Biophysical Chemistry 186 (2014) 31–45

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Review

Exploring the role of receptor flexibility in structure-based drug discovery

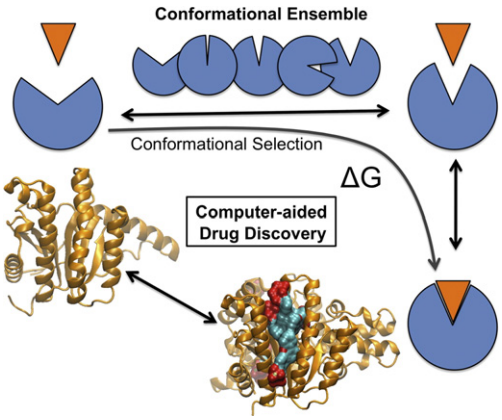
Ferran Feixas [a,b,*], Steffen Lindert [a,b], William Sinko [a,b], J. Andrew McCammon [a,b,c,d]

[a] Deparment of Chemistry & Biochemistry, University of California San Diego, La Jolla, CA 92093, United States
[b] Department of Pharmacology, University of California San Diego, La Jolla, CA 92093, United States
[c] NSF Center for Theoretical Biological Physics, La Jolla, CA, United States
[d] Howard Hughes Medical Institute, University of California San Diego, La Jolla, CA, United States

HIGHLIGHTS

- Receptor flexibility plays a key role in structure-based drug design.
- Receptor ensemble-based methods improve predictive power of virtual screening.
- MD and enhanced sampling techniques are useful tools to explore conformational space.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 2 September 2013
Received in revised form 31 October 2013
Accepted 31 October 2013
Available online 9 November 2013

Keywords:
Conformational selection
Allostery
Molecular dynamics
Receptor flexibility
Computer-aided drug design
Accelerated molecular dynamics

ABSTRACT

The proper understanding of biomolecular recognition mechanisms that take place in a drug target is of paramount importance to improve the efficiency of drug discovery and development. The intrinsic dynamic character of proteins has a strong influence on biomolecular recognition mechanisms and models such as conformational selection have been widely used to account for this dynamic association process. However, conformational changes occurring in the receptor prior and upon association with other molecules are diverse and not obvious to predict when only a few structures of the receptor are available. In view of the prominent role of protein flexibility in ligand binding and its implications for drug discovery, it is of great interest to identify receptor conformations that play a major role in biomolecular recognition before starting rational drug design efforts. In this review, we discuss a number of recent advances in computer-aided drug discovery techniques that have been proposed to incorporate receptor flexibility into structure-based drug design. The allowance for receptor flexibility provided by computational techniques such as molecular dynamics simulations or enhanced sampling techniques helps to improve the accuracy of methods used to estimate binding affinities and, thus, such methods can contribute to the discovery of novel drug leads.

* Corresponding author at: Department of Chemistry & Biochemistry, University of California San Diego/HHMI, 9500 Gilman Drive, Mail Code 0365, La Jolla, CA 92093-0365, United States. Tel.: +1 858 534 2913 (Office); fax: +1 858 534 4974.
E-mail address: ffeixas@ucsd.edu (F. Feixas).

0301-4622/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bpc.2013.10.007

Contents

1. Biomolecular recognition mechanisms

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. In the last decades, the interpretation of mechanisms describing biomolecular recognition has been the focus of a passionate debate that has contributed to push forward the research in many fields such as biophysics and pharmacology among others [1–3]. More than 50 years ago, our view of binding events underwent a Copernican turn evolving from an idea based on rigid lock-and-key like models to be described as a dynamic and flexible process [4,5]. All these findings not only served to advance the field towards a better understanding of protein–ligand binding but also introduced an extra degree of complexity to the description of biomolecular recognition processes. Biomolecular recognition is an intricate process of orchestrated and random motions, where the ligand from one side and the receptor from the other seek for complementary conformations to improve the binding affinity with its partner along this fascinating biomolecular dance.

The description of protein–ligand interactions is not a simple task due to the variety of motions and mechanisms interplaying in this complex but vital process. To comprehend how biomolecular recognition occurs, we first need to understand the role of all different partners involved in this association process. One of the main centers of attention has been to elucidate the role played by the ligand during the binding event. In particular, whether it is directly responsible for inducing a conformational change to the biological receptor upon binding or whether it stabilizes specific preexistent conformational states displayed by the dynamic protein. In other words, by which mechanisms do ligands such as substrates or synthetic drugs regulate biochemical reactions? In the last decades, the concepts of induced fit and conformational selection emerged as the most popular mechanisms to explain the intricate biomolecular recognition process. The idea of induced fit, introduced by Koshland more than fifty years ago, 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 [4]. This model implies that interacting biomolecules do not necessarily have a complementary shape prior the binding event because it is induced by the ligand. However, experimental evidences based on kinetic studies proved that the induced fit hypothesis was not able to describe all the variety of binding scenarios [6]. In 1999, Nussinov and coworkers coined the term conformational selection, also known as population shift, 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 towards this specific conformational state (see Fig. 1) [7–10]. Both theories, although they appear to be antagonistic, are not necessarily mutually exclusive. Recent studies show that conformational selection is usually followed by a conformational adjustment [11]. In this line, extended models that combine characteristics of conformational selection, induced fit and classical lock-and-key mechanisms have been reported [3]. Despite being often disregarded, water plays a crucial role in molecular association. In the last years, great efforts have been put to determine the nature of the hydrophobic effect and its implications for biomolecular recognition. 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 [12–15]. Computer-aided drug design techniques try to incorporate some of the main features of biomolecular recognition process to improve the accuracy and predictive power of these computational methods. For example, a plethora of techniques have been proposed to account for conformational selection and induced fit during the estimation of binding affinities in structure-based virtual screening [16–19].

The debate on mechanisms underlying biomolecular recognition has been always strongly linked to the study of allosteric effects. Allostery is a phenomenon that describes the interaction occurring between a regulatory site, also called allosteric site, and another site of the protein, usually the active site, that gives rise to a functional change on the latter [5,20]. This process is mediated by an effector that binds to the allosteric site, which induces a conformational change to the protein that affects the activity of another site, altering protein function. Thus, the allosteric effector is responsible for regulating the biological activity of the protein. The allosteric term was coined and popularized in the early 1960s by Changeux, Jacob and Monod from their studies of conformational changes mediated by signal transduction in several enzymes, where they tried to initially explain allosteric effects from the induced fit perspective [21,22]. Despite the youth of the term allostery, this concept underwent a rapid revolution when the Monod–Wyman–Changeux (MWC) model was proposed to account for positive cooperativity and allosteric effects of oxygen binding in myoglobin [5]. This model states that when an allosteric binding event occurs, a shift of the equilibrium of two pre-existing conformational states is observed. Consequently, the early works of Changeux and coworkers laid the foundations of some of the ideas that would eventually lead to the introduction of the conformational selection biomolecular recognition mechanism. The MWC theory of allostery was opposed to the Koshland–Némethy–Filmer (KNF) model, which explained the conformational transitions observed as a consequence of allosteric binding, in the same terms as the induced fit theory [23]. The KNF theory also incorporated some of the ideas introduced by Pauling on the study of cooperativity in oxygen binding in hemoglobin [24]. After several years of discussion, the MWC model and its subsequent generalizations [3,25,26] remained as the most widely used theories to account for allosteric effects. A third model of allostery, referred to us as entropic allostery, pictures the remote effects of ligand binding to have a purely

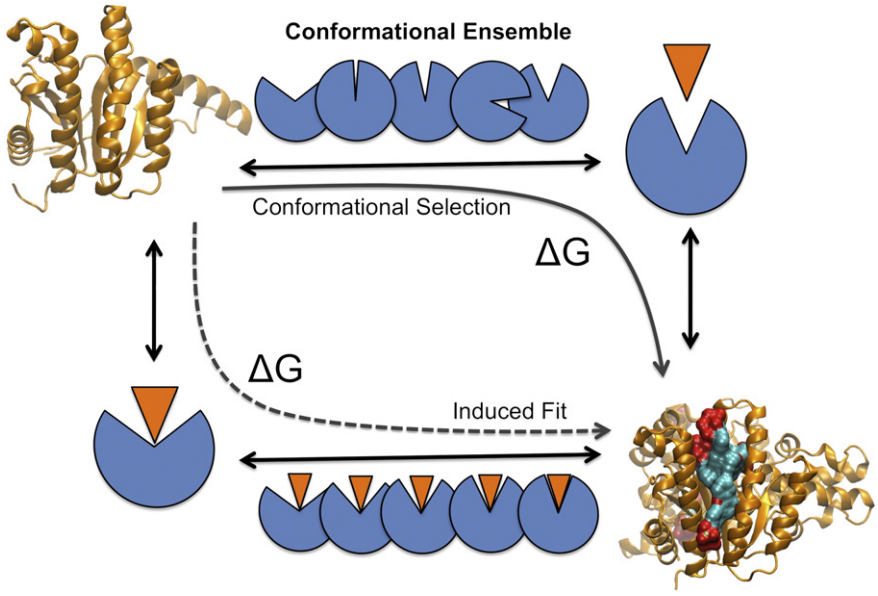


Fig. 1. Schematic pathways of biomolecular recognition. Conformational selection and induced fit mechanisms are depicted in solid and dashed lines respectively.

dynamical character [27], and some evidence for this model has been seen in experimental and computational work [28,29]. Allosteric transitions have been proven to be of great importance to explain signal transduction mediated by G-protein coupled receptors (GPCR) [30,31]. Depending on the nature of the ligand bound to the orthosteric or allosteric sites, some GPCRs are able to assume different conformations that may lead to the activation of different pathways. For example, β_2-adrenergic receptor (β_2AR), which activates several G-proteins, adopts different conformations and binds to a large diversity of ligands that are able to trigger different signaling pathways [32].

Allosteric effects alter the shape and/or dynamics of the protein. These changes cause a set of responses to the protein that affect not only the function of the protein itself but also its cellular pathway, or they may even produce a large-scale response in the organism. Consequently, the attention has been focused on the design of allosteric drugs and the study of how these drugs are able to alter protein network pathways [33–35]. This strategy offers a wide range of possibilities for the synthesis of new drugs. For example, the chemokine CCR5 can be modulated by the approved allosteric drug maraviroc, which acts as a negative modulator [36]. However, the identification of allosteric sites and allosteric mechanisms is not a straightforward task because conformational states associated with this process may be less populated in the unbound receptor ensemble and can be difficult to trap by X-ray crystallography. Therefore, the study of such processes at molecular level is still a challenging task. In this review, we analyze some of the computational tools designed to help with the exploration of the free energy landscape of proteins that one can use to identify biologically relevant conformational states or to locate potential druggable binding sites in different drug targets. In particular, we will focus on how conformational selection and allostery features can be incorporated in the structure-based drug design process. We address all of these methodologies from the computer-aided drug design perspective with special focus on their applications. To this end, we selected some examples that illustrate not only the potential, but also the current limitations and challenges, of computational methods, these examples include a number of GPCRs and some highly flexible antibacterial drug targets involved in the isoprenoid biosynthesis.

2. Introduction to receptor flexibility

In parallel to the extensive debate on biomolecular recognition mechanisms, the fast progress of experimental and computational techniques has led to a better understanding of biomolecular interactions and ligand binding events, providing better tools to interpret the ligand recognition process [2,3,37–39]. The picture of a protein changed from a rigid and inflexible structure to an intrinsically dynamic and flexible body that displays a wide spectrum of motions. Those motions take place on a broad range of time scales that span from ultrafast bond vibrations occurring on the femtosecond time scale to large conformational changes that require milliseconds to even seconds to be completed. Flexibility has been shown to be a concept inherent to proteins that gives them the ability to adopt multiple conformations by generating what is known as a conformational ensemble. This plasticity results in continuous changes of the shape of the protein, for example, by creating transient cavities with functional properties or revealing transitions between conformational states that may open or close the gate for the interaction with endogenous or exogenous molecules. Protein flexibility is crucial for biomolecular recognition processes and it is directly linked to protein dynamics. The understanding of the variety of motions and dynamic processes interplaying in the protein ensemble is relevant in rational drug design [17,40,41]. To this end, it is valuable to find ways not only of analyzing protein motions and protein responses upon binding, but also of accounting for the inherent receptor flexibility when assessing the binding affinity between a potential therapeutic drug and its target.

