

# Computational structure-based drug design: predicting target flexibility

## Advanced Review

Jelisa Iglesias<sup>1</sup>, Suwipa Saen-oon<sup>2</sup>, Robert Soliva<sup>2</sup> and Victor Guallar<sup>\*1,3</sup>

<sup>1</sup>Barcelona Supercomputing Center (BSC), Jordi Girona 29, E-08034 Barcelona, Spain

<sup>2</sup>Nostrum Biodiscovery, Jordi Girona 29, D128, E-08034 Barcelona, Spain

<sup>3</sup>ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain

Correspondence to: [rsoliva@nostrumbiodiscovery.com](mailto:rsoliva@nostrumbiodiscovery.com), [victor.guallar@bsc.es](mailto:victor.guallar@bsc.es)

### Abstract

The role of molecular modeling in drug design has experienced a significant revamp in the last decade. The increase in computational resources and molecular models, along with software developments, is finally introducing a competitive advantage in early phases of drug discovery. Medium and small companies with strong focus on computational chemistry are being created, some of them having introduced important leads on drug design pipelines. An important source for this success is the extraordinary development of faster and more efficient techniques for describing flexibility in 3D structural molecular modeling. At different levels, from docking techniques to atomistic molecular dynamics, conformational sampling between receptor and drug results in improved predictions, such as screening enrichment, discovery of transient cavities, etc. In this review article we perform an extensive analysis of these modeling techniques, dividing them into high and low throughput screening techniques and emphasizing in their application to drug design studies. We finalize the review with a section describing our Monte Carlo method, PELE, recently highlighted as an outstanding advance in an international blind competition and industrial benchmarks.

This is the pre-peer reviewed version of the following article: [Iglesias J, Saen-oon S, Soliva R, Guallar V. Computational structure-based drug design: Predicting target flexibility. WIREs Comput Mol Sci. 2018;e1367], which has been published in final form at [<https://doi.org/10.1002/wcms.1367>]. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions

## INTRODUCTION

Drugs can be discovered from a variety of sources and with varying strategies. Amongst the most important sources we find endogenous metabolites, natural products, chemical synthesis compounds and recombinant DNA proteins. The strategies can range from phenotypic screening, where compounds are prioritized based on phenotype alterations (even if the underlying molecular mechanisms are not known), to target-centric approaches whereby the onset or progression of a disease is directly linked to the function of a particular macromolecular receptor, usually a protein.<sup>1</sup> In this latter approach, drugs are targeted molecular agents (ligands) that are prioritized to tightly bind to the receptor to alter its function.

The first target-ligand theoretical models were based on a lock and key concept, describing two interacting molecules as rigid particles perfectly matching one another (Figure 1a).<sup>2</sup> However, this oversimplified view was quickly challenged by two alternative theories, namely induced fit and conformational selection,<sup>3, 4</sup> both of which already describe molecules as flexible entities. The induced fit theory argues that both receptor and ligand change conformation to maximize complementarity only when they come into close contact (Figure 1b). The conformational selection theory, in contrast, argues that ligand and receptor are both constantly visiting a range of conformers that includes the bound conformation, complex formation merely shifting both equilibria towards a bound state (Figure 1c). These two theories or their combination are now widely accepted to describe molecular recognition. Hence, it becomes key for the target-centric approach to gain insight on, not only the structure of the biomolecular target, but crucially its flexibility.<sup>5-7</sup>

Biomolecules undergo conformational transitions due to a variety of factors such as temperature, ionic strength, pH or the presence of other molecules. Structural transitions take many forms, from subtle movements of a single side chain, through concerted motions of a few residues, to significant secondary element rearrangements or complete domain changes.<sup>8</sup> All of them can be critical for drug design. Receptors and enzymes usually change their conformation when isolated (apo) or bound to ligands (holo). Importantly, holo structures of the same receptor with different ligands can also differ significantly.<sup>9</sup> Further, the binding of ligands on one site can shift conformational equilibria on other binding sites, causing allosteric effects (Figure 1d).<sup>10</sup> These phenomena impose serious difficulties in terms of predicting from scratch the universe of accessible conformations of a particular biomolecule.<sup>11</sup>

The structure of biomolecular targets is usually solved by X-ray crystallography, NMR and cryo-EM, and, increasingly, characterized by other experimental techniques such as circular dichroism or small angle X-ray scattering techniques. Their phenomenal evolution has led to a remarkable increase in throughput, applicability and resolution, so that the Protein Data Bank currently holds 3D structures of ca. 135K macromolecules, for which nearly ¾ are complexes.<sup>12</sup> The dynamic behavior of biomolecules can be partly derived at 3D structure determination. X-ray crystallography gives the displacement of atoms from their mean positions in the crystal by obtaining temperature factors, pointing at what segments of a 3D structure are more mobile.<sup>13</sup> NMR generates an ensemble of alternative structures that conform to the restraints derived from the chemical shift experiment.<sup>14</sup> However, these are only crude approximations to the myriad of accessible conformations of a particular receptor.

The field of target flexibility is now being propelled by the application of computational chemistry and simulation methods.<sup>15</sup> Its evolution over the past two decades has been guided by improvements in algorithms, computer power, as well as a growing body of experimental data. Computational methods are now routinely applied in the first phases of a drug discovery project, from the screening for new chemical matter to the optimization of its potency and

ADMET (absorption, distribution, metabolism, excretion and toxicity) properties; all along this timeline, target dynamics must be factored in. This rapid evolution has led to increasing the size of the systems under study, the level of theoretical detail in their representation (not only of the target but also of its environment, especially the solvent) and the simulated time per system. Computational modeling must move up and down these coordinates (Figure 2) depending on the problem to be solved and this choice has a direct impact on how target flexibility is modeled. Typically, the goal is to quickly and reliably predict: i) geometries of ligand-receptor interactions; ii) the energetics (thermodynamics) of complex formation; and only recently iii) the prediction of the kinetics of binding.<sup>16</sup>

At the lowest end of the scale, there are static methods that describe the system in a simplified form, neglecting target dynamics (conformational sampling) completely. This is the case for the rigid docking (rigid receptor and flexible ligand) approach. Although it is one of the most widely used techniques for structure-based pose prediction and for virtual screening (VS) of compound catalogues, it handles the receptor as a rigid object. This leads to the appearance of many false negatives, as a particular target conformation can only match certain ligands while others require rearrangements at the binding site. This can partly be alleviated by the so-called “soft-docking” approach, where the van der Waals radii of target atoms, ligand atoms or both are reduced to diminish steric clashes. Scaling the van der Waals, however, comes with the drawback of generating more false positives, as the receptor-binding site then becomes artificially too slack. Rigid docking can be applied not to one, but to many structures such as an ensemble derived from modeling, NMR or several X-ray (hence its name “ensemble docking”).<sup>17</sup> However, target flexibility is only modeled before the ligand is docked. Of note, rigid docking only makes crude estimates of the thermodynamics of binding, enabling, however, a high throughput performance.

At the middle of the complexity scale, there are methods that take into account the flexibility of targets and ligands simultaneously, but which typically represent the solvent only implicitly. One such method relies on the iterative combination of rigid docking with side chain and (slight) backbone sampling.<sup>18</sup> Medusadock is an example of a docking program that samples the conformations of both interacting partners on the fly,<sup>19</sup> which has given good results. RosettaLigand also samples ligand and target side chain and backbone flexibility simultaneously via a Monte Carlo (MC) algorithm.<sup>20</sup> Another such method is CDOCKER,<sup>21</sup> which has been shown to outperform rigid receptor docking tools. The MC program PELE (Protein Energy Landscape Exploration) also explores all degrees of freedom simultaneously while simulating a binding event.<sup>22</sup> It has been benchmarked as a very efficient tool for binding mode prediction, both in the CSAR competition<sup>23</sup> and in an industrial setting.<sup>24</sup> These methods, although representing bulk solvent implicitly, can handle a few explicit water molecules in an active site, if they are bridging target and ligand via H-bonds. In general, this group of techniques is more reliable at pose prediction and ranking than rigid (receptor) docking approaches, as they account for target flexibility. However, they are computationally more demanding, so that their throughput is lower.

Arguably, the gold standard for computational exploration of target flexibility is molecular dynamics (MD) and all its varying forms and derivatives.<sup>25</sup> Depending on the force field used, system description can be at the coarse-grained or atomic level. The system can even be partitioned such that a small portion is studied at the quantum mechanics (QM) level while the rest is modeled via molecular mechanics, opening a way to study covalent docking<sup>26</sup> and ligand polarization.<sup>27</sup> The solvent can be represented implicitly or explicitly, consisting of water or a combination of different solvents. Given enough computational resources, MD simulations can in principle be capable of fully describing target-ligand binding in explicit solvent. In fact, this has been the case for the most powerful MD machine described (Anton),<sup>28</sup> which has been

used to simulate from scratch the binding of drugs to receptors.<sup>29</sup> However, limitations inherent to the force fields as well as restrictions on the simulated times (as of today, usually from hundreds to just a few thousand nanoseconds) often preclude sampling transitions over longer timescales. Because of this, a series of MD-based enhanced sampling techniques have been developed, such as steered MD (sMD),<sup>30</sup> replica-exchange MD (REMD)<sup>31</sup> and metadynamics.<sup>32</sup> These developments, along with high performance computing in the form of, for example, graphical processing units (GPUs) and cloud computing, open great perspectives to map target flexibility.

Computational chemists have found a way to exploit and combine MD-based techniques for studying conformational transitions of targets when engaged in drug recognition.<sup>33</sup> The particular choice of system description, level of theory and simulation time dictates their eventual throughput, which is usually rather low for MD-based techniques. However, they are routinely applied for the prediction of ligand binding thermodynamics with greater success than low level flexibility descriptions. In fact, atomistic simulations can be used for estimating both absolute and relative binding free energies (DGs).<sup>34</sup> Absolute DGs can be calculated by applying “end-point” methods, where receptor and ligand are only simulated in isolation or in closed complex, such as MMPB(SA) and MMGB(SA),<sup>35, 36</sup> or they can be calculated based on reproducing the whole binding event, such as those that apply Markov State Models (MSM).<sup>37</sup> In the former methods, flexibility is explored partially, as the intermediate states of recognition are neglected, while the latter exhaustively explore how the conformational landscape of target and ligand vary all along binding, opening up a way to study association mechanisms and kinetics. Relative DGs can be calculated thanks to the application of thermodynamic cycles by way of applying Free Energy Perturbation (FEP) or Thermodynamic integration (TI),<sup>38</sup> which are considered some of the most reliable methods.<sup>39</sup>

The next sections will review recent studies involving receptor flexibility in 3D structural drug design. We divide them in i) high-medium throughput flexible techniques, ii) low throughput flexible techniques, and iii) the PELE technique. While we discuss success cases, it should be mentioned that introducing protein flexibility might sometimes degrade the modeling, as a result of incomplete sampling, poor energy functions, inaccuracy of solvent models, etc. Assessment of the modeling results with higher-level techniques and more sampling should be a common practice.

## **HIGH-MEDIUM THROUGHPUT FLEXIBLE PROTEIN-LIGAND TECHNIQUES**

In this section we will review methods and applications capable of screening thousands of compounds in a fast manner. As stated, the lack of flexibility in the receptor is one of the main sources of error. In parallel, however, there is a real interest in developing faster approaches capable of screening millions of compounds.<sup>40, 41</sup> Thus, a compromise between speed (number of compounds) and accuracy is necessary, which is forcing, more and more, the use of hierarchical approaches: an overall technique combining methods that increase conformational sampling only for a selected subset of the screening space.<sup>42-45</sup>

The first important addition of flexibility in 3D docking was introduced at the ligand level; today almost all docking techniques include ligand flexibility to some degree. Ligand flexibility is mainly introduced by an exhaustive search of possible conformations of the ligand, using stochastic or deterministic approaches, where possible conformations are pre-generated and selected before the docking step.<sup>46</sup> Such deterministic generation has recently included mixed classic/quantum chemistry explorations, providing, not only an improved list of conformations for the ligand, but also a better estimation for the strain energy of the bound complex.<sup>47</sup> While this review focuses on structural methods, we note also the importance in sampling ligand

conformations for quantitative structure–activity relationship (QSAR) descriptors, as shown recently with the xMaP (flexible MaP) descriptor.<sup>48</sup>

Sampling receptor flexibility is a significantly more complex process, which has centered a significant amount of methods development. The approaches for high-medium throughput VS can be divided in two main blocks: the ensemble docking technique and a plethora of protein rearrangement algorithms aiming at quickly sampling side chains and, less often, add into it the protein backbone as well. Here we will not consider soft-docking as a flexible method since it only allows for a deeper overlap of molecules, not describing conformation rearrangements. Nevertheless, its importance/use in the early steps of hierarchical VS should be seriously considered.

