

Understanding the Challenges of Protein Flexibility in Drug Design

Dinler A Antunes, Didier Devaurs, and Lydia E Kaviraki

ABSTRACT

Introduction: Protein-ligand interactions play key roles in various metabolic pathways, and the proteins involved in these interactions represent major targets for drug discovery. Molecular docking is widely used to predict the structure of protein-ligand complexes, and protein flexibility stands out as one of the most important and challenging issues for binding mode prediction. Various docking methods accounting for protein flexibility have been proposed, tackling problems of ever-increasing dimensionality.

Areas Covered: This paper presents an overview of conformational sampling methods treating target flexibility during molecular docking. Special attention is given to approaches considering full protein flexibility. Contrary to what is frequently done, this review does not rely on classical biomolecular recognition models to classify existing docking methods. Instead, it applies algorithmic considerations, focusing on the level of flexibility accounted for. This review also discusses the diversity of docking applications, from virtual screening of small drug-like compounds to geometry prediction of protein-peptide complexes.

Expert Opinion: Considering the diversity of docking methods presented here, deciding which one is the best at treating protein flexibility depends on the system under study and the research application. In virtual screening experiments, ensemble docking can be used to implicitly account for large-scale conformational changes, and selective docking can additionally consider local binding-site rearrangements. In other cases, on-the-fly exploration of the whole protein-ligand complex might be needed for accurate geometry prediction of the binding mode. Among other things, future methods are expected to provide alternative binding modes, which will better reflect the dynamic nature of protein-ligand interactions.

Keywords: molecular docking, conformational sampling, protein flexibility, geometry prediction, virtual screening

ARTICLE HIGHLIGHTS

- Flexibility is an intrinsic feature of proteins that plays an essential role in protein-ligand binding.

This work has been supported in part by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil), NSF CCF 1423304, NSF ABI 0960612 and NSF ABI 1262491.

All the authors are with the Department of Computer Science, Rice University, Houston, TX, USA.

D. Antunes and D. Devaurs have equally contributed to this work.

Corresponding author: Lydia E Kaviraki (kavraki@rice.edu).

- There is an extreme diversity of molecular docking methods accounting for target flexibility, which reflects the broad range of their applications.
- The dichotomy between induced fit and conformational selection is not always adapted to understand and classify these methods.
- Virtual screening (VS) and geometry prediction (GP) are two different applications that require considering different levels of protein flexibility.
- VS can benefit from combining ensemble docking and selective docking.
- GP is pushing the boundaries of molecular docking, focusing on complex systems and realistic binding predictions.

I. INTRODUCTION

Proteins are the main effectors of genomic information. They are involved in a wide range of essential tasks and mediate both physiological and pathological processes. In many cases, a protein's function is influenced by interactions with other molecules. In this context, the molecular understanding of proteins and their interactions with specific partners can pave the way for rational drug design, allowing interfering with specific protein targets. Over the past decades, the fast-growing rate of experimentally-determined protein structures and the impressive advances in computational resources have fueled the development of computer-aided strategies for drug development [1]. Molecular docking, which aims at predicting and analyzing molecular complexes (see Fig. 1 for an example), is one of the most popular and diversified of these approaches [2]–[4].

Despite having different goals and requirements, all docking applications build on two basic components: *sampling* and *scoring*. Sampling consists of exploring (some of) the structural degrees of freedom (DoFs) of a protein-ligand complex and predicting its *binding mode* [5]. Scoring consists of estimating the strength of the binding interaction (*binding energy*) for a given binding mode [6]. Sampling and scoring engender different challenges, but have a combined influence on docking accuracy and performance [7], [8]. Issues related to scoring have been examined in numerous reviews [4]–[6], [9], [10], and will not be discussed here. Instead, the present paper focuses on sampling methods.

With the exception of early rigid-docking methods, ligand flexibility has always been an essential part of molecular docking [2]. Even small ligands with only few rotatable bonds can adopt alternative low-energy conformations. Various conformational sampling techniques have been proposed to

explore ligand flexibility. The most common ones involve (i) a systematic search (e.g., exhaustive search, ensemble-based search, fragment-based search) or (ii) a stochastic search (e.g., Monte Carlo, evolutionary algorithm, swarm optimization, tabu search) [2], [5], [10]. Large ligands and peptides represent bigger challenges for docking, which might exceed the capabilities of many current approaches [11]–[15]. The treatment of ligand flexibility has been extensively reviewed elsewhere [2], [5], [10] and will not be specifically discussed here. Instead, this review focuses on strategies used to account for protein flexibility (in addition to ligand flexibility).