In the last few years, the improvement of experimental techniques triggered a large number of advances in the field of protein dynamics. Techniques such as nuclear magnetic resonance (NMR) have been proven to provide invaluable information on the understanding of protein motions and the generation of conformational ensembles [38,42,43]. In terms of biomolecular recognition, NMR data allows the visualization of heterogeneous protein ensembles where bound and unbound receptor conformations are represented. These observations are in line with the conformational selection view of biomolecular recognition [44]. On the other hand, advances in specialized computer hardware and software have brought computational methods to a status where they can provide answers at the atomic level to diverse phenomena such as protein folding [45] or biomolecular recognition [46], as well as play a relevant role in the structure-based drug design process [41,47]. Particularly in the last decade, molecular dynamics (MD) simulations have undergone a step forward because of the increase in computational power translating to longer and more accurate simulations going beyond the microsecond time scale [37,48–50]. Similarly, the advances in enhanced sampling techniques allow us to capture slow

conformational changes that remain hidden in conventional molecular dynamics simulations [18,51,52]. The combination of experimental techniques, such as NMR, with molecular simulations has represented a step forward to comprehend how proteins move and interact with their partners, providing relevant information towards a better understanding of mechanistic details in biomolecular recognition [32,53–55].

Intra- and inter-molecular interactions interplaying among ligand, receptor, and water molecules are the driving force of protein dynamics and recognition processes. To understand how molecules interact and, thus, their affinity, it is of great interest to improve the efficiency and accuracy of rational drug design. Affinity is strongly related to the concept of free energy, a quantity that measures the favorability of one state over another, for example, between a ligand bound to its target compared to the unbound situation. In biomolecular recognition, orchestrated enthalpy and entropy changes determine the favorability of the binding event. The ability to predict such a property is of great interest for drug discovery and, consequently, several methods with different levels of efficiency and accuracy have been proposed to estimate binding free energies. To correctly assess binding affinities, it is convenient to know the three dimensional orientation of ligands interacting with their receptors. Molecular docking and scoring functions have been proven useful in assigning rankings and scores to different poses that can be used to predict relative ligand orientations with respect to a receptor. Docking techniques are fast and efficient, presenting a wide range of applications in the early stages of virtual screening when large libraries of compounds are explored [56–58]. Docking techniques are fast but the large number of approximations taken into account limits their applicability beyond pose prediction and very rough ranking of compounds [59]. Despite some well-known limitations such as system dependency, docking methods and scoring functions are key techniques in hit identification. When more precise binding affinities are needed, alchemical free energy methods represent a more robust and accurate way to compute binding affinities [60]. The continuous improvements made in the last years are leading towards a greater applicability of free energy methods in the lead-optimization stage of the rational drug design process [61–65].

It is crucial to identify the key conformations the drug target before starting with structure-based drug design efforts because the inaccurate description of the binding site region directly affects the correct estimation of binding affinities. Unbound receptors exist in an equilibrium conformational ensemble characterized by a set of conformational states and populations. During the binding event, either to a catalytic or an allosteric site, the relative distribution of states changes while the set of conformational states remains the same. Thus, it may be possible to design drugs that stabilize specific conformations of the ensemble by targeting these sites and trapping the active or inactive states of the protein. To this end, methods that efficiently explore the conformational space and techniques that provide accurate and fast calculation of binding affinities are crucial to improve the predictive power of virtual screening protocols (see Fig. 2). In the last part of this review, we focus our attention on the estimation of binding energies with special emphasis on the incorporation of receptor flexibility to account for different conformational states in the drug design process.

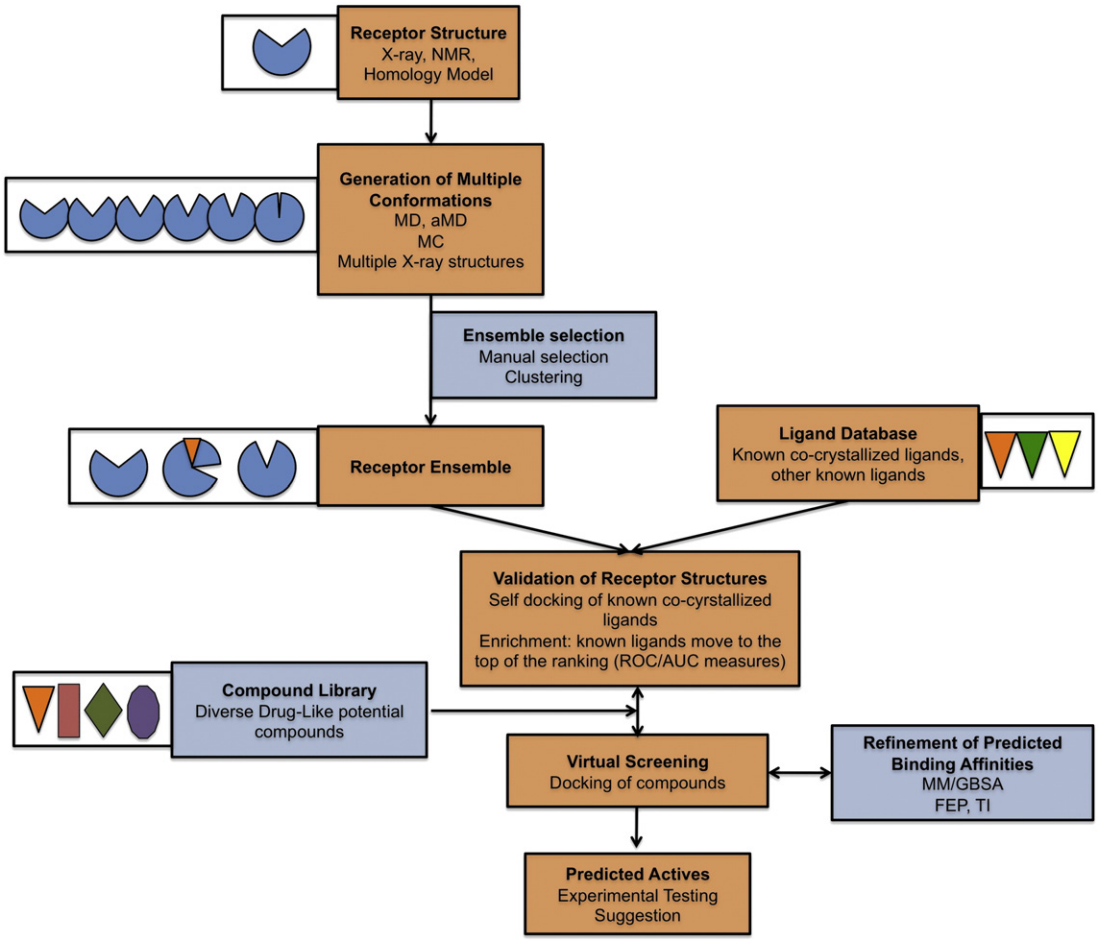


Fig. 2. Accounting for receptor flexibility in structure-based virtual screening. Examples of methods used at each step that are discussed in the present review.

3. Receptor ensemble-based screening methods

Structure-based screening methods require an initial receptor structure of a drug target, either obtained experimentally or through molecular modeling, to start with the rational drug design efforts (see Fig. 2). Usually these receptor structures correspond to crystallographic and NMR structures or can be generated from computer modeling, molecular dynamics simulations, or from enhanced sampling methods. The use of high-resolution crystal structures (if available) has long been the established approach to rationally design small drug molecules. In the framework of conformational selection, ligands act to selectively stabilize specific protein conformations and, thus, proteins can be co-crystallized in alternative conformations depending on the nature of the ligand. The use of only one receptor conformation limits the chemical space of potential ligands for a specific drug target. To improve the predictive power of receptor-based methods, it is useful to generate an ensemble of receptor structures where the most relevant conformations of the receptor are taken into account in the structure-based drug design process.

Ensemble-based screening methods aim to account for receptor flexibility and are based on using several receptor structures in the docking phase of the virtual screening protocol [66]. Ensemble-based methods represent an indirect way of accounting for conformational selection in structure-based drug design and have been widely used to improve binding pose prediction and enrichment factors in virtual screening [67–69]. To illustrate the success of ensemble-based screening methods in drug lead identification, we selected undecaprenyl diphosphate synthase enzyme (UPPS) as an example. This enzyme involved in isoprenoid biosynthesis has been the focus of several studies and a large number of crystal structures co-crystallized with ligands of different natures have been reported in the last decade [70–72]. In 2013, Zhu and coworkers reported the discovery and design of a large set of new chemically diverse inhibitors for UPPS [70]. UPPS is an essential enzyme for the biosynthesis of the bacterial cell wall in most bacteria, such as *Staphylococcus aureus* or *Escherichia coli*, and it has been shown to be an interesting antibacterial drug target [73,74]. According to the available set of crystal structures, UPPS is a reasonably flexible enzyme that can be found in three different conformations (closed, ajar, and open), depending on the nature of the substrate or the ligand bound to the different binding sites of the enzyme. In this case, the variety of X-ray crystal structures available is sufficient to build a representative receptor ensemble that can be used for structure-based drug design efforts. The receiver operating characteristic/area under the curve (ROC/AUC) approach can be used to validate the predictive power of each structure of the ensemble for the virtual screening. In the ROC/AUC procedure a set of known active compounds and presumed inactives are docked into each structure [75]. The AUC is associated with how well a ranking algorithm will rank and separate actives from inactives, and it is used to assess the performance that different conformations of the same receptor may have in virtual screening. Prediction and performance of receptor ensembles in virtual screening have been discussed at length in the literature [41]. Using a 112-compound screening dataset for UPPS, the best enrichment was observed for UPPS crystal structures that belong to open and ajar states, with AUC values close to 0.8. Then, the best predictive structures were selected and used as receptor structures to computationally screen large databases of compounds. Some of the computationally predicted compounds using ensemble-based docking methods led to the discovery of UPPS inhibitors. The reported antibacterial drug leads show therapeutic activity in animal models and have also been shown to restore the sensitivity of antibiotics such as methicillin, which make them promising leads for further antibiotic development [70]. In many cases, the use of an ensemble of conformations enhances the predictive power of virtual screening. As was shown for UPPS, crystal structure diversity is often enough to generate an ensemble that describes the most relevant receptor conformations to rational design of active compounds. However, an extensive set of crystal structures of different and relevant conformations is only available for a very limited number of proteins.