### Ensemble methods

One straightforward approach to introduce flexibility into docking is by means of ensemble docking, that is, to use several (different) structures of the same target to dock ligands. These structures can be obtained from modeling and experimental sources: (i) receptors co-crystallized with different ligands that induce distinctive conformational states; (ii) different models from an NMR ensemble; (iii) snapshots from an atomistic MD or MC simulation;<sup>49</sup> (iv) snapshots from simpler modeling approaches such as using a normal mode technique (anisotropic network modeling, etc.) or homology modeling. Once the ensemble is generated, one can readily use rigid docking methods turning the overall VS simulation time directly proportional to the number of receptor conformations (Figure 3).

Several software packages provide means to perform ensemble docking. The OEDocking suite, for example, includes different options. Its HYBRID module is capable of using different receptors structures as input.<sup>50</sup> A different module, POSIT, has an automatic procedure to select from a list of ligand-protein structures the best suited to guide ligand docking. They found its performance to be very close to that found by using ensemble docking, with the corresponding significant gain in performance (using one structure instead of many). We find multiple application studies of this suite of programs in the discovery of new anti-cancer molecules,<sup>51</sup> in the development of new models to find new antiviral molecules against HIV<sup>52</sup> and antibacterial activity.<sup>53</sup>

A 4D docking approach (the 4<sup>th</sup> dimension referring to the multiple receptor variable) is also implemented in ICM.<sup>54</sup> In addition, different methods for the ensemble generation are provided, including a normal modes and a “fumigation” technique, which samples pocket torsion angles in the presence of a repulsive generic ligand. The 4D docking methodology developed by Bottegoni et al.<sup>55</sup> optimized the efficiency of ensemble docking by integrating the procedure in one step. The user defines one structure from the ensemble as a template, to which all the other structures will be aligned, creating a grid with multiple conformations of the receptor that is sampled along rotations and translations of the ligand. This methodology has been used in the discovery of anticancer drugs, in particular cGMP efflux inhibitors.<sup>56</sup>

While some of the top commercial vendors, such as Schrodinger with the Glide rigid docking software, do not offer a ready to use ensemble docking tools, their graphical interfaces provide workflow solutions, like the Virtual Screening Workflow in Maestro, that largely facilitate this task. MOE, for example, also provides MD or normal modes tools to generate different receptor structures. Titan et al. used Schrodinger’s workflow in order to improve the docking results over one single structure-docking in kinase targets.<sup>57</sup> Middendorp et al. used FlexX from MOE to dock several ligands into different homology models of GABA, finding new leads for this receptor.<sup>58</sup> The ICM workflow was used to find novel inhibitors of the SERT protein, which could act as antidepressants or psychostimulants.<sup>59</sup> The ensemble docking

methodology, however, is in high demand and many groups use their own approach to implement this technique manually, by combining different software for different steps.<sup>60</sup>

In some cases it is necessary to study the influence of large conformational changes such as the differences between active and inactive conformations. The presence of multiple protein-ligand activation states opens new inhibition possibilities; this is the case, for example, of kinases and their activation loop, where computational modeling has difficulties to predict such large scale (and slow) conformational change. While below we review some attempts to such a modeling effort, the existence of crystallographic structures in different states introduces a ready to use ensemble. Meirson et al used 4 crystal structures of Pyk2 with different conformations induced by ligands with different selectivity profiles. To each of them they performed rigid docking, ADMET efficiency metrics filtering, and refined the best candidates using MD (a nice example of hierarchical VS).<sup>61</sup> Another example of the potential of this technique (and an example of the heavy use of soft-docking techniques) is its use in the development of a new consensus scoring model for mTOR by Li et al., where they use six different crystal structures to perform soft docking with Glide in order to account for induced fit effects.<sup>62</sup>

### **Protein rearrangement algorithms**

One of the most used approaches to provide induced fit conformational explorations is to perform binding pocket side chain sampling. Different variations of this approach are incorporated, for example, into Medusadock,<sup>19</sup> GOLD,<sup>63</sup> Molegro Virtual Docker (MVD),<sup>64</sup> IFD,<sup>18</sup> PLANTS,<sup>65</sup> RosettaLigand,<sup>20</sup> AutoDock4,<sup>66</sup> CDOCKER,<sup>21</sup> IMGdock<sup>67</sup> and ICM<sup>68</sup>; these induced fit techniques have shown significant improvement in pose prediction when compared to rigid docking. In addition, some tools like IMGDock and PLANTS, also have algorithms that optimize crystal water molecules within the binding pocket along the flexible docking.

Matijssen et al. were able to define binding modes of benzimidazole to the CHK2 kinase thanks to the exploration of side chain flexibility. They combined the use of ensemble docking with a rigid body docking methodology and the flexible docking of GOLD.<sup>69</sup> Gupta et al used a combination of rigid docking and flexible docking to study new putative drugs against malaria. They observed that flexible docking results differed from rigid docking ones in the same fashion for both MVD and Autodock, and that for both programs the flexible docking had a better performance than the rigid docking.<sup>70</sup>

Rosetta's flexible docking was used by Brüser et al. to explain the differences in binding for UDP and PGE2-Gf to P2Y6 receptor.<sup>71</sup> Luo et al. also used Rosetta's flexible docking (along with MD simulations) to model the binding modes of allosteric inhibitors to an extracellular cap of the KP2 channel;<sup>72</sup> the high flexibility of this region mandated the use of flexible docking.

Ding et al. stressed the importance of taking care of the receptor flexibility for the prediction of compounds during their participation on the DRD3 2 challenge.<sup>73</sup> They found that the efficacy of rigid docking with CDOCKER was impaired by conformational changes of the protein upon binding of different ligands, and by steric clashes between the ligand and some residues with alternate positions. These problems could be solved by using a receptor with the right conformation or by using the flexible docking protocol.

Flexible docking with ICM was used to model the binding mode and induced-fit effect of a new drug against KEAP1, which has been patented in USA by the General Hospital Corp, University of California.<sup>74</sup> Lane et al. also used ICM flexible docking to model dopamine binding to the apo dopamine D3 receptor, using the resulting model to perform ICM rigid docking of 4.1 million

compounds.<sup>75</sup> Flexible docking with Autodock has been used to explain the inhibition mechanism of laccase by medicarpin.<sup>76</sup> It also has been key in the study of hydrophobic surfaces on the cataract-related G18V variant of human  $\gamma$ S-crystallin.<sup>77</sup> IFD was used in the identification of active compounds to develop new kind of drugs against chronic infections.<sup>78</sup> In the work of Chatzileontiadou et al. the IFD protocol was key for finding binding modes in human Angiogenin in agreement with the NMR structures.<sup>79</sup>

Adding protein backbone flexibility is a more complex and demanding (CPU time) task. Some of the programs providing side-chain rotamers treat backbone sampling by means of coupling the (side-chain) induced fit procedure with ensemble docking or with minimizations, as shown in Medusadock.<sup>80</sup> The newest Schrodinger IFD protocol combines Glide rigid body docking with soft potentials and protein sampling using Prime. It is based on 4 steps: (i) a rigid docking of the ligand into the protein using soft-potentials; (ii) sampling protein conformations that may accommodate the ligand; (iii) rigid body docking of the ligand poses obtained in the first step with the protein conformations from the second step; (iv) scoring the receptor/ligand poses using a combination of Glide's score and Prime's energy. This protocol has been used to study new phitoconstituents derived from *Silybum marianum*, which can have an anti-amnesic effect.<sup>81</sup> Most of these methods, however, are not used as a stand-alone approach, the lack of a robust backbone sampling drives users to combine them with more computationally expensive techniques, such as metadynamics, as in the work of Clark et al..<sup>82</sup>

## LOW THROUGHPUT FLEXIBLE PROTEIN-LIGAND TECHNIQUES

In this section we turn into techniques that are capable of introducing flexibility in a more accurate manner, but at the expense of screening only dozens (low hundreds in the best case scenario) compounds. We first center on MD techniques analyzing, afterwards, MC contributions.

### Molecular Dynamics

As stated, MD simulations are currently the gold standard for exploring target flexibility in drug design.<sup>43, 83, 84</sup> We explore their contribution by grouping them in three main areas: target treatment, mechanistic studies, and pose refinement (Figure 4); the separation between these groups, however, is often narrow and the same simulation might address multiple aspects.

#### Target treatment

Besides providing diversity for the ensemble docking procedure,<sup>85</sup> adding flexibility to a target by means of MD simulations can reveal surface properties not observed in the available crystal structures, such as the presence of cryptic transient open binding pockets or subpockets,<sup>86-88</sup> or the role of solvation effects through identifying/mapping key water molecules.<sup>89</sup> For example, Wassman et al. described a transiently binding pocket between loop L1 and sheet S3 of the tumor suppressor p53 core domain. Virtual screening against this revealed pocket identified a compound capable of reactivating mutated forms of p53 in human cell, demonstrating its potential as a pharmaceutical target.<sup>90</sup> Related MD work also reported a narrow crevice on the surface of the p53 Y220C mutant, pinpointing key interactions for developing stabilizing small molecules.<sup>91, 92</sup>

Hagler et al. studied the effect of protein plasticity on the ability to identify active compounds through VS. They performed small-molecule docking on the ensemble structures generated by MD and REMD on the androgen receptor (AR), the HIV protease, and CDK2. Their results showed that flexibility increased significantly the enrichment and enhanced the diversity of hits.<sup>93, 94</sup> Similarly, Hou et al. reported better predictive accuracies when using MD structures for three kinases: ALK, CDK2 and VEGFR2.<sup>57</sup> Antolin et al. applied REMD simulations to sample the conformational space of the catalytic domain of PARP-1 in the ligand-bound and unbound

forms, assessing how enzyme flexibility affects the docking of a library of PARP-1 inhibitors. They pinpointed a key role of Leu324, Tyr325 and Lys242 in opening an additional binding site pocket with implications on ligand binding enrichment factors.<sup>95</sup>

McCammon and co-workers introduced the "relaxed complex scheme" (RCS) for receptor flexibility, which combines all-atom nanosecond MD simulations with small molecule docking to representative snapshots. RCS has shown to be successful to a variety of docking studies, demonstrating its potential for discovering new inhibitors and characterization of local-induced and global effects on ligand binding.<sup>96-98</sup> For instance, Schames et al. identified a novel binding cavity in HIV integrase using RCS with the 5CITEP inhibitor,<sup>99</sup> which inspired the discovery of FDA-approved drug raltegravir.<sup>100</sup> Cheng et al. have extended RCS by the efficient use of RMSD-based clustering on the study of avian influenza N1 neuraminidase in the apo form and in complex with oseltamivir. They showed a wide opening from the closed crystallographic structure (PDB:2HU4) and, by docking the National Cancer Institute Diversity Set 1 (NCIDS1), identified new hits.<sup>101</sup>

Recently, Valant et al. applied accelerated MD (aMD) to generate a receptor ensemble of the M2 muscarinic acetylcholine receptor (mAChR). Then, through iterative molecular docking and experimental testing, they successfully identified positive and negative allosteric modulators of M2 mAChR with remarkable chemical diversity, showing that aMD simulations combined with Glide IFD provided much-improved enrichment factors when compared with standard Glide.<sup>102</sup> The mixed Essential Dynamics/Molecular dynamics technique has also recently been described for quickly generating alternative images of a receptor cavity.<sup>103</sup> This technique uses a previously existing protein ensemble to describe its essential space of deformation. This space is perturbed by the presence of a small molecule, so that new, perturbed images of the receptor cavity can be used in the rigid body approximation. The technique was shown to outperform not only docking based on a single conformation, but also docking based on an unperturbed ensemble of the same protein.

