



# Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview

Veronica Salmaso and Stefano Moro\*

Molecular Modeling Section, Department of Pharmaceutical and Pharmacological Sciences, University of Padova, Padova, Italy

## OPEN ACCESS

### Edited by:

Adriano D. Andricopulo,  
Universidade de São Paulo, Brazil

### Reviewed by:

José Pedro Cerón-Carrasco,  
Universidad Católica San Antonio de  
Murcia, Spain  
Andrea Mozzarelli,  
Università degli Studi di Parma, Italy

### \*Correspondence:

Stefano Moro  
stefano.moro@unipd.it

### Specialty section:

This article was submitted to  
Experimental Pharmacology and Drug  
Discovery,  
a section of the journal  
Frontiers in Pharmacology

**Received:** 04 May 2018

**Accepted:** 26 July 2018

**Published:** 22 August 2018

### Citation:

Salmaso V and Moro S (2018)  
Bridging Molecular Docking to  
Molecular Dynamics in Exploring  
Ligand-Protein Recognition Process:  
An Overview.  
Front. Pharmacol. 9:923.  
doi: 10.3389/fphar.2018.00923

**Keywords:** ligand-protein binding, molecular docking, molecular dynamics, enhanced sampling, protein flexibility, molecular recognition

## INTRODUCTION

No protein is an island but exerts its function through the recognition of other molecular partners (Salmaso, 2018). Ligand-protein interactions are involved in many biological processes with consequent pharmaceutical implications. Thus, the scientific community has been putting a great effort into the investigation of the binding phenomenon during the years, leading to the proposal of several theories characterized by an increasing emphasis on the degree of flexibility of the ligand and protein counterparts.

The first explanation of binding was provided by Emil Fischer in 1894 (Fischer, 1894) with the “lock-key” model to interpret enzyme specificity: the ligand rigidly recognizes and occupies the protein binding site like a key to its lock, because of their native shape complementary. Since this model could not explain either the behavior of enzyme noncompetitive inhibition or allosteric modulation, different modifications have been proposed. Koshland (1958) introduced the “induced-fit” theory: according to his observations on enzyme-substrate interactions, the ligand is able to induce conformational changes to the protein, optimizing ligand-target interactions. Later works suggested that proteins naturally exist as an ensemble of conformations (Monod et al., 1965), described by an energy landscape (Frauenfelder et al., 1991), and ligands preferentially bind to one of them (Austin et al., 1975; Foote and Milstein, 1994). According to this interpretation of binding, known as “conformational selection,” the ligand stabilizes one of the protein conformations with a consequent shift of the protein population equilibrium (Kumar et al., 2000). These two apparently contrasting theories have simply different ranges of applicability, and the descriptions they provide of molecular binding differ for the chronological sequence of events in which the binding process is decomposed (Kobilka and Deupi, 2007; Okazaki and Takada, 2008; Zhou, 2010). New theories are emerging, making a compromise between the aforementioned ones: according to

the extended conformational selection model, for example, the conformational selection is followed by a conformational adjustment (induced fit) (Csermely et al., 2010).

The evolution of binding models has practical relevance besides an epistemological significance; the knowledge of ligand-target binding is at the basis of rational drug design but understanding this complex process on a mechanistic level may open new scenarios. In addition, to suggest ligand modification meant to optimize the final bound state, the medicinal chemist may look at kinetically relevant intermediate states and try to affect them.

## COMPUTATIONAL METHODS TO STUDY LIGAND-PROTEIN BINDING

Since the 1980s, computer technologies have been applied to the drug discovery process (Van Drie, 2007), giving rise to Computer-Aided Drug Design (CADD). This technique earned soon great interest and deserved a cover article on October 5, 1981, Fortune magazine, entitled “Next Industrial Revolution: Designing Drugs by Computer at Merck” (Van Drie, 2007). CADD techniques are used principally for three reasons: virtual screening *hit/lead* optimization and design of novel compounds. In virtual screening a huge database of compounds is examined searching for binding capacity for a target and a subset of compounds is picked out and suggested for *in vitro* testing; the purpose is to increase the *hit* rate of novel drugs by reducing the number of compounds to test experimentally. The second application of CADD is the optimization of a *hit/lead* compound driven by the rationalization of a structure-activity relationship. After the individuation of key elements for binding, the design of new compounds can be attempted (Salmaso, 2018).

CADD methods may be classified as ligand-based (LB) and structure-based (SB), depending on the availability and employment of the target structure (Sliwoski et al., 2014). In the framework of CADD, structure-based drug design (SBDD) methods take advantage of the abundance of experimentally solved structures in the Protein Data Bank (Berman et al., 2000), which can possibly be used also as templates for homology models if the structure of interest is lacking. SBDD is based on the premise that the knowledge of the target structure can help to rationalize and optimize binding since ligand-target interactions are mediated by their complementarity. With the evolution of the binding models, it is clear that speaking of “target structure” is an approximation, given that proteins fluctuate among an ensemble of structures (Miller and Dill, 1997).

The possibility to predict ligand binding modes and to interpret binding processes is valuable to individuate, optimize and suggest novel ligands, and for this reason, the scientific community has been putting great efforts in developing new computational techniques.

In the following paragraphs, we will present an excursus over the main structure-based computational techniques employed in drug discovery. An urgency to simulate protein flexibility throughout binding has been experienced over the years, arising from the evolution of the binding models

from static to dynamic. The inclusion of flexibility features in conformational sampling entails an increase in the number of degrees of freedom of the system, and consequently in the computational effort. For this reason, the development of computational tools has been occurring in parallel and thanks to the continuous improvement of hardware technologies.

## Molecular Docking

Molecular docking techniques aim to predict the best matching binding mode of a ligand to a macromolecular partner (here just proteins are considered). It consists in the generation of a number of possible conformations/orientations, i.e., poses, of the ligand within the protein binding site. For this reason, the availability of the three-dimensional structure of the molecular target is a necessary condition; it can be an experimentally solved structure (such as by X-ray crystallography or NMR) or a structure obtained by computational techniques (such as homology modeling) (Salmaso, 2018).

Molecular docking is composed mainly by two stages: an engine for conformations/orientations sampling and a scoring function, which associates a score to each predicted pose (Abagyan and Totrov, 2001; Kitchen et al., 2004; Huang and Zou, 2010). The sampling process should effectively search the conformational space described by the free energy landscape, where energy, in docking, is approximated by the scoring function. The scoring function should be able to associate the native bound-conformation to the global minimum of the energy hypersurface.

## Scoring Functions

Scoring functions play the role of poses selector, used to discriminate putative correct binding modes and binders from non-binders in the pool of poses generated by the sampling engine.

There are essentially three types of scoring functions:

### 1. Force-field based scoring functions:

Force-field is a concept typical of molecular mechanics (see **Box 1**) which approximates the potential energy of a system with a combination of bonded (intramolecular) and nonbonded (intermolecular) components. In molecular docking, the nonbonded components are generally taken into account, with possibly the addition of the ligand-bonded terms, especially the torsional components. Intermolecular components include the van der Waals term, described by the Lennard-Jones potential, and the electrostatic potential, described by the Coulomb function, where a distance-dependent dielectric may be introduced to mimic the solvent effect. However, additional terms have been added to the force-field scoring functions, such as solvation terms (Brooijmans and Kuntz, 2003).

Examples of force field based scoring functions are GoldScore (Verdonk et al., 2003), AutoDock (Morris et al., 1998) (improved as a semiempirical version in AutoDock4, Huey et al., 2007), GBVI/WSA (Corbeil et al., 2012).

## 2. Empirical scoring functions:

These functions are the sum of various empirical energy terms such as van der Waals, electrostatic, hydrogen bond, desolvation, entropy, hydrophobicity, etc., which are weighted by coefficients optimized to reproduce binding affinity data of a training set by least squares fitting (Huang and Zou, 2010).

The LUDI (Böhm, 1994) scoring function was the first example of an empirical one. Other empirical scoring functions are GlideScore (Halgren et al., 2004; Friesner et al., 2006), ChemScore (Eldridge et al., 1997), PLANTS<sub>CHEMPLP</sub> (Korb et al., 2009).

## 3. Knowledge-based scoring functions:

### Box 1 | Molecular mechanics.

Molecular mechanics is a method which approximates the treatment of molecules with the laws of classical mechanics, in order to limit the computational cost required for quantum mechanical calculations (Vanommeslaeghe et al., 2014). Atoms are considered as charged spheres connected by springs, neglecting the presence of electrons, in accordance with Born-Oppenheimer approximation (Born and Oppenheimer, 1927). The potential energy is approximated by a simple function which is called force-field; it is the sum of bonded (intramolecular) and nonbonded energy terms. The basic form of the function comprise bond stretching and bending described by harmonic potential, and torsional potential described by a trigonometric function, in the bonded portion. Nonbonded terms consist of van der Waals and Coulomb electrostatic interactions between couples of atoms.

As an example, these basic components of the CHARMM [78] force field are reported in the following equations

$$\begin{aligned}
 V &= V_{\text{bonded}} + V_{\text{nonbonded}} \\
 V_{\text{bonded}} &= \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \\
 &\quad + \sum_{\text{dihedrals}} K_\chi(1 + \cos(n\chi - \delta)) \\
 V_{\text{nonbonded}} &= \sum_{\text{pairs } ij} \frac{q_i q_j}{\epsilon r_{ij}} \\
 &\quad + \sum_{\text{pairs } ij} \epsilon_{ij} \left[ \left( \frac{R_{\text{min}, ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{\text{min}, ij}}{r_{ij}} \right)^6 \right]
 \end{aligned}$$

where  $K_b$ ,  $K_\theta$ , and  $K_\chi$  are the bond, angle and torsional force constants;  $b$ ,  $\theta$  and  $\chi$  are bond length, bond angle and dihedral angle (those with the 0-subscript are the equilibrium values);  $n$  is multiplicity and  $\delta$  the phase of the torsional periodic function;  $r_{ij}$  is the distance between atoms  $i$  and  $j$ ;  $q_i$  and  $q_j$  are the partial charges of atoms  $i$  and  $j$ ;  $\epsilon$  is the effective dielectric constant;  $\epsilon_{ij}$  is the Lennard-Jones well depth and  $R_{\text{min}, ij}$  is the distance between atoms at Lennard-Jones minimum.

These terms may appear slightly different in different force-fields, and anharmonicity and cross-terms are generally added.

The parameters of the force field are obtained by fitting quantum mechanical or experimental values.

These methods assume that ligand-protein contacts statistically more explored are correlated with favorable interactions. Starting from a database of structures, the frequencies of ligand-protein atom pairs contacts are computed and converted into an energy component. When evaluating a pose, the aforementioned tabulated energy components are summed up for all ligand-protein atom pairs, giving the score of the pose.

DrugScore (Gohlke et al., 2000; Velec et al., 2005) and GOLD/ASP (Mooij and Verdonk, 2005) are examples of knowledge-based scoring functions.

Another strategy consists in the combination of multiple scoring functions leading to the so-called consensus scoring (Charifson et al., 1999).

In addition, new scoring functions have been developed: for example, based on machine learning technologies, interaction fingerprints and attempts with quantum mechanical scores (Yuriev et al., 2015).

## Sampling

The first molecular docking algorithm was developed in the 1980s by Kuntz et al. (1982); the receptor was approximated by a series of spheres filling its surface clefts, and the ligand by another set of spheres defining its volume. A search was made to find the best steric overlap between binding site and receptor spheres, neglecting any kind of conformational movement.