Proteins are flexible and dynamic entities. There exist numerous examples showing differences between a protein's unbound (*apo*) and bound (*holo*) structures, ranging from local side-chain rearrangements [9], [16], [17] to large-scale conformational changes, such as domain motions, hinge-bending motions, or (un)folding of entire segments [9], [18]–[21]. Considering target flexibility has a direct impact on docking experiments, potentially leading to different results, and consequently having major implications for drug discovery [22], [23]. Therefore, protein flexibility has become a major goal in molecular docking, as acknowledged by several recent reviews [5], [24], [25]. Because of the diversity of docking methods accounting for protein flexibility, general classifications are useful to provide overviews of shared concepts and applications. In this context, analogies with the major theories underlying biomolecular recognition (i.e., *induced fit* [26] and *conformational selection* [27]) are sometimes used to classify these docking methods [9], [25]. Although convenient, this classification can become arbitrary, as several methods are clearly blurring the boundary between both categories [25].

This paper presents an overview of sampling methods considering protein flexibility, in light of an alternative (more algorithmic) classification strategy: the level of flexibility accounted for (see Section II). Differences between four major categories are highlighted, and special attention is given to methods considering full receptor flexibility. This review also discusses how the application context influences the choice of a docking method dealing with protein flexibility (see Section III): e.g., virtual screening vs. accurate prediction of binding geometry, small ligand vs. large peptide. Note, however, that protein-protein docking is out of the scope of this review.

II. SAMPLING METHODS ACCOUNTING FOR PROTEIN FLEXIBILITY

For decades, the concepts of *induced fit* and *conformational selection* (also known as population selection, population shift, or selected fit) have been competing to model the mechanisms explaining biomolecular recognition [29]. In the context of protein-ligand binding, both models account for protein flexibility, but in different ways. According to the induced-fit model, a protein may experience conformational changes upon ligand binding [26]. These conformational rearrangements result in a protein-ligand complex with tighter binding. This model implies that no mutual fit is required between the ligand and the protein prior to binding, as it is effectively

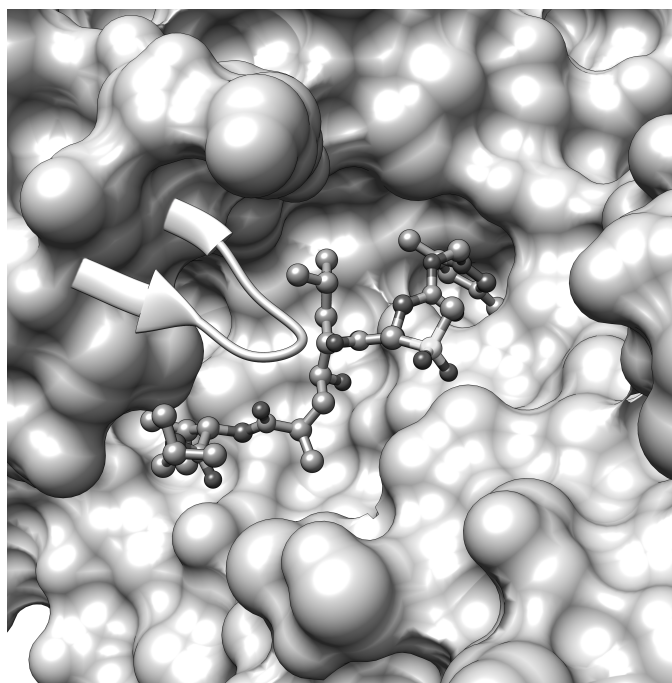


Fig. 1. Example of a protein-ligand complex (PDB access code 2G94). The binding site of the human beta-secretase protein is depicted using a “surface” representation, while a selective inhibitor (ligand) is depicted using a “ball-and-sticks” representation. Protein residues Y68 to K75 are depicted using a “ribbon” representation because they encompass a flexible loop that visually occludes the ligand. Picture produced with the UCSF Chimera package [28].

induced by the binding event. The conformational-selection model relies on the fact that, in solution, proteins exist as an ensemble of conformations rather than as a unique native state. According to this model, the ligand selectively stabilizes specific unbound-receptor conformations, causing a shift in the population toward a particular bound-receptor conformation [27]. In other words, changes in the population are dictated by the laws of statistical mechanics, and the underlying energy landscape is reshaped upon ligand binding [25], [30].