Zacharias performed principal component analysis (PCA) on a MD trajectory and extracted its "soft" flexible modes that were used later in the docking process, implemented in the "PCRELAX" program.<sup>104</sup> They showed that rigid docking of the FK506 ligand to an unbound FKBP conformation failed to identify a pose close to the experimental structure; inclusion of the flexible soft modes provided a native-like structure as the lowest energy conformation. Louet et al. applied Normal Mode Analysis (NMA) with conventional MD simulation to investigate the conformational dynamics of the hetero-trimeric G- protein. Their results showed that G-proteins undergo large conformational changes without energy penalties. More interestingly, one of the lowest-frequency representative motions was able to open the GDP binding site and was consistent with experimental data. They proposed that GDP release and subsequent GDP/GTP exchange mainly involves an inter-domain motion between the ras-like and the helical domains of G $\alpha$ , together with an uprising of both  $\alpha$ G and  $\alpha$ 4 helices.<sup>105, 106</sup>

Pande et al. introduced a computational paradigm coupling transition pathway with MSM techniques in the frame of massively distributed simulations, mapping the conformational landscape of a c-Src tyrosine kinase. This approach modeled the thermodynamics and kinetics of kinase activation for the first time, identifying key structural intermediates, and providing a database of c-Src conformations for future design of novel kinase inhibitors.<sup>107</sup>

Importantly, Barril et al. raised concern on the extent at which protein flexibility should be considered. Using MD simulations of hen egg-white lysozyme (HEWL) with explicit aqueous/organic solvent mixtures (MDmix method) and a range of restraint conditions, they observed how artificially restricted mobility affects binding hot spots. The authors suggested



that using carefully selected experimental structures may be more realistic and productive than a MD ensemble. Ultimately, choosing the right level of flexibility will depend on the goal of each individual investigation and the nature of each particular system.<sup>108</sup>

Finally, within the target treatment, we want to underline MD studies addressing the role of water molecules<sup>109</sup> which, on top of mediating in the protein-ligand interactions, might affect the receptor flexibility. Different techniques, mostly based in inhomogeneous solvation theory, have been recently developed for this purpose, including WaterMap<sup>89</sup> and STOW.<sup>110</sup> A deeper analysis of these methods is out of the scope of this review; we refer the reader to recent studies benchmarking these techniques and comparing them with quicker grid based methods.<sup>111, 112</sup>

#### *Mechanistic studies*

Flexibility studies through MD have also emphasized on mapping protein-ligand association mechanisms, ligand entry and exit and kinetics; all these being key aspects for rational drug design. We refer here to the term “dynamic docking” introduced by De Vivo et al. in their recent review.<sup>113</sup> Due to the (very) large timescale associated to some of these processes, the use of enhanced techniques has been significant. Parrinello and co-workers, for example, applied metadynamics,<sup>32, 114</sup> a technique based on penalizing visited phase space points along a collective variable, to investigate the dissociation process of a nonsteroidal anti-inflammatory ligand (SC-558) to COX-1 and COX-2 isoforms. In this way, they found a novel alternative binding mode to COX-2, identified key residues along the binding path and built a kinetic model for the binding mechanism.<sup>115</sup> Incerti et al. also used metadynamics for building the unbinding free energy potential surface of the antagonist UniPR129 from its EphA2 receptor. They reported a binding mode that agreed with structure-activity relationship data and proposed, synthesized and tested new compounds for their ability to displace ephrin-A1 from the EphA2 receptor.<sup>116</sup> Recently, Berne and co-workers have combined metadynamics with IFD aiming for accuracy and efficiency (reducing the overall computational cost). The method significantly increased the performance of the underlying IFD protocol across a large data set comprising 42 different ligand–receptor complexes.<sup>82</sup>

sMD, based on imposing an external potential to the ligand, have made important contributions to study binding/unbinding pathways, including its free energy estimation by means of the Jarzynski equality. Thus, one might use it for ranking of ligand affinities as well as elucidation of key residues along the exit/entrance pathway. For instance, we find research involving ligands dissociation from malaria parasite *Plasmodium falciparum* FabZ protein,<sup>117</sup> CDK5 protein kinase,<sup>118</sup> and GPCRs.<sup>119, 120</sup> Moreover, sMD was used to study the conformational transition on GC-T4P using the 18 subunits long cryo-EM reconstruction to probe dynamics under tension, and to gain insights about the response of GC-T4P to external force at atomistic detail.<sup>121</sup>

REMD uses several copies of the system evolving in parallel under different simulating conditions, such as temperature or even the Hamiltonian, exchanging coordinates among the replicas at regular intervals with a MC criterion.<sup>122, 123</sup> In several studies with disordered proteins, associated with diseases such as diabetes, Parkinson's and Alzheimer's, REMD was used to characterize the conformational states.<sup>124</sup> In Alzheimer's disease, disordered  $\beta$  amyloid monomers are postulated to form toxic fibrils. While fibrils can be resolved in X-ray crystallography, the conformation of amyloid monomers and their oligomerization process remain elusive. REMD simulations studied this transition and addressed if the disordered monomer structure is pre-folded and contains information on the well defined oligomeric state.<sup>125</sup> Kokubo et al. applied their replica-exchange based technique for exploring the ligand binding to two kinase systems (p38 and JNK3) with two different ligand molecules for each

kinase. They illustrated that protein flexibility was essential to predict the correct binding structure for one of the systems, where dihydroquinolinone was bound to p38 alpha kinase.<sup>126</sup>

In order to enhance the cavity opening and its exploration, some methods include a biasing repulsive potential (repulsive probes, etc.) in the binding site, in a similar fashion to the ICM fumigation technique for docking (Figure 5). This bias tends to perturb the active site, allowing it to readapt better to a given ligand. The Laughton group introduced in 2008 the Active Site Pressurization (ASP) technique,<sup>127</sup> where uncharged Lennard-Jones particles are injected in the active site, and applied it later on to map kinases<sup>128</sup> and GPCR.<sup>129</sup> More recently, Zacharias et al. introduced the repulsive potential as a simulation condition in REMD.<sup>123</sup>

Finally, hardware and software development has allowed, in the past few years, to perform complex protein-ligand binding mechanism studies using standard MD approaches. Besides the seminal work by the Shaw lab, partly introduced above,<sup>29, 130</sup> we find several studies using GPU computing power.<sup>131, 132</sup> Buch et al. performed the full binding mechanism of the trypsin-benzamidine complex, including a kinetic model by means of a MSM analysis. Decherchi et al. combined microsecond-long MD with machine learning algorithms to estimate the thermodynamics and kinetic values of a transition state analogue molecule to PNP.<sup>133</sup> WExplore, a technique based on ensemble trajectories combined with residence time calculations, has also been recently introduced by Dickson et al. to describe ligand binding kinetics on the TPPU receptor.<sup>134</sup> While still being slightly off from experimental values, these studies constitute impressive simulations aiming at modeling ligand dynamics and kinetics.<sup>135</sup> Alternatively, computer power (including GPUs) can be provided by distributed computing, such as in the Folding@home project. Multiple studies from the Pande group, present clear examples of how by adding flexibility we have advanced in the understanding of biomedical research, including glycan binding in the NMDA receptors,<sup>136</sup> kinases,<sup>107, 137</sup> and GPCRs.<sup>138</sup> In this regard, distribute and cloud computing have the potential to turn low into med/high throughput techniques.

#### *Pose refinement*

MD can be used as a post processing tool to validate and/or refine docking results, typically implemented in the drug design lead optimization phase.<sup>139</sup> From MD refinement, the docking poses should display more stable and specific interactions, as the result of the induced-fit effect. Moreover, MD postprocessing can provide clues on the specific role played by solvent and allow for better estimates to the binding free energy, both in absolute and relative terms.

MM/GB(PB)SA is probably the most practical and widely used approach to estimate the binding energy of small molecules in low throughput virtual screening. It basically rescores multiple complex snapshots, typically derived from an MD simulation, by means of single point all-atom force field interaction energies with the addition of implicit solvent (GB or PB). Numerous, and somehow contradictory, benchmark studies have assessed its performance. Based on the PDBbind data set several authors have shown its ability to yield higher enrichment factors and to discriminate/sort by affinity a series of small molecules.<sup>140-142</sup> However, the success of this approach seems to be system dependent and the entropic term to be crucial for an accurate prediction (these are normally estimated from normal modes). For further detail, a nice review from Ulf Ryde on the MMGB/SA method is recommended.<sup>143</sup> The message from the author concludes that, while this approach is useful and easy to implement, one should be careful when applying it. Higher accuracy techniques, involving more expensive quantum mechanics are under investigation.<sup>144, 145</sup> Nevertheless, these deal with rescoring matters rather than adding additional flexibility to the system.

The Linear Interaction Energy (LIE) method, proposed by Åqvist, has shown good compromise between speed and accuracy when estimating absolute ligand-binding affinities.<sup>146, 147</sup> This method is based on force field estimations of the receptor-ligand interactions and thermal conformational sampling. A notable feature is that the binding energy can be predicted from intermolecular interactions (electrostatic and van der Waals interaction energies) ensemble differences.<sup>148</sup> Vermeulen, et al. described the combination of rigid docking, MD simulations and LIE, to quantitatively evaluate the set of small molecules binding affinity to human Cytochrome P450 with estimated errors of less than 1 kcal/mol.<sup>149</sup> LIE performance however, has been shown to be slightly worst than MM/GBSA in some benchmarks.<sup>150</sup> Finally, we should state that the main limitations of the LIE and MMGBSA techniques, when considering the number of compounds that are capable of studying, resides in the length of the MD trajectory performed for each compound. While performing longer simulations might seem an advantage (in terms of adding conformational sampling), it could degrade numerical convergence, as discovering new conformations might change significantly the interaction energies.<sup>143, 151</sup>

FEP methods are the standard when aiming for accuracy in low-throughput VS refinement. While this group of studies could be introduced in the previous mechanistic block, we prefer to add them here since they are typically used as a post docking refinement phase. FEP convergence requires extensive sampling and, in contrast to the previous methods, it requires pair-wise interactions to be calculated with a higher frequency. This fact together with the requirement that the transformations must be “smooth”, i.e., small perturbations, often make the method quite computationally demanding. In early 1990, Merz and Kollman employed FEP within an MD framework to correctly predict the binding free energy of a previously unreported inhibitor of the thermolysin endopeptidase, and to predict the affinity of a novel HIV-1 peptide inhibitor.<sup>152, 153</sup> Today, FEP calculations have demonstrated astonishing potential for driving lead optimization campaigns. Especially in the recent year, the implementation of FEP+ from Schrödinger has made the method accessible and attractive for the pharmaceutical industry.<sup>39, 154</sup> When combined with an extensive ligand dihedral parameterization, Wang et al. demonstrated that FEP+ enables highly accurate affinity predictions across a broad range of target classes and ligands, many of which involve significant changes in ligand chemical structures.<sup>39</sup> In a recent study, Lenselink et al. predicted relative binding free energies of congeneric ligands binding to GPCRs using the FEP+ package, with successful predictions of binding affinities for 39 of 45 compounds (investigated on four GPCRs).<sup>155</sup> Moreover, FEP+ can also predict the affinity of novel and potent adenosine A2A receptor (A2AR) antagonists. Four novel compounds were synthesized and tested and the affinity of two out of the four was correctly predicted (within 1 kcal/mol), including one compound with approximately a tenfold increase in affinity compared to the starting compound.<sup>155</sup> Extensive test by pharmaceutical companies, however, have raised concerns on the efficacy of the technique, which still seems to be case dependent.<sup>156</sup>

### Monte Carlo

Stochastic MC techniques offer a valuable alternative to deterministic MD methods. Since the study of bovine pancreatic trypsin by Wako et al. in 1981,<sup>157</sup> multiple methods aimed at mapping protein (and ligand) flexibility by MC techniques. The difficulties in combining meaningful perturbations with a large enough acceptance ratio, however, drove the development of the MC-minimization technique by Li and Scheraga.<sup>158</sup> An important contribution was then introduced by Still and coworkers: the MacroModel program,<sup>159</sup> combining random moves, minimizations and a GBSA implicit solvent model; this technique is still part of the Schrödinger suite, being one of the most used MC software for drug design. Recent implementations of a MC minimization scheme include our PELE software, which combines random movements with protein structure prediction algorithms to enhance the exploration.<sup>22</sup>

An important group of MC techniques centered on obtaining protein-ligand binding free energies,<sup>160, 161</sup> mainly using alchemical methods such as FEP. An example is the pioneering BOSS (Biomolecular and Organic Simulation System) and MCPRO (MC for Proteins)<sup>162</sup> software from the Jorgensen group. Other free energy methods, such as thermodynamic integration, have also been used with MC, as seen in different works of the Essex' group<sup>163</sup> with their software package, ProtoMS.<sup>164</sup> Application studies of MCPRO include the discovery of anti-HIV agents<sup>165</sup> and the inhibitor design on tyrosine kinase.<sup>166</sup> In addition, MCPRO+, a version under Schrödinger commercial use, has been used in the drug design studies for the acetylcholine receptor<sup>167</sup> and the Bcr-Abl kinase.<sup>168</sup> We also find several studies using ProtoMS, more focused in obtaining free energies<sup>163</sup> and the role of water molecules in drug recognition and binding.<sup>169, 170</sup>

Efforts have also been made to combine MC with MD. Recently, Chodera and coworkers presented Nonequilibrium candidate Monte Carlo (NCCMC),<sup>171</sup> where proposal moves are constructed via nonequilibrium dynamics, in a series of perturbation mixed with propagation moves. This technique showed an enhancing of two orders of magnitude in finding binding modes of toluene in a mutated T4 lysozyme.<sup>172</sup>

As a rule of thumb, MC techniques have a harder time mapping backbone reorganization than MD, but can perform local side-chains adjustments more efficiently.