This method belongs to the group of fully-rigid docking techniques, according to the classification which divides docking methods according to the degrees of flexibility of the molecules involved in the calculation Halperin et al., 2002 (**Figure 1**):

### 1. Rigid docking:

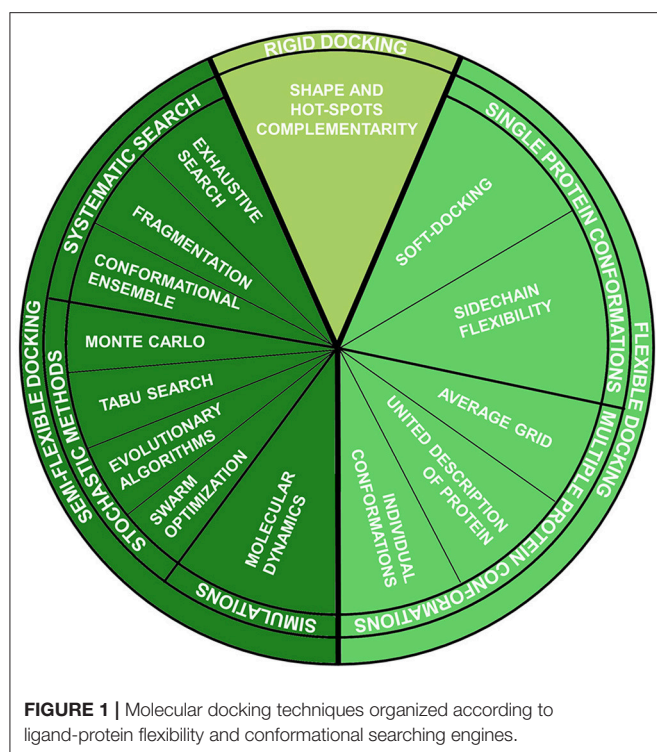
Both ligand and protein are considered rigid entities, and just the three translational and three rotational degrees of freedom are considered during sampling. This approximation is analogous to the “lock-key” binding model and is mainly used for protein-protein docking, where the number of conformational degrees of freedom is too high to be sampled. Generally, in these methods, the binding site and the ligand are approximated by “hot” points and the superposition of matching point is evaluated (Taylor et al., 2002).

### 2. Semi-flexible docking:

Just one of the molecules, the ligand, is flexible, while the protein is rigid. Thus, the conformational degrees of freedom of the ligand are sampled, in addition to the six translational plus rotational ones. These methods assume that a fixed conformation of a protein may correspond to the one able to recognize the ligands to be docked. This assumption, as already reported, is not always verified.

### 3. Flexible docking:

It is based on the concept that a protein is not a passive rigid entity during binding and considers both ligand and protein as flexible counterparts. Different methods have been introduced during the years, some rested on the induced fit binding model and others on conformational selection.



The great number of degrees of freedom introduced by flexible docking makes the potential energy surface to be a function of numerous coordinates. Consequently, the computational effort required to perform a docking calculation is augmented, but both sampling and scoring should be optimized to give a good balance between accuracy and speed. In fact, virtual screening campaign of millions of compounds depends on the velocity of docking calculations. For this reason, more and more improvements have been made in the development of the new algorithm, able to deeply search the phase space but not at the expense of velocity.

### Semi-flexible Docking

Numerous docking algorithms have been developed since the 1980s. Often it is difficult to classify clearly each docking software, because different algorithms may be integrated into a multi-phase approach. However, docking algorithms can be classified as follows (Kitchen et al., 2004; Huang and Zou, 2010):

#### 1. Systematic search techniques:

In a systematic search, a set of discretized values is associated with each degree of freedom, and all the values of each coordinate are explored in a combinatorial way (Brooijmans and Kuntz, 2003). These methods are subdivided into:

- a. *Exhaustive search* - it is a systematic search in the strict sense since all the rotatable bonds of the ligands are examined in a systematic way. A number of constraints and termination criteria is generally established to limit the search space and to avoid a combinatorial explosion. The docking pipeline of the software Glide (Friesner et al., 2004; Halgren et al., 2004) involves a stage of the exhaustive search.

- b. *Fragmentation* - the first implementation of ligand flexibility into docking was introduced by DesJarlais et al. (1986), who proposed a method made of fragmentation of the ligand, rigid docking of the fragments into the binding site, and subsequent linking of the fragments. In this way, partial flexibility is implemented at the joints between the fragments. Other methods, defined as incremental construction, dock one fragment first and then attach incrementally the others. Examples of methods utilizing fragmentation are FlexX (Rarey et al., 1996) and Hammerhead (Welch et al., 1996).
- c. *Conformational Ensemble* - rigid docking algorithms can be easily enriched by a sort of flexibility if an ensemble of previously generated conformers of the ligand is docked to the target, in a sort of conformational selection fashion on the ligand counterpart. Examples are offered by FLOG (Miller et al., 1994), EUDOC (Pang et al., 2001), MS-DOCK (Sauton et al., 2008).

#### 2. Stochastic methods:

Stochastic algorithms change randomly, instead of systematically, the values of the degrees of freedom of the system. The advantage of these techniques is the speed, so they could potentially find the optimal solution really fast. As a drawback, they do not ensure a full search of the conformational space, so the true solution may be missed. The lack of convergence is partially solved by increasing the number of iterations of the algorithm. The most famous stochastic algorithms are (Huang and Zou, 2010):

- a. *Monte Carlo (MC) methods* - Monte Carlo methods are based on the Metropolis Monte Carlo algorithm, which introduces an acceptance criterion in the evolution of the docking search. In particular, at every iteration of the algorithm, a random modification of the ligand degrees of freedom is performed. Then, if the energy score of the pose is improved, the change is accepted, otherwise, it is accepted according to the probability expressed in the following equation:

$$P \sim \exp \left[ \frac{-(E_1 - E_0)}{k_B T} \right]$$

where  $E_1$  and  $E_0$  are the energy score before and after the modification,  $k_B$  the Boltzmann constant, and  $T$  the temperature of the system.

This is the original form of the Metropolis algorithm, but it is implemented in different variants within docking software. Some example are provided by the earlier versions of AutoDock (Goodsell and Olson, 1990; Morris et al., 1996), ICM (Abagyan et al., 1994), QXP (McMartin and Bohacek, 1997), MCDock (Liu and Wang, 1999), AutoDock Vina (Trott and Olson, 2010), ROSETTALIGAND (Meiler and Baker, 2006).

- b. *Tabu search methods* - the aim of these algorithms is to prevent the exploration of already sampled zones of the conformational/positional space. Random modifications are performed on the degrees of freedom of the ligand at each iteration. The already sampled conformations are registered, and when a new pose is obtained, it is accepted only if not similar to any previously explored pose. PRO\_LEADS (Baxter



- et al., 1998) and PSI-DOCK (Pei et al., 2006) are two examples of this category.
- c. *Evolutionary Algorithms* (EA) - these algorithms are based on the idea of biological evolution, with the most famous Genetic Algorithms (GAs). The concept of the gene, chromosome, mutation, and crossover are borrowed from biology. In particular, the degrees of freedom are encoded into genes, and each conformation of the ligand is described by a chromosome (collection of genes), which is assigned a fitness score. Mutations and crossovers occur within a population of chromosomes, and chromosomes with higher fitness survive and replace the worst ones. The most famous examples are GOLD (Jones et al., 1995, 1997), AutoDock 3 & 4 (which implement a different version of GA, the Lamarckian GA) (Morris et al., 1998), PSI-DOCK (Pei et al., 2006), rDock (Ruiz-Carmona et al., 2014).
  - d. *Swarm optimization (SO) methods* - these methods take inspiration from swarm behavior. The sampling of the degrees of freedom of a ligand is guided by the information deposited by already sampling good poses. For example, PLANTS (Korb et al., 2006) adopts an ACO (Ant Colony Optimization) algorithm, which mimics the behavior of ants, who communicate the easiest way to reach a source of food through the deposition of pheromone. Here, each degree of freedom is associated with a pheromone. Virtual ants choose conformations considering the values of pheromones, and successful ants contribute to pheromone deposition.

Other examples of SOs are SODOCK (Chen et al., 2007), pso@autodock (Namasivayam and Günther, 2007), PSOVina (Ng et al., 2015).

### 3. Simulation methods:

The most famous example of this category is Molecular Dynamics, a method that describes the time evolution of a system. A wider explanation will be given in section Molecular Dynamics.

Energy minimization methods can be inserted in this category, but generally, they are not used as stand-alone search engines (Kitchen et al., 2004). Energy minimization is a local optimization technique, used to bring the system to the closest minimum on the potential energy surface.

## Flexible Docking

Some attempts have been made to introduce protein flexibility into docking calculations. These methods take advantage of different degrees of approximation and can be divided into approaches that consider single protein or multiple protein conformations (Alonso et al., 2006).

### 1. Single Protein Conformation:

#### a. Soft docking:

This method, firstly described by Jiang and Kim (1991), consists of an implicit and rough treatment of protein flexibility. The van der Waals repulsion term employed in force field scoring functions is reduced, allowing small clashes that permit a closer ligand-protein packing. In this way, a sort of induced-fit is

simulated. As a drawback, this approach approximates just feeble protein movements and could implicate unreal poses (Apostolakis et al., 1998; Vieth et al., 1999).

#### b. Sidechain flexibility:

This strategy introduces alternative conformations for some protein side chains (Leach, 1994). This is generally done exploiting databases of rotamer libraries. Some docking methods, such as GOLD, sample some degrees of freedom within their own search engine. Obviously, considering side chain flexibility, huge conformational variations of the protein are neglected by these methods.

### 2. Multiple Protein Conformations:

Multiple experimental structures may be available for the same target. Moreover, an ensemble of protein conformations can be obtained via computational techniques, such as Monte Carlo or Molecular Dynamics simulations. The idea of multiple protein conformations docking is to take into account all the diverse structures, following different possible strategies:

#### a. Average grid:

The structures of the ensemble are used to construct a single average-grid, which can be either a simple or weighted average combination of them (Knegt et al., 1997).

#### b. United description of the protein:

In this case, the structures do not collapse into an average grid but are used to construct the best performing “chimera” protein. For example, FlexE (Rarey et al., 1996) extracts the structurally conserved portions from the structures of the ensemble and uses them to construct an average rigid structure. This portion is fused to the flexible parts of the ensemble in a combinatorial fashion, giving a pool of “chimeras” that are used for docking.

#### c. Individual conformations:

The structures of the ensemble are considered as conformations that can possibly be bound by the ligand, so various docking runs are performed, evaluating the ligands of interest on all the target conformations (Huang and Zou, 2007). Moreover, a preliminary benchmark assessing the performance of different target structures in a cross-docking experiment may be employed to filter the ensemble of structures (Salmaso et al., 2016, 2018).

Among the drugs approved by the Food and Drug Administration, few examples of successful applications of CADD are available (Talele et al., 2010). Among them, the renin-inhibitor Aliskiren was developed by means of a combination of molecular modeling and crystallographic structure analysis (Wood et al., 2003). However, the binding of non-peptidomimetic ligands to renin has shown huge structural rearrangement of the protein (Teague, 2003), addressing the problem of considering protein flexibility in drug design campaigns. Recently, a comparative study evaluating the performance of ensemble docking and individual crystal structure docking has been proposed for renin (Strecker and Meyer, 2018). An ensemble of 4 crystal structures outperformed the mean results of individual crystal structures in terms of

binding mode prediction and screening utility. The ensemble gave worse results than the best performing crystal structure, which though is not known a priori. Not as good results were obtained through a Molecular Dynamics ensemble when compared to crystallographic structures, as confirmed in other cases reported in the literature (Osguthorpe et al., 2012; Ganser et al., 2018). However, Molecular Dynamics has proven to be effective as a tool to explore molecular conformations and as a docking method itself, as reported in the following paragraphs.

## Molecular Dynamics

Molecular dynamics (MD) is a computational technique which simulates the dynamic behavior of molecular systems as a function of time, treating all the entities in the simulation box (ligand, protein, as long as waters if explicit) as flexible (Salmaso, 2018).

It was developed to simulate simple systems, with the first application to study collisions among hard spheres, in 1957 (Alder and Wainwright, 1957). The first MD simulation of a biomolecule was accomplished in 1977 by McCammon et al. (McCammon et al., 1977); it was a 9.2 ps simulation of a 58-residues Bovine Pancreatic Trypsin Inhibitor (BPTI), performed in vacuum with a crude molecular mechanics potential.

