: 789–796.
3 Ma, B., Kumar, S., Tsai, C.J., and Nussinov, R. (1999). Folding funnels and binding mechanisms. *Protein Eng.* 12 (9): 713–720.
4 Breiten, B., Lockett, M.R., Sherman, W. et al. (2013). Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *J. Am. Chem. Soc.* 135: 15579–15584.
5 Jorgensen, W.L. (2004). The many roles of computation in drug discovery. *Science* 303: 1813–1818.
6 Borhani, D.W. and Shaw, D.E. (2012). The future of molecular dynamics simulations in drug discovery. *J. Comput.-Aided Mol. Des.* 26: 15–26.
7 Karplus, M. and Kuriyan, J. (2005). Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* 102: 6679–6685.
8 Shaw, D.E., Chao, J.C., Eastwood, M.P. et al. (2008). Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* 51: 91.
9 Harvey, M.J., Giupponi, G., and Fabritiis, G.D. (2009). ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* 5: 1632–1639.
10 Pronk, S., Pall, S., Schulz, R. et al. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29: 845–854.

11 Dror, R.O., Pan, A.C., Arlow, D.H. et al. (2011). Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* 108: 13118–13123.

12 Shan, Y., Kim, E.T., Eastwood, M.P. et al. (2011). How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* 133: 9181–9183.

13 Kirkpatrick, S., Gelatt, C., D. Jr., and Vecchi, M.P. (1983). Optimization by simulated annealing. *Science* 220: 671–680.

14 Sugita, Y. and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314: 141–151.

15 Ostermeir, K. and Zacharias, M. (2013). Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins. *Biochim. Biophys. Acta* 1834: 847–853.

16 Mitsutake, A. and Mori, Y. (2013). *Enhanced Sampling Algorithms*. Springer.

17 Huber, T., Torda, A.E., and van Gunsteren, W.F. (1994). Local elevation: a method for improving the searching properties of molecular dynamics simulation. *J. Comput.- Aided Mol. Des.* 8: 695–708.

18 Cvijovicacute, D. and Klinowski, J. (1995). Taboo search: an approach to the multiple minima problem. *Science* 267: 664–666.

19 Wang, F. and Landau, D.P. (2001). Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* 86: 2050–2053.

20 Darve, E., Rodriguez-Gomez, D., and Pohorille, A. (2008). Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* 128: 144120.

21 Grubmuller, H. (1995). Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdisciplin. Top.* 52: 2893–2906.

22 Patey, G.N. and Valleau, J.P. (1975). Monte-Carlo method for obtaining interionic potential of mean force in ionic solution. *J. Chem. Phys.* 63: 2334–2339.

23 Ferrenberg, A.M. and Swendsen, R.H. (1988). New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* 61: 2635–2638.

24 Bolhuis, P.G., Chandler, D., Dellago, C. et al. (2002). Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* 53: 291–318.

25 Grubmüller, H., Heymann, B., and Tavan, P. (1996). Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* 271: 997–999.

26 Zhou, T. and Caflisch, A. (2012). Free energy guided sampling. *J. Chem. Theory Comput.* 8: 2134–2140.

27 Laio, A. and Parrinello, M. (2002). Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* 99: 12562–12566.

28 Hamelberg, D., Mongan, J., and McCammon, J.A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120: 11919–11929.

29 Mark, A.E., van Gunsteren, W.F., and Berendsen, H.J.C. (1991). Calculation of relative free energy via indirect pathways. *J. Chem. Phys.* 94: 3808–3816.

30 De Vivo, M., Masetti, M., Bottegoni, G., and Cavalli, A. (2016). Role of molecular dynamics and related methods in drug discovery. *J. Med. Chem.* 59: 4035–4061.

31 Stumpf, M.P.H. et al. (2008). Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U.S.A.* 105: 6959–6964.

32 Venkatesan, K. et al. (2009). An empirical framework for binary interactome mapping. *Nat. Methods* 6: 83–90.

33 Cunningham, A.D. et al. (2017). Peptides and peptidomimetics as regulators of protein–protein interactions. *Curr. Opin. Struct. Biol.* 44: 59–66.

34 Szilagyi, A. and Zhang, Y. (2014). Template-based structure modeling of protein–protein interactions. *Curr. Opin. Struct. Biol.* 24: 10–23.

35 Wells, J.A. and McClendon, C.L. (2007). Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* 450: 1001–1009.

36 Arkin, M.R. and Wells, J.A. (2004). Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nat. Rev. Drug. Discov.* 3: 301–317.

37 Shan, Y., Gnanasambandan, K., Ungureanu, D. et al. (2014). Molecular basis for pseudokinase-dependent autoinhibition of JAK2 tyrosine kinase. *Nat. Struct. Mol. Biol.* 21: 579–584.