## THE PELE TOOL FOR FLEXIBLE PROTEIN-LIGAND SAMPLING

In this section, first we give a brief introduction to the method and then we summarize representative applications in drug design. As stated, the method was highlighted as a "outstanding breakthrough" in the latest CSAR blind competition.<sup>23</sup>

### The PELE Method

PELE follows a heuristic MC approach, generating conformational proposals by means of protein structure prediction techniques coupled to a system perturbation, so that the probability of acceptance in the Metropolis test remains high. The procedure is divided in two blocks: i) a ligand plus receptor perturbation, aimed at providing a conformational change in the overall system; ii) a relaxation step, comprising a side-chain sampling and a minimization step, aimed at driving the system to a local minima. A typical simulation involves tens to hundreds of processors (explorers) for hundreds of MC steps.

*Ligand Perturbation.* The ligand is perturbed by translating and rotating it within a user-defined box (limiting the exploration space). After several perturbation trials (typically between 1 and 20), PELE chooses the one with the lowest system total energy. Each trial involves clash relieving using internal ligand degrees of freedom and rotamers from side chains around the ligand. When searching for the binding site, large translations are usually selected (up to 6 Å), while once in the binding site, small adjustments of ~1Å (or less) are used; a combination of translation ranges is often applied depending on the ligand solvent accessible surface area (SASA). In addition, to enhance the sampling of rare events, the (random) translational vector may be kept for a given number of MC simulation steps.

*Receptor Perturbation.* The receptor backbone is perturbed following normal modes calculated using the Anisotropic Network Model (ANM, see<sup>173</sup>) or PCA from different structures. It is possible to use a single mode, or to mix them randomly, with a preferred mode being given more weight, or weighting each mode according to its frequency (simulations typically use only the 6 lowest modes). As in the ligand translation, a perturbation direction can be kept during

several simulation steps, changing sense periodically, to allow a more thorough sampling of a given vibrational movement.

*Side Chain Adjustment.* The side chain relaxation is done for all side chains within a user-defined distance from the ligand (and for the ligand itself). The problem becomes manageable by considering only rotamers as possible conformations for the residue side chains, and by placing the best rotamer for each side chain (after clustering) in a residue by residue way, keeping all the other residues fixed at their current rotamers.<sup>174</sup> Once the side chains have been readjusted, a local side-chain minimization is performed.

*Minimization.* The relaxation step ends with a multi-scale Truncated Newton minimization.<sup>175</sup> Along the minimization, the atoms representing the nodes in the ANM phase are (typically) weakly constrained, so that this minimization does not undo the perturbation phase. Convergence criterion is significantly relaxed,  $\sim 0.1 \text{ kcal}/(\text{mol}\cdot\text{\AA})^2$ , allowing to partly recover the detailed balance condition (lost when introducing the relaxation step).<sup>176</sup>

In addition, PELE introduces several techniques to enhance conformational sampling. In the *spawning approach*, introduced in Borrelli et al.,<sup>177</sup> the simulation is biased along a reaction coordinate (RMSD, distances, etc.). More recently, *adaptive-PELE*<sup>178</sup> has been developed, a strategy that shifts the exploration (trajectories) towards poorly explored regions.

By default, PELE uses the OPLS 2005 energy function and parameters,<sup>179, 180</sup> but also implements the AMBER99sbBSC0 force field,<sup>181</sup> more suitable for nucleic acids. While PELE uses implicit solvent models, it allows placing discrete water molecules that are deemed important, which will be moved during the different energy minimizations in a simulation step (the next release will include a water perturbation/relaxation step).

### PELE's application studies

PELE was first designed to map ligands' exit/entrance migration pathways. In few hours of a modest computational cluster ( $\sim 8/16$  computing cores), its first application study could map exit pathways in myoglobin, cytochrome P450cam and a fatty acid binding protein without any additional knowledge nor bias.<sup>177</sup> In all cases, ligand migration highly depended on protein conformational reorganization. The fast performance of PELE, for example, allowed the first migration study on a (tetramer) human hemoglobin, identifying differences in ligand migration between the  $\alpha$  and  $\beta$  subunits and between their tensed and relaxed states.<sup>182</sup>

By expanding the exploration to tens/hundreds of computing cores, we can perform full non-biased explorations, mapping the entire receptor surface in search for binding sites, in a similar fashion to the pioneer studies of the Shaw group; Figure 6 shows PELE's exploration of the Src kinase, reproducing the dynamic binding observed by the Shaw group with unbiased MD. Several systems have been modeled in this manner, including quite complex ones such as the small molecule and peptide binding to prolyl oligopeptidase,<sup>183</sup> the binding of a porphyrin ligand into Gun4,<sup>184</sup> mTOR<sup>185</sup> and BCL-2 apoptotic targets,<sup>186</sup> the human phosphomannomutase2 receptor,<sup>187</sup> and on diverse nuclear hormone receptors (NHRs) in collaboration with AstraZeneca.<sup>188, 189</sup> In all of these systems, coupling ligand dynamics to backbone conformational changes was key for effective binding; constraining the alpha carbons to the apo structures, for example, resulted in unsuccessful searches. Analogous simulations have also been performed for a DNA-cisplatin complex.<sup>190</sup> Moreover, to enhance ligand screening, we have coupled PELE with MSM, approximating absolute binding free energies in a fast manner.<sup>189, 191</sup>

In most cases, however, modeling ligand binding does not require such extensive study; local rearrangements of the ligand at the (already known) binding site suffice. Therefore, we centered on developing induced fit protocols. Our initial benchmark showed significant improvement over that of IFD, providing accurate cross docking and apo docking.<sup>192</sup> Applied biomedical studies in this line included the challenging prediction of drug resistance in the HIV-1 protease receptor, where, in collaboration with the Aids IrsiCaixa institute, we were able to blindly identify high resistance patients using only their viruses' sequence data.<sup>193</sup> We should underline that this effort involved modeling 15-30 mutations in each of the 42 patients receptor. Interestingly, resistance mechanism was the result of a delicate steric (flexibility) network, communicating the binding site with distant protein regions.

In the light of the recent upsurge of fragment-based techniques, PELE was also probed for small-fragment pose prediction in a challenging case, the enzyme epoxyde hydrolase, whose active site is not only highly flexible, but extremely hydrophobic and composed of three sub-sites. The cross docking exercises with an efficient docking tool such as Glide were notoriously unsuccessful, meaning that the enzyme highly adapts in each case to the presence of a specific chemical scaffold, precluding the binding of others. The small fragments, whose molecular weights were as low as 125Da., were successfully dynamically docked with PELE yielding RMSDs with respect to the corresponding crystal structures on the order of 1.5-2 Å, in all cases. Especially encouraging was the fact that PELE even found alternative simultaneous binding modes for the fragments that had been revealed by X-ray crystallography to bind to more than one sub-site.<sup>194</sup>

In terms of efficiency, a recent enhanced conformational sampling has been introduced by means of an adaptive sampling.<sup>178</sup> This procedure improves ~1 order of magnitude the exploration performed by standard PELE. Such enhancement allowed mapping complex dynamic docking, in GPCRs and NHRs, in only three hours using a commodity computing cluster (~32 computing cores), and achieving (almost) interactive rates (~5 minutes) in difficult local induced fit cases. This performance will place PELE into medium/large VS refinement methods in the near future.

## CONCLUSION

Advances in multiple disciplines enable today the exploration of biomolecular target flexibility to a degree that was not conceivable only a couple of decades ago. This has had a direct impact on the quality of molecular modeling and computational chemistry predictions, which, by applying the cited methodologies, have gained a space of their own both in academia and in industry. In fact, the pharmaceutical industry has recently backed up computational technologies with high volume deals such as the one signed by Sanofi with Schrodinger<sup>a</sup>. Moreover, some last-generation companies have been founded with a strong computational rationale, such as Nimbus or Relay Therapeutics. Clearly, recent success stories of computational chemistry applied to the discovery of novel molecules have been achieved thanks to modeling (to some degree) flexibility. The number of cases in the future is bound to increase, as techniques will continue their evolution. In this evolution, hierarchical methods, mixing different levels of theory and flexibility treatment, will have an important role.

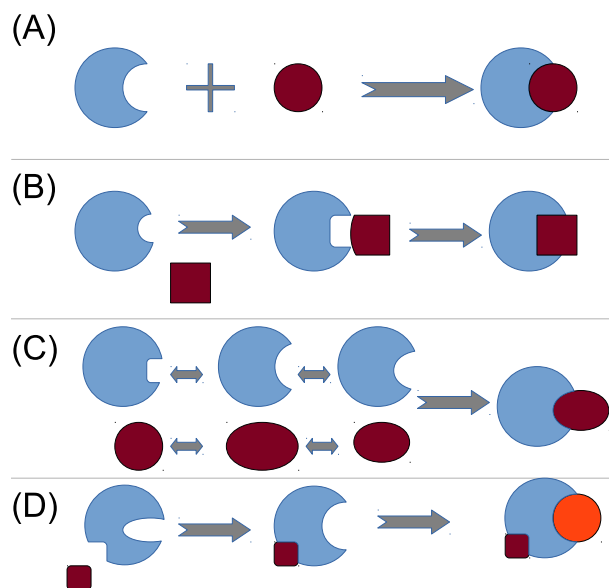
## ACKNOWLEDGMENTS

We acknowledge the BSC-CRG-IRB Joint Research Program in Computational Biology. This work was supported by a grant from the Spanish Government CTQ2016-79138-R.

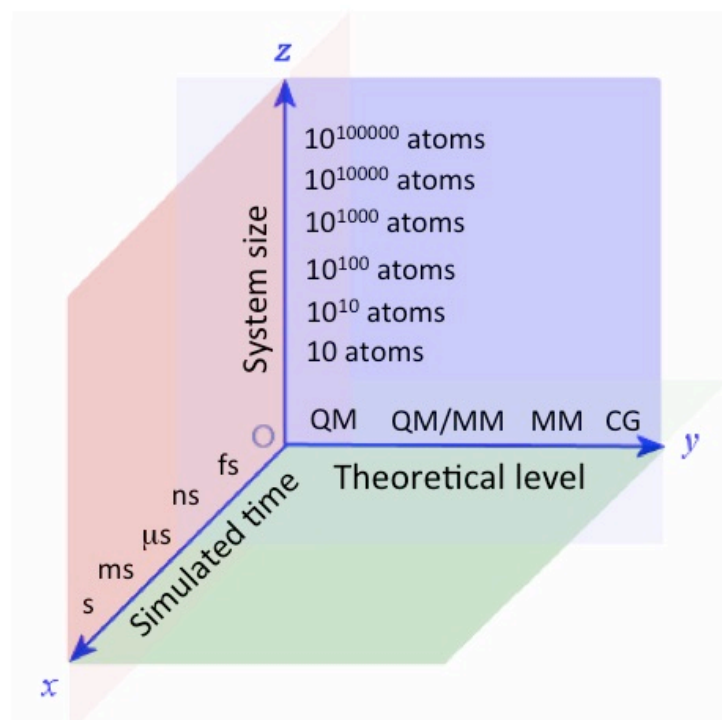
---

<sup>a</sup> See, for example:

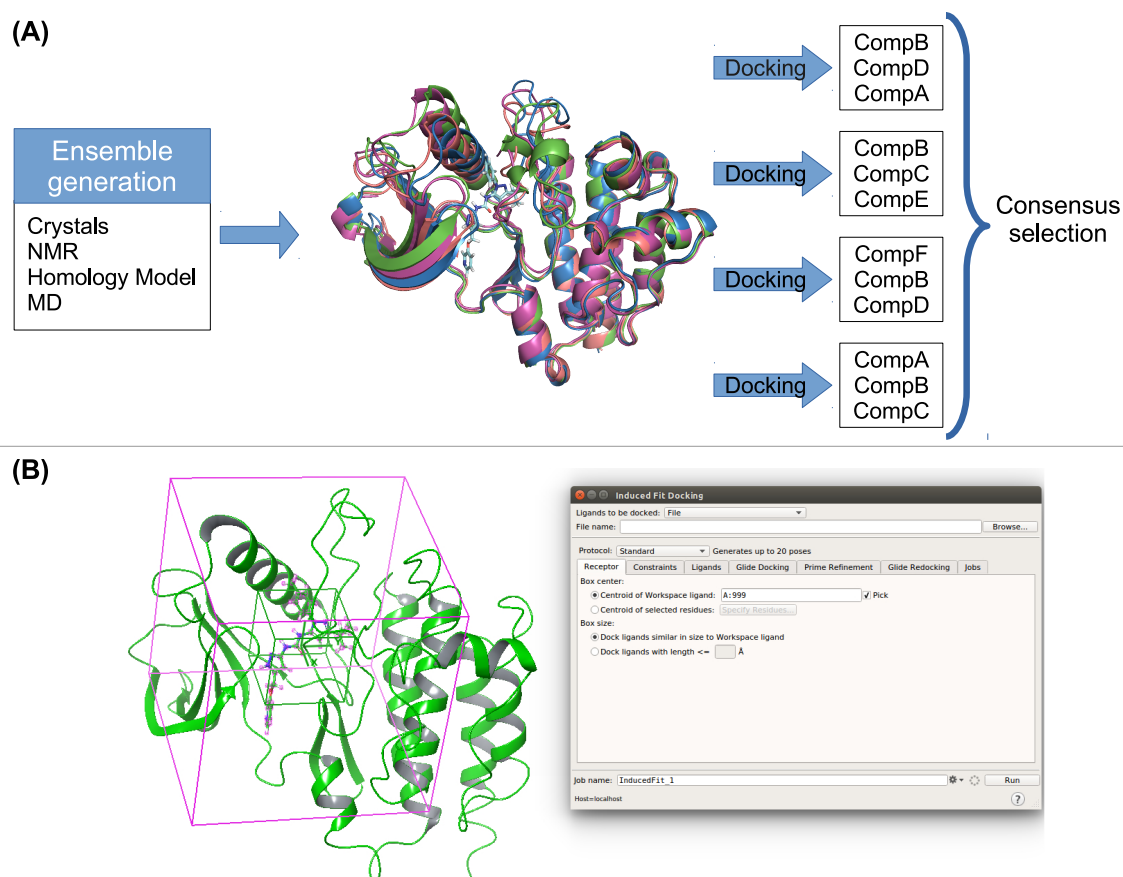
[http://blogs.sciencemag.org/pipeline/archives/2015/04/03/sanofi\\_bets\\_on\\_schrodinger](http://blogs.sciencemag.org/pipeline/archives/2015/04/03/sanofi_bets_on_schrodinger)



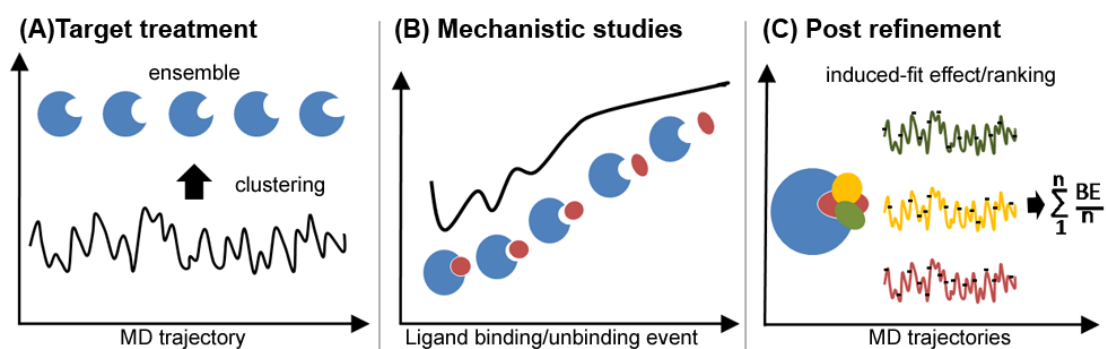
**Figure 1** Different protein-ligand interaction mechanisms: Lock-key (a), induced fit (b), conformational selection (c) and allosteric models (d).



**Figure 2** A quick simplified view of what can be done today in molecular modeling. QM dynamics, for example, can only be applied to small systems and for few femtoseconds. On the other site, huge systems and large propagation times require a CG approach.

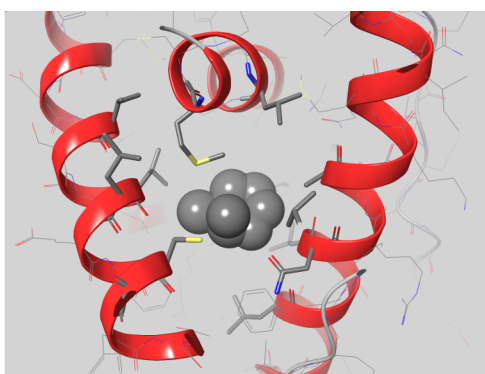


**Figure 3.** (A) Scheme illustrating the idea behind an ensemble docking approach: generation of receptor snapshots, docking, and consensus (or ranking) selection. (B) Screen capture of an IFD job setup with the Maestro graphical interface developed by Schrodinger.



**Figure 4.** Different MD contributions for exploring target flexibility in drug design. (A) Target treatment: to generate a receptor conformational ensemble from MD trajectories. (B) Mechanistic studies: to investigate the ligand binding/unbinding pathway and to derive binding free energies. (C) Post refinement: to simulate the stability, induced-fit and interaction energies of docking poses, in order to improve pose/ligand ranking and for a better structural characterization.





Representative biasing repulsive *docking* techniques:

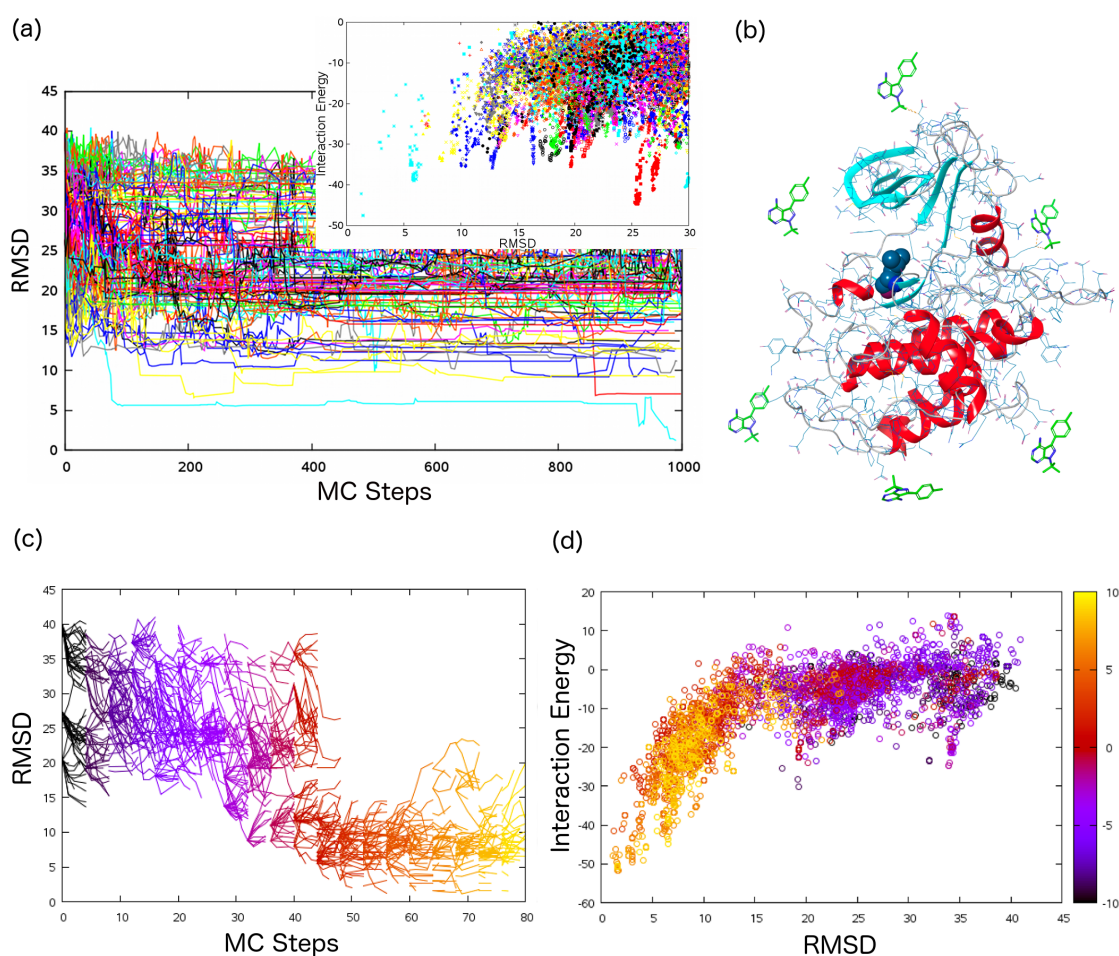
**Fumigation** (ICM):  
Repulsive generic ligand

**Essential Dynamics** (Orozco Lab.):  
Perturbation based on a probe ligand

**Active Site Pressurization** (Laughton Lab.):  
Lennard-Jones particle

**RP-REMD-DOCK** (Zacharias Lab.):  
Hamiltonian REMD with repulsive biasing potential

**Figure 5.** Several docking and MD-based techniques use the addition of repulsive forces in the active site in order to map its flexibility. In the left illustration we grew some Lennard-Jones particles into the mineral corticoid nuclear hormone receptor active site, aiming at opening it for ligand docking.



**Figure 6.** Binding site search and docking simulation for the Src kinase with the PP1 inhibitor (PDB entry 1QCF). (a) Ligand RMSD evolution to the bound crystal along 1000 MC steps for a

64-processor standard PELE job. Notice how the red and light blue trajectories find the active site at the end of the simulation (the blue one reaching  $\sim 1\text{\AA}$  heavy atom ligand RMSD). The inset shows the interaction energies with respect to the RMSD, where we see the light blue processor reaching the best energies. (b) The kinase structure showing the 6 different initial positions of the PP1 ligand in the bulk solvent. The active site is highlighted with Leu354 space fill representation. Note that we superimpose here all 6 positions, but each trajectory has only one ligand. (c) Ligand RMSD evolution when using the new Adaptive PELE for the same system and trajectories (64). Notice how the adaptive scheme allows reaching the active site in  $\sim 50$  MC steps (less than one hour!), improving almost 20x the efficiency of standard PELE. (d) Interaction energy for the adaptive simulation where we can clearly identify the bound pose as the best one. The color scheme (black to yellow) describes the epochs' evolution in the adaptive procedure.

## REFERENCES

1. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* 2017, 16:19-34.
2. Fischer E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft* 1894, 27:2985-2993.
3. Koshland DE. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* 1958, 44:98-104.
4. Frauenfelder H, Parak F, Young RD. Conformational substates in proteins. *Annu Rev Biophys Biophys Chem* 1988, 17:451-479.
5. Sliwoski G, Kothiwale S, Meiler J, Lowe EW, Jr. Computational methods in drug discovery. *Pharmacol Rev* 2014, 66:334-395.
6. Henzler AM, Rarey M. Protein Flexibility in Structure-Based Virtual Screening: From Models to Algorithms. In: *Virtual Screening*: Wiley-VCH Verlag GmbH & Co. KGaA; 2011, 223-244.
7. Carlson HA, McCammon JA. Accommodating Protein Flexibility in Computational Drug Design &span class="xref-sep">&lt;/span>. *Mol Pharmacol* 2000, 57:213.
8. Teague SJ. Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov* 2003, 2:527-541.
9. Cozzini P, Kellogg GE, Spyraakis F, Abraham DJ, Costantino G, Emerson A, Fanelli F, Gohlke H, Kuhn LA, Morris GM, et al. Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem* 2008, 51:6237-6255.
10. Wagner JR, Lee CT, Durrant JD, Malmstrom RD, Feher VA, Amaro RE. Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem Rev* 2016, 116:6370-6390.
11. Carlson HA. Protein flexibility and drug design: how to hit a moving target. *Curr Opin Chem Biol* 2002, 6:447-452.
12. Burley SK, Berman HM, Christie C, Duarte JM, Feng Z, Westbrook J, Young J, Zardecki C. RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Sci* 2018, 27:316-330.
13. Davis AM, Teague SJ, Kleywegt GJ. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew Chem Int Ed Engl* 2003, 42:2718-2736.
14. Li Y, Kang C. Solution NMR Spectroscopy in Target-Based Drug Discovery. *Molecules* 2017, 22.
15. Buonfiglio R, Recanatini M, Masetti M. Protein Flexibility in Drug Discovery: From Theory to Computation. *ChemMedChem* 2015, 10:1141-1148.
16. Schuetz DA, de Witte WEA, Wong YC, Knasmueller B, Richter L, Kokh DB, Sadiq SK, Bosma R, Nederpelt I, Heitman LH, et al. Kinetics for Drug Discovery: an industry-driven effort to target drug residence time. *Drug Discov Today* 2017, 22:896-911.
17. Huang SY, Zou X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* 2007, 66:399-421.

18. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J Med Chem* 2006, 49:534-553.
19. Nedumpully-Govindan P, Jemec DB, Ding F. CSAR Benchmark of Flexible MedusaDock in Affinity Prediction and Nativelike Binding Pose Selection. *J Chem Inf Model* 2016, 56:1042-1052.
20. Davis IW, Baker D. RosettaLigand docking with full ligand and receptor flexibility. *J Mol Biol* 2009, 385:381-392.
21. Gagnon JK, Law SM, Brooks CL, 3rd. Flexible CDOCKER: Development and application of a pseudo-explicit structure-based docking method within CHARMM. *J Comput Chem* 2016, 37:753-762.
22. Borrelli KW, Vitalis A, Alcantara R, Guallar V. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J Chem Theory Comput* 2005, 1:1304-1311.
23. Carlson HA, Smith RD, Damm-Ganamet KL, Stuckey JA, Ahmed A, Convery MA, Somers DO, Kranz M, Elkins PA, Cui G, et al. CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma. *J Chem Inf Model* 2016, 56:1063.
24. Grebner C, Iegre J, Ulander J, Edman K, Hogner A, Tyrchan C. Binding Mode and Induced Fit Predictions for Prospective Computational Drug Design. *J Chem Inf Model* 2016.
25. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* 2002, 9:646-652.
26. Artola M, Wu L, Ferraz MJ, Kuo C-L, Raich L, Breen IZ, Offen WA, Codée JDC, van der Marel GA, Rovira C, et al. 1,6-Cyclophellitol Cyclosulfates: A New Class of Irreversible Glycosidase Inhibitor. *ACS Central Science* 2017, 3:784-793.
27. Cho AE, Guallar V, Berne BJ, Friesner RA. Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) approach. *J Comput Chem* 2005, 26:915-931.
28. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, et al. Anton, a special-purpose machine for molecular dynamics simulation. *Proceedings of the 34th annual international symposium on Computer architecture* 2007, Pages 1-12.
29. Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. How Does a Drug Molecule Find Its Target Binding Site? *JACS* 2011, 133:9181-9183.
30. Isralewitz B, Gao M, Schulten K. Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol* 2001, 11:224-230.
31. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 1999, 314:141-151.
32. Leone V, Marinelli F, Carloni P, Parrinello M. Targeting biomolecular flexibility with metadynamics. *Curr Opin Struct Biol* 2010, 20:148-154.
33. Orozco M. A theoretical view of protein dynamics. *Chem Soc Rev* 2014.
34. Shirts MR. Best practices in free energy calculations for drug design. *Methods Mol Biol* 2012, 819:425-467.
35. Wang C, Nguyen PH, Pham K, Huynh D, Le TB, Wang H, Ren P, Luo R. Calculating protein-ligand binding affinities with MMPBSA: Method and error analysis. *J Comput Chem* 2016, 37:2436-2446.

36. Wang W, Cao S, Zhu L, Huang X. Constructing Markov State Models to elucidate the functional conformational changes of complex biomolecules. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2018, 8:n/a-n/a.
37. Plattner N, Noé F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nature Communications* 2015, 6:7653.
38. Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS. Alchemical free energy methods for drug discovery: progress and challenges. *Curr Opin Struct Biol* 2011, 21:150-160.
39. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *JACS* 2015, 137:2695-2703.
40. Allen BK, Mehta S, Ember SW, Schonbrunn E, Ayad N, Schurer SC. Large-Scale Computational Screening Identifies First in Class Multitarget Inhibitor of EGFR Kinase and BRD4. *Sci Rep* 2015, 5:16924.
41. Mullard A. The drug-maker's guide to the galaxy. *Nature* 2017, 549:445-447.
42. Cho AE, Wendel JA, Vaidehi N, Keken-Huskey PM, Floriano WB, Maiti PK, Goddard WA, 3rd. The MPSim-Dock hierarchical docking algorithm: application to the eight trypsin inhibitor cocrystals. *J Comput Chem* 2005, 26:48-71.
43. De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem* 2016, 59:4035-4061.
44. Siddiqi AR, Nioche P, Siddiqui AB, Rauf SA, Waseem A, Villoutreix BO. EFFICIENCY OF A HIERARCHICAL DOCKING PROTOCOL FOR COMPUTATIONAL LIGAND SCREENING AGAINST HOMOLGY MODELS. *Biomedical Engineering: Applications, Basis and Communications* 2014, 26:1450024.
45. Kumar A, Zhang KY. Hierarchical virtual screening approaches in small molecule drug discovery. *Methods* 2015, 71:26-37.
46. Guedes IA, de Magalhães CS, Dardenne LE. Receptor–ligand molecular docking. *Biophysical Reviews* 2014, 6:75-87.
47. Forti F, Cavasotto CN, Orozco M, Barril X, Luque FJ. A Multilevel Strategy for the Exploration of the Conformational Flexibility of Small Molecules. *J Chem Theory Comput* 2012, 8:1808-1819.
48. Dreher J, Scheiber J, Stiefl N, Baumann K. xMaP—An Interpretable Alignment-Free Four-Dimensional Quantitative Structure–Activity Relationship Technique Based on Molecular Surface Properties and Conformer Ensembles. *J Chem Inf Model* 2017.
49. Ferrari AM, Wei BQ, Costantino L, Shoichet BK. Soft docking and multiple receptor conformations in virtual screening. *J Med Chem* 2004, 47:5076-5084.
50. McGann M. FRED and HYBRID docking performance on standardized datasets. *J Comput Aided Mol Des* 2012, 26:897-906.
51. Evain-Bana E, Schiavo L, Bour C, Lanfranchi DA, Berardozi S, Ghirga F, Bagrel D, Botta B, Hanquet G, Mori M. Synthesis, biological evaluation and

- molecular modeling studies on novel quinonoid inhibitors of CDC25 phosphatases. *J Enzyme Inhib Med Chem* 2017, 32:113-118.
52. Quevedo MA, Ribone SR, Brinon MC, Dehaen W. Development of a receptor model for efficient in silico screening of HIV-1 integrase inhibitors. *J Mol Graph Model* 2014, 52:82-90.
  53. Kahlous NA, Bawarish MAM, Sarhan MA, Kupper M, Hasaba A, Rajab M. Using Chemoinformatics, Bioinformatics, and Bioassay to Predict and Explain the Antibacterial Activity of Nonantibiotic Food and Drug Administration Drugs. *Assay Drug Dev Technol* 2017, 15:89-105.
  54. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol* 2008, 18:178-184.
  55. Bottegoni G, Kufareva I, Totrov M, Abagyan R. Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking. *J Med Chem* 2009, 52:397-406.
  56. Sager G, Ørvoll EØ, Lysaa RA, Kufareva I, Abagyan R, Ravna AW. Novel cGMP Efflux Inhibitors Identified by Virtual Ligand Screening (VLS) and Confirmed by Experimental Studies. *J Med Chem* 2012, 55:3049-3057.
  57. Tian S, Sun H, Pan P, Li D, Zhen X, Li Y, Hou T. Assessing an Ensemble Docking-Based Virtual Screening Strategy for Kinase Targets by Considering Protein Flexibility. *J Chem Inf Model* 2014, 54:2664-2679.
  58. Middendorp SJ, Puthenkalam R, Baur R, Ernst M, Sigel E. Accelerated discovery of novel benzodiazepine ligands by experiment-guided virtual screening. *ACS Chem Biol* 2014, 9:1854-1859.
  59. Gabrielsen M, Kurczab R, Siwek A, Wolak M, Ravna AW, Kristiansen K, Kufareva I, Abagyan R, Nowak G, Chilmonczyk Z, et al. Identification of novel serotonin transporter compounds by virtual screening. *J Chem Inf Model* 2014, 54:933-943.
  60. Hou X, Li K, Yu X, Sun JP, Fang H. Protein Flexibility in Docking-Based Virtual Screening: Discovery of Novel Lymphoid-Specific Tyrosine Phosphatase Inhibitors Using Multiple Crystal Structures. *J Chem Inf Model* 2015, 55:1973-1983.
  61. Meirson T, Samson AO, Gil-Henn H. An in silico high-throughput screen identifies potential selective inhibitors for the non-receptor tyrosine kinase Pyk2. *Drug Des Devel Ther* 2017, 11:1535-1557.
  62. Li DD, Meng XF, Wang Q, Yu P, Zhao LG, Zhang ZP, Wang ZZ, Xiao W. Consensus scoring model for the molecular docking study of mTOR kinase inhibitor. *J Mol Graph Model* 2017, 79:81-87.
  63. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. *Proteins* 2003, 52:609-623.
  64. Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. *J Med Chem* 2006, 49:3315-3321.
  65. Korb O, Stützle T, Exner TE. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In: Dorigo M, Gambardella LM, Birattari M, Martinoli A, Poli R, Stützle T, eds. *Ant Colony Optimization and Swarm Intelligence: 5th International Workshop, ANTS 2006, Brussels, Belgium, September 4-7, 2006. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006, 247-258.

66. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 2009, 30:2785-2791.
67. Schumann M, Armen RS. Systematic and efficient side chain optimization for molecular docking using a cheapest-path procedure. *J Comput Chem* 2013, 34:1258-1269.
68. Neves MA, Totrov M, Abagyan R. Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J Comput Aided Mol Des* 2012, 26:675-686.
69. Matijssen C, Silva-Santisteban MC, Westwood IM, Siddique S, Choi V, Sheldrake P, van Montfort RL, Blagg J. Benzimidazole inhibitors of the protein kinase CHK2: clarification of the binding mode by flexible side chain docking and protein-ligand crystallography. *Bioorg Med Chem* 2012, 20:6630-6639.
70. Gupta S, Jadaun A, Kumar H, Raj U, Varadwaj PK, Rao AR. Exploration of new drug-like inhibitors for serine/threonine protein phosphatase 5 of *Plasmodium falciparum*: a docking and simulation study. *J Biomol Struct Dyn* 2015, 33:2421-2441.
71. Bruser A, Zimmermann A, Crews BC, Sliwoski G, Meiler J, König GM, Kostenis E, Lede V, Marnett LJ, Schöneberg T. Prostaglandin E2 glyceryl ester is an endogenous agonist of the nucleotide receptor P2Y6. *Sci Rep* 2017, 7:2380.
72. Luo Q, Chen L, Cheng X, Ma Y, Li X, Zhang B, Li L, Zhang S, Guo F, Li Y, et al. An allosteric ligand-binding site in the extracellular cap of K2P channels. *Nat Commun* 2017, 8:378.
73. Ding X, Hayes RL, Vilseck JZ, Charles MK, Brooks CL, 3rd. CDOCKER and [Formula: see text]-dynamics for prospective prediction in D3R Grand Challenge 2. *J Comput Aided Mol Des* 2017.
74. Kazantsev AG, Thompson LM, Abagyan R, Casale M. Small molecule activators of NRF2 pathway. 2017. Vol. US9737525B2.
75. Lane JR, Chubukov P, Liu W, Canals M, Cherezov V, Abagyan R, Stevens RC, Katritch V. Structure-based ligand discovery targeting orthosteric and allosteric pockets of dopamine receptors. *Mol Pharmacol* 2013, 84:794-807.
76. Martínez-Sotres C, Rutiaga-Quñones JG, Herrera-Bucio R, Gallo M, López-Albarrán P. Molecular docking insights into the inhibition of laccase activity by medicarpin. *Wood Science and Technology* 2015, 49:857-868.
77. Khago D, Wong EK, Kingsley CN, Freitas JA, Tobias DJ, Martin RW. Increased hydrophobic surface exposure in the cataract-related G18V variant of human gammaS-crystallin. *Biochim Biophys Acta* 2016, 1860:325-332.
78. Kordbacheh H, Eftekhari F, Ebrahimi SN. Anti-quorum sensing activity of *Pistacia atlantica* against *Pseudomonas aeruginosa* PAO1 and identification of its bioactive compounds. *Microb Pathog* 2017, 110:390-398.
79. Chatzileontiadou DS, Tsika AC, Diamantopoulou Z, Delbe J, Badet J, Courty J, Skamnaki VT, Parmenopoulou V, Komiotis D, Hayes JM, et al. Evidence for novel action at the cell binding site of human Angiogenin revealed by heteronuclear NMR spectroscopy, in silico and in vivo studies. *ChemMedChem* 2018.