These two biomolecular recognition models have often been used to provide protein-ligand docking methods with some theoretical justification, and to classify these methods [9], [25]. In fact, direct analogies can be drawn between these theoretical models and some of the docking approaches accounting for protein flexibility. On one hand, *selective docking* methods (see Section II-B), which account for local flexibility, have usually been referring to the induced-fit model. On the other hand, earlier *ensemble docking* methods (see Section II-C), which only relied on an ensemble of rigid protein conformations, could clearly be related to the conformational-selection model. However, because of the considerable diversity of docking methods accounting for protein flexibility, any classification based on a dichotomy between these two theoretical models is bound to fail. Indeed, several methods clearly blur the boundary between these models; that is the case for some recent *ensemble docking* methods (see Section II-C) and several *on-the-fly docking* methods (see Section II-D).

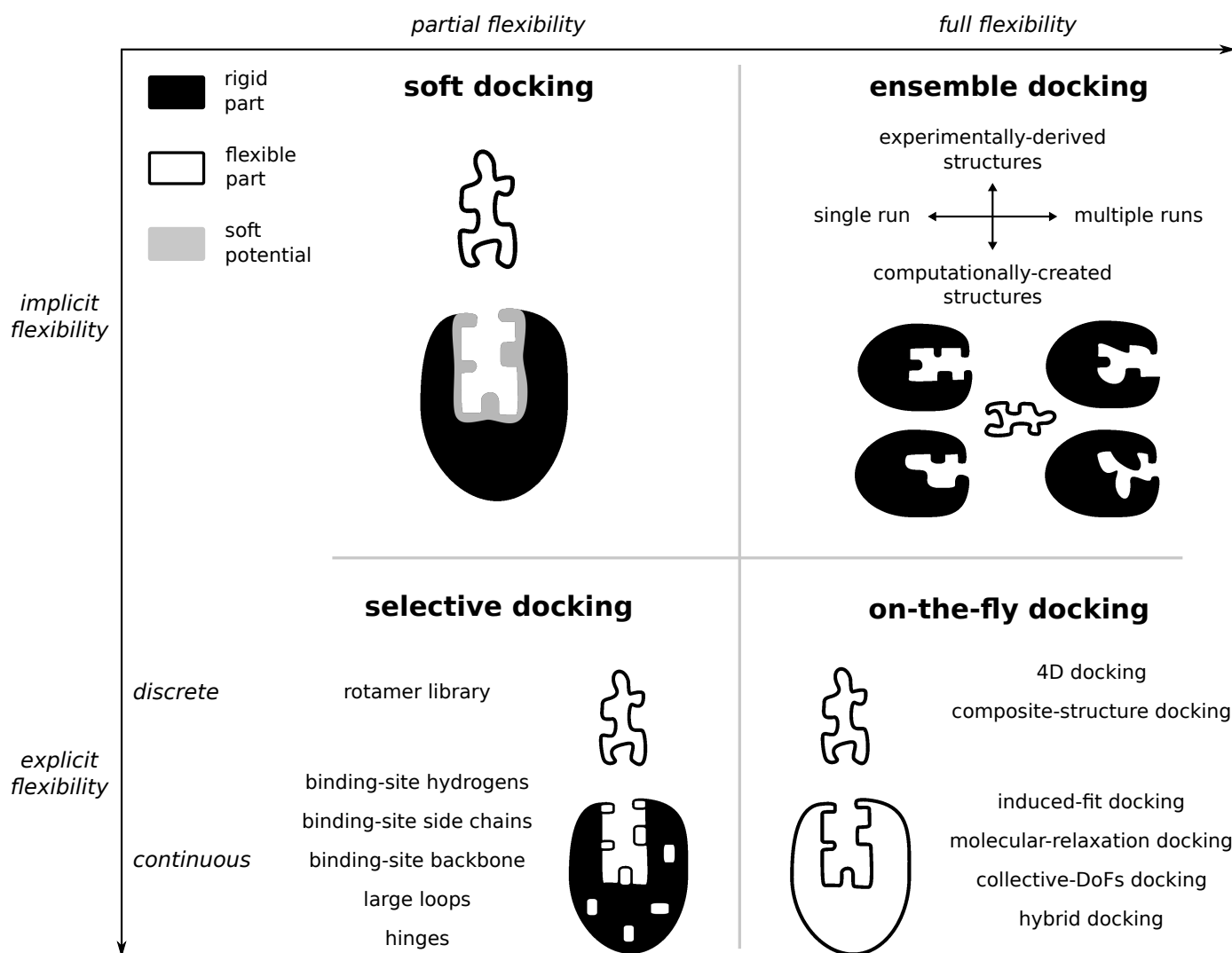


Fig. 2. Classification of molecular docking methods accounting for receptor flexibility. The schematic representations of proteins (large objects) and ligands (small objects) illustrate the level of flexibility considered in each docking category.

This paper introduces a different classification of the docking methods considering protein flexibility, which relies only on the level of flexibility accounted for. Focusing on the algorithmic side, this classification is decomposed along two orthogonal dimensions that relate to two kinds of evolution of docking methods (see Fig. 2). In practice, the constraints imposed by computational resources have always been conflicting with the dimensionality of the problems that practitioners wanted to address. As considering a fully-flexible protein was computationally prohibitive, earlier methods treated only partial target flexibility, focusing on the binding site. Since then, we have witnessed a transition from *partial* flexibility to *full* flexibility, which corresponds to the first dimension of the classification. The second dimension relates to the transition from *implicit* flexibility to *explicit* flexibility. Performing conformational sampling by explicitly exploring the protein's

DoFs during the docking procedure is computationally expensive; therefore, many methods involve an implicit treatment of protein flexibility. These two dimensions lead to a classification encompassing the following categories (see Fig. 2): soft docking (partial, implicit flexibility), selective docking (partial, explicit flexibility), ensemble docking (full, implicit flexibility), and on-the-fly docking (full, explicit flexibility).

A. Soft Docking

The simplest method accounting for protein flexibility is *soft docking* [31]. The idea is to allow for small overlaps between the ligand and the receptor by “softening” the van der Waals potentials. The Lennard-Jones potential, typically used to compute van der Waals interactions, increases rapidly to infinity at short inter-atomic distances, thus producing large steric penalties. Minor steric clashes between the ligand and