4. Exploration of the conformational space

Crystal structures that capture pharmacologically relevant binding conformations may not be available or are difficult to obtain and, occasionally, the bound crystal structures available for a drug target do not represent the conformation of interest. In particular, conformations associated with important states may be transient and, thus, trapping these particular conformations with experimental techniques can be a tedious task. For instance, allosteric sites are particularly difficult to capture in crystal structures due to their less conserved character with respect to catalytic sites. Protein function is only superficially understood from a single structure because proteins are inherently dynamic and display a wide range of motions that span from simple side chain rotations to accommodate a substrate in the catalytic site to large backbone rearrangements that may even alter the secondary structure. Molecular simulations have been proven as a useful tool to explore the conformational space of proteins and can overcome the lack of receptor structures by generating new alternative conformations [47,76]. In addition, molecular simulations can sample conformational states that could be important to characterize allosteric sites that are not evident from the crystal structures available [77]. Protein motions directly affect the association between the ligand and the receptor but a single structure does not tell much about the intricate motions of protein dynamics. Molecular dynamics (MD) simulations are among the most widely used methods to study protein flexibility from the computational perspective [78,79]. Since the first MD simulation of a protein performed more than thirty years ago [80], MD has been used in a wide range of applications in the field of biomolecular recognition, however, short time-scale simulation are often not capable of capturing important conformational changes. In the last years, the significant increase in computational power has broadened the applicability of all-atom molecular dynamics; longer and longer simulations were produced and this had significant implications for the interpretation of biomolecular recognition at the molecular level. Besides special purpose hardware [49], graphics-processing-units (GPUs) have been used to speed up molecular dynamics simulations by an order of magnitude compared to the central-processing-units (CPUs) [48]. Particularly interesting was the interplay between individual users and researchers in the folding-project that resulted in the reconstruction of the free-energy surface of a protein folding event by means of Markov state models and the theory of exponential kinetics [81]. In a recent example, NMR techniques and 550 μs of all-atom MD simulations were used in conjunction to characterize the dynamic activation process of the β_2-adregeneric receptor, identifying conformational states that were not observed in the crystal structures available [32]. The role of different ligands on the stabilization of selected conformational states was also explored showing that this GPCR is highly dynamic and adopts a large number of different conformations. The understanding of GPCR dynamics and how different ligands trigger the association with different signaling proteins paved the way towards structure-based drug design efforts on these particularly interesting receptors. In addition, the analysis of binding modes and intermolecular interactions observed during the binding event may lead to the design of new allosteric modulators that will be able to modulate the activity of GPCRs. However, all methods used for molecular simulations have their limitations and it is important to be conscious of the advantages and drawbacks of each technique [82]. Some examples of the limitations that have been associated with molecular dynamics simulations could be: 1) instabilities associated with force fields in simulations that exceed the microsecond time-scale; or 2) poor description of quantum effects that, for example, are particularly important when transition metals take part in ligand binding, among others. Protonation states of certain residues are also a key property to consider in MD simulations and binding affinity

calculations. A change of the protonation state can lead to an overestimation or underestimation of binding affinities, decreasing the success rate of structure-based screening methods. Running simulations with different protonation states or constant-pH simulations are tools that can help to ameliorate these limitations [83,84].

An interesting question is how the use of MD structures affects the quality of the structure-based virtual screening procedure. In a recent example, Sinko and coworkers studied the influence of protein flexibility on the design of UPPS inhibitors [85]. To this end, they ran long MD simulations and analyzed the performance of high- and low-populated conformations on the docking of known inhibitors. UPPS is a highly flexible enzyme as shown by the variety of different crystallized conformations. The conformational changes displayed along the unbound MD trajectory were similar to those seen in the ensemble of bound and unbound crystal structures, in line with the conformational selection idea of biomolecular recognition. In this case, MD simulations are capable of sampling some of the most relevant conformational states of UPPS. Particularly interesting was the identification of a rarely sampled conformational state with an expanded pocket that is significantly important to properly describe ligand binding in UPPS. The results obtained from virtual screening suggested that different classes of known inhibitors recognize different active conformational states of UPPS. Only when this rarely sampled conformation with an expanded pocket was used as receptor conformation in the virtual screening procedure, the poses obtained with docking methods mimicked those observed in the open bisphosphonate-bound UPPS crystal structures. This is in contrast to other inhibitor chemotypes, which require a less expanded active site conformation. Consequently, it is relevant for drug design to identify the conformational states where a specific inhibitor binds. Interestingly, apo MD simulations were able to capture the most relevant conformational changes that would be involved in the accommodation of the ligand in the biomolecular recognition process of UPPS.

The ROC and AUC analysis can be also used to quantitatively assess the performance of MD structures in receptor-based virtual screening. Nichols et al. used this strategy to assess the performance of MD structures in HIV reverse transcriptase (RT-HIV), another popular disease target with multiple experimentally determined crystal structures available [86]. ROC/AUC analysis can be used to accurately predict the level of enrichment of a virtual screening run by evaluating the predictive power of different conformations of the same receptor. In this case, a total of 200 ns of MD simulations for two bound and two unbound RT-HIV receptors were used to generate the conformational ensemble for the virtual screening. The attention was focused on the NNRTI binding pocket that has been shown to be highly flexible, changing from a "collapsed" inhibitor-free state to an "open" inhibitor-bound state. The results obtained from the virtual screening of the NNRTI pocket were compared with an ensemble of 15 experimentally determined structures that contain both unbound and bound structures. First, they found that bound receptors improve virtual screening results compared to unbound structures. Second, ROC/AUC results showed that the performance of nearly 20% of the MD structures studied was superior to the available crystal structures.

MD trajectories can be used to interpret and identify conformational changes that not only play a critical role in the biomolecular recognition process, but also are a useful tool to improve the predictive power of virtual screening by generating new structures that broaden the conformational ensemble. The increasing recognition of the importance of target flexibility culminated in the definition of the relaxed complex scheme (RCS), which is an ensemble-based docking method that accounts for receptor flexibility to perform docking studies of compound libraries [87,88]. RCS relies on the use of previously determined conformations with molecular dynamics simulations that are used as receptor structures to screen chemical compounds with docking techniques. The idea behind RCS is to enrich the variety of low-energy conformations present in the ensemble in order to increase the diversity of ligands that bind to a receptor and, ideally, identify a larger number of hits obtained from compound libraries. RCS has been successfully applied to find compounds for several targets. For example, Schames and coworkers identified a novel binding cavity in HIV integrase using RCS in conjunction with docking [89], which helped to inspire the discovery of FDA-approved drug raltegravir [90]. More recently, Wassman et al. observed by means of MD simulations a transiently open binding pocket in tumor suppressor p53 [91]. Applying the RSC virtual screening procedure on this novel site, they identified a compound that is potentially able to reactivate mutated forms of p53 in human cells. These examples highlight the importance of MD simulations and ensemble-based screening methods on the identification of new druggable pockets and the design of potential active compounds.

However, care must be taken when using MD structures for virtual screening because we do not know a priori if a specific MD structure will improve the estimation of binding affinities [86]. In general, enrichment may be better for ensembles of crystal structures than for ensemble of MD simulations structure [92], however, some MD structures can enhance the prediction power compared to experimental structures. Occasionally, MD trajectories are not long enough to identify relevant conformational transitions that may lead to low-energy configurations of interest for drug design. The time-scales reached using MD simulations are typically the order of nanoseconds to even sometimes microseconds. However, many interesting processes take place on the timescales of milliseconds to seconds, which may reveal new binding sites important for structure-based drug design. These binding sites would be missed by conventional MD simulations. Thus, new strategies are needed to overcome such high-energy barriers associated with slow motions that connect low-energy states. Methods for identifying the most predictive structures and methods for sampling a greater part of biomolecular phase space would be useful for structure-based drug design and they will be the focus of the next sections.

5. Enhanced sampling methods

Some important processes such as biomolecular recognition, allosteric regulation, or signal transduction, usually take place on the micro- to millisecond or even longer time scales. Low-energy states relevant for these processes may be separated by high-energy barriers, which are rarely crossed over the course of conventional MD simulations, unless the simulation is really long. Such conformational changes associated with slow motions, may play a critical role in biomolecular recognition and their description is of capital importance to identify relevant conformations for rational drug design. Moreover, if one wishes to perform accurate free energy calculations by recovering the Boltzmann ensemble of structures, the crossing of high-energy barriers should be observed multiple times to obtain converged statistics. In the direction of improving the exploration of the conformational space, new strategies have been proposed to overcome the, sometimes, scarce conformational sampling associated with standard molecular dynamics simulations, and also to speed up the crossing of high energy barriers. Besides the aforementioned specialized computer hardware improvements and the increasing popularity of multi-scale techniques [93], a lot of attention has been paid to simulation techniques that speed up and improve the efficiency of conformational sampling while keeping the atomistic description of the system. These methods can be encompassed in the group of enhanced sampling techniques.

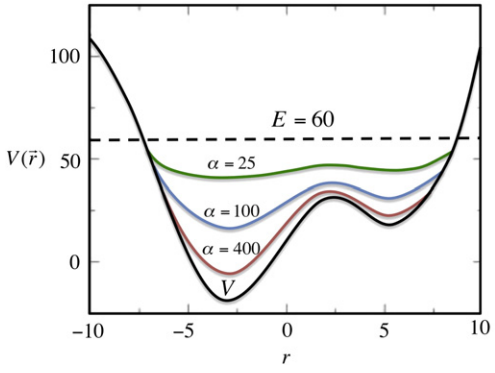
The basis of speeding up conformational sampling is the introduction of an artificial bias into the model upon which the simulations are based. These methods go from the simple raising of the temperature of the system to methods that display different levels of sophistication. Temperature accelerated replica exchange [94], umbrella sampling [95–97], metadynamics [98,99], or accelerated molecular dynamics [100,101] are among the most widely used methods to enhance conformational sampling in all-atom simulations. Some of these methods require an a priori definition of a reaction coordinate: either a transition

pathway between known initial and final states or a set of collective variables (CVs) are defined a priori to drive the course of the simulation. The calculation of free energy differences between two states connected by a reaction coordinate requires adequate sampling of both low- and high-energy regions found along the reaction path. The umbrella sampling method introduces a bias potential to facilitate the transition over energy barriers, and is an efficient technique to sample high-energy regions. To this end, separate simulations, which overlap, are run in a series of windows along the reaction path to connect the initial and final states. It is important to ensure that the sampling along the reaction coordinate is as uniform as possible. The main difference between umbrella sampling and metadynamics is that in the latter a non-systematic sampling along a set of collective variables is performed. For this, a history-dependent bias potential is introduced to the Hamiltonian. This bias potential enhances sampling by adding Gaussian contributions to the potential energy along the sampled trajectory to prevent the system from visiting regions that have already been sampled. Then, the free-energy surface of the process can be accurately reconstructed as a function of the chosen set of collective variables. In a recent example, that comprehensively shows the potential of metadynamics to study ligand binding events, Limongelli and coworkers described the full unbinding pathway of inhibitor SC-558 of cyclooxygenase-2 and identified an alternative binding-mode with similar thermodynamic stability to the one found in experiments that could help to explain the long occupancy of this inhibitor in the binding site [102]. However, the results obtained by means of metadynamics strongly depend on the set of CVs used for the simulations [99]. These techniques work remarkably well in analyzing free energy changes between known conformations but present some handicaps when looking for unknown conformations, for instance, when a drug-like compound binds to the active site of its protein target but there is no knowledge of the final conformational state.

In this review, we focus our attention on accelerated molecular dynamics (aMD), an enhanced sampling technique that does not rely on the a priori definition of reaction coordinates. Extensive reviews on the theory underlying aMD simulations, its distinct flavors, and a wide range of applications can be found elsewhere in the literature [101,103]. Here we focus on the potential of aMD as a tool to efficiently explore the rough free energy landscape of proteins and its direct contribution to structure-based drug design by providing new structures that may reveal new binding and allosteric sites relevant for biomolecular recognition.

Accelerated molecular dynamics enhances sampling through modification of the system's Hamiltonian in a relatively simple way (only two to four parameters are required). In addition, it does not rely on the definition of a reaction coordinate or a set of collective variables (a priori knowledge of the underlying free energy landscape is not needed), and it conserves the essential details of the free-energy landscape. AMD typically modifies the underlying potential energy surface by applying a boost potential $\Delta V\left(\vec{r}\right)$ at each point of the MD trajectory according to the equation in Fig. 3 [100]. The value of $\Delta V\left(\vec{r}\right)$ depends on the difference between a reference, 'threshold', or 'boost' energy, E, and the actual potential energy. Based on initial conventional MD simulations, the reference energy is chosen to lie above the minimum of the potential energy surface (PES). All states with energy above E are not modified in the standard aMD method. The larger the difference, the greater the modification of the PES becomes, pushing up low-energy valleys and in effect decreasing the magnitude of energy barriers.