80. Ding F, Dokholyan NV. Incorporating backbone flexibility in MedusaDock improves ligand-binding pose prediction in the CSAR2011 docking benchmark. *J Chem Inf Model* 2013, 53:1871-1879.
81. Nazir N, Karim N, Abdel-Halim H, Khan I, Wadood SF, Nisar M. Phytochemical analysis, molecular docking and antiamnesic effects of methanolic extract of *Silybum marianum* (L.) Gaertn seeds in scopolamine induced memory impairment in mice. *J Ethnopharmacol* 2018, 210:198-208.
82. Clark AJ, Tiwary P, Borrelli K, Feng S, Miller EB, Abel R, Friesner RA, Berne BJ. Prediction of Protein–Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations. *J Chem Theory Comput* 2016, 12:2990-2998.
83. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biol* 2011, 9:71.
84. Ganesan A, Coote ML, Barakat K. Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discovery Today* 2017, 22:249-269.
85. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. *Med Res Rev* 2006, 26:531-568.
86. Zhao H, Huang D, Caflisch A. Discovery of Tyrosine Kinase Inhibitors by Docking into an Inactive Kinase Conformation Generated by Molecular Dynamics. *ChemMedChem* 2012, 7:1983-1990.
87. Kokh DB, Richter S, Henrich S, Czodrowski P, Rippmann F, Wade RC. TRAPP: A Tool for Analysis of Transient Binding Pockets in Proteins. *J Chem Inf Model* 2013, 53:1235-1252.
88. Bernini A, Henrici De Angelis L, Morandi E, Spiga O, Santucci A, Assfalg M, Molinari H, Pillozzi S, Arcangeli A, Niccolai N. Searching for protein binding sites from Molecular Dynamics simulations and paramagnetic fragment-based NMR studies. *Biochim Biophys Acta* 2014, 1844:561-566.
89. Abel R, Young T, Farid R, Berne BJ, Friesner RA. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *JACS* 2008, 130:2817-2831.
90. Wassman CD, Baronio R, Demir Ö, Wallentine BD, Chen C-K, Hall LV, Salehi F, Lin D-W, Chung BP, Wesley Hatfield G, et al. Computational identification of a transiently open L1/S3 pocket for reactivation of mutant p53. *Nature Communications* 2013, 4:1407.
91. Joerger AC, Ang HC, Fersht AR. Structural basis for understanding oncogenic p53 mutations and designing rescue drugs. *Proceedings of the National Academy of Sciences* 2006, 103:15056-15061.
92. Joerger AC, Bauer MR, Wilcken R, Baud MGJ, Harbrecht H, Exner TE, Boeckler FM, Spencer J, Fersht AR. Exploiting Transient Protein States for the Design of Small-Molecule Stabilizers of Mutant p53. *Structure* 2015, 23:2246-2255.
93. Osguthorpe DJ, Sherman W, Hagler AT. Exploring Protein Flexibility: Incorporating Structural Ensembles From Crystal Structures and Simulation into Virtual Screening Protocols. *The Journal of Physical Chemistry B* 2012, 116:6952-6959.



94. Osguthorpe DJ, Sherman W, Hagler AT. Generation of Receptor Structural Ensembles for Virtual Screening Using Binding Site Shape Analysis and Clustering. *Chem Biol Drug Des* 2012, 80:182-193.
95. Antolin AA, Carotti A, Nuti R, Hakkaya A, Camaioni E, Mestres J, Pellicciari R, Macchiarulo A. Exploring the effect of PARP-1 flexibility in docking studies. *J Mol Graphics Modell* 2013, 45:192-201.
96. Cole JC, Murray CW, Nissink JWM, Taylor RD, Taylor R. Comparing protein-ligand docking programs is difficult. *Proteins: Structure, Function, and Bioinformatics* 2005, 60:325-332.
97. Amaro RE, Baron R, McCammon JA. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J Comput Aided Mol Des* 2008, 22:693-705.
98. Nichols SE, Baron R, McCammon JA. On the Use of Molecular Dynamics Receptor Conformations for Virtual Screening. In: Baron R, ed. *Computational Drug Discovery and Design*. New York, NY: Springer New York; 2012, 93-103.
99. Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA. Discovery of a Novel Binding Trench in HIV Integrase. *J Med Chem* 2004, 47:1879-1881.
100. Summa V, Petrocchi A, Bonelli F, Crescenzi B, Donghi M, Ferrara M, Fiore F, Gardelli C, Gonzalez Paz O, Hazuda DJ, et al. Discovery of Raltegravir, a Potent, Selective Orally Bioavailable HIV-Integrase Inhibitor for the Treatment of HIV-AIDS Infection. *J Med Chem* 2008, 51:5843-5855.
101. Cheng LS, Amaro RE, Xu D, Li WW, Arzberger PW, McCammon JA. Ensemble-Based Virtual Screening Reveals Potential Novel Antiviral Compounds for Avian Influenza Neuraminidase. *J Med Chem* 2008, 51:3878-3894.
102. Miao Y, Goldfeld DA, Moo EV, Sexton PM, Christopoulos A, McCammon JA, Valant C. Accelerated structure-based design of chemically diverse allosteric modulators of a muscarinic G protein-coupled receptor. *Proceedings of the National Academy of Sciences* 2016, 113:E5675-E5684.
103. Chaudhuri R, Carrillo O, Laughton CA, Orozco M. Application of Drug-Perturbed Essential Dynamics/Molecular Dynamics (ED/MD) to Virtual Screening and Rational Drug Design. *J Chem Theory Comput* 2012, 8:2204-2214.
104. Zacharias M. Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: Binding of FK506 to FKBP. *Proteins: Structure, Function, and Bioinformatics* 2004, 54:759-767.
105. Louet M, Perahia D, Martinez J, Floquet N. A Concerted Mechanism for Opening the GDP Binding Pocket and Release of the Nucleotide in Hetero-Trimeric G-Proteins. *J Mol Biol* 2011, 411:298-312.
106. Preininger AM, Meiler J, Hamm HE. Conformational Flexibility and Structural Dynamics in GPCR-Mediated G Protein Activation: A Perspective. *J Mol Biol* 2013, 425:2288-2298.
107. Shukla D, Meng Y, Roux B, Pande VS. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nature Communications* 2014, 5:3397.

108. Alvarez-Garcia D, Barril X. Relationship between Protein Flexibility and Binding: Lessons for Structure-Based Drug Design. *J Chem Theory Comput* 2014, 10:2608-2614.
109. de Beer SB, Vermeulen NP, Oostenbrink C. The role of water molecules in computational drug design. *Curr Top Med Chem* 2010, 10:55-66.
110. Li Z, Lazaridis T. Computing the thermodynamic contributions of interfacial water.
111. Bucher D, Stouten P, Triballeau N. Shedding Light on Important Waters for Drug Design: Simulations versus Grid-Based Methods. *J Chem Inf Model* 2018.
112. Balias TE, Fischer M, Stein RM, Adler TB, Nguyen CN, Cruz A, Gilson MK, Kurtzman T, Shoichet BK. Testing inhomogeneous solvation theory in structure-based ligand discovery. *Proc Natl Acad Sci USA* 2017, 114:E6839-E6846.
113. De Vivo M, Cavalli A. Recent advances in dynamic docking for drug discovery. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2017, 7:n/a-n/a.
114. Laio A, Parrinello M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* 2002, 99:12562-12566.
115. Limongelli V, Bonomi M, Marinelli L, Gervasio FL, Cavalli A, Novellino E, Parrinello M. Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition. *Proceedings of the National Academy of Sciences* 2010, 107:5411-5416.
116. Incerti M, Russo S, Callegari D, Pala D, Giorgio C, Zanotti I, Barocelli E, Vicini P, Vacondio F, Rivara S, et al. Metadynamics for Perspective Drug Design: Computationally Driven Synthesis of New Protein–Protein Interaction Inhibitors Targeting the EphA2 Receptor. *J Med Chem* 2017, 60:787-796.
117. Colizzi F, Perozzo R, Scapozza L, Recanatini M, Cavalli A. Single-Molecule Pulling Simulations Can Discern Active from Inactive Enzyme Inhibitors. *JACS* 2010, 132:7361-7371.
118. Patel JS, Berteotti A, Ronsisvalle S, Rocchia W, Cavalli A. Steered Molecular Dynamics Simulations for Studying Protein–Ligand Interaction in Cyclin-Dependent Kinase 5. *J Chem Inf Model* 2014, 54:470-480.
119. Vuong QV, Nguyen TT, Li MS. 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* 2015, 55:2731-2738.
120. Okimoto N, Suenaga A, Taiji M. Evaluation of protein–ligand affinity prediction using steered molecular dynamics simulations. *J Biomol Struct Dyn* 2017, 35:3221-3231.
121. Baker JL, Biais N, Tama F. Steered Molecular Dynamics Simulations of a Type IV Pilus Probe Initial Stages of a Force-Induced Conformational Transition. *PLoS Comp Biol* 2013, 9:e1003032.
122. Ostermeir K, Zacharias M. Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins. *Biochim Biophys Acta* 2013, 1834:847-853.
123. Ostermeir K, Zacharias M. Accelerated flexible protein-ligand docking using Hamiltonian replica exchange with a repulsive biasing potential. *PLoS One* 2017, 12:e0172072.

124. Fisher CK, Stultz CM. Constructing ensembles for intrinsically disordered proteins. *Curr Opin Struct Biol* 2011, 21:426-431.
125. Cecchini M, Rao F, Seeber M, Caflisch A. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *The Journal of Chemical Physics* 2004, 121:10748-10756.
126. Kokubo H, Tanaka T, Okamoto Y. Prediction of Protein-Ligand Binding Structures by Replica-Exchange Umbrella Sampling Simulations: Application to Kinase Systems. *J Chem Theory Comput* 2013, 9:4660-4671.
127. Withers IM, Mazanetz MP, Wang H, Fischer PM, Laughton CA. Active site pressurization: a new tool for structure-guided drug design and other studies of protein flexibility. *J Chem Inf Model* 2008, 48:1448-1454.
128. Mazanetz MP, Laughton CA, Fischer PM. Investigation of the flexibility of protein kinases implicated in the pathology of Alzheimer's disease. *Molecules* 2014, 19:9134-9159.
129. Emtage AL, Mistry SN, Fischer PM, Kellam B, Laughton CA. GPCRs through the keyhole: the role of protein flexibility in ligand binding to beta-adrenoceptors. *J Biomol Struct Dyn* 2017, 35:2604-2619.
130. Pan AC, Xu H, Palpant T, Shaw DE. Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. *J Chem Theory Comput* 2017, 13:3372-3377.
131. Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* 2013, 9:3878-3888.
132. Doerr S, Harvey MJ, Noe F, De Fabritiis G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J Chem Theory Comput* 2016, 12:1845-1852.
133. Decherchi S, Berteotti A, Bottegoni G, Rocchia W, Cavalli A. The ligand binding mechanism to purine nucleoside phosphorylase elucidated via molecular dynamics and machine learning. *Nat Commun* 2015, 6:6155.
134. Lotz SD, Dickson A. Unbiased Molecular Dynamics of 11 min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. *J Am Chem Soc* 2018.
135. Bruce NJ, Ganotra GK, Kokh DB, Sadiq SK, Wade RC. New approaches for computing ligand-receptor binding kinetics. *Curr Opin Struct Biol* 2017, 49:1-10.
136. Sinitskiy AV, Pande VS. Simulated Dynamics of Glycans on Ligand-Binding Domain of NMDA Receptors Reveals Strong Dynamic Coupling between Glycans and Protein Core. *J Chem Theory Comput* 2017, 13:5496-5505.
137. Sultan MM, Denny RA, Unwalla R, Lovering F, Pande VS. Millisecond dynamics of BTK reveal kinome-wide conformational plasticity within the apo kinase domain. *Sci Rep* 2017, 7:15604.
138. Shukla D, Lawrenz M, Pande VS. Elucidating Ligand-Modulated Conformational Landscape of GPCRs Using Cloud-Computing Approaches. *Methods Enzymol* 2015, 557:551-572.
139. Chen Y-C. Beware of docking! *Trends Pharmacol Sci* 2015, 36:78-95.
140. Sgobba M, Caporuscio F, Anighoro A, Portioli C, Rastelli G. Application of a post-docking procedure based on MM-PBSA and MM-GBSA on single and multiple protein conformations. *Eur J Med Chem* 2012, 58:431-440.