the protein can be tolerated by using a more permissive repulsive term in the Lennard-Jones potential. This adjustment of the potential energy function simulates a slightly larger binding site and mimics some level of conformational plasticity. Therefore, soft docking is an *implicit* way to consider *partial* protein flexibility (see Fig. 2).

Historically, the advantage of soft docking was that it introduced no additional computational cost compared to rigid-protein docking. It could accommodate small local conformational changes in the protein (on the order of 1 Å), which could have a significant influence on docking results [32]. Some variations of this technique, referred to as “vdW-offset grid”, have been used more recently [33]. Unfortunately, the scope of soft docking is limited to small-scale rearrangements associated to side-chain plasticity. It does not allow for larger conformational changes, such as large side-chain rotations or backbone motions [34], [35]. Another drawback of soft docking is its tendency to increase the rate of false positives [25], [36]. Nowadays, soft docking is used only rarely on its own, but sometimes as a first docking step within more complex methods (see Section II-D) [37], [38].

B. Selective Docking

To address larger conformational changes, another simple method considering only *partial* flexibility of the protein is *selective docking* (see Fig. 2). It consists of selecting a few “critical” DoFs of the protein, additionally to those of the ligand, and *explicitly* exploring their variability. Therefore, selective docking provides a clear analogy with the induced-fit model: conformational rearrangements are allowed in the protein to better accommodate the binding. Selective docking can be useful when a good structural knowledge of the receptor is available, as well as some expert knowledge on its function. For example, if structural changes between the unbound and bound states are expected to be limited and to involve only active-site residues, this approach can ensure an efficient treatment of side-chain flexibility [39], [40].

Performing selective docking increases the dimensionality of the docking problem, as compared to dealing with a rigid protein, and therefore its computational cost. To mitigate this computational overhead, earlier methods were varying only the essential torsional DoFs of a few side chains in the active site of the protein, while keeping bond lengths and bond angles fixed. Additionally, they were following a *discrete* approach: side-chain rotatable bonds were explored using a rotamer library [41]. Such libraries typically contain experimentally observed and preferred rotamers (i.e., low-energy conformations) of amino acid side-chains. Thanks to its simplicity and low computational cost, this discrete approach is still in use today [42]–[44]. However, its drawback is that sampling may be restricted or biased by the content of the rotamer library. Therefore, it has been extended into a *continuous* approach where the rotatable bonds of the selected side chains are fully explored [45], [46]. Another early continuous paradigm was based on explicitly treating the reorientation of hydrogen atoms only, to optimize hydrogen bonds between the protein and the ligand [47], [48].