Acting as an MD simulation catalyst, aMD speeds sampling by lowering the size of energy barriers and smoothening the potential energy landscape as a function of parameters E and α, which regulate the level of acceleration, and their optimal values are system specific. Since the system moves in a smoother potential energy surface, high-energy barriers can be more easily conquered, making possible multiple transitions over these previously impassable barriers and, then, unexplored low-energy states are visited multiple times along the aMD trajectory. aMD is a flexible technique that allows different variants and extensions [101]. The most popular is the dual boost approach, which combines two different levels of acceleration, a more aggressive one applied to only the torsional angles of the system, and a less vigorous acceleration applied to all elements of the force field including explicit solvent, which is to sample diffusive motions in the solvent. The modifications introduced to the Hamiltonian bias the actual potential energy surface; the low-energy conformations become less stable and the populations are altered with respect to the original PES and, consequently, the system moves over a non-Boltzmann energy surface. Since we know the modification of the potential energy introduced at each point, a corresponding reweighting function is used to recover the Boltzmann statistics, and then thermodynamic properties may be accurately determined. However, recovering the canonical ensemble from non-Boltzmann aMD simulations is not a straightforward task, which complicates obtaining accurate free energy statistics When longer timescales are reached with aMD, the reweighting procedure is subject to statistical errors in the estimate of the weighting factor at each point [104]. A sufficiently long simulation is required to observe a proper reweighting of the Boltzmann canonical ensemble, which may limit the applicability and efficiency of free energy aMD-based calculations for certain large bio-molecular systems. Several alternative solutions have been proposed to improve reweighting process. In the case of drug design, selective aMD is a particularly interesting technique, which limits acceleration to a few dihedrals angles [105]. This reduces the amount of statistical error and the free-energy statistics can be properly recovered using the re-weighting factor.



$$V^*(\vec{r}) = \begin{cases} V(\vec{r}), & V(\vec{r}) \geq E \\ V(\vec{r}) + \Delta V(\vec{r}), & V(\vec{r}) < E \end{cases}$$

$$\Delta V(\vec{r}) = \frac{(E - V(\vec{r}))^2}{\alpha + (E - V(\vec{r}))}$$

Fig. 3. Accelerated molecular dynamics. Equations to calculate boost energy, $V^*(\vec{r})$, and boost potential, $\Delta V\left(\vec{r}\right)$. The true potential energy function is shown as a solid black line, $\Delta V\left(\vec{r}\right)$. A series of modified potential energy functions are represented in different colors for various values of α as shown in the plot while E was always fixed at 60 (black dashed line).

aMD has been applied to several systems and has also been validated with experiments. Recently, Gasper and coworkers found a good correlation between experimental NMR order parameters and order parameters which were computationally predicted with aMD simulations. There was a significant improvement over standard MD predictions [106]. They made use of aMD simulations to describe the correlation of the structural fluctuations of thrombin in great detail. In addition, two allosteric pathways that mediate the activity of thrombin were identified [29]. From the perspective of drug design, GPCRs are at the epicenter of experimental and computational efforts. It is of great importance to be able to understand the activation mechanism of such complex drug targets prior to starting with structure-based drug design efforts, but most of these mechanisms remain unclear due to the lack of experimental structures. AMD is a powerful tool to explore conformational space and can be used to elucidate the activation pathways of these important receptors that may take on the order of milliseconds to be completed. In a recent example, Miao et al. focused their attention on the activation and dynamics of M2 muscarinic receptor, which regulates heart rate and contractile forces of cardiomyocytes [107]. The crystal structure of M2 receptor was recently determined in the inactive antagonist bound state. Interestingly, aMD was capable of capturing the activation process that was not observed in previous microsecond time-scale conventional MD simulations. It was found that the activation takes place through a series of events that trigger the formation of a Tyr206–Tyr440 hydrogen bond and relocation of alpha helix 6. Moreover, these results identified a dynamic network for allosteric regulation of M2 receptor that may open the way towards structure-based design of allosteric drugs [108]. AMD can also be used in combinations with other enhanced sampling techniques such as the adaptive biasing force method to determine the energetics of conformational changes upon the biomolecular recognition process. Recently, Wereszczynski and McCammon probed the conformational space of Get3 protein, by means of aMD simulations and analyzed conformational changes undergone in the presence of various bound nucleotides [109]. The calculation of an accurate potential of the mean force was used to compute the free-energy landscape of the Get3 opening/closing pathway. In addition, it was found in the apo aMD simulation that a semi-open conformation might be sampled when Get3 is free in solution, as well as play a crucial role on nucleotide recognition and be important for drug discovery.

Improvements in hardware, algorithms and methodological developments can be combined to perform high-throughput simulations to access the millisecond time scale. For instance, Pierce et al. amalgamated aMD simulations and GPUs to routinely study millisecond events from a desktop computer [110]. In particular, calculations were done on the bovine pancreatic trypsin inhibitor (BPTI) to show that 500 ns of aMD simulation sample the same conformational space as a previously performed millisecond conventional MD simulation on the same protein. In another recent example, Buch and coworkers reconstructed the enzyme-inhibitor binding process of the complex trypsin–benzamidine, including the description of its diffusion pathway, surface exploration and final binding [50]. The binding paths obtained from almost 500 molecular dynamic simulations of 100 ns length were used to reconstruct the kinetic pathway of the inhibitor benzamidine from the solvent to the bound state passing through two different metastable states. To this end, GPUs, the ACEMD software, and Markov state models were used to describe the drug binding pathway to the drug target. The absolute binding free energy of the process was also estimated and showed a good agreement with the experimental value. Since enhanced sampling techniques probe a vast variety of protein conformational states, it is of great importance to make use of methods that are able to extract the most relevant information from these molecular simulations.

6. Extraction of the most relevant protein motions

Protein dynamics is a key concept in conformational selection theory where the ligand selectively stabilizes certain conformational states that preexist in the unbound ensemble of protein conformations. In other words, a population shift towards the conformations stabilized by the ligand is observed upon the ligand binding event [2]. How can we evaluate this population shift that takes place in biomolecular recognition from a conformational ensemble that could be either generated by a MD trajectory or from a set of experimentally determined structures? MD trajectories contain a large amount of information that may obscure conformational changes relevant for function and rational drug design. Several methods were proposed with the aim to reduce the amount of MD trajectory data and analyze collective motions in proteins [111–113]. The concept of “essential dynamics” introduced by Berendsen and coworkers symbolized one of the first popular ways of extracting the most relevant motions from molecular dynamics simulations [114]. In essential dynamics, the conformational space is divided into different subspaces, the “essential” subspace that contain the relevant collective degrees of freedom or principal components of protein motion, and the remaining space. In general, principal component analysis (PCA) is a tool to extract the most important motions of a protein conformational ensemble and it is frequently used to describe important conformational phenomena [115,116]. In the case of molecular dynamics simulations, the complex motions associated with an MD trajectory are split into just a few variables giving an idea of the regions of the conformational space sampled during the MD simulation. Thus, it is a useful tool to study conformational selection by assessing a shift on the equilibrium distributions of ligand bound and unbound receptors. Sinko and coworkers made use of PCA to study the population shift mechanism in changing the equilibrium towards other conformations upon inhibitor binding in the UPPS enzyme. As we have mentioned above, a considerable expansion of the active pocket size is observed in UPPS upon binding of bisphosphonates inhibitors, which stabilize open structures that are only occasionally sampled conformations in the apo-simulation [85]. Thus, a shift in the populations towards a markedly different conformation was observed in the bisphosphonate bound UPPS structures. PC analysis shows how after the binding event the populations are shifted from the center of PC plot corresponding to the apo structure towards other regions of the conformational space (see Fig. 5 in reference [85]). Principal component analysis of the MD trajectory showed that inhibitors recognize different sets of conformations, which can vary significantly between families of ligands.

The principal component space can also be built from a set of NMR or crystal structures in order to discern between different conformational states associated with different substrates or inhibitors bound to the receptor. Fig. 4 shows the PCA of some of the *E. Coli* UPPS crystal structures available [70]. The analysis of the principal components clearly separates the three binding conformations of UPPS: substrates are bound to closed structures, non-bisphosphonate inhibitors belong to the group of slightly open conformations, and bisphosphonates are part of the widely open conformational states. It can be seen that the apo-crystal structure also belongs to the ‘ajar’ group of crystal structures, which suggests that non-bisphosphonate inhibitors require less energetic costs to accommodate the ligand, because the equilibrium distribution is not strongly shifted from the apo structure. Open and ajar crystal structures were used as UPPS receptor structures in virtual screening that led to the discovery of new inhibitors with potent activity in cells and animal models [70].

The principal component space built from the available crystal or NMR structures can be used in conjunction with molecular dynamics simulations to assess the conformational space explored during the MD trajectory and to compare the population shift of ligand-bound trajectories with respect to experimentally obtained structures. The extraction of the most relevant motions may also reveal new binding or allosteric sites. Multidimensional nuclear magnetic resonance and other advanced experimental techniques have been used to demonstrate the predominance of the conformational selection mechanism in a wide range of proteins [2]. From the computational perspective, the role of conformational selection in the formation of ligand–receptor

complexes can be studied through molecular dynamics simulations by evaluating how the equilibrium distribution is shifted towards different distributions in the ligand-bound receptor simulations compared to the unbound receptor trajectory. Grant and coworkers used this methodology to study conformational selection in G-proteins [117]. In particular, they focused their attention on the activation mechanism of Ras and Rho G-proteins through GTP binding that was traditionally described in terms of induced fit theory. To this end, they performed a series of conventional and accelerated MD simulations of free Ras and Rho proteins. Interestingly, free-nucleotide aMD simulations sampled multiple conformations, including regions populated by GTP and GDP crystal structures. These results show the ability of unbound G-proteins to interconvert between different conformations in the absence of a ligand. On the other hand, the conformational space explored by GTP and GDP aMD simulations was found to be more restricted to the region of available nucleotide-bound crystal structures. The influence of correlated motions in the aMD simulations was also studied by means of correlation diagrams that provide information about significant correlated motions between residues. These results predicted the coupling between nucleotide-site and the C-terminus via highly flexible Loop 3, suggesting that Loop 3 may represent a potential allosteric site present in G-proteins. PC analysis helped to classify the ensemble of structures in terms of the most important motions and to assess the enhanced sampling obtained by aMD. Then, MD trajectories are useful to connect these low-energy regions by describing the transitions between them. This information is useful for structure-based drug design because it can help to group the most relevant structures for drug discovery.

7. Selection of biologically relevant structures for ensemble-based methods

An intrinsically dynamic receptor samples a substantial number of conformations, but only a subset of them are stabilized by the ligand upon binding, producing a shift of the population towards the conformations that favor binding. Consequently, it is of great importance to find ways of extracting these biologically relevant conformations from a molecular dynamics trajectory in order to identify the best set of structures to use in virtual screening. One option is to pick MD snapshots at regular time intervals from the MD trajectory. However, all proteins display different flexibility patterns and binding properties that vary with time and have an impact on the formation of favorable receptor–ligand complexes. The regular extraction of frames may contain redundant information and may not reflect the variety of the ensemble. Clustering techniques have been proven to be useful for generating representative structures for virtual screening [88,118]. Among the variety of clustering techniques, particularly popular is the RMSD-based clustering that groups structures from the MD trajectory based on mutual structure similarity criteria. The idea behind clustering is to avoid the loss of ensemble information that may otherwise be lost in the selection of single snapshots. To identify relevant conformations, subsets of representative coordinates are chosen for the RMSD-based clustering calculation. For example, relevant binding conformations can be captured using the set of residues that embrace the binding site. More general information about the protein motions can be obtained using all alpha carbons of the protein of interest. Other ensemble selection methods have been proposed such as the QR-factorization technique, which relies on the reordering of the redundant data in terms of increasing linear dependence [88,119]. In a recent example, Durrant et al. ran MD simulations of the drug target Uridine Diphosphate Galactose 4′-Epimerase found in *Trypanosoma brucei* and involved in the African sleeping sickness [120]. They successfully applied RMSD-based clustering of the active site that led to the identification of 14 low-micromolar inhibitors with an impressive hit rate of 62%. A total of 24 cluster structures were used in combination with the AutoDock Vina docking scoring function [121] and a population-weighted ensemble-based docking score to rank the compounds obtained from the screening of the National Cancer Institute Diversity Set II.