141. Greenidge PA, Kramer C, Mozziconacci J-C, Wolf RM. MM/GBSA Binding Energy Prediction on the PDBbind Data Set: Successes, Failures, and Directions for Further Improvement. *J Chem Inf Model* 2013, 53:201-209.
142. Greenidge PA, Kramer C, Mozziconacci JC, Sherman W. Improving Docking Results via Reranking of Ensembles of Ligand Poses in Multiple X-ray Protein Conformations with MM-GBSA. *J Chem Inf Model* 2014, 54:2697-2717.
143. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* 2015, 10:449-461.
144. Wichapong K, Rohe A, Platzer C, Slynko I, Erdmann F, Schmidt M, Sippl W. Application of Docking and QM/MM-GBSA Rescoring to Screen for Novel Myt1 Kinase Inhibitors. *J Chem Inf Model* 2014, 54:881-893.
145. Ryde U, Söderhjelm P. Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chem Rev* 2016, 116:5520-5566.
146. Johan A, John M. The Linear Interaction Energy Method for Predicting Ligand Binding Free Energies. *Combinatorial Chem High Throughput Screening* 2001, 4:613-626.
147. Bjelic S, Nervall M, Gutiérrez-de-Terán H, Ersmark K, Hallberg A, Åqvist J. Computational inhibitor design against malaria plasmepsins. *Cell Mol Life Sci* 2007, 64:2285-2305.
148. Gutiérrez-de-Terán H, Åqvist J. Linear Interaction Energy: Method and Applications in Drug Design. In: Baron R, ed. *Computational Drug Discovery and Design*. New York, NY: Springer New York; 2012, 305-323.
149. Capoferri L, Verkade-Vreeker MCA, Buitenhuis D, Commandeur JNM, Pastor M, Vermeulen NPE, Geerke DP. Linear Interaction Energy Based Prediction of Cytochrome P450 1A2 Binding Affinities with Reliability Estimation. *PLoS One* 2015, 10:e0142232.
150. Genheden S, Ryde U. Comparison of the Efficiency of the LIE and MM/GBSA Methods to Calculate Ligand-Binding Energies. *J Chem Theory Comput* 2011, 7:3768-3778.
151. Xu L, Sun H, Li Y, Wang J, Hou T. Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models. *J Phys Chem B* 2013, 117:8408-8421.
152. Merz KM, Kollman PA. Free energy perturbation simulations of the inhibition of thermolysin: prediction of the free energy of binding of a new inhibitor. *JACS* 1989, 111:5649-5658.
153. Ferguson DM, Radmer RJ, Kollman PA. Determination of the relative binding free energies of peptide inhibitors to the HIV-1 protease. *J Med Chem* 1991, 34:2654-2659.
154. Sherborne B, Shanmugasundaram V, Cheng AC, Christ CD, DesJarlais RL, Duca JS, Lewis RA, Loughney DA, Manas ES, McGaughey GB, et al. Collaborating to improve the use of free-energy and other quantitative methods in drug discovery. *J Comput Aided Mol Des* 2016, 30:1139-1141.
155. Lenselink EB, Louvel J, Forti AF, van Veldhoven JPD, de Vries H, Mulder-Krieger T, McRobb FM, Negri A, Goose J, Abel R, et al. Predicting Binding Affinities for GPCR Ligands Using Free-Energy Perturbation. *ACS Omega* 2016, 1:293-304.

156. Schmidt TC, Eriksson P-O, Gustafsson D, Cosgrove D, Frølund B, Boström J. Discovery and Evaluation of Anti-Fibrinolytic Plasmin Inhibitors Derived from 5-(4-Piperidyl)isoxazol-3-ol (4-PIOL). *J Chem Inf Model* 2017, 57:1703-1714.
157. Wako H, Scheraga HA. Use of distance constraints to fold a protein. *Macromolecules* 1981, 14:961-969.
158. Li ZQ, Scheraga HA. Monte Carlo -minimization approach to the multiple-minima problem in protein folding. *Proc Natl Acad Sci USA* 1987, 84:2633-2636.
159. Chang G, Guida WC, Still WC. An internal-coordinate Monte Carlo method for searching conformational space. *JACS* 1989, 111:4379-4386.
160. Lee W-G, Gallardo-Macias R, Frey KM, Spasov KA, Bollini M, Anderson KS, Jorgensen WL. Picomolar Inhibitors of HIV Reverse Transcriptase Featuring Bicyclic Replacement of a Cyanovinylphenyl Group. *JACS* 2013, 135:16705-16713.
161. Michel J, Essex JW. Hit Identification and Binding Mode Predictions by Rigorous Free Energy Simulations. *J Med Chem* 2008, 51:6654-6664.
162. Jorgensen WL, Tirado-Rives J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J Comput Chem* 2005, 26:1689-1700.
163. Michel J, Essex JW. Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J Comput Aided Mol Des* 2010, 24:639-658.
164. C J Woods JM, M Bodnarchuk, S Genheden, R Bradshaw, G A Ross, C Cave-Ayland, H Bruce-Macdonald, A I Cabedo Martinez and J Graham. ProtoMS 3.3. <http://protoms.org>. 2017.
165. Jorgensen WL. Computer-aided discovery of anti-HIV agents. *Bioorg Med Chem* 2016, 24:4768-4778.
166. Lovering F, Aevazelis C, Chang J, Dehnhardt C, Fitz L, Han S, Janz K, Lee J, Kaila N, McDonald J, et al. Imidazotriazines: Spleen Tyrosine Kinase (Syk) Inhibitors Identified by Free-Energy Perturbation (FEP). *ChemMedChem* 2016, 11:217-233.
167. Wang NX, Watson GB, Loso MR, Sparks TC. Molecular modeling of sulfoxaflor and neonicotinoid binding in insect nicotinic acetylcholine receptors: impact of the Myzus beta1 R81T mutation.
168. Peng Z, Maxwell DS, Sun D, Bhanu Prasad BA, Pal A, Wang S, Balatoni J, Ghosh P, Lim ST, Volgin A, et al. Imatinib analogs as potential agents for PET imaging of Bcr-Abl and c-KIT expression at a kinase level.
169. Ross GA, Bodnarchuk MS, Essex JW. Water Sites, Networks, And Free Energies with Grand Canonical Monte Carlo. *JACS* 2015, 137:14930-14943.
170. Bodnarchuk MS, Viner R, Michel J, Essex JW. Strategies to Calculate Water Binding Free Energies in Protein–Ligand Complexes. *J Chem Inf Model* 2014.
171. Nilmeier JP, Crooks GE, Minh DDL, Chodera JD. Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation. *Proceedings of the National Academy of Sciences* 2011, 108:E1009-E1018.
172. Samuel G, Nathan M. L, Patrick G, Ariën S. R, Josh F, Gregory R, John D. C, David M. *Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes Using Nonequilibrium Candidate Monte Carlo*; 2017.

173. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys J* 2001, 80:505-515.
174. Jacobson MP, Kaminski GA, Friesner RA, Rapp CS. Force field validation using protein side chain prediction. *J Phys Chem B* 2002, 106:11673-11680.
175. Zhu K, Shirts MR, Friesner RA, Jacobson MP. Multiscale optimization of a truncated Newton minimization algorithm and application to proteins and protein-ligand complexes. *J Chem Theory Comput* 2007, 3:640-648.
176. Lecina D. A study on protein-ligand interactions using a Monte Carlo procedure. *Physics* 2017. Vol. PhD.
177. Borrelli KW, Vitalis A, Alcantara R, Guallar V. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J Chem Theory Comput* 2005, 1:1304-1311.
178. Lecina D, Gilabert JF, Guallar V. Adaptive simulations, towards interactive protein-ligand modeling. *Sci Rep* 2017, 7:8466.
179. Banks JL, Beard HS, Cao Y, Cho AE, Damm W, Farid R, Felts AK, Halgren TA, Mainz DT, Maple JR, et al. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J Comput Chem* 2005, 26:1752-1780.
180. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and Reparametrization of the Opls-Aa Force Field for Proteins Via Comparison With Accurate Quantum Chemical Calculations on Peptides. *J Phys Chem B* 2001, 105:6474-6487.
181. Pérez A, Marchán I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M. Refinement of the AMBER force field for nucleic acids: improving the description of  $\alpha/\gamma$  conformers. *Biophys J* 2007, 92:3817-3829.
182. Lucas MF, Guallar V. An Atomistic View on Human Hemoglobin Carbon Monoxide Migration Processes. *Biophys J* 2012, 102:887-896.
183. Kotev M, Lecina D, Tarragó T, Giralt E, Guallar V. Unveiling Prolyl Oligopeptidase Ligand Migration by Comprehensive Computational Techniques. *Biophys J* 2015, 108:116-125.
184. Kopečná J, Cabeza de Vaca I, Adams NBP, Davison PA, Brindley AA, Hunter CN, Guallar V, Sobotka R. Porphyrin Binding to Gun4 Protein, Facilitated by a Flexible Loop, Controls Metabolite Flow through the Chlorophyll Biosynthetic Pathway. *J Biol Chem* 2015, 290:28477-28488.
185. Espona-Fiedler M, Soto-Cerrato V, Hosseini A, Lizcano JM, Guallar V, Quesada R, Gao T, Perez-Tomas R. Identification of dual mTORC1 and mTORC2 inhibitors in melanoma cells: prodigiosin vs. obatoclax. *Biochem Pharmacol* 2012, 83:489-496.
186. Hosseini A, Espona-Fiedler M, Soto-Cerrato V, Quesada R, Pérez-Tomás R, Guallar V. Molecular interactions of prodiginines with the BH3 domain of anti-apoptotic Bcl-2 family members. *PLoS One* 2013, 8.
187. Andreotti G, Cabeza de Vaca I, Poziello A, Monti MC, Guallar V, Cubellis MV. Conformational response to ligand binding in phosphomannomutase2: insights into inborn glycosylation disorder. *J Biol Chem* 2014, 289:34900-34910.
188. Edman K, Hosseini A, Bjursell Magnus K, Aagaard A, Wissler L, Gunnarsson A, Kaminski T, Köhler C, Bäckström S, Jensen Tina J, et al. Ligand Binding Mechanism in Steroid Receptors: From Conserved Plasticity to Differential Evolutionary Constraints. *Structure*.

189. Grebner C, Lecina D, Gil VA, Ulander J, Hansson P, Dellsen C, Tyrchan C, Edman K, Hogner Anders C, Guallar V. Exploration of binding mechanisms in nuclear hormone receptors by Monte Carlo simulations and X-ray derived motion modes. *Biophysical Journal* 2017, ASAP.
190. Cabeza de Vaca I, Lucas MF, Guallar V. New Monte Carlo Based Technique To Study DNA–Ligand Interactions. *J Chem Theory Comput* 2015, 11:5598-5605.
191. Takahashi R, Gil VA, Guallar V. Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations. *J Chem Theory Comput* 2014, 10:282-288.
192. Borrelli KW, Cossins B, Guallar V. Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J Comput Chem* 2010, 31:1224-1235.
193. Hosseini A, Alibés A, Noguera-Julian M, Gil VA, Paredes R, Soliva R, Orozco M, Guallar V. Computational Prediction of HIV-1 Resistance to Protease Inhibitors. *J Chem Inf Model* 2016, 56:915–923.
194. Kotev M, Soliva R, Orozco M. Challenges of docking in large, flexible and promiscuous binding sites. *Bior Med Chem* 2016, 24:4961-4969.