Numerous molecular docking programs have incorporated the continuous selective docking approach, allowing users to define a set of flexible side-chains in the active site of the protein. In some methods, the ligand’s pose is optimized before varying the protein side-chains [49], while in more recent methods, both the ligand and side-chain conformations are explored simultaneously [50], [51]. The latter methods have demonstrated significant improvement over rigid-protein docking [52], [53]. Unfortunately, limiting protein flexibility to side chains can sometimes produce worse results than keeping the protein rigid [54]. That is why some methods also account for the flexibility of the backbone around the active site [55], in hinges [18], or in large loop regions of the receptor [56]. However, even these approaches allow for only limited-scale conformational changes.

C. Ensemble Docking

Simulating large-scale conformational changes in the receptor requires considering *full* flexibility of the protein. This can be achieved by adopting the *ensemble docking* approach, which attempts to dock the ligand to an ensemble of receptor conformations instead of a single one [57], [58]. A direct analogy with the conformational-selection model arises from the fact that receptor conformations are generated before docking, usually considering a fully-flexible protein. However, from a docking perspective, protein flexibility is accounted for only *implicitly* because these conformations are not allowed to vary during the docking procedure. Ensemble docking quickly became popular because it allowed considering protein flexibility without introducing any significant changes into existing molecular docking programs.

Ensemble docking methods differ with respect to how the ensemble is used (see Fig. 2): the receptor conformations can be treated independently or simultaneously, which results in two distinct paradigms [59], [60]. The “multiple-run” paradigm consists of performing a series of independent runs for each conformation, using traditional rigid-protein docking programs [57], [61], [62]. The “single-run” paradigm consists of performing a single docking run by combining all the conformations into a single representation, such as a grid-based average of the ensemble [58], [63] or a dynamic pharmacophore model [64], [65]. In all cases, receptor conformations can be provided by experimental techniques, such as X-ray crystallography [66]–[68] or NMR spectroscopy [69], [70]. Alternatively, they can be produced by computational techniques, such as homology modeling [71], molecular dynamics (MD) [62], [72], [73], Monte Carlo simulations [74], [75], or normal mode analysis (NMA) [61], [76], the latter technique sometimes involving elastic network models [74], [77], [78].

Numerous studies have shown that ensemble docking provides great improvement compared with rigid-protein docking [68], [79]–[81]. The dominant view is that docking to an ensemble of protein conformations is superior to docking to a single one [35], but things are not so clear. When experimentally-derived structures are available, using an ensemble of conformations performs better than using a single conformer in many cases, but not all the time [67], [82], [83].

It appears that the composition of the ensemble has a great impact on docking results. More precisely, an anti-cooperative behavior can be observed, where adding some conformations to the ensemble actually deteriorates the docking outcome [82]. In fact, the noise introduced by each extra conformation may obscure the additional information it provides, typically by generating false positives [66]. This problem highlights the need for clear guidelines to select the experimental structures that should compose the ensemble [24]. To this end, various selection strategies have been suggested [68], [82], [84]–[86], but no consensus has been reached.

When using computational simulation techniques, it becomes even more difficult to select representative conformations and build a good ensemble. Indeed, in this case too, using an ensemble can sometimes provide worse results than using a single structure [73]. A strategy has been suggested to generate an ensemble using NMA [76]; however, applying NMA has raised issues such as how many modes have to be used [61]. Several selection techniques have been proposed to extract promising protein conformations from MD simulations [87]–[89]. Unfortunately, MD is strongly limited by the available computational power. Additionally, MD is known for often being unable to cross high-energy barriers, and exploring only local minima of the rugged energy landscape [36], [81]. Therefore, performing a reasonable exploration of the protein’s conformational space can be computationally extremely high. A solution is to introduce artificial biases into MD simulations by applying enhanced-sampling techniques (e.g., temperature-accelerated replica exchange, umbrella sampling, metadynamics) [90] or using accelerated MD [36], [81]. Other methods have been suggested, combining MD with essential dynamics [91] or with molecular mechanics and quantum mechanics [92].