The same protocol was used by Durrant et al. to rationally design inhibitors for the anticancer and antibacterial drug target farnesyl diphosphate synthase (FPPS) [122]. Bisphosphonates are known to tightly bind the active site of isoprenoid biosynthesis diphosphate synthases such as FPPS and UPPS, two enzymes that work in series in

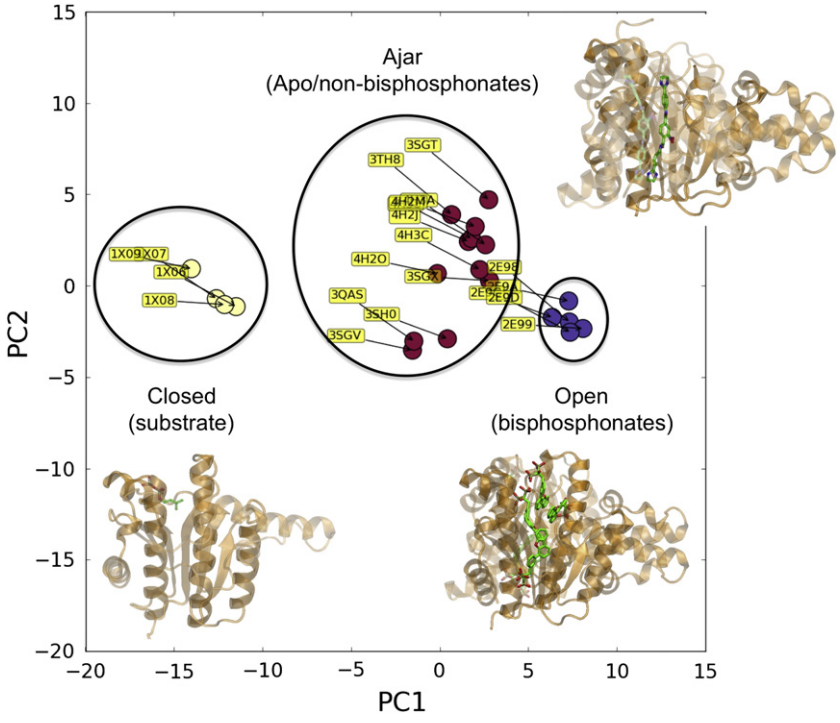


Fig. 4. Principal component analysis. PCA build with *E. coli* UPPS crystal structures. Substrate bound structures in yellow are closed; apo (PDB ID code 3QAS) and non-bisphosphonate inhibitors in red are in slightly open conformation; bisphosphonate inhibitors in blue are crystallized in open conformations.

the same pathway. As bisphosphonates present several undesired features due to their high polarity, it is strongly desirable to find alternative compounds for these interesting drug targets. To achieve this, Durrant and coworkers made use of the relaxed complex scheme, with an MD structure-based ensemble of protein conformations, to identify low-micromolar non-bisphosphonate inhibitors for FPPS. In 2010, several non-bisphosphonate inhibitors that bind to a FPPS allosteric site were identified to block the synthesis of farnesyl diphosphate [123]. In a recent publication, Lindert and coworkers identified, by means of RCS virtual screening protocol, a number of leads for non-bisphosphonate inhibitors that target the allosteric site of FPPS [124]. These compounds are classified as bisamidines and are chemically different from the compounds found to target the active site of FPPS [122]. As we have seen above, the best scoring receptors for some compounds were obtained from MD structures. NMR and X-ray crystallographic structures may not reveal allosteric sites that are less conserved due to protein dynamics. MD simulations in combination with clustering techniques can overcome this limitation, generating conformations where allosteric sites are well defined providing an enrichment in the virtual screening process. Interestingly, some of the compounds found by Durrant et al. [122] and Lindert et al. [124] also bind and inhibit UPPS, opening up the possibility of developing dual FPPS/UPPS inhibitors that target isoprenoid biosynthesis in bacteria, which may help to tackle future problems associated with drug resistance.

8. Mapping of druggable binding sites

One of the first steps in structure-based drug design is to identify where a drug could possibly bind. The most common targeted sites are: active sites, allosteric sites, and protein–protein interaction sites [125,126]. These binding sites present different features, shape and dynamics that sometimes make them difficult to target. A growing number of promising allosteric drugs are showing that structure-based drug design efforts can usefully go beyond the active site region of the protein of interest. Some of these allosteric compounds exhibit higher target specificity or reduced toxicity and, thus, they also open the way for the exploration of new regions of the chemical space [127]. Allosteric effects are often related to protein conformational changes induced by a population shift between conformational states that belong to the conformational ensemble and, thus, allosteric sites may not be evident from the unbound X-ray crystal structures. The transient character associated with allosteric pockets make them difficult to predict when no bound crystal structure is available. Thus, it is important to find methods that account for protein flexibility in order to reveal druggable binding sites on the protein surface that are not evident on the initial structure. Ivetac and McCammon described a protocol to identify druggable binding sites that takes into account the receptor flexibility by means of molecular dynamic simulations [128]. MD simulations are used to sample multiple protein conformations and also for identifying novel structures different from the experimental structure, some of which may expose druggable binding sites not observed in the initial conformation. Once the MD ensemble was generated, the most representative conformations were selected using a RMSD-based clustering method. Then, they performed solvent mapping analysis on each structure obtained from the MD ensemble using the FTMAP algorithm developed by Vadja and coworkers (http://ftmap.bu.edu) [129]. In the FTMAP algorithm, a set of small probe molecules corresponding to drug fragments is docked to the protein surface using empirical scoring functions in order to identify the so-called "consensus sites" that are represented by the accumulation of probes in certain regions of the protein surface and could be associated with potential druggable binding sites. In addition, residues can be ranked according to non-bonded interactions with probe molecules in order to identify the most favorable binding sites.

Ivetac and McCammon used this method to identify potential allosteric binding sites on the protein surface of two well-known GPCRs, β_1 and β_2 adrenergic receptors. The combination of an ensemble of 15 MD receptor structures and surface mapping analysis led to the detection of five potentially druggable allosteric sites on β_1 and β_2. The results were compared with available experimental data to confirm the existence of these druggable pockets. Two sites were found to be solvent-exposed corresponding to the extracellular and intracellular mouths of the GPCR. The extracellular site coincides with a well-known region of allosteric-binding activity, which may block the entrance of ligands to the orthosteric site of GPCR. Then, the other three pockets were found in the lipid–protein interface. Particularly interesting is the site that corresponds to the cholesterol-binding site, which has been found to be important to stabilize distinct states of β_2-AR and may be important from the structure-based drug design perspective. GPCR activity can be regulated through allosteric modulation; the identification of potential allosteric sites on GPCRs without the knowledge of crystal structures opens the way towards the identification of new therapeutic agents for such important drug targets. In another illustrative example, Zhu and coworkers used a similar protocol to identify potential binding sites on their study of undecaprenyl diphosphate synthase (UPPS), a potent antibacterial drug target [70]. The variety of inhibitor-bound crystal structures shows four different binding sites on the UPPS surface. In the case of bisphosphonate structures, the substrate site (site 1 in Fig. 5) is always occupied. Surprisingly, the most potent inhibitors were found to bind in site 4, which is quite far from the flexible region of the substrate site. Interestingly, it is the site where potent non-bisphosphonate inhibitors are mainly bound. All of these sites were also predicted to be druggable by solvent mapping program FTMAP.

9. Accounting for receptor flexibility in the estimation of binding affinities

Once the conformational ensemble, containing the most relevant and predictive structures for virtual screening, is defined for the different binding sites that one wants to target, the focus is shifted to how one can predict reliable binding affinities to rank and predict the most suitable compounds for interacting with each drug target. The level of accuracy of the prediction of binding affinities will depend on the stage of the drug design process. For instance, in the lead-optimization stage where accurate binding affinities are necessary, alchemical free energy methods are one of the most precise choices, while in the early-stages of drug discovery docking methods combined with scoring functions are of great use to rank large libraries of chemical compounds

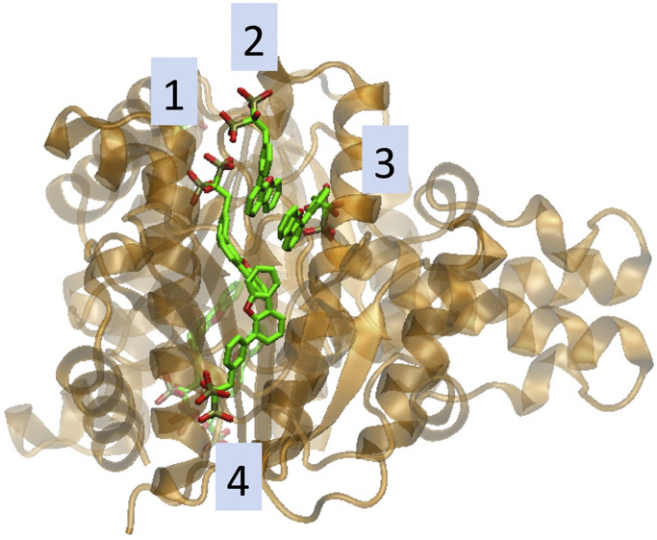


Fig. 5. Binding sites of UPPS. Site 1 (substrate site) and sites 2–4 are binding sites where inhibitors can bind. Bisphosphonates can bind to all sites as shown in PDB ID code 2E98. All of these sites were predicted by FTMAP to be druggable.

and reduce the search space because of their simplicity and speed. All of these methods contribute in different capacities to the identification and design of new candidate compounds. It is of great importance to understand the current limits of applicability and different sources of error in the estimation of binding affinities; several reviews extensively discuss the best practices in virtual screening and free energy calculations with particular focus on rational drug design [60–63,130,131]. In this section, we will briefly discuss the inclusion of receptor flexibility in the estimation of binding affinities using different methods. Protein flexibility and receptor conformational changes upon binding strongly affect the calculation of ligand-binding affinities and, consequently, the predictive power of virtual screening.

Docking methods in combination with scoring functions are used for defining rankings of compounds based on specific binding modes and affinities by performing three-dimensional searches of the best ligand pose and using a wide spectrum of different scoring algorithms [58,121,132]. One of the reasons for the relative speed of docking methods is that often a rigid receptor conformation is used, in a way that the ligand just needs to be accommodated in a fixed structure. However, docking and scoring functions can account for receptor flexibility in different ways [133–135]. As we have seen, an indirect way is to use a receptor ensemble where different conformations of the active site are included [88]. Using techniques such as soft docking, or the softening of van der Waals potentials, that allow for certain overlap between the ligand and the receptor, one can also introduce some flexibility [136,137]. The main drawback of this method is that it may increase the rate of false positives, and the flexibility of the receptor is not fully taken into account. On the other hand, some docking algorithms have been developed to explicitly account for protein flexibility in the estimation of binding modes and affinities. A few of them allow selected side chains to rotate and account partially for the receptor flexibility in the active region [138]. A different approach is introduced by induced fit docking algorithms, where the induced fit rearrangements associated with ligand binding are also considered in the docking procedure [139,140]. However, it is still a challenge to predict large conformational changes that may lead to different binding modes. Rosetta Ligand offers a different perspective on ligand docking by accounting for both receptor and ligand flexibility during the docking stage [141]. Currently, this method allows for full protein backbone and side-chain flexibility [142]. Finally, there are also strategies used to account for conformational selection in ligand docking, where methods used for the prediction of protein–protein interactions are combined with ensemble-based docking methods [19]. The study of docking and scoring functions that account for receptor flexibility is an active field in constant evolution and to improve the performance and transferability of scoring functions among different systems is one of the major challenges of current research. The description of large-scale conformational motions is still challenging and may represent a step forward for docking techniques. In summary, accounting for receptor flexibility in docking protocols often improves the prediction of binding affinities that may lead to the selection of more adequate compounds for experimental validation.