Molecular dynamics compute the movements of atoms along time by the integration of Newton's equations of motions (classical mechanics), reported in the following equation (Leach, 2001; Adcock and McCammon, 2006).

$$\frac{d^2 r_i(t)}{dt^2} = \frac{F_i(t)}{m_i}$$

with  $F_i(t)$  force exerted on atom  $i$  at time  $t$ ,  $r_i(t)$  vector position of the atom  $i$  at time  $t$ ,  $m_i$  mass of the atom (Figure 2).

In particular, time is partitioned into time steps ( $\delta t$ ), which are used to propagate the system forward in time. Several integration algorithms are available, which derive Newton's equations by a discrete-time numerical approximation. The velocity-Verlet integrator is reported in the following equations as an example to compute position and velocity of an atom  $i$  at the time step  $t + \delta t$ , starting from step  $t$ .

$$\begin{aligned} r_i(t + \delta t) &= r_i(t) + v_i(t) \delta t + \frac{1}{2} a_i(t) \delta t^2 \\ v_i(t + \delta t) &= v_i(t) + \frac{1}{2} [a_i(t) + a_i(t + \delta t)] \delta t \end{aligned}$$

where  $r_i(t)$ ,  $v_i(t)$  and  $a_i(t)$  are respectively position, velocity and acceleration of atom  $i$  at time  $t$ , and  $r_i(t + \delta t)$ ,  $v_i(t + \delta t)$  and  $a_i(t + \delta t)$  are respectively position, velocity and acceleration of atom  $i$  at time  $t + \delta t$ .

Acceleration is calculated from the forces acting on atom  $i$  according to Newton's second law, and forces are computed from the force field, according to the following equation:

$$a_i(t) = \frac{d^2 r_i(t)}{dt^2} = \frac{F_i(t)}{m_i} = - \frac{dV(r(t))}{m_i dr_i(t)}$$

where  $V(r(t))$  is the potential energy function retrieved by the force field (see Box 1).

The most used force fields in molecular dynamics are CHARMM (MacKerell et al., 1998), AMBER (Cornell et al., 1995), OPLS (Jorgensen and Tirado-Rives, 1988) and GROMOS (Oostenbrink et al., 2004).

## Molecular Dynamics and Exploration of the Phase Space

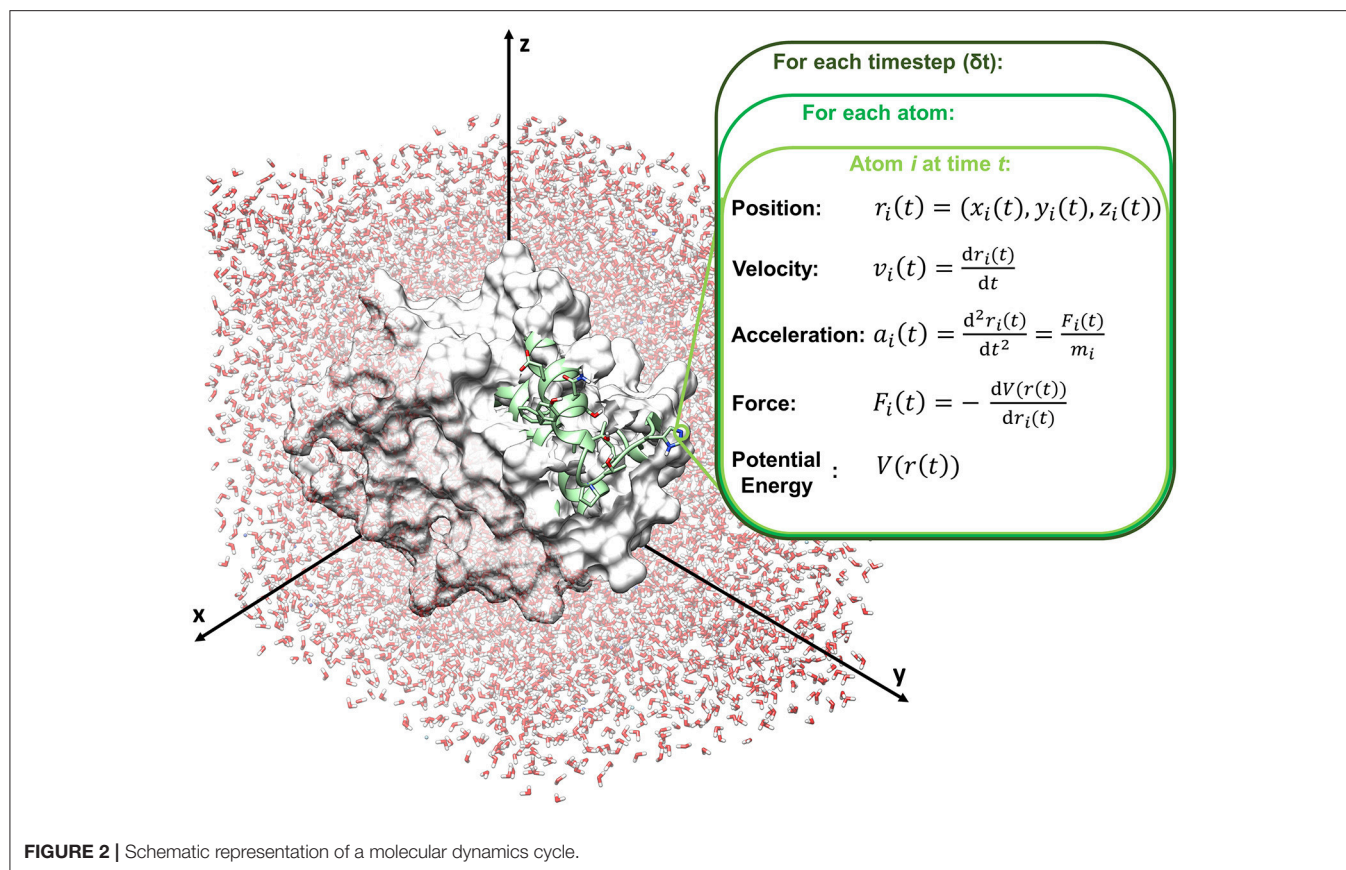
MD trajectories can be used as sampling engines; in fact, they produce protein conformations usable for Multiple Protein Conformations docking applications. In particular, McCammon et al. developed the so-called Relaxed-Complex Scheme (RCS), consisting in docking mini-libraries of compounds with AutoDock (Morris et al., 1998) against a large ensemble of snapshots derived from unliganded protein MD trajectories (Lin et al., 2002, 2003; Amaro et al., 2008). This approach is based on the conformational selection binding model, disregarding any influence of the ligand on the receptor. The application of the RCS to the UDP-galactose 4'-epimerase (*TbGalE*), for example, led to the identification of 14 low-micromolar inhibitors (Durrant et al., 2010). Another computational pipeline integrating MD simulations and virtual screening has proved to be effective: the coupling of MD, clustering, and choice of the target structure through fingerprints for ligand and proteins (MD-FLAP) improved VS performance (Spyrakis et al., 2015).

MD has further applications as a docking-coupled technique (Alonso et al., 2006) more anchored to the induced-fit model, as it can be used to assess stability (Sabbadin et al., 2014; Yu et al., 2018), to refine and to rescore docking poses (Rastelli et al., 2009).

The relevance of MD simulations as source of target conformational profusion can be exploited to retrieve insights into cryptic pockets or allosteric binding sites (Durrant and McCammon, 2011), as reported by Schame et al., who identified an alternative binding site, named "trench," close to the active site of the HIV-1 integrase (Schames et al., 2004). Moreover, simulations in the explicit solvent may give information on water molecules, that can be classified as "cold" or stable and "hot" or unstable (for a recent and comprehensive overview on the role of water in SBDD; see Spyris et al., 2017). In particular, MD may enable to individuate relevant water molecules, according to their order (Li and Lazaridis, 2003) and stationarity (Cuzzolin et al., 2018), and to estimate their contribution in modulating ligand binding (Bortolato et al., 2013; Betz et al., 2016).

All the aforementioned applications of MD are used as a complement to classic molecular docking techniques. However, the simulation of the complete binding process of a ligand, from the unbound state in bulk solvent to the bound state, be considered a fully-flexible docking in explicit solvent. The possibility to investigate the whole binding process could give insights into metastable states reached by the ligand during the simulation, alternative binding sites, the role of water during binding and conformational rearrangements preceding, concurrent or consecutive to binding.

However, the observation of a binding event during a classical MD simulation is very rare, raising the timescale problem. The timestep in molecular dynamics has to be compatible with the



fastest motion in the system; in particular, a timestep of 1–2 fs, corresponding to bond vibrations, has to be used. Thus, a high number of MD steps is required to simulate slow processes, such as large domain motions and binding ( $\mu$ s–ms) (Henzler-Wildman and Kern, 2007), making the computational effort really hard. In particular, slow timescale are linked to processes that require the overcoming of a high energy barrier (Henzler-Wildman and Kern, 2007), corresponding to low populated states in the conformational energy landscape; in this case the simulated system gets trapped in a local minimum, making classical MD inadequate to explore largely the conformational space.

### Advances in Classical MD Simulations

In 1998 Duan and Kollman performed the first 1  $\mu$ s simulation of a protein in explicit solvent, observing the folding of a 36-residue villin headpiece subdomain from a fully unfolded state. This simulation was two orders of magnitude longer than a state-of-the-art simulation of that period, and it was made possible by advances in massively parallel supercomputers and efficient parallelized codes, but still required 2 months of CPU (Central Processing Units) time (Duan and Kollman, 1998).

Specialized informatic infrastructures have also been designed specifically for MD calculations; for example, a supercomputer named Anton was conceived as a “computational microscope” and was developed with the idea to reach previously inaccessible simulation timescales within a reasonable computation time (Shaw et al., 2008). This machine allowed Shaw et al. to

characterize the folding of Fip35 WW domain from a fully extended state in a 100  $\mu$ s simulation and, in addition, to reach the millisecond timescale in a single simulation of BPTI in the folded-state (Shaw et al., 2010), followed recently by ubiquitin (Lindorff-Larsen et al., 2016). Moreover, with unbiased simulations in the order of ten microseconds, Shaw’s group could simulate the complete binding process of beta blockers and agonists to the  $\beta_2$ -adrenergic receptor (Dror et al., 2011) and kinase inhibitors to Src kinase (Shan et al., 2011).

As a drawback, the utilization of supercomputer is an expense that not many research groups can afford. Fortunately, the recent years have been characterized by the development of code able to exploit the speed of GPUs (Graphics Processing Units), which has given access to tera-scale performances with the use of a common workstation, and a consequent relatively low cost (Van Meel et al., 2008; Friedrichs et al., 2009; Harvey et al., 2009; Nobile et al., 2017). The architecture of a GPU is meant to parallelize a computation over thousands of cores, with all cores executing the same instructions on different data (“Same Instruction Multiple Data,” SIM) (Nobile et al., 2017). For this reason, together with few preliminary applications in the field of molecular docking (Korb et al., 2011; Khar et al., 2013), GPUs have been mainly exploited for MD simulations, which can be parallelized at the level of atoms. In fact, nowadays, simulations of hundreds of nanoseconds are easily performed, and reaching the microsecond timescale is an affordable issue on a GPU-equipped workstation (Harvey and De Fabritiis, 2012). In addition, cloud computing



has been emerging nowadays, not just through the use of web-servers intended to make molecular modeling accessible to a community of non-developers users, but also with the provision of computation power scalable and on-demand (Ebejer et al., 2013). As an example, AceCloud is an on-demand service for MD simulations, which is accessed through an extension of the ACEMD MD code (Harvey and De Fabritiis, 2015).