To sum up, despite its success and popularity, ensemble docking suffers from several drawbacks. First, flexibility is only implicit: it is not possible to generate new protein conformations during docking. Second, the docking outcome is highly biased by the composition of the ensemble; clear selection strategies are still needed to alleviate this problem. Finally, some studies suggest that ensemble docking might not be the best way to consider full protein flexibility. Indeed, better docking results can sometimes be obtained by restricting protein flexibility to only relevant side chains and loops [93].

An important remark is that various ensemble docking methods are blurring the boundary between the induced-fit and conformational-selection models. This is the case for several approaches where the ensemble of protein conformations is generated using a knowledge-based strategy accounting for receptor flexibility. For example, ensembles derived from NMA or Monte Carlo can be filtered using ligand information to select representative protein conformations [75]. Using a dynamic virtual ligand (represented by a collection of functional groups), an MD simulation of the binding event can generate promising protein conformations [94], [95]. Another method generates an ensemble of protein conformations by docking flexible ligands to a flexible receptor, using known active compounds [96]. Finally, several MD simulations can be conducted on protein-ligand complexes involving known active

compounds; the numerous protein conformations produced this way can then be filtered to obtain a small working ensemble [97].

D. On-the-fly Docking

Contrary to ensemble docking, the *on-the-fly docking* approach treats protein flexibility *explicitly* by exploring the protein’s DoFs during the docking procedure. In other words, these approaches generate new protein conformations “on the fly” during docking by altering the protein’s structure. This is achieved using various conformational sampling techniques and/or optimization algorithms. Because on-the-fly docking can account for *full* protein flexibility, suitable strategies or heuristics are often applied to keep the problem computationally manageable given its extremely high dimensionality. Moreover, complexity increases even further when dealing with ligand flexibility and protein flexibility simultaneously, as some methods do.

Some on-the-fly docking methods follow a *discrete* approach, while more sophisticated ones follow a *continuous* approach (see Fig. 2). Many on-the-fly docking methods can be related to the induced-fit model of biomolecular recognition. This is the case for the appropriately-named *induced-fit docking* approach (see Section II-D1), but also for *molecular-relaxation docking* (see Section II-D2) and *collective-degrees-of-freedom docking* (see Section II-D3). However, many other methods, such as *four-dimensional docking* (see Section II-D4), *composite-structure docking* (see Section II-D5), or *hybrid docking* (see Section II-D6) are blurring the boundary between induced fit and conformational selection.

1) *Induced-Fit Docking*: Building heavily on the induced-fit concept, the idea behind the *induced-fit docking* approach is to optimize the conformation of the receptor after having docked the ligand into it [37]. More precisely, the ligand is first docked to a rigid protein using soft docking; then, conformations of the protein side-chains are explored using a rotamer library; finally, a short minimization of the protein-ligand complex is performed. Therefore, this method can account for significant side-chain conformational changes, but only small backbone relaxations, in the receptor [37]. Another limitation is that the ligand’s and protein’s DoFs are not sampled simultaneously, which is reasonable only when induced-fit effects are expected to be relatively limited [37]. In spite of these drawbacks, this approach has been successfully used in several studies [98], [99].

As an example of a molecular docking program underpinned by the induced-fit concept, RosettaLigand is a perfect illustration of the increasing level of flexibility that has been accounted for over the years. In its initial implementation, based on Rosetta [100], RosettaLigand only allowed for side-chain flexibility in the protein’s active site [101]. Later, it was augmented with the incorporation of full side-chain and backbone flexibility, as well as with the simultaneous treatment of the ligand and the protein [55]. The latest release includes additional features tackling problems of even higher dimensionality, such as the simultaneous docking of multiple flexible ligands, accounting for protein flexibility [102].

2) *Molecular-Relaxation Docking*: Exploring the flexibility of the ligand and the protein simultaneously represents a significant increase in complexity, but allows considering larger-scale induced-fit effects. This idea has been leveraged in various ways by the *molecular-relaxation docking* approach. Its most physicochemically-accurate implementation consists of performing the docking procedure as an optimization of the conformation of the protein-ligand complex using energy minimization [103], Monte Carlo methods [104], or MD [105]. As these optimization algorithms are computationally expensive, they have often been confined to a post-processing refinement step following rigid-protein docking [106] or selective docking [107]. To refine the binding pose, some studies have involved energy minimization [107], but most of them rely on MD simulations [19], [106], [108]–[112].