Docking methods have been shown to be extremely useful to sort out a large number of chemical compounds and identify potential hits in the very first stages of structure-based drug design processes, when a large high-throughput virtual screening of large libraries of compounds is performed. Once the initial hits are confirmed by experiments, some of the chemical features of these compounds are finely tuned up to improve binding affinities in the so-called lead-optimization stage. Scoring methods are still quite limited beyond the hit identification phase, but may be used to visualize the binding mode associated with small compound modifications. Thus, less approximate and more robust techniques are needed for further refinement of the prediction of binding affinities for potential drug candidates. An intermediate method between the efficiency of docking and the accuracy of free energy perturbations (FEP) and thermodynamic integration (TI) methods are the molecular mechanics/generalized Born surface area (MM/GBSA) [143] and the molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) [144]. These methods, that try to find the balance between efficiency and accuracy, consist of calculating binding free energies from molecular mechanics force fields and continuum solvent models. In contrast to alchemical free energy calculations, only the states corresponding to the ligand bound and unbound are simulated using MD. To improve the accuracy of the method, binding free energies are averaged through multiple conformations. These techniques fail to achieve convergence when large conformational changes are observed in protein dynamics and then, multiple conformations or MD snapshots are required. However, the calculation of the entropy term becomes remarkably costly when the number of conformations taken into account increases, which prevents its application. Then, this method may suffer from insufficient sampling in some cases due to difficulties in achieving convergence. Some truncation methods have been proposed to reduce the computational cost using a certain cutoff [145]. Other alternatives are focused on just taking into account the protein–ligand MD simulation, reducing computational cost but this does not account for conformational changes in the free and bound receptor [146]. In a recent example Rastelli et al. used a single-energy minimized ligand–receptor complex in MM/GB(PB)SA to successfully estimate binding free energies of a set of inhibitors of *Plasmodium falciparum* DHFR [147]. Both MM/GBSA and MM/PBSA exhibited good correlations with experimental values and with binding affinities obtained after averaging over multiple MD snapshots. The accuracy of MM/PBSA was found to be higher than that of MM/GBSA for the calculation of absolute binding free energies but the performance was similar for relative free energy calculations [148].

In the lead-optimization phase, it is of great importance to understand the relative differences between binding affinities of related ligands to the same drug target. Currently, the most accurate ways of computing relative binding free energies are the FEP and TI techniques [61]. In contrast to endpoint techniques, FEP and TI belong to the group of pathway methods, where the system is transformed from one state to the other by means of alchemical changes introduced to the Hamiltonian in combination with MD or Monte Carlo simulations in explicit solvent water molecules. These methods rely on the definition of a thermodynamic cycle to calculate absolute or relative binding free energies [149]. Alchemical transformations between two states are possible with the fine-tuning of ligand–receptor interactions. Free energy differences can be calculated between the bound and unbound states of the ligand and its receptor, however, alternate cycles can be defined to account for free energy of solvation or relative binding free energy between two ligands. Using FEP, Bollini and coworkers were able to optimize a number of docking hits to obtain potent anti-HIV inhibitors with EC_{50} values in the range of 55 to 320 pM in human T-cells [150]. Further chemical modifications of these inhibitors proposed by means of FEP calculations have been shown to improve the solubility and activity against other HIV variants with respect to previously approved FDA drugs. However, alchemical free energy calculations are associated with an elevated computational cost because they require sufficient sampling to obtain suitable overlap between the phase space of the successive states of the reaction coordinate that connects the initial and final states. Consequently, to achieve convergence of free energy calculations, it is of great importance to sufficiently sample the conformational changes that the system undergoes under alchemical conditions, as well as converge the energy of those conformations, which can require large amounts of sampling. Since the value of the binding free energy relies on sampling, all simulations will be different, leading to different values. This problem can be partially offset by running multiple independent simulations to get better statistics and an estimation of the error. For instance, methods such as the independent-trajectory thermodynamic integration (IT-TI) that take into account multiple independent simulations have been used to calculate accurate relative binding free energies for some inhibitors of the H5N1 avian influenza virus neuramidase [151]. Alchemical free

energy methods are accurate but time consuming and the analysis of several ligands that bind to the same receptor still requires large amounts of computational time. Additionally, the best accuracy is often achieved within a congeneric series of compounds. Enhanced sampling methods can be used to address the sampling limitation associated with free energy calculations by exploring the free energy landscape more effectively, as discussed below.

Biomolecular recognition is an intricate process that takes place in a series of orchestrated ligand and receptor motions and conformational changes. In this conformational dance, the ligand acts to selectively pick some preexistent receptor conformation, which leads to the stabilization of certain conformations. Consequently, it is crucial to identify the conformations that will improve the estimation of binding affinities. As we have seen above, enhanced sampling techniques perform remarkably well to rapidly identify low energy configurations. If these methods are combined with existent alchemical free-energy methods, the convergence of these calculations may be more quickly achieved, decreasing computational cost as well as making free energy methods more applicable for the rational drug design process [52]. Accelerated molecular dynamics have been applied in a number of different ways to free-energy calculations. To overcome the issues related to reweighting and improve free-energy convergence, Oliveira and coworkers proposed the upside down aMD method [152]. This method makes the energy barriers more accessible by lowering high-energy barriers and keeping low-energy configurations unchanged. This technique improves population statistics of low energy minima, while accelerating the transition between energy barriers and facilitates the exploration of the conformational space. However, when the system of interest has a large number of degrees of freedom and a complex free energy landscape the method is difficult to parameterize. To extend the application of the upside down approach to biomolecules, Sinko et al. proposed what they called windowed aMD [153]. This method requires more parameters than original aMD, but has been shown to achieve a rapid convergence of free energy calculations. However, the improvement in reweighting efficiency stemmed from the frequent transitions between the normal potential energy surface and the modified potential energy surface. These transitions can be harder to parameterize with increasing system complexity, but windowed aMD was successfully used to calculate binding free energies between the antibiotic vancomycin and two small glycopeptide-binding partners. Finally, one of the main sources of error associated with reweighting is the level of acceleration and the size of the system. Selectively applied aMD restrains the acceleration to only a portion of the system, more precisely, to a set of predefined dihedral angles in order to overcome the issues associated with reweighting [105]. Free energy calculations on the decoupling of oseltamivir's binding to neuromidase were performed for a set of twenty dihedrals. However, while this technique provides a better statistical recovery it requires the manual selection of dihedral angles that are important for the biomolecular recognition process.

10. Conclusions

Over the last decades, several biomolecular recognition mechanisms have been proposed that try to explain how ligand binding occurs and how such a binding event triggers a set of responses in the receptor. Proteins are inherently flexible bodies displaying a wide range of motions that span from local side-chain rotations to global conformational changes. In the course of this permanent motion, proteins are capable of adopting multiple conformations generating an ensemble of structures that may accommodate a wide variety of ligands. In the framework of the conformational selection mechanism, during the process of biomolecular recognition some of these conformations are selectively stabilized when the ligand binds to a specific binding site of the receptor and, thus, such conformations are particularly relevant for structure-based drug design. Ensemble-based screening methods aim to account for receptor flexibility, helping to improve the predictive power of receptor-based drug discovery. However, an extensive set of crystal structures of different and relevant conformations of the bound and unbound receptor is only available for a very limited number of proteins. The computational techniques that we have described in this review offer a way to extensively explore the conformational space of proteins and can help to identify biologically and pharmacologically relevant states that are difficult to trap with experimental techniques. The allowance for receptor flexibility improves the accuracy of algorithms used to estimate binding affinities between a potential therapeutic drug and its target. Methods that efficiently explore the conformational space and techniques that provide accurate and fast calculation of binding affinities are important to improve the predictive power of virtual screening protocols. In addition, accounting for receptor flexibility also can aid in the identification of cryptic binding sites that may remain hidden in the unbound receptor structures. This is of particular interest for recognizing potential allosteric sites. Allosteric transitions open a new broad range of possibilities in the field of drug design. Thus, the identification of allosteric sites is of paramount importance towards the discovery of new drugs that can target such relevant allosteric pathways.

Acknowledgments

The authors would like to thank the members of the McCammon group for fruitful discussions. This work was supported by the National Institutes of Health, the National Science Foundation, the Howard Hughes Medical Institute, the National Biomedical Computation Resource, and the NSF Supercomputer Centers. F. F. acknowledges financial support of the Beatriu de Pinós program from AGAUR for the postdoctoral grant (2010 BP_A 00339). S. L. was supported by the American Heart Association and the Center for Theoretical Biological Physics.

References

[1] J.P. Changeux, S. Edelstein, Conformational selection or induced fit? 50 years of debate resolved, F1000 Biol. Rep. 3 (2011).
[2] D.D. Boehr, R. Nussinov, P.E. Wright, The role of dynamic conformational ensembles in biomolecular recognition, Nat. Chem. Biol. 5 (2009) 789–796.
[3] P. Csermely, R. Palotai, R. Nussinov, Induced fit, conformational selection and independent dynamic segments: an extended view of binding events, Trends Biochem. Sci. 35 (2010) 539–546.
[4] D. Koshland, Application of a theory of enzyme specificity to protein synthesis, Proc. Natl. Acad. Sci. U. S. A. 44 (1958) 98.
[5] J. Monod, J. Wyman, J.P. Changeux, On the nature of allosteric transitions: a plausible model, J. Mol. Biol. 12 (1965) 88–118.
[6] H.R. Bosshard, Molecular recognition by induced fit: how fit is the concept? Physiology 16 (2001) 171–173.
[7] B. Ma, S. Kumar, C.-J. Tsai, R. Nussinov, Folding funnels and binding mechanisms, Protein Eng. 12 (1999) 713–720.
[8] C.-J. Tsai, B. Ma, R. Nussinov, Folding and binding cascades: shifts in energy landscapes, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 9970–9972.
[9] C.-J. Tsai, S. Kumar, B. Ma, R. Nussinov, Folding funnels, binding funnels, and protein function, Protein Sci. 8 (1999) 1181–1190.
[10] D. Tobi, I. Bahar, Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 18908–18913.
[11] T. Wlodarski, B. Zagrovic, Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 19346–19351.
[12] P.W. Snyder, M.R. Lockett, D.T. Moustakas, G.M. Whitesides, Is it the shape of the cavity, or the shape of the water in the cavity? Eur. Phys. J. Spec. Top. (2013) 1–39.
[13] B. Breiten, M.R. Lockett, W. Sherman, S. Fujita, M.H. Al-Sayah, H. Lange, C.M. Bowers, A. Heroux, G. Krilov, G.M. Whitesides, Water networks contribute to enthalpy/entropy compensation in protein–ligand binding, J. Am. Chem. Soc. 135 (2013) 15579–15584.
[14] R. Baron, P. Setny, J.A. McCammon, Water in cavity–ligand recognition, J. Am. Chem. Soc. 132 (2010) 12091–12097.
[15] A. Biela, N.N. Nasief, M. Betz, A. Heine, D. Hangauer, G. Klebe, Dissecting the hydrophobic effect on the molecular level: The role of water, enthalpy, and entropy in ligand binding to thermolysin, Angew. Chem. Int. Ed. 52 (2013) 1822–1828.
[16] J.A. McCammon, Theory of biomolecular recognition, Curr. Opin. Struct. Biol. 8 (1998) 245–249.
[17] S.J. Teague, Implications of protein flexibility for drug discovery, Nat. Rev. Drug Discov. 2 (2003) 527–541.