Moreover, a paradigm shift seems to have been spreading, that is the possibility to simulate long processes using numerous trajectories shorter than the process itself instead of a single long trajectory. This idea has been exploited by the folding@home project, a worldwide distributed computing environment benefitting from the computers of private citizens, when not in use (Shirts and Pande, 2000). Since during a classical MD simulation, the system is stuck in a minimum, waiting for the fortunate event that triggers the overcoming of an energy barrier, the simulation of many trajectories in parallel would increase the probability to meet the lucky event. Thus, numerous simulations are started from the same initial condition and run in parallel on different computers, and when one escapes from the energy minimum, all the simulations are stopped and started from the new productive configuration (Pande et al., 2003).

The new paradigm has found its best application in the use of Markov State Models (MSMs) and adaptive sampling. In fact, MSMs are based on an ensemble view of the dynamics, from which statistical properties, such as the probability to occupy a state and the probability to jump from one state to another, are computed. The construction of a Markov model is made of the discretization and projection of a trajectory into microstates, and of a transition probability matrix  $T(\tau)$  computation at a given time, the lag-time  $\tau$ , chosen in a way that the transition is memory-less (Markovian). Each element  $T_{ij}(\tau)$  of the transition matrix represents the conditional probability to find the system in state  $j$  at time  $t + \tau$  while being in state  $i$  at time  $t$ . The transition matrix approximates the dynamic of the system and enables to extrapolate the free energy from the equilibrium probability distribution of the system and the timescale of the slowest processes, even if they are not directly explored. In a qualitative fashion, the MSM may individuate diverse metastable states and construct multi-states models of the processes (Prinz et al., 2011). As an example, an MSM was constructed on an aggregate of nearly 500 100 ns-trajectories describing benzamidine-trypsin binding (with 37% productive trajectories); this enabled to characterize the binding process individuating three transition states, and to estimate binding free energy with 1 kcal/mol difference from the experimental one (while a higher deviation from experiment was associated with the extrapolated  $k_{on}$  and  $k_{off}$ ) (Buch et al., 2011). Moreover, the computation of MSM on the collected data can give a feedback about undersampled zones of the phase space, suggesting where to focus further simulation, adapting the sampling (adaptive sampling methods) and increasing the efficiency of simulations (Bowman et al., 2010; Doerr and De Fabritiis, 2014). Currently, the major difficulties of this technique are related to the trajectory partition into discrete states, the choice of the lagtime and sufficient sampling to guarantee statistical significance (Pande et al., 2010).

Several alternative techniques have been developed during the years to overcome the time limitation imposed by classical MD simulations. A first example consists of the Coarse-Grained MD simulations, in which groups of atoms are condensed into spheres, reducing the degrees of freedom of the system (Kmiciek et al., 2016). This simplifies the conformational landscape of the system, but, as a drawback, the information on the all-atom simulations, that are precious for drug-discovery aim, are lost.

Additional strategies consist of enhanced sampling techniques that apply a bias to molecular dynamics simulations to increase the accessible timescale, enabling the simulation of slow processes like binding, unbinding and folding processes in a reduced amount of time.

## Enhanced Sampling Techniques

These methods add a bias force/potential to the system to increase the rate of escape from local minima, entailing an acceleration of conformational sampling. They have been conceived primarily to study either folding or binding or unbinding processes, sharing the underlying idea of enhancement of sampling and overcoming high energy barriers.

Enhanced sampling techniques can be divided into methods that make use of collective variables to introduce the bias and methods that do not (De Vivo et al., 2016) (Figure 3).

The employment of a collective variable (CV) is based on the idea that a complex system can be decomposed into one or a combination of reaction coordinates describing the process of interest. These coordinates are named as collective variables since it is assumed they can summarize the behavior of the entire system. After a careful choice of the CVs, the bias is added on these coordinates during the simulation enhancing sampling along the CVs. The phase space is reduced to the space of the collective variables, since the conformational space is projected to the selected CVs, with a consequent dimensional reduction of the free energy surface.

In the following paragraphs, few representative enhanced sampling techniques are reported as an example, focusing on their application in binding and unbinding and going toward a fully dynamic docking (De Vivo and Cavalli, 2017).

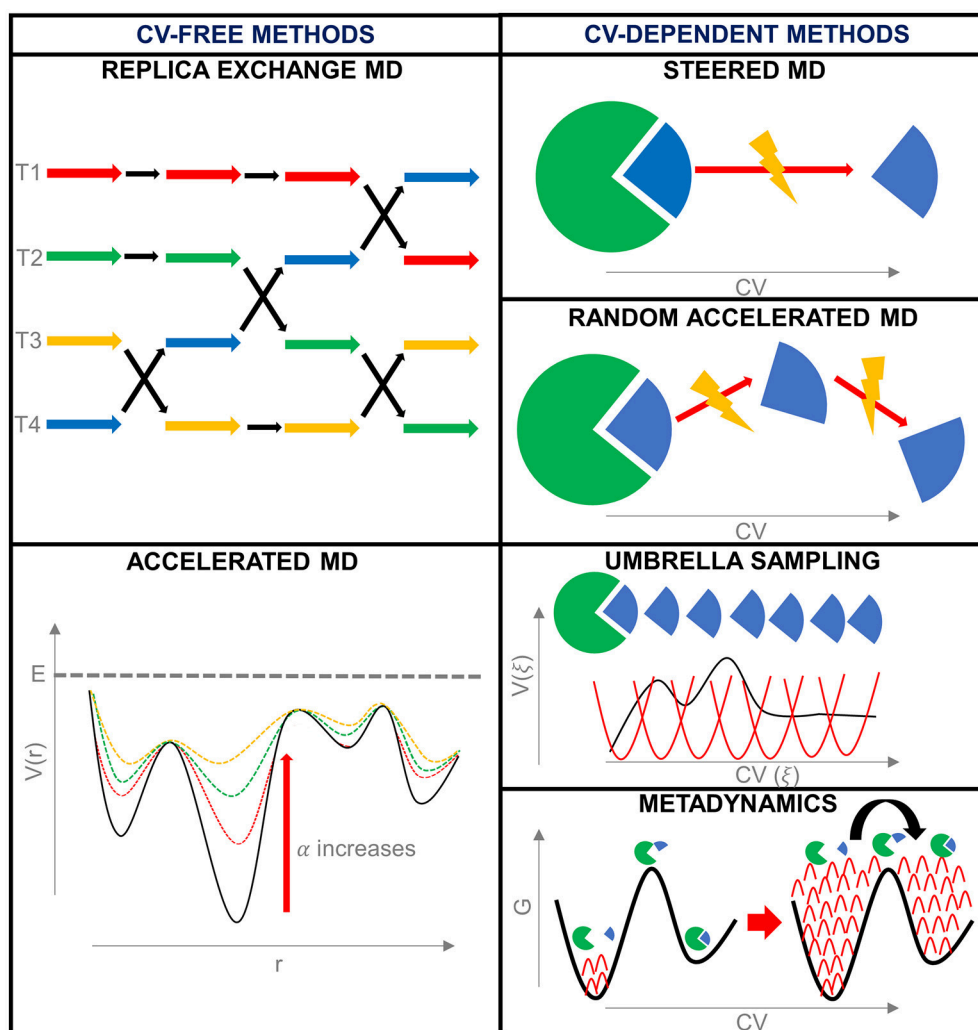
### Collective variables-free methods

**Replica Exchange Molecular Dynamics (REMD)** This method adopts an increase in temperature to accelerate the conformational sampling. The first formulation of Replica Exchange MD (Sugita and Okamoto, 1999), also known as Parallel Tempering (PT), consists of the parallel simulation of a number of independent and simultaneous replicas of the same system, starting from the same configuration, but at different temperatures. At regular time intervals, two replicas characterized by neighbor temperatures are switched, or, in other terms, their temperatures are exchanged, with a probability determined by the energy ( $E$ ) and temperature ( $T$ ) of the system. In particular, the transition probability between simulations at temperature  $T_1$  and  $T_2$  is determined by the Metropolis criterion:

$$P(T_1 \rightarrow T_2) = \begin{cases} 1 & \text{for } [\beta_2 - \beta_1](E_1 - E_2) \leq 0 \\ e^{-[\beta_2 - \beta_1](E_1 - E_2)} & \text{for } [\beta_2 - \beta_1](E_1 - E_2) > 0 \end{cases}$$

where  $\beta = 1/k_B T$  (with  $k_B$  the Boltzmann constant).





**FIGURE 3 |** Summary of the enhanced sampling techniques described during this work.

Temperatures are updated by rescaling the velocities of the parent simulations ( $v_1$  and  $v_2$  to  $v_1'$  and  $v_2'$ ) according to the following equation:

$$\begin{cases} v_1' = \sqrt{\frac{T_2}{T_1}} v_1 \\ v_2' = \sqrt{\frac{T_1}{T_2}} v_2 \end{cases}$$

The choice of the panel of temperatures is critical, and various strategies have been proposed to guide the selection (Patriksson and van der Spoel, 2008).

Further development of REMD has been introduced, such as the Hamiltonian Replica Exchange (H-REMD), where Hamiltonians are exchanged among replicas instead of temperatures (Fukunishi et al., 2002), and Replica Exchange with Solute Tempering, where a different treatment of the central group and the solvent buffer is performed (Liu et al., 2005). HREMD has been recently combined to conventional MD simulations using multi-ensemble Markov models (MEMMs) (Wu et al.,

2016) to investigate the multistate kinetics of Mdm2 and its inhibitor peptide PMI (Paul et al., 2017). An ensemble of 500  $\mu$ s unbiased MD simulations conducted from different initial states, especially dissociated, were combined to HREMD simulations (6 simulations of 1  $\mu$ s and with 14 replicas) to enhance sampling of rare dissociation events; the results were analyzed through the TRAMMBAR estimator, leading to the prediction of a residence time beyond the second timescale, despite a sub-millisecond simulation time. Moreover, the trajectories were further analyzed to investigate the binding mechanism and binding-induced folding of PMI (Paul et al., 2018). It appeared that a multitude of parallel pathways is possible and that binding and folding are coupled, while not temporarily ordered and separated.

**Accelerated Molecular Dynamics (aMD)** Accelerated MD (aMD) facilitates the egress from a low energy basin by adding a bias potential function ( $\Delta V(r)$ ) when the system is entrapped in an energy minimum. In particular, when the potential energy ( $V(r)$ ) is lower

than a certain cut-off ( $E$ ), the bias is added giving a modified potential ( $V^*(r)=V(r)+\Delta V(r)$ ); otherwise the simulation continues in the true-unbiased potential ( $V^*(r)=V(r)$ ).

The bias function is reported in the following equation:

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

where  $E$  is the potential energy cut-off and  $\alpha$  is a tuning parameter determining the depth of the modified potential energy basin.

$E$  has to be at least greater than  $V_{min}$  (the minimum potential energy, close to the starting configuration), while  $\alpha = E - V_{min}$  will allow maintaining the underlying shape of the landscape (Hamelberg et al., 2004).

As an example, aMD showed qualitatively similar results to classical MD with fewer computational effort in the simulation of tiotropium- $M_3$  Muscarinic Acetylcholine Receptor binding: tiotropium was observed to recognize the extracellular vestibule of the receptor, as in a previously reported long (16  $\mu$ s) classical MD simulation (Kruse et al., 2012), by accelerating the process of about one order of magnitude (three aMD replicas of 200 ns, 500 ns, and 1  $\mu$ s) (Kappel et al., 2015).

### Collective Variables-dependent methods

**Steered Molecular Dynamics (SMD)** Taking inspiration from atomic force microscopy experiments, in Steered MD (SMD) an external force is applied to a ligand to drive it out of the target binding site (Isralewitz et al., 1997, 2001; Izrailev et al., 1997). Other possibilities involve the application of forces on different CVs, such as nonlinear coordinates that can help to explore the conformational rearrangement of protein domains (Izrailev et al., 1999).

SMD gives insights into the ligand-target unbinding mechanism, which can be investigated through the dynamical evolution of the ligand-target pattern of interactions, as reported for a series of Cyclin-Dependent Kinase 5 (CDK5) inhibitors (Patel et al., 2014). In the same work, the second application of SMD in drug discovery is highlighted: since the bias force added during an SMD simulation is assumed to be related to the binding strength, the binding force profile can be used to discriminate binders from non-binders.