38 Eyrisch, S. and Helms, V. (2007). Transient pockets on protein surfaces involved in protein–protein interaction. *J. Med. Chem.* 50: 3457–3464.

39 Eyrisch, S., Medina-Franco, J.L., and Helms, V. (2012). Transient pockets on XIAP-BIR2: toward the characterization of putative binding sites of small-molecule XIAP inhibitors. *J. Mol. Model.* 18: 2031–2042.

40 Arkin, M.R., Randal, M., DeLano, W.L. et al. (2003). Binding of small molecules to an adaptive protein–protein interface. *Proc. Natl. Acad. Sci. U.S.A.* 100: 1603–1608.

41 Dixit, A. and Verkhivker, G.M. (2012). Probing molecular mechanisms of the Hsp90 chaperone: biophysical modeling identifies key regulators of functional dynamics. *PLoS One* 7: e37605.

42 Bastianelli, G., Bouillon, A., Nguyen, C. et al. (2011). Computational reverse-engineering of a spider-venom derived peptide active against Plasmodium falciparum SUB1. *PLoS One* 6: e21812.

43 Moreira, I.S., Fernandes, P.A., and Ramos, M.J. (2007). Hot spots-a review of the protein–protein interface determinant amino-acid residues. *Proteins* 68: 803–812.

44 Sousa, S.F., Tamames, B., Fernandes, P.A., and Ramos, M.J. (2011). Detailed atomistic analysis of the HIV-1 protease interface. *J. Phys. Chem. B* 115: 7045–7057.

45 London, N., Movshovitz-Attias, D., and Schueler-Furman, O. (2010). The structural basis of peptide-protein binding strategies. *Structure* 18 (2): 188–199.

46 Hardy, J. and Allsop, D. (1991). Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends Pharmacol. Sci.* 12: 383–388.

47 Pino-Angeles, A., Leveritt, J.M. III, and Lazaridis, T. (2016). Pore structure and synergy in antimicrobial peptides of the magainin family. *PLoS Comput. Biol.* 12 (1).

48 Antunes, D.A., Devaurs, D., and Kavraki, L.E. (2015). Understanding the challenges of protein flexibility in drug design. *Expert Opin. Drug Discovery* 10: 1301–1313.

49 Pietrucci, F., Marinelli, F., Carloni, P., and Laio, A. (2009). Substrate binding mechanism of HIV-1 protease from explicit-solvent atomistic simulations. *J. Am. Chem. Soc.* 131: 11811–11818.

50 Tortorici, M., Borrello, M.T., Tardugno, M. et al. (2013). Protein recognition by short peptide reversible inhibitors of the chromatin-modifying LSD1/CoREST lysine demethylase. *ACS Chem. Biol.* 8 (8): 1677–1682.

51 Mena-Ulecia, K., Vergara-Jaque, A., Poblete, H. et al. (2014). Study of the affinity between the protein kinase PKA and peptide substrates derived from kemptide using molecular dynamics simulations and MM/GBSA. *PLoS One* 9 (10): e109639.

52 Spiliotopoulos, D., Spitaleri, A., and Musco, G. Exploring PHD fingers and H3K4me0 interactions with molecular dynamics simulations and binding free energy calculations: AIRE-PHD1, a comparative study. *PLoS One* 7 (10): e46902.

53 de Ruiter, A. and Oostenbrink, C. (2011). Free energy calculations of protein–ligand interactions. *Curr. Opin. Chem. Biol.* 15 (4): 547–552.

54 Gumbart, J.C., Roux, B., and Chipot, C. (2012). Standard binding free energies from computer simulations: what is the best strategy? *J. Chem. Theory Comput.* 9 (1): 794–802.

55 Klepeis, J.L., Lindorff-Larsen, K., Dror, R.O., and Shaw, D.E. (2009). Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* 19: 120–127.

56 Buch, I. and Giorgino, T. (2011). Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* 108 (25): 10184–10189.

57 Decherchi, S., Berteotti, A., Bottegoni, G. et al. (2015). The ligand binding mechanism to purine nucleoside phosphorylase elucidated via molecular dynamics and machine learning. *Nat. Commun.* 6: 6155.

58 Cavalli, A., Spitaleri, A., Saladino, G., and Gervasio, F.L. (2015). Investigating drug–target association and dissociation mechanisms using metadynamics-based algorithms. *Acc. Chem. Res.* 48 (2): 277–285.

59 Gervasio, F.L., Laio, A., and Parrinello, M. (2005). Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* 127: 2600–2607.

“keywords/abstract

Dear Author,

Keywords and abstracts will normally not be included in the print version of your chapter but only in the online version (if not decided differently by Wiley-VCH).

Thank you!”

Abstract

Keywords