[18] W. Sinko, S. Lindert, J.A. McCammon, Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design, Chem. Biol. Drug Des. 81 (2013) 41–49.
[19] G. Kuzu, O. Keskin, A. Gursoy, R. Nussinov, Expanding the conformational selection paradigm in protein–ligand docking, Computational Drug Discovery and Design, Springer, 2012, pp. 59–74.
[20] J.P. Changeux, Allostery and the Monod–Wyman–Changeux model after 50 years, Annu. Rev. Biophys. 41 (2012) 103–133.
[21] J. Monod, J.P. Changeux, F. Jacob, Allosteric proteins and cellular control systems, J. Mol. Biol. 6 (1963) 306–329.
[22] J.P. Changeux, 50th anniversary of the word "allosteric", Protein Sci. 20 (2011) 1119–1124.
[23] D. Koshland Jr., G. Nemethy, D. Filmer, Comparison of experimental binding data and theoretical models in proteins containing subunits, Biochemistry 5 (1966) 365–385.
[24] L. Pauling, The oxygen equilibrium of hemoglobin and its structural interpretation, Proc. Natl. Acad. Sci. U. S. A. 21 (1935) 186.
[25] J.P. Changeux, J. Thiéry, Y. Tung, C. Kittel, On the cooperativity of biological membranes, Proc. Natl. Acad. Sci. U. S. A. 57 (1967) 335.
[26] J. Kister, C. Poyart, S. Edelstein, Oxygen–organophosphate linkage in hemoglobin A. The double hump effect, Biophys. J. 52 (1987) 527–535.
[27] A. Cooper, D. Dryden, Allostery without conformational change, Eur. Biophys. J. 11 (1984) 103–109.
[28] D. Long, R. Brüschweiler, Structural and entropic allosteric signal transduction strength via correlated motions, J. Phys. Chem. Lett. 3 (2012) 1722–1726.
[29] P.M. Gasper, B. Fuglestad, E.A. Komives, P.R. Markwick, J.A. McCammon, Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 21216–21222.
[30] P.J. Conn, A. Christopoulos, C.W. Lindsley, Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders, Nat. Rev. Drug Discov. 8 (2009) 41–54.
[31] L.T. May, K. Leach, P.M. Sexton, A. Christopoulos, Allosteric modulation of G protein-coupled receptors, Annu. Rev. Pharmacol. Toxicol. 47 (2007) 1–51.
[32] R. Nygaard, Y. Zou, R.O. Dror, T.J. Mildorf, D.H. Arlow, A. Manglik, A.C. Pan, C.W. Liu, J.J. Fung, M.P. Bokoch, F.S. Thian, T.S. Kobilka, D.E. Shaw, L. Mueller, R.S. Prosser, B.K. Kobilka, The dynamic process of β2-adrenergic receptor activation, Cell 152 (2013) 532–542.
[33] G. Kar, O. Keskin, A. Gursoy, R. Nussinov, Allostery and population shift in drug discovery, Curr. Opin. Pharmacol. 10 (2010) 715–722.
[34] R. Nussinov, C.-J. Tsai, Allostery in disease and in drug discovery, Cell 153 (2013) 293–305.
[35] A. Szilágyi, R. Nussinov, P. Csermely, Allo-network drugs: extension of the allosteric drug concept to protein–protein interaction and signaling networks, Curr. Top. Med. Chem. 13 (2013) 64–77.
[36] N.J. Smith, G. Milligan, Allostery at G protein-coupled receptor homo-and heteromers: uncharted pharmacological landscapes, Pharmacol. Rev. 62 (2010) 701–725.
[37] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, W. Wriggers, Atomic-level characterization of the structural dynamics of proteins, Science 330 (2010) 341–346.
[38] A. Mittermaier, L.E. Kay, New tools provide new insights in NMR studies of protein dynamics, Science 312 (2006) 224–228.
[39] S. Esteban-Martín, R.B. Fenwick, X. Salvatella, Synergistic use of NMR and MD simulations to study the structural heterogeneity of proteins, Wiley Interdiscip. Rev. Comput. Mol. Sci. 2 (2012) 466–478.
[40] H.A. Carlson, Protein flexibility and drug design: how to hit a moving target, Curr. Opin. Chem. Biol. 6 (2002) 447–452.
[41] S.E. Nichols, R.V. Swift, R.E. Amaro, Rational prediction with molecular dynamics for hit identification, Curr. Top. Med. Chem. 12 (2012) 2002–2012.
[42] H.J. Dyson, P.E. Wright, Unfolded proteins and protein folding studied by NMR, Chem. Rev. 104 (2004) 3607–3622.
[43] R.B. Fenwick, S. Esteban-Martín, X. Salvatella, Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles, Eur. Biophys. J. 40 (2011) 1339–1355.
[44] O.F. Lange, N.-A. Lakomek, C. Farès, G.F. Schröder, K.F. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, B.L. de Groot, Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution, Science 320 (2008) 1471–1475.
[45] K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw, How fast-folding proteins fold, Science 334 (2011) 517–520.
[46] Y. Shan, E.T. Kim, M.P. Eastwood, R.O. Dror, M.A. Seeliger, D.E. Shaw, How does a drug molecule find its target binding site? J. Am. Chem. Soc. 133 (2011) 9181–9183.
[47] W.L. Jorgensen, The many roles of computation in drug discovery, Science 303 (2004) 1813–1818.
[48] M.S. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. Legrand, A.L. Beberg, D.L. Ensign, C.M. Bruns, V.S. Pande, Accelerating molecular dynamic simulation on graphics processing units, J. Comput. Chem. 30 (2009) 864–872.
[49] J.L. Klepeis, K. Lindorff-Larsen, R.O. Dror, D.E. Shaw, Long-timescale molecular dynamics simulations of protein structure and function, Curr. Opin. Struct. Biol. 19 (2009) 120–127.
[50] I. Buch, T. Giorgino, G. De Fabritiis, Complete reconstruction of an enzyme–inhibitor binding process by molecular dynamics simulations, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 10184–10189.
[51] W. Berhanu, P. Jiang, U.H. Hansmann, Enhanced sampling for biomolecular simulations, Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes, Springer, 2014, pp. 249–267.
[52] E. Gallicchio, R.M. Levy, Advances in all atom sampling methods for modeling protein–ligand binding affinities, Curr. Opin. Struct. Biol. 21 (2011) 161–166.
[53] W.F. van Gunsteren, J. Dolenc, A.E. Mark, Molecular simulation as an aid to experimentalists, Curr. Opin. Struct. Biol. 18 (2008) 149–153.
[54] M. Candotti, S. Esteban-Martín, X. Salvatella, M. Orozco, Toward an atomistic description of the urea-denatured state of proteins, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 5933–5938.
[55] P.R. Markwick, G. Bouvignies, M. Blackledge, Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy, J. Am. Chem. Soc. 129 (2007) 4724–4730.
[56] B.K. Shoichet, Virtual screening of chemical libraries, Nature 432 (2004) 862–865.
[57] C. McInnes, Virtual screening strategies in drug discovery, Curr. Opin. Chem. Biol. 11 (2007) 494–502.
[58] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications, Nat. Rev. Drug Discov. 3 (2004) 935–949.
[59] G.L. Warren, C.W. Andrews, A.-M. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head, A critical assessment of docking programs and scoring functions, J. Med. Chem. 49 (2006) 5912–5931.
[60] C.D. Christ, A.E. Mark, W.F. van Gunsteren, Basic ingredients of free energy calculations: a review, J. Comput. Chem. 31 (2010) 1569–1582.
[61] W.L. Jorgensen, Efficient drug lead discovery and optimization, Acc. Chem. Res. 42 (2009) 724–733.
[62] J.D. Chodera, D.L. Mobley, M.R. Shirts, R.W. Dixon, K. Branson, V.S. Pande, Alchemical free energy methods for drug discovery: progress and challenges, Curr. Opin. Chem. Biol. 21 (2011) 150–160.
[63] D.L. Mobley, P.V. Klimovich, Perspective: alchemical free energy calculations for drug discovery, J. Chem. Phys. 137 (2012) 230901.
[64] R. Baron, J.A. McCammon, Molecular recognition and ligand association, Annu. Rev. Phys. Chem. 64 (2013) 151–175.
[65] J. Wereszczynski, J.A. McCammon, Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition, Q. Rev. Biophys. 45 (2012) 1–25.
[66] R.E. Amaro, W.W. Li, Emerging methods for ensemble-based virtual screening, Curr. Top. Med. Chem. 10 (2010) 3.
[67] R. Knegtel, I.D. Kuntz, C. Oshiro, Molecular docking to ensembles of protein structures, J. Mol. Biol. 266 (1997) 424–440.
[68] X. Barril, S.D. Morley, Unveiling the full potential of flexible receptor docking using multiple crystallographic structures, J. Med. Chem. 48 (2005) 4432–4443.
[69] M. Totrov, R. Abagyan, Flexible ligand docking to multiple receptor conformations: a practical alternative, Curr. Opin. Chem. Biol. 18 (2008) 178–184.
[70] W. Zhu, Y.H. Zhang, W. Sinko, M.E. Hensler, J. Olson, K.J. Molohon, S. Lindert, R. Cao, K. Li, K. Wang, Y. Wang, Y.L. Liu, A. Sankovsky, C.A.F. de Oliveira, D.A. Mitchell, V. Nizet, J.A. McCammon, E. Oldfield, Antibacterial drug leads targeting isoprenoid biosynthesis, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 123–128.
[71] M. Fujihashi, Y.-W. Zhang, Y. Higuchi, X.-Y. Li, T. Koyama, K. Miki, Crystal structure of cis-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase, Proc. Natl. Acad. Sci. 98 (2001) 4337–4342.
[72] R.-T. Guo, R. Cao, P.-H. Liang, T.-P. Ko, T.-H. Chang, M.P. Hudock, W.-Y. Jeng, C.K.-M. Chen, Y. Zhang, Y. Song, C.J. Kuo, F. Yin, E. Oldfield, A.H. Wang, Bisphosphonates target multiple sites in both cis-and trans-prenyltransferases, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 10022–10027.
[73] E. Oldfield, Targeting isoprenoid biosynthesis for drug discovery: bench to bedside, Acc. Chem. Res. 43 (2010) 1216–1226.
[74] E. Oldfield, F.-Y. Lin, Terpene biosynthesis: modularity rules, Angew. Chem. Int. Ed. 51 (2012) 1124–1137.
[75] H.S. Lee, J. Choi, I. Kufareva, R. Abagyan, A. Filikov, Y. Yang, S. Yoon, Optimization of high throughput virtual screening by combining shape-matching and docking methods, J. Chem. Inf. Model. 48 (2008) 489–497.
[76] J. Durrant, J.A. McCammon, Molecular dynamics simulations and drug discovery, BMC Biol. 9 (2011) 71.
[77] A.C. Kruse, J. Hu, A.C. Pan, D.H. Arlow, D.M. Rosenbaum, E. Rosemond, H.F. Green, T. Liu, P.S. Chae, R.O. Dror, D.E. Shaw, W.I. Weis, J. Wess, B.K. Kobilka, Structure and dynamics of the M3 muscarinic acetylcholine receptor, Nature 482 (2012) 552–556.
[78] M. Karplus, J.A. McCammon, Molecular dynamics simulations of biomolecules, Nat. Struct. Mol. Biol. 9 (2002) 646–652.
[79] S.A. Adcock, J.A. McCammon, Molecular dynamics: survey of methods for simulating the activity of proteins, Chem. Rev. 106 (2006) 1589–1615.
[80] J.A. McCammon, B.R. Gelin, M. Karplus, Dynamics of folded proteins, Nature 267 (1977) 6.
[81] G.R. Bowman, V.A. Voelz, V.S. Pande, Taming the complexity of protein folding, Curr. Opin. Struct. Biol. 21 (2011) 4–11.
[82] W.F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D.P. Geerke, A. Glättli, P.H. Hünenberger, M.A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N.F. van der Vegt, H. Yu, Biomolecular modeling: goals, problems, perspectives, Angew. Chem. Int. Ed. 45 (2006) 4064–4092.
[83] J. Mongan, D.A. Case, Biomolecular simulations at constant pH, Curr. Opin. Struct. Biol. 15 (2005) 157–163.
[84] S.L. Williams, P.G. Blachly, J.A. McCammon, Measuring the successes and deficiencies of constant pH molecular dynamics: a blind prediction study, Proteins Struct. Funct. Bioinform. 79 (2011) 3381–3388.
[85] W. Sinko, C.A.F. de Oliveira, S. Williams, A. Van Wynsberghe, J.D. Durrant, R. Cao, E. Oldfield, J.A. McCammon, Applying molecular dynamics simulations to identify