SMD relies on an a priori definition of the applied force direction, which can be fixed (for example a simple straight line) or can change during the simulation. The choice of the direction is not trivial, because a ligand may bump into obstructions during its way out of the protein, but a method evaluating the minimal steric hindrance has been reported (Vuong et al., 2015). Moreover, integration with the targeted molecular dynamics (TMD) are reported: in TMD a bias force is applied to conduct the system from an initial to a desired final configuration (Schlitter et al., 1993), leading to the individuation of a path that can be used as set of directions for an SMD simulation (Isralewitz et al., 2001).

**Random Acceleration Molecular Dynamics (RAMD)** Random Acceleration MD (RAMD), also defined Random Expulsion MD, is an extension of SMD, and, like this, was developed to study the egress of a ligand from its target binding site. It consists of the application of an artificial randomly-directed force on a ligand to accelerate its unbinding. In this way, in comparison with SMD, RAMD avoids the preliminary choice of the force direction; consequently, if some obstructions are found during the exit pathway, the escape direction is switched.

In particular, the direction of the force is chosen stochastically and maintained for a number of MD steps. If during this time interval the average velocity of the ligand is lower than a specified cut-off (or, in other terms, if the distance covered by the ligand is lower than a cut-off distance,  $r_{min}$ ), meaning that probably a rigid obstruction has been met, a new force direction is assigned to allow the ligand to search for alternative exit pathways (Lüdemann et al., 2000).

As SMD, RAMD is predominantly used to simulate ligand unbinding from a molecular target. The egress of carazolol from  $\beta_2$  Adrenergic Receptor was for example described thanks to an ensemble of RAMD simulations (100 simulation, with a variable length of maximum 1 ns): the extracellular surface opening of the receptor was individuated as the predominant exit root, entailing the rupture of a salt bridge linking extracellular loop 2 to transmembrane helix 7 (Wang and Duan, 2009).

**Umbrella Sampling (US)** Umbrella Sampling (US) (Torrie and Valleau, 1977) consists of restraining the system along one or a combination of CVs. Commonly, the range of interest of the CV is divided into windows, each characterized by a reference value of the CV ( $\xi_{ref}$ ). The bias potential enhances sampling in each window by forcing the system to stay close to the respective CV reference value. The bias is a function of the reaction coordinate, and can have different shapes, but generally consists of a simple harmonic, as in the following equation:

$$V(\xi) = \frac{k}{2} (\xi - \xi_{ref})^2$$

Where  $k$  is the strength of the potential and  $\xi$  is the value of the CV.

The strength of the bias has to be high enough to let energy barriers crossing, but sufficiently low to enable the overlapping of system distributions of different windows, as required for post-processing analysis.

The aim of US is to force sampling in each window to collect sufficient statistics along with the whole reaction coordinate. Then the distribution of the system and consequently the free energy is calculated along the CV (Kästner, 2011). Different post-processing methods can be used to perform combination and analysis of the data coming from the different US windows; the most famous is umbrella integration (Kästner and Thiel, 2005), the weighted histogram analysis method (WHAM) (Kumar et al., 1992), and the more recent Dynamic Weighted Histogram Analysis (DHAM) (Rosta and Hummer, 2015), which can be used also to derive kinetic parameters.

Integrations of US with other enhanced sampling techniques are reported in the literature, such as the replica-exchange umbrella sampling method (REUS), where an umbrella potential is exchanged among replicas (Sugita et al., 2000; Kokubo et al., 2011). This technique was applied to the prediction of ligand-protein binding structures, starting from unbound initial states and employing as CV  $\xi$  the distance between the centers of mass of the ligand and of the backbone of two selected residues. This technique resulted to be effective in the prediction of the binding mode of a couple of ligands on p38 and JNK3 kinases (RMSD minor than 1.7 Å), and outperformed a cross-docking experiment, highlighting the importance of considering protein flexibility to accurately predict the coordinates of a complex (Kokubo et al., 2013).

**Metadynamics** Metadynamics (Laio and Parrinello, 2002) introduces a bias potential to the Hamiltonian of the system in the form of a Gaussian-shaped function of one or more CVs. In this case, the bias does not restrain or constrain the system, neither force the system along with a preferred direction in the CV space. The bias is used to keep the memory of the already explored zones of the phase space, and to discourage the system to visit them again (Laio and Gervasio, 2008).

At time  $t$ , the bias potential ( $V_G(S, t)$ ) is reported in the following equation:

$$V_G(S, t) = \int_0^t dt' \omega \exp \left( - \sum_{i=1}^d \frac{(S_i(R) - S_i(R(t')))^2}{2\sigma_i^2} \right)$$

where  $S(R) = (S_1(R), \dots, S_d(R))$  is a set of  $d$  CVs (which are functions of the coordinates  $R$  of the system),  $S_i(R(t))$  is the value of the  $i$ th CV at time  $t$ ,  $\sigma_i$  is the Gaussian width for the  $i$ th CV, and  $\omega$  is the energy rate, given by:

$$\omega = \frac{W}{\tau_G}$$

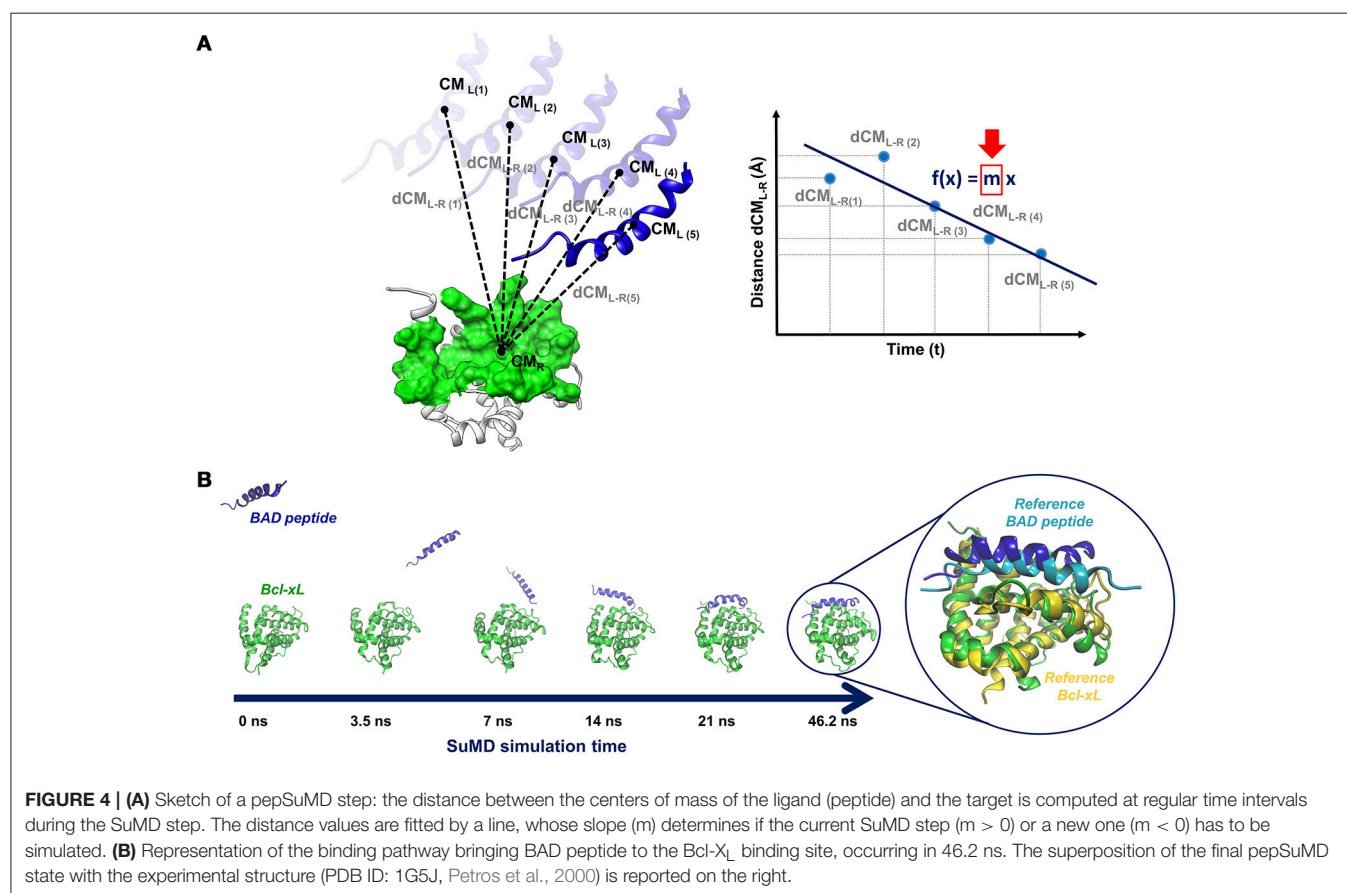
with  $W$  the Gaussian height and  $\tau_G$  the deposition rate.

Thus, the bias is “history-dependent,” because it is the sum of the Gaussians that have already been deposited in the CV space during the time.

The free energy landscape is explored, starting from the bottom of a well, by a random walk; bias-Gaussians are deposited in the CV space with a given frequency, and at each iteration, the bias is given by the sum of the already deposited Gaussians. As time goes by, the system, instead of being trapped in the bottom of a well, is pushed out by the hill of deposited Gaussians and enters a new minimum. The process continues until all the minima are compensated by the bias potential (Barducci et al., 2011).

Metadynamics in this way enables to enhance sampling and to reconstruct the free energy surface; this can be used to explore binding/unbinding processes (Gervasio et al., 2005), and, with the application of funnel metadynamics (Limongelli et al., 2013), to the estimation of binding free energy.

Unfortunately, it may occur that the free energy surface is overfilled, but this has been partially solved by well-tempered metadynamics, in which the height of the added Gaussian is rescaled by the already deposited bias (Barducci et al., 2008). Another issue with metadynamics is the choice of the CVs, which should describe the slowest motions of the system and the initial-final-relevant intermediates. Moreover, a small number of CVs has to be used, and a good strategy is a combination with other techniques able to enhanced sampling along a great number of transverse coordinates (Barducci et al., 2011), such as with parallel tempering (Bussi et al., 2006). Using a well-tempered multiple-walker funnel-restrained metadynamics, the binding pathway of several ligands to 5 G-protein-coupled receptors (including X-ray crystal structures and homology models) has been recently explored, resulting in the prediction of





binding free energies with a root-mean-square error minor than 1 kcal mol<sup>-1</sup> (Saleh et al., 2017).

## Supervised Molecular Dynamics

In the last years, a new method, called Supervised Molecular Dynamics (SuMD), has been introduced to accelerate the binding process (Sabbadin and Moro, 2014; Cuzzolin et al., 2016). SuMD is distinguished from enhanced sampling simulations since it does not affect the energy profile of the system.

A SuMD simulation consists of a series of small MD windows (hundreds of picoseconds), called SuMD steps, where step  $n+1$  is run after the evaluation of step  $n$  in terms of ligand-target approaching. During each SuMD step, the distance between the centers of mass of the ligand and of the target binding site (few selected residues) is computed; distance values are collected at regular intervals during the simulation and are fitted by a line (Figure 4A). If the slope of the line is negative, it means that the ligand is approaching the binding site, the SuMD step (step  $n$ ) is considered productive, and a new step (step  $n+1$ ) is started from the last coordinates and velocities of the current step. Otherwise, if the slope is positive, it means that the SuMD step is unproductive, thus the current SuMD step simulation is deleted and restarted from its initial coordinates (starting configuration of step  $n$ ). The simulation is concluded after that the distance between the centers of mass of ligand and target fall under a certain cut-off. Finally, the consecutive SuMD steps are merged together providing the SuMD trajectory.

In this way, SuMD enables to observe a binding event in a reduced timescale, in the orders of tens to hundreds of nanoseconds, without the introduction of any energetic bias. Indeed, SuMD simply focuses sampling by the introduction of a tabu-like algorithm which favors the progress of a simulation toward productive events and avoids wasting simulation time in uninteresting portions of the search space.