The reason hindering the use of MD as a molecular docking tool *per se* is the limited extent of simulations allowed by today’s computational power. As already mentioned in Section II-C, performing an acceptable exploration of the conformational space of a protein-ligand complex would be computationally prohibitive. In most cases, simulating a complete docking event with MD is currently out of reach, except if some approximations are used. Some adjustments to MD simulations have been suggested to make them more computationally efficient, using different temperatures [113], [114], implicit solvent models [115], geometric constraints [116], or a combination with molecular mechanics and quantum mechanics [92]. Even in methods where MD is only used as a refinement tool, it has been sometimes enhanced (e.g., using a torsion-angle approach [93], an optimized potential [117], or steered MD [118]) to improve efficiency.

3) *Collective-Degrees-of-Freedom Docking*: Another kind of approximation, which allows simulating large-scale conformational changes in the protein-ligand complex, relies on using a lower-dimensional representation of this complex that can capture the dominant modes of the protein’s motion. This reduction of the high-dimensionality of the original problem is the concept underlying the *collective-degrees-of-freedom (collective-DoFs) docking* approach. This approximation allows accounting for realistic protein plasticity in a computationally-tractable way [119]. Several dimensionality-reduction techniques have been suggested to define heuristics guiding conformational sampling in molecular docking. The most popular one is NMA [104], [120]–[122], which sometimes involves applying elastic networks [20], [38], but principal component analysis (PCA) [123] and MD [124] have also been used. Note that collective-DoFs docking can be restricted to backbone motions and combined with selective docking, that can better describe local, typically anharmonic, side-chain motions [125]. The main drawback of collective-DoFs docking is that using collective motion modes instead of native DoFs can potentially lead to inaccuracies [119].

4) *Four-dimensional Docking*: The simplest way to combine induced fit and conformational selection is to perform *four-dimensional (4D) docking* [126], [127]. This approach involves an ensemble of protein conformations that are not used sequentially, but simultaneously. The term “4D docking” does not refer to the dimensionality of the docking problem itself (which

depends on the level of flexibility considered for the ligand, in this case). Rather, it refers to the fact that, additionally to the three dimensions of space in which the ligand resides, the 4D docking approach involves an extra dimension through a *discrete* variable corresponding to the index of the protein conformation in the ensemble. In fact, this variable is explored during the search in the same way as the ligand’s DoFs are. The drawbacks of this method are the same as for ensemble docking: sampling is strongly dependent on the composition of the ensemble, and no new conformation of the receptor can be created.

5) *Composite-Structure Docking*: A similar approach, unifying induced fit and conformational selection, but providing an additional level of complexity is *composite-structure docking* [128], [129]. Composite structures are conformations of the protein constructed using parts of the conformations present in the initial ensemble. Each part can be as small as a single side chain or as big as the whole backbone [129]. Additionally, some methods restrict these parts within the regions of the protein showing high variability (because of high disorder or point mutations) in the initial ensemble [128]. The interesting aspect of this approach, as opposed to 4D docking, is that new conformations are created during the search, thus increasing structural variation. Several studies have reported good results with this approach [130], [131]. However, it remains a *discrete* method, strongly influenced by the content of the ensemble.

6) *Hybrid Docking Approaches*: Several docking methods are based on hybrid approaches combining docking paradigms, and therefore blurring the boundary between induced fit and conformational selection. For example, practitioners often implicitly use such hybrid approaches when they perform ensemble docking using popular software allowing for the selection of flexible side-chains in the protein (i.e., for selective docking) [50], [51]. Some methods explicitly perform a similar connection, mixing ensemble docking and induced-fit docking [67], [132]. Other examples are approaches that combine ensemble docking and molecular-relaxation docking [118], [133]. In this case, each binding pose of the protein-ligand complex is refined using an optimization algorithm.

E. Conclusion

Proteins are dynamic entities that are believed to exist as an ensemble of conformational states in solution. Interactions with a ligand may impact these pre-existing states, by inducing conformational changes in proteins or stabilizing a particular conformer [25], [29]. Although understanding and modeling the mechanisms of biomolecular recognition is still a topic of research, receptor flexibility is recognized as essential for protein-ligand interactions and, therefore, drug discovery.