rarely sampled ligand-bound conformational states of undecaprenyl pyrophosphate synthase, an antibacterial target, Chem. Biol. Drug Des. 77 (2011) 412–420.
[86] S.E. Nichols, R. Baron, A. Ivetac, J.A. McCammon, Predictive power of molecular dynamics receptor structures in virtual screening, J. Chem. Inf. Model. 51 (2011) 1439–1446.
[87] J.-H. Lin, A.L. Perryman, J.R. Schames, J.A. McCammon, Computational drug design accommodating receptor flexibility: the relaxed complex scheme, J. Am. Chem. Soc. 124 (2002) 5632–5633.
[88] R.E. Amaro, R. Baron, J.A. McCammon, An improved relaxed complex scheme for receptor flexibility in computer-aided drug design, J. Comput. Aided Mol. Des. 22 (2008) 693–705.
[89] J.R. Schames, R.H. Henchman, J.S. Siegel, C.A. Sotriffer, H. Ni, J.A. McCammon, Discovery of a novel binding trench in HIV integrase, J. Med. Chem. 47 (2004) 1879–1881.
[90] V. Summa, A. Petrocchi, F. Bonelli, B. Crescenzi, M. Donghi, M. Ferrara, F. Fiore, C. Gardelli, O. Gonzalez Paz, D.J. Hazuda, P. Jones, O. Kinzel, R. Laufer, E. Monteguedo, E. Muraglia, E. Nizi, F. Orvieto, P. Pace, G. Pescatore, R. Scarpelli, K. Stillmock, M.V. Witmer, M. Rowley, Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection, J. Med. Chem. 51 (2008) 5843–5855.
[91] C.D. Wassman, R. Baronio, Ö. Demir, B.D. Wallentine, C.-K. Chen, L.V. Hall, F. Salehi, D.-W. Lin, B.P. Chung, G.W. Hatfield, A.R. Chamberlin, H. Luecke, R.H. Lathrop, P. Kaiser, R.E. Amaro, Computational identification of a transiently open L1/S3 pocket for reactivation of mutant p53, Nat. Commun. 4 (2013) 1407.
[92] D.J. Osguthorpe, W. Sherman, A.T. Hagler, Exploring protein flexibility: incorporating structural ensembles from crystal structures and simulation into virtual screening protocols, J. Phys. Chem. B 116 (2012) 6952–6959.
[93] G.S. Ayton, W.G. Noid, G.A. Voth, Multiscale modeling of biomolecular systems: in serial and in parallel, Curr. Opin. Struct. Biol. 17 (2007) 192–198.
[94] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chem. Phys. Lett. 314 (1999) 141–151.
[95] G.M. Torrie, J.P. Valleau, Monte Carlo free energy estimates using non-Boltzmann sampling: application to the sub-critical Lennard–Jones fluid, Chem. Phys. Lett. 28 (1974) 578–581.
[96] G.M. Torrie, J.P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling, J. Comput. Phys. 23 (1977) 187–199.
[97] G. Patey, J. Valleau, A Monte Carlo method for obtaining the interionic potential of mean force in ionic solution, J. Chem. Phys. 63 (1975) 2334.
[98] A. Laio, M. Parrinello, Escaping free-energy minima, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 12562–12566.
[99] A. Barducci, M. Bonomi, M. Parrinello, Metadynamics, WIREs Comput. Mol. Sci. 1 (2011) 826–843.
[100] D. Hamelberg, J. Mongan, J.A. McCammon, Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules, J. Chem. Phys. 120 (2004) 11919.
[101] P.R. Markwick, J.A. McCammon, Studying functional dynamics in bio-molecules using accelerated molecular dynamics, Phys. Chem. Chem. Phys. 13 (2011) 20053–20065.
[102] V. Limongelli, M. Bonomi, L. Marinelli, F.L. Gervasio, A. Cavalli, E. Novellino, M. Parrinello, Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 5411–5416.
[103] J. Wereszczynski, J.A. McCammon, Accelerated molecular dynamics in computational drug design, Computational Drug Discovery and Design, Springer, 2012, pp. 515–524.
[104] T. Shen, D. Hamelberg, A statistical analysis of the precision of reweighting-based simulations, J. Chem. Phys. 129 (2008) 034103.
[105] J. Wereszczynski, J.A. McCammon, Using selectively applied accelerated molecular dynamics to enhance free energy calculations, J. Chem. Theory Comput. 6 (2010) 3285–3292.
[106] B. Fuglestad, P.M. Gasper, M. Tonelli, J.A. McCammon, P.R. Markwick, E.A. Komives, The dynamic structure of thrombin in solution, Biophys. J. 103 (2012) 79–88.
[107] Y. Miao, S.E. Nichols, P.M. Gasper, V.T. Metzger, J.A. McCammon, Activation and dynamic network of the M2 muscarinic receptor, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 10982–10987.
[108] Y. Miao, J.A. McCammon, Enhanced conformational sampling of M2 muscarinic acetylcholine receptor for designing selective allosteric drugs, Biophys. J. 104 (2013) 26.
[109] J. Wereszczynski, J.A. McCammon, Nucleotide-dependent mechanism of Get3 as elucidated from free energy calculations, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 7759–7764.
[110] L.C. Pierce, R. Salomon-Ferrer, C.A.F. de Oliveira, J.A. McCammon, R.C. Walker, Routine access to millisecond time scale events with accelerated molecular dynamics, J. Chem. Theory Comput. 8 (2012) 2997–3002.
[111] R. Elber, M. Karplus, Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin, Science 235 (1987) 318–321.
[112] A. Kitao, F. Hirata, N. Gö, The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum, Chem. Phys. 158 (1991) 447–472.
[113] A.E. Garcia, Large-amplitude nonlinear motions in proteins, Phys. Rev. Lett. 68 (1992) 2696.
[114] A. Amadei, A. Linssen, H.J. Berendsen, Essential dynamics of proteins, Proteins Struct. Funct. Bioinform. 17 (1993) 412–425.
[115] H.J. Berendsen, S. Hayward, Collective protein dynamics in relation to function, Curr. Opin. Chem. Biol. 10 (2000) 165–169.
[116] M.A. Balsera, W. Wriggers, Y. Oono, K. Schulten, Principal component analysis and long time protein dynamics, J. Phys. Chem. 100 (1996) 2567–2572.
[117] B.J. Grant, J.A. McCammon, A.A. Gorfe, Conformational selection in G-proteins: lessons from Ras and Rho, Biophys. J. 99 (2010) L87–L89.
[118] M. Rueda, G. Bottegoni, R. Abagyan, Recipes for the selection of experimental protein conformations for virtual screening, J. Chem. Inf. Model. 50 (2009) 186–193.
[119] A. Sethi, P. O'Donoghue, Z. Luthey-Schulten, Evolutionary profiles from the QR factorization of multiple sequence alignments, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 4045–4050.
[120] J.D. Durrant, M.D. Urbaniak, M.A. Ferguson, J.A. McCammon, Computer-aided identification of *Trypanosoma brucei* uridine diphosphate galactose 4,Ä≤-epimerase inhibitors: toward the development of novel therapies for African sleeping sickness, J. Med. Chem. 53 (2010) 5025–5032.
[121] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010) 455–461.
[122] J.D. Durrant, R. Cao, A.A. Gorfe, W. Zhu, J. Li, A. Sankovsky, E. Oldfield, J.A. McCammon, Non-bisphosphonate inhibitors of isoprenoid biosynthesis identified via computer-aided drug design, Chem. Biol. Drug Des. 78 (2011) 323–332.
[123] W. Jahnke, J.-M. Rondeau, S. Cotesta, A. Marzinzik, X. Pellé, M. Geiser, A. Strauss, M. Götte, F. Bitsch, R. Hemmig, C. Henry, S. Lehmann, J.F. Glickman, T.P. Roddy, S.J. Stout, J.R. Green, Allosteric non-bisphosphonate FPPS inhibitors identified by fragment-based discovery, Nat. Chem. Biol. 6 (2010) 660–666.
[124] S. Lindert, W. Zhu, Y. Liu, R. Pang, E. Oldfield, J.A. McCammon, Farnesyl diphosphate synthase inhibitors from in silico screening, Chem. Biol. Drug Des. 81 (2013) 742–748.
[125] B. Nisius, F. Sha, H. Gohlke, Structure-based computational analysis of protein binding sites for function and druggability prediction, J. Biotechnol. 159 (2012) 123–134.
[126] J. Kenneth Morrow, S. Zhang, Computational prediction of protein hot spot residues, Curr. Pharm. Des. 18 (2012) 1255–1265.
[127] A. Christopoulos, Allosteric binding sites on cell-surface receptors: novel targets for drug discovery, Nat. Rev. Drug Discov. 1 (2002) 198–210.
[128] A. Ivetac, J.A. McCammon, Mapping the druggable allosteric space of G-protein coupled receptors: a fragment-based molecular dynamics approach, Chem. Biol. Drug Des. 76 (2010) 201–217.
[129] R. Brenke, D. Kozakov, G.-Y. Chuang, D. Beglov, D. Hall, M.R. Landon, C. Mattos, S. Vajda, Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques, Bioinformatics 25 (2009) 621–627.
[130] G. Klebe, Virtual ligand screening: strategies, perspectives and limitations, Drug Discov. Today 11 (2006) 580–594.
[131] M.R. Shirts, D.L. Mobley, J.D. Chodera, Alchemical free energy calculations: ready for prime time? Annu. Rep. Comput. Chem. 3 (2007) 41–59.
[132] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, J. Med. Chem. 47 (2004) 1739–1749.
[133] C. B-Rao, J. Subramanian, S.D. Sharma, Managing protein flexibility in docking and its applications, Drug Discov. Today 14 (2009) 394–400.
[134] K.W. Lexa, H.A. Carlson, Protein flexibility in docking and surface mapping, Q. Rev. Biophys. 45 (2012) 301–343.
[135] D.B. Kokh, R.C. Wade, W. Wenzel, Receptor flexibility in small-molecule docking calculations, WIREs Comput. Mol. Sci. 1 (2011) 298–314.
[136] F. Jiang, S.-H. Kim, "Soft docking": matching of molecular surface cubes, J. Mol. Biol. 219 (1991) 79–102.
[137] A.M. Ferrari, B.Q. Wei, L. Costantino, B.K. Shoichet, Soft docking and multiple receptor conformations in virtual screening, J. Med. Chem. 47 (2004) 5076–5084.
[138] A.R. Leach, Ligand docking to proteins with discrete side-chain flexibility, J. Mol. Biol. 235 (1994) 345–356.
[139] W. Sherman, H.S. Beard, R. Farid, Use of an induced fit receptor structure in virtual screening, Chem. Biol. Drug Des. 67 (2006) 83–84.
[140] S.B. Nabuurs, M. Wagener, J. De Vlieg, A flexible approach to induced fit docking, J. Med. Chem. 50 (2007) 6507–6518.
[141] J. Meiler, D. Baker, ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility, Proteins Struct. Funct. Bioinform. 65 (2006) 538–548.
[142] I.W. Davis, D. Baker, RosettaLigand docking with full ligand and receptor flexibility, J. Mol. Biol. 385 (2009) 381–392.
[143] H. Gohlke, C. Kiel, D.A. Case, Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes, J. Mol. Biol. 330 (2003) 891–913.
[144] I. Massova, P.A. Kollman, Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding, Perspect. Drug Discov. Des. 18 (2000) 113–135.
[145] J. Kongsted, U. Ryde, An improved method to predict the entropy term with the MM/PBSA approach, J. Comput. Aided Mol. Des. 23 (2009) 63–71.
[146] H. Gohlke, D.A. Case, Converging free energy estimates: MM/PB(GB)SA studies on the protein–protein complex Ras–Raf, J. Comput. Chem. 25 (2004) 238–250.
[147] G. Rastelli, A.D. Rio, G. Degliesposti, M. Sgobba, Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA, J. Comput. Chem. 31 (2010) 797–810.
[148] T. Hou, J. Wang, Y. Li, W. Wang, 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 (2010) 69–82.
[149] B.L. Tembe, J.A. McCammon, Ligand–receptor interactions, Comput. Chem. 8 (1984) 281–283.
[150] M. Bollini, R.A. Domaoal, V.V. Thakur, R. Gallardo-Macias, K.A. Spasov, K.S. Anderson, W.L. Jorgensen, Computationally-guided optimization of a docking hit

to yield catechol diethers as potent anti-HIV agents, J. Med. Chem. 54 (2011) 8582–8591.
[151] M. Lawrenz, R. Baron, J.A. McCammon, Independent-trajectories thermodynamic-integration free-energy changes for biomolecular systems: determinants of H5N1 avian influenza virus neuraminidase inhibition by peramivir, J. Chem. Theory Comput. 5 (2009) 1106–1116.
[152] C.A.F. de Oliveira, D. Hamelberg, J.A. McCammon, Coupling accelerated molecular dynamics methods with thermodynamic integration simulations, J. Chem. Theory Comput. 4 (2008) 1516–1525.
[153] W. Sinko, C.A.F. de Oliveira, L.C. Pierce, J.A. McCammon, Protecting high energy barriers: a new equation to regulate boost energy in accelerated molecular dynamics simulations, J. Chem. Theory Comput. 8 (2011) 17–23.