Certainly, a single SuMD trajectory is not sufficient to explain the complex binding process, and the retrieval of thermodynamic quantities from a single simulation must be avoided. Nevertheless, a SuMD trajectory depicts one of the possible binding pathways leading a ligand to reach the target, so it can be useful to propose a mechanistic hypothesis.

The technique was first applied to Adenosine Receptors, where it facilitated the characterization of the binding pathways of several ligands toward the receptor, with the exploration of metabinding sites (Sabbadin and Moro, 2014; Sabbadin et al., 2015). In this context, SuMD can be useful in the interpretation of allosteric interactions (Deganutti et al., 2015) and has proved to be supportive to the identification of fragment-like positive allosteric modulators (Deganutti and Moro, 2017). In fact, SuMD turned out to be effective in simulating fragment compounds, as shown by the accurate prediction of the binding mode of a catechol fragment to human peroxiredoxin 5 (PRDX5), reaching a minimum RMSD of 0.7 Å from the crystallographic pose.

The applicability spectrum of SuMD has been furtherly enlarged, till the development of pepSuMD, a revised version of the technique

able to simulate the binding pathway of a peptide ligand toward its protein binding site (Salmaso et al., 2017). The recognition process of the BAD peptide to Bcl-X<sub>L</sub> protein (Figure 4B) and of the p53 peptide to MDM2 has been recently reported, with the achievement of an RMSD less than 5 Å from the experimental conformation in tens of nanoseconds in both cases (46.2 and 23.40 ns, respectively). During the BAD/Bcl-X<sub>L</sub> simulation, the C-terminal helix explored different conformations, meaning that peptide and protein conformational rearrangements can be observed during a SuMD simulation when occurring in the same time scale of the SuMD-accelerated binding.

## CONCLUSIONS AND PERSPECTIVES

In this review, an excursus over some relevant computational techniques in drug discovery has been performed, highlighting how protein flexibility has been introduced into the simulations during the years. Starting from simple rigid docking strategies justified by the lock-key model, it was soon necessary to consider conformational degrees of freedom of ligands during docking. Experimental data proving the existence of different conformations of protein structures has made the molecular models to face the problem of interpreting and simulating conformational transitions of macromolecules.

From rough attempts to include protein flexibility during classical molecular docking, the development of hardware technologies and of novel MD computational techniques has been allowing more and more to simulate huge conformational movements. The possibility to simulate contemporary folding and binding phenomena can be exploited to answer the long-standing debate about “induced-fit” and “conformational selection” binding models, by giving a mechanistic interpretation of binding pathways.

Moreover, some of the enhanced sampling techniques are no more an exclusive methodological exercise, but has become within reach of many research groups, with a consequent real applicability in drug discovery.

## AUTHOR CONTRIBUTIONS

VS and SM devised the organization, the main conceptual ideas, proof outline and wrote the review. The content of the present work has been largely taken from the PhD thesis entitled Exploring protein flexibility during docking to investigate ligand-target recognition written by VS under the supervision of SM.

## ACKNOWLEDGMENTS

MMS lab is very grateful to Chemical Computing Group, OpenEye, and Acellera for the scientific and technical partnership. MMS lab gratefully acknowledges the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

## REFERENCES

- Abagyan, R., and Totrov, M. (2001). High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* 5, 375–382. doi: 10.1016/S1367-5931(00)00217-9
- Abagyan, R., Totrov, M., and Kuznetsov, D. (1994). ICM? A new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* 15, 488–506. doi: 10.1002/jcc.540150503
- Adcock, S. A., and McCammon, J. A. (2006). Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* 106, 1589–1615. doi: 10.1021/cr040426m

- Alder, B. J., and Wainwright, T. E. (1957). Phase transition for a hard sphere system. *J. Chem. Phys.* 27, 1208–1209. doi: 10.1063/1.1743957
- Alonso, H., Bliznyuk, A. A., and Gready, J. E. (2006). Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* 26, 531–568. doi: 10.1002/med.20067
- Amaro, R. E., Baron, R., and McCammon, J. A. (2008). An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided Mol. Des.* 22, 693–705. doi: 10.1007/s10822-007-9159-2
- Apostolakis, J., Pluckthun, A., and Caflisch, A. (1998). Docking small ligands in flexible binding sites. *J. Comput. Chem.* 19, 21–37. doi: 10.1002/(SICI)1096-987X(19980115)19:1<21::AID-JCC2>3.0.CO;2-0
- Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H., and Gunsalus, I. C. (1975). Dynamics of ligand binding to myoglobin. *Biochemistry* 14, 5355–5373. doi: 10.1021/bi00695a021
- Barducci, A., Bonomi, M., and Parrinello, M. (2011). Metadynamics. *WIREs Comput. Mol. Sci.* 1, 826–843. doi: 10.1002/wcms.31
- Barducci, A., Bussi, G., and Parrinello, M. (2008). Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* 100:020603. doi: 10.1103/PhysRevLett.100.020603
- Baxter, C. A., Murray, C. W., Clark, D. E., Westhead, D. R., and Eldridge, M. D. (1998). Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins* 33, 367–382. doi: 10.1002/(SICI)1097-0134(19981115)33:3<367::AID-PROTE6>3.0.CO;2-W
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235
- Betz, M., Wulsdorf, T., Krimmer, S. G., and Klebe, G. (2016). Impact of surface water layers on protein–ligand binding: how well are experimental data reproduced by molecular dynamics simulations in a thermolysin test case?. *J. Chem. Inf. Model.* 56, 223–233. doi: 10.1021/acs.jcim.5b00621
- Böhm, H. J. (1994). The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* 8, 243–256. doi: 10.1007/BF00126743
- Born, M., and Oppenheimer, R. (1927). Zur quantentheorie der molekeln. *Ann. Phys.* 389, 457–484. doi: 10.1002/andp.19273892002
- Bortolato, A., Tehan, B. G., Bodnarchuk, M. S., Essex, J. W., and Mason, J. S. (2013). Water network perturbation in ligand binding: adenosine A(2A) antagonists as a case study. *J. Chem. Inf. Model.* 53, 1700–1713. doi: 10.1021/ci4001458
- Bowman, G. R., Ensign, D. L., and Pande, V. S. (2010). Enhanced modeling via network theory: adaptive sampling of Markov state models. *J. Chem. Theory Comput.* 6, 787–794. doi: 10.1021/ct900620b
- Brooijmans, N., and Kuntz, I. D. (2003). Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* 32, 335–373. doi: 10.1146/annurev.biophys.32.110601.142532
- Buch, I., Giorgino, T., and De Fabritiis, G. (2011). Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10184–10189. doi: 10.1073/pnas.1103547108
- Bussi, G., Gervasio, F. L., Laio, A., and Parrinello, M. (2006). Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* 128, 13435–13441. doi: 10.1021/ja062463w
- Charifson, P. S., Corkery, J. J., Murcko, M. A., and Walters, W. P. (1999). Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* 42, 5100–5109. doi: 10.1021/jm990352k
- Chen, H.-M., Liu, B.-F., Huang, H.-L., Hwang, S.-F., and Ho, S.-Y. (2007). SODOCK: swarm optimization for highly flexible protein-ligand docking. *J. Comput. Chem.* 28, 612–623. doi: 10.1002/jcc.20542
- Corbeil, C. R., Williams, C. I., and Labute, P. (2012). Variability in docking success rates due to dataset preparation. *J. Comput. Aided Mol. Des.* 26, 775–786. doi: 10.1007/s10822-012-9570-1
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., et al. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117, 5179–5197. doi: 10.1021/ja00124a002
- Csermely, P., Palotai, R., and Nussinov, R. (2010). Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* 35, 539–546. doi: 10.1016/j.tibs.2010.04.009
- Cuzzolin, A., Deganutti, G., Salmasso, V., Sturlese, M., and Moro, S. (2018). AquaMMaP: an alternative tool to monitor the role of water molecules during protein-ligand association. *ChemMedChem* 13, 522–531. doi: 10.1002/cmdc.201700564
- Cuzzolin, A., Sturlese, M., Deganutti, G., Salmasso, V., Sabbadin, D., Ciancetta, A., et al. (2016). Deciphering the complexity of ligand-protein recognition pathways using supervised molecular dynamics (SuMD) simulations. *J. Chem. Inf. Model.* 56, 687–705. doi: 10.1021/acs.jcim.5b00702
- De Vivo, M., and Cavalli, A. (2017). Recent advances in dynamic docking for drug discovery. *WIREs Comput. Mol. Sci.* 7:e1320. doi: 10.1002/wcms.1320
- De Vivo, M., Masetti, M., Bottegoni, G., and Cavalli, A. (2016). Role of molecular dynamics and related methods in drug discovery. *J. Med. Chem.* 59, 4035–4061. doi: 10.1021/acs.jmedchem.5b01684
- Deganutti, G., Cuzzolin, A., Ciancetta, A., and Moro, S. (2015). Understanding allosteric interactions in G protein-coupled receptors using supervised molecular dynamics: a prototype study analysing the human A3 adenosine receptor positive allosteric modulator LUF6000. *Bioorg. Med. Chem.* 23, 4065–4071. doi: 10.1016/j.bmc.2015.03.039
- Deganutti, G., and Moro, S. (2017). Supporting the identification of novel fragment-based positive allosteric modulators using a supervised molecular dynamics approach: a retrospective analysis considering the human A2A adenosine receptor as a key example. *Molecules* 22:818. doi: 10.3390/molecules22050818
- Desjarlais, R. L., Sheridan, R. P., Dixon, J. S., Kuntz, I. D., and Venkataraghavan, R. (1986). Docking flexible ligands to macromolecular receptors by molecular shape. *J. Med. Chem.* 29, 2149–2153. doi: 10.1021/jm00161a004
- Doerr, S., and De Fabritiis, G. (2014). On-the-fly learning and sampling of ligand binding by high-throughput molecular simulations. *J. Chem. Theory Comput.* 10, 2064–2069. doi: 10.1021/ct400919u
- Dror, R. O., Pan, A. C., Arlow, D. H., Borhani, D. W., Maragakis, P., Shan, Y., et al. (2011). Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13118–13123. doi: 10.1073/pnas.1104614108
- Duan, Y., and Kollman, P. A. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282, 740–744. doi: 10.1126/science.282.5389.740
- Durrant, J. D., and McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol.* 9:71. doi: 10.1186/1741-7007-9-71
- Durrant, J. D., Urbaniak, M. D., Ferguson, M. A., and McCammon, J. A. (2010). Computer-aided identification of Trypanosoma brucei uridine diphosphate galactose 4'-epimerase inhibitors: toward the development of novel therapies for African sleeping sickness. *J. Med. Chem.* 53, 5025–5032. doi: 10.1021/jm100456a
- Ebejer, J.-P., Fulle, S., Morris, G. M., and Finn, P. W. (2013). The emerging role of cloud computing in molecular modelling. *J. Mol. Graph. Model.* 44, 177–187. doi: 10.1016/j.jmgm.2013.06.002
- Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., and Mee, R. P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* 11, 425–445. doi: 10.1023/A:1007996124545
- Fischer, E. (1894). Einfluss der configuration auf die wirkung der enzyme. *Ber. Dtsch. Chemischen Ges.* 27, 2985–2993. doi: 10.1002/cber.18940270364
- Foot, J., and Milstein, C. (1994). Conformational isomerism and the diversity of antibodies. *Proc. Natl. Acad. Sci. U.S.A.* 91, 10370–10374. doi: 10.1073/pnas.91.22.10370
- Frauenfelder, H., Sligar, S. G., and Wolynes, P. G. (1991). The energy landscapes and motions of proteins. *Science* 254, 1598–1603. doi: 10.1126/science.1749933
- Friedrichs, M. S., Eastman, P., Vaidyanathan, V., Houston, M., Legrand, S., Beberg, A. L., et al. (2009). Accelerating molecular dynamic simulation on graphics processing units. *J. Comput. Chem.* 30, 864–872. doi: 10.1002/jcc.21209
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749. doi: 10.1021/jm0306430

- Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., et al. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* 49, 6177–6196. doi: 10.1021/jm051256o
- Fukunishi, H., Watanabe, O., and Takada, S. (2002). On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction. *J. Chem. Phys.* 116:9058. doi: 10.1063/1.1472510
- Ganser, L. R., Lee, J., Rangadurai, A., Merriman, D. K., Kelly, M. L., Kansal, A. D., et al. (2018). High-performance virtual screening by targeting a high-resolution RNA dynamic ensemble. *Nat. Struct. Mol. Biol.* 25, 425–434. doi: 10.1038/s41594-018-0062-4
- Gervasio, F. L., Laio, A., and Parrinello, M. (2005). Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* 127, 2600–2607. doi: 10.1021/ja0445950
- Gohlke, H., Hendlich, M., and Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* 295, 337–356. doi: 10.1006/jmbi.1999.3371
- Goodsell, D. S., and Olson, A. J. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins* 8, 195–202. doi: 10.1002/prot.340080302
- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* 47, 1750–1759. doi: 10.1021/jm030644s
- Halperin, I., Ma, B., Wolfson, H., and Nussinov, R. (2002). Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins* 47, 409–443. doi: 10.1002/prot.10115
- Hamelberg, D., Mongan, J., and McCammon, J. A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120, 11919–11929. doi: 10.1063/1.1755656
- Harvey, M. J., and De Fabritiis, G. (2012). High-throughput molecular dynamics: the powerful new tool for drug discovery. *Drug Discov. Today* 17, 1059–1062. doi: 10.1016/j.drudis.2012.03.017
- Harvey, M. J., and De Fabritiis, G. (2015). Acecloud: molecular dynamics simulations in the cloud. *J. Chem. Inf. Model.* 55, 909–914. doi: 10.1021/acs.jcim.5b00086
- Harvey, M. J., Giupponi, G., and Fabritiis, G. D. (2009). ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* 5, 1632–1639. doi: 10.1021/ct9000685
- Henzler-Wildman, K., and Kern, D. (2007). Dynamic personalities of proteins. *Nature* 450, 964–972. doi: 10.1038/nature06522
- Huang, S.-Y., and Zou, X. (2007). Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* 66, 399–421. doi: 10.1002/prot.21214
- Huang, S.-Y., and Zou, X. (2010). Advances and challenges in protein-ligand docking. *Int. J. Mol. Sci.* 11, 3016–3034. doi: 10.3390/ijms11083016
- Huey, R., Morris, G. M., Olson, A. J., and Goodsell, D. S. (2007). A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* 28, 1145–1152. doi: 10.1002/jcc.20634
- Isralewitz, B., Gao, M., and Schulten, K. (2001). Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* 11, 224–230. doi: 10.1016/S0959-440X(00)00194-9
- Isralewitz, B., Izrailev, S., and Schulten, K. (1997). Binding pathway of retinal to bacterio-opsin: a prediction by molecular dynamics simulations. *Biophys. J.* 73, 2972–2979. doi: 10.1016/S0006-3495(97)78326-7
- Izrailev, S., Crofts, A. R., Berry, E. A., and Schulten, K. (1999). Steered molecular dynamics simulation of the Rieske subunit motion in the cytochrome bc(1) complex. *Biophys. J.* 77, 1753–1768. doi: 10.1016/S0006-3495(99)77022-0
- Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y., and Schulten, K. (1997). Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* 72, 1568–1581. doi: 10.1016/S0006-3495(97)78804-0
- Jiang, F., and Kim, S. H. (1991). “Soft docking”: matching of molecular surface cubes. *J. Mol. Biol.* 219, 79–102. doi: 10.1016/0022-2836(91)90859-5
- Jones, G., Willett, P., and Glen, R. C. (1995). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* 245, 43–53. doi: 10.1016/S0022-2836(95)80037-9
- Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267, 727–748. doi: 10.1006/jmbi.1996.0897
- Jorgensen, W. L., and Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110, 1657–1666. doi: 10.1021/ja00214a001
- Kappel, K., Miao, Y., and McCammon, J. A. (2015). Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor. *Q. Rev. Biophys.* 48, 479–487. doi: 10.1017/S0033583515000153
- Kästner, J. (2011). Umbrella sampling. *WIREs Comput. Mol. Sci.* 1, 932–942. doi: 10.1002/wcms.66
- Kästner, J., and Thiel, W. (2005). Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: “Umbrella integration”. *J. Chem. Phys.* 123:144104. doi: 10.1063/1.2052648
- Khar, K. R., Goldschmidt, L., and Karanickolas, J. (2013). Fast docking on graphics processing units via Ray-Casting. *PLoS ONE* 8:e70661. doi: 10.1371/journal.pone.0070661
- Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 3, 935–949. doi: 10.1038/nrd1549
- Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., and Kolinski, A. (2016). Coarse-grained protein models and their applications. *Chem. Rev.* 116, 7898–7936. doi: 10.1021/acs.chemrev.6b00163
- Knegtel, R. M., Kuntz, I. D., and Oshiro, C. M. (1997). Molecular docking to ensembles of protein structures. *J. Mol. Biol.* 266, 424–440. doi: 10.1006/jmbi.1996.0776
- Kobilka, B. K., and Deupi, X. (2007). Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* 28, 397–406. doi: 10.1016/j.tips.2007.06.003
- Kokubo, H., Tanaka, T., and Okamoto, Y. (2011). Ab initio prediction of protein-ligand binding structures by replica-exchange umbrella sampling simulations. *J. Comput. Chem.* 32, 2810–2821. doi: 10.1002/jcc.21860
- Kokubo, H., Tanaka, T., and Okamoto, Y. (2013). Prediction of protein-ligand binding structures by replica-exchange umbrella sampling simulations: application to kinase systems. *J. Chem. Theory Comput.* 9, 4660–4671. doi: 10.1021/ct4004383
- Korb, O., Stützle, T., and Exner, T. E. (2006). “PLANTS: application of ant colony optimization to structure-based drug design,” in *Ant Colony Optimization and Swarm Intelligence*, eds M. Dorigo, L. M. Gambardella, M. Birattari, A. Martinoli, R. Poli, and T. Stützle (Berlin; Heidelberg: Springer), 247–258.
- Korb, O., Stützle, T., and Exner, T. E. (2009). Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J. Chem. Inf. Model.* 49, 84–96. doi: 10.1021/ci800298z
- Korb, O., Stützle, T., and Exner, T. E. (2011). Accelerating molecular docking calculations using graphics processing units. *J. Chem. Inf. Model.* 51, 865–876. doi: 10.1021/ci100459b
- Koshland, D. E. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 44, 98–104. doi: 10.1073/pnas.44.2.98
- Kruse, A. C., Hu, J., Pan, A. C., Arlow, D. H., Rosenbaum, D. M., Rosemond, E., et al. (2012). Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* 482, 552–556. doi: 10.1038/nature10867
- Kumar, S., Ma, B., Tsai, C. J., Sinha, N., and Nussinov, R. (2000). Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.* 9, 10–19. doi: 10.1110/ps.9.1.10
- Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1992). THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13, 1011–1021. doi: 10.1002/jcc.540130812
- Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., and Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* 161, 269–288. doi: 10.1016/0022-2836(82)90153-X
- Laio, A., and Gervasio, F. L. (2008). Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* 71:126601. doi: 10.1088/0034-4885/71/12/126601
- Laio, A., and Parrinello, M. (2002). Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12562–12566. doi: 10.1073/pnas.202427399
- Leach, A. R. (1994). Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* 235, 345–356. doi: 10.1016/S0022-2836(05)80038-5



- Leach, A. R. (2001). *Molecular Modelling: Principles and Applications*. illustrated. Pearson Education. Available online at: [https://books.google.it/books/about/Molecular\\_Modelling.html?id=kB7jsbV-uhkC&redir\\_esc=y](https://books.google.it/books/about/Molecular_Modelling.html?id=kB7jsbV-uhkC&redir_esc=y)
- Li, Z., and Lazaridis, T. (2003). Thermodynamic contributions of the ordered water molecule in HIV-1 protease. *J. Am. Chem. Soc.* 125, 6636–6637. doi: 10.1021/ja0299203
- Limongelli, V., Bonomi, M., and Parrinello, M. (2013). Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6358–6363. doi: 10.1073/pnas.1303186110
- Lin, J.-H., Perryman, A. L., Schames, J. R., and McCammon, J. A. (2002). Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J. Am. Chem. Soc.* 124, 5632–5633. doi: 10.1021/ja0260162
- Lin, J.-H., Perryman, A. L., Schames, J. R., and McCammon, J. A. (2003). The relaxed complex method: accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers* 68, 47–62. doi: 10.1002/bip.10218
- Lindorff-Larsen, K., Maragakis, P., Piana, S., and Shaw, D. E. (2016). Picosecond to millisecond structural dynamics in human ubiquitin. *J. Phys. Chem. B* 120, 8313–8320. doi: 10.1021/acs.jpcb.6b02024
- Liu, M., and Wang, S. (1999). MCDock: a Monte Carlo simulation approach to the molecular docking problem. *J. Comput. Aided Mol. Des.* 13, 435–451. doi: 10.1023/A:1008005918983
- Liu, P., Kim, B., Friesner, R. A., and Berne, B. J. (2005). Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13749–13754. doi: 10.1073/pnas.0506346102
- Lüdemann, S. K., Lounnas, V., and Wade, R. C. (2000). How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J. Mol. Biol.* 303, 797–811. doi: 10.1006/jmbi.2000.4154
- MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616. doi: 10.1021/jp973084f
- McCammon, J. A., Gelin, B. R., and Karplus, M. (1977). Dynamics of folded proteins. *Nature* 267, 585–590. doi: 10.1038/267585a0
- McMartin, C., and Bohacek, R. S. (1997). QXP: powerful, rapid computer algorithms for structure-based drug design. *J. Comput. Aided Mol. Des.* 11, 333–344. doi: 10.1023/A:1007907728892
- Meiler, J., and Baker, D. (2006). ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins* 65, 538–548. doi: 10.1002/prot.21086
- Miller, D. W., and Dill, K. A. (1997). Ligand binding to proteins: the binding landscape model. *Protein Sci.* 6, 2166–2179. doi: 10.1002/pro.5560061011
- Miller, M. D., Kearsley, S. K., Underwood, D. J., and Sheridan, R. P. (1994). FLOG: a system to select “quasi-flexible” ligands complementary to a receptor of known three-dimensional structure. *J. Comput. Aided Mol. Des.* 8, 153–174.
- Monod, J., Wyman, J., and Changeux, J. P. (1965). ON THE NATURE OF ALLOSTERIC TRANSITIONS: A PLAUSIBLE MODEL. *J. Mol. Biol.* 12, 88–118. doi: 10.1016/S0022-2836(65)80285-6
- Mooij, W. T., and Verdonk, M. L. (2005). General and targeted statistical potentials for protein-ligand interactions. *Proteins* 61, 272–287. doi: 10.1002/prot.20588
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., et al. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662. doi: 10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B
- Morris, G. M., Goodsell, D. S., Huey, R., and Olson, A. J. (1996). Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J. Comput. Aided Mol. Des.* 10, 293–304. doi: 10.1007/BF00124499
- Namasivayam, V., and Günther, R. (2007). pso@autodock: a fast flexible molecular docking program based on Swarm intelligence. *Chem. Biol. Drug Des.* 70, 475–484. doi: 10.1111/j.1747-0285.2007.00588.x
- Ng, M. C., Fong, S., and Siu, S. W. (2015). PSOVina: the hybrid particle swarm optimization algorithm for protein-ligand docking. *J. Bioinform. Comput. Biol.* 13:1541007. doi: 10.1142/S0219720015410073
- Nobile, M. S., Cazzaniga, P., Tangherloni, A., and Besozzi, D. (2017). Graphics processing units in bioinformatics, computational biology and systems biology. *Brief Bioinform.* 18, 870–885. doi: 10.1093/bib/bbw058
- Okazaki, K., and Takada, S. (2008). Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11182–11187. doi: 10.1073/pnas.0802524105
- Oostenbrink, C., Villa, A., Mark, A. E., and van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* 25, 1656–1676. doi: 10.1002/jcc.20090
- Osguthorpe, D. J., Sherman, W., and Hagler, A. T. (2012). Exploring protein flexibility: incorporating structural ensembles from crystal structures and simulation into virtual screening protocols. *J. Phys. Chem. B* 116, 6952–6959. doi: 10.1021/jp3003992
- Pande, V. S., Baker, I., Chapman, J., Elmer, S. P., Khaliq, S., Larson, S. M., et al. (2003). Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers* 68, 91–109. doi: 10.1002/bip.10219
- Pande, V. S., Beauchamp, K., and Bowman, G. R. (2010). Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* 52, 99–105. doi: 10.1016/j.ymeth.2010.06.002
- Pang, Y.-P., Perola, E., Xu, K., and Prendergast, F. G. (2001). EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J. Comput. Chem.* 22, 1750–1771. doi: 10.1002/jcc.1129
- Patel, J. S., Berteotti, A., Ronsisvalle, S., Rocchia, W., and Cavalli, A. (2014). Steered molecular dynamics simulations for studying protein-ligand interaction in cyclin-dependent kinase 5. *J. Chem. Inf. Model.* 54, 470–480. doi: 10.1021/ci4003574
- Patriksson, A., and van der Spoel, D. (2008). A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* 10, 2073–2077. doi: 10.1039/b716554d
- Paul, F., Noé, F., and Weikl, T. R. (2018). Identifying conformational-selection and induced-fit aspects in the binding-induced folding of PMI from Markov state modeling of atomistic simulations. *J. Phys. Chem. B* 122, 5649–5656. doi: 10.1021/acs.jpcb.7b12146
- Paul, F., Wehmeyer, C., Abualrous, E. T., Wu, H., Crabtree, M. D., Schöneberg, J., et al. (2017). Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations. *Nat. Commun.* 8:1095. doi: 10.1038/s41467-017-01163-6
- Pei, J., Wang, Q., Liu, Z., Li, Q., Yang, K., and Lai, L. (2006). PSI-DOCK: towards highly efficient and accurate flexible ligand docking. *Proteins* 62, 934–946. doi: 10.1002/prot.20790
- Petros, A. M., Nettesheim, D. G., Wang, Y., Olejniczak, E. T., Meadows, R. P., Mack, J., et al. (2000). Rationale for Bcl-xL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. *Protein Sci.* 9, 2528–2534. doi: 10.1110/ps.9.12.2528
- Prinz, J.-H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., et al. (2011). Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* 134:174105. doi: 10.1063/1.3565032
- Rarey, M., Kramer, B., Lengauer, T., and Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 261, 470–489. doi: 10.1006/jmbi.1996.0477
- Rastelli, G., Degliesposti, G., Del Rio, A., and Sgobba, M. (2009). Binding estimation after refinement, a new automated procedure for the refinement and rescoring of docked ligands in virtual screening. *Chem. Biol. Drug Des.* 73, 283–286. doi: 10.1111/j.1747-0285.2009.00780.x
- Rosta, E., and Hummer, G. (2015). Free energies from dynamic weighted histogram analysis using unbiased Markov state model. *J. Chem. Theory Comput.* 11, 276–285. doi: 10.1021/ct500719p
- Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., Garmendia-Doval, A. B., Juhos, S., Schmidtke, P., et al. (2014). rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput. Biol.* 10:e1003571. doi: 10.1371/journal.pcbi.1003571
- Sabbadin, D., Cianchetta, A., Deganutti, G., Cuzzolin, A., and Moro, S. (2015). Exploring the recognition pathway at the human A2A adenosine receptor

- of the endogenous agonist adenosine using supervised molecular dynamics simulations. *Medchemcomm* 6, 1081–1085. doi: 10.1039/C5MD00016E
- Sabbadin, D., Ciancetta, A., and Moro, S. (2014). Bridging molecular docking to membrane molecular dynamics to investigate GPCR-ligand recognition: the human A2A adenosine receptor as a key study. *J. Chem. Inf. Model.* 54, 169–183. doi: 10.1021/ci400532b
- Sabbadin, D., and Moro, S. (2014). Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale. *J. Chem. Inf. Model.* 54, 372–376. doi: 10.1021/ci400766b
- Saleh, N., Ibrahim, P., Saladino, G., Gervasio, F. L., and Clark, T. (2017). An efficient metadynamics-based protocol to model the binding affinity and the transition state ensemble of G-protein-coupled receptor ligands. *J. Chem. Inf. Model.* 57, 1210–1217. doi: 10.1021/acs.jcim.6b00772
- Salmaso, V. (2018) *Exploring Protein Flexibility During Docking to Investigate ligand-Target Recognition*. Ph.D. thesis, University of Padova, Padova.
- Salmaso, V., Sturlese, M., Cuzzolin, A., and Moro, S. (2016). DockBench as docking selector tool: the lesson learned from D3R grand challenge 2015. *J. Comput. Aided Mol. Des.* 30, 773–789. doi: 10.1007/s10822-016-9966-4
- Salmaso, V., Sturlese, M., Cuzzolin, A., and Moro, S. (2017). Exploring protein-peptide recognition pathways using a supervised molecular dynamics approach. *Structure* 25, 655.e2–662.e2. doi: 10.1016/j.str.2017.02.009
- Salmaso, V., Sturlese, M., Cuzzolin, A., and Moro, S. (2018). Combining self- and cross-docking as benchmark tools: the performance of DockBench in the D3R grand challenge 2. *J. Comput. Aided Mol. Des.* 32, 251–264. doi: 10.1007/s10822-017-0051-4
- Sauton, N., Lagorce, D., Villoutreix, B. O., and Miteva, M. A. (2008). MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. *BMC Bioinformatics* 9:184. doi: 10.1186/1471-2105-9-184
- Schames, J. R., Henchman, R. H., Siegel, J. S., Sotriffer, C. A., Ni, H., and McCammon, J. A. (2004). Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* 47, 1879–1881. doi: 10.1021/jm0341913
- Schlitter, J., Engels, M., Krüger, P., Jacoby, E., and Wollmer, A. (1993). Targeted molecular dynamics simulation of conformational change-application to the T ↔ R transition in insulin. *Mol. Simul.* 10, 291–308. doi: 10.1080/08927029308022170
- Shan, Y., Kim, E. T., Eastwood, M. P., Dror, R. O., Seeliger, M. A., and Shaw, D. E. (2011). How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* 133, 9181–9183. doi: 10.1021/ja202726y
- Shaw, D. E., Chao, J. C., Eastwood, M. P., Gagliardo, J., Grossman, J. P., Ho, C. R., et al. (2008). Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* 51, 91–97. doi: 10.1145/1364782.1364802
- Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., et al. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science* 330, 341–346. doi: 10.1126/science.1187409
- Shirts, M., and Pande, V. S. (2000). COMPUTING: screen savers of the world unite! *Science* 290, 1903–1904. doi: 10.1126/science.290.5498.1903
- Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacol. Rev.* 66, 334–395. doi: 10.1124/pr.112.007336
- Spyrakakis, F., Ahmed, M. H., Bayden, A. S., Cozzini, P., Mozzarelli, A., and Kellogg, G. E. (2017). The roles of water in the protein matrix: a largely untapped resource for drug discovery. *J. Med. Chem.* 60, 6781–6827. doi: 10.1021/acs.jmedchem.7b00057
- Spyrakakis, F., Benedetti, P., Decherchi, S., Rocchia, W., Cavalli, A., Alcaro, S., et al. (2015). A pipeline to enhance ligand virtual screening: integrating molecular dynamics and fingerprints for ligand and proteins. *J. Chem. Inf. Model.* 55, 2256–2274. doi: 10.1021/acs.jcim.5b00169
- Strecker, C., and Meyer, B. (2018). Plasticity of the binding site of renin: optimized selection of protein structures for ensemble docking. *J. Chem. Inf. Model.* 58, 1121–1131. doi: 10.1021/acs.jcim.8b00010
- Sugita, Y., Kitao, A., and Okamoto, Y. (2000). Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* 113, 6042–6051. doi: 10.1063/1.1308516
- Sugita, Y., and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314, 141–151. doi: 10.1016/S0009-2614(99)01123-9
- Talele, T. T., Khedkar, S. A., and Rigby, A. C. (2010). Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Curr. Top. Med. Chem.* 10, 127–141. doi: 10.2174/156802610790232251
- Taylor, R. D., Jewsbury, P. J., and Essex, J. W. (2002). A review of protein-small molecule docking methods. *J. Comput. Aided Mol. Des.* 16, 151–166. doi: 10.1023/A:1020155510718
- Teague, S. J. (2003). Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.* 2, 527–541. doi: 10.1038/nrd1129
- Torrie, G. M., and Valleau, J. P. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* 23, 187–199. doi: 10.1016/0021-9991(77)90121-8
- Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334
- Van Drie, J. H. (2007). Computer-aided drug design: the next 20 years. *J. Comput. Aided Mol. Des.* 21, 591–601. doi: 10.1007/s10822-007-9142-y
- Van Meel, J. A., Arnold, A., Frenkel, D., Portegies Zwart, S. F., and Belleman, R. G. (2008). Harvesting graphics power for MD simulations. *Mol. Simul.* 34, 259–266. doi: 10.1080/08927020701744295
- Vanommeslaeghe, K., Guvench, O., and MacKerell, A. D. (2014). Molecular mechanics. *Curr. Pharm. Des.* 20, 3281–3292. doi: 10.2174/13816128113199990600
- Veale, H. F. G., Gohlke, H., and Klebe, G. (2005). DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* 48, 6296–6303. doi: 10.1021/jm050436v
- Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. *Proteins* 52, 609–623. doi: 10.1002/prot.10465
- Vieth, M., Hirst, J. D., Kolinski, A., and Brooks, C. L. III. (1999). Assessing energy functions for flexible docking. *J. Comput. Chem.* 19, 1612–1622.
- Vuong, Q. V., Nguyen, T. T., and Li, M. S. (2015). A new method for navigating optimal direction for pulling ligand from binding pocket: application to ranking binding affinity by steered molecular dynamics. *J. Chem. Inf. Model.* 55, 2731–2738. doi: 10.1021/acs.jcim.5b00386
- Wang, T., and Duan, Y. (2009). Ligand entry and exit pathways in the beta2-adrenergic receptor. *J. Mol. Biol.* 392, 1102–1115. doi: 10.1016/j.jmb.2009.07.093
- Welch, W., Ruppert, J., and Jain, A. N. (1996). Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* 3, 449–462. doi: 10.1016/S1074-5521(96)90093-9
- Wood, J. M., Maibaum, J., Rahuel, J., Grütter, M. G., Cohen, N.-C., Rasetti, V., et al. (2003). Structure-based design of aliskiren, a novel orally effective renin inhibitor. *Biochem. Biophys. Res. Commun.* 308, 698–705. doi: 10.1016/S0006-291X(03)01451-7
- Wu, H., Paul, F., Wehmeyer, C., and Noé, F. (2016). Multiensemble Markov models of molecular thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U.S.A.* 113, E3221–E3230. doi: 10.1073/pnas.1525092113
- Yu, J., Ciancetta, A., Dudas, S., Duca, S., Lottermoser, J., and Jacobson, K. A. (2018). Structure-guided modification of heterocyclic antagonists of the P2Y14 receptor. *J. Med. Chem.* 61, 4860–4882. doi: 10.1021/acs.jmedchem.8b00168
- Yuriev, E., Holien, J., and Ramsland, P. A. (2015). Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. *J. Mol. Recognit.* 28, 581–604. doi: 10.1002/jmr.2471
- Zhou, H.-X. (2010). From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys. J.* 98, L15–L17. doi: 10.1016/j.bpj.2009.11.029

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Salmaso and Moro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.