In the context of molecular docking, a great diversity of sampling methods have been proposed to account for protein flexibility. In an attempt to better comprehend such diversity, these methods have often been classified using analogies to the main models of biomolecular recognition: induced fit and conformational selection [9], [25]. However, as highlighted in this review, many sampling methods are clearly blurring the boundary between these two theories. This goes along

the lines of recent advances in the theory of biomolecular recognition. Indeed, after recognizing that classical models could not describe all binding mechanisms [25], [134], theoreticians are now trying to reconcile both models of binding into a unified one [135]–[137], in accordance with recent experimental evidence [138], [139]. Such work could lead to a new understanding of protein-ligand binding.

III. EXPERT OPINION

Accounting for protein flexibility in molecular docking is a challenging task, especially because it should come additionally to the treatment of ligand flexibility. This increase in the dimensionality of the docking problem literally translates into an explosion of computational costs, which has always constrained the development of docking software. Consequently, a huge variety of molecular docking methods have been proposed over the years, continuously increasing the level of protein flexibility considered during conformational sampling. This diversity of methods also reflects the broad range of their proposed applications. For instance, virtual screening (VS) and geometry prediction (GP) are very different goals in molecular docking, with totally opposed computational constraints; this requires choosing adapted strategies when tackling protein flexibility. While VS focuses on identifying a short list of likely binders out of a large dataset of ligands (*classification problem*), GP focuses on finding the best possible geometry for a given protein-ligand complex (*optimization problem*).

VS is arguably the most common application of molecular docking. Since the docking procedure is repeated multiple times when screening a large library of ligands, VS imposes a trade-off between the speed of the docking method and its accuracy [6]. Consequently, exploring large-scale structural changes and folding events during conformational sampling can represent a detrimental computational burden. Instead, building a compact ensemble of representative protein conformations, as in ensemble docking, can be a useful pre-processing step. As this is done only once, this can involve more sophisticated and computationally-expensive approaches to sample large-scale conformational changes [36], [140]–[144]. Additionally, selective docking can be used for local refinement of the binding mode for each ligand in the dataset. Therefore, combining these two docking approaches represents an efficient way to consider protein flexibility in VS experiments.

Top-ranking binders identified by VS can be further analyzed with other tools or become the input of GP experiments, to optimize the corresponding binding modes. In this context, the combination of ensemble and selective docking can show limitations. First, no matter how good the conformational ensemble is, it is only a discrete collection of states. A local exploration of the binding site might not be enough to correct these initial structures and discover an accurate bioactive pose. Second, combining global protein conformational sampling with local binding site refinement in a *sequential* way is not always likely to predict major ligand-induced rearrangements or allosteric effects. GP requires greater confidence in the binding mode description, which can benefit from a *simultaneous* exploration of protein and ligand flexibility. Being less

affected by computational constraints, GP can fully leverage the capabilities of on-the-fly docking. Therefore, GP will continue to push the boundaries of molecular docking, by tackling more complex systems and yielding more realistic predictions of protein-ligand interactions.

Despite its impressive advances, on-the-fly docking is still facing many challenges, such as dealing with large-scale conformational changes in the receptor [9], [19]–[21] or treating multiple flexible ligands while considering protein flexibility [102]. In turn, this directly relates to the issue of cryptic allosteric sites—i.e., transient pockets that are usually not observable from the crystal structure but that can influence protein function by allosteric communication with the active site. The so-called “drugability” of these sites is an active field of research [145], [146]. In some cases, the binding of one drug at a cryptic allosteric site might trigger conformational changes in the protein, allowing a second drug to bind in another active site. Such mechanisms can be accurately modeled only by docking methods that simultaneously account for multiple ligands and large-scale conformational changes in the protein.

An additional challenge for on-the-fly docking relates to predicting binding modes of peptides [147]. As opposed to small drug-like ligands, large peptides have a complex dynamic behavior: they can establish intricate networks of hydrogen bonds and even fold into secondary structures [147]. Such folding properties are difficult to simulate using standard docking protocols. More generally, the challenge of correctly predicting the structure of a protein-peptide complex is becoming more and more related to *ab initio* protein modeling and protein-protein docking [9]. Accordingly, the biggest claims regarding large target flexibility in docking are made using methods based on protein prediction platforms, such as Rosetta [100] or CABS [148]. Both platforms have been used for on-the-fly docking, considering peptides and large-scale conformational changes in the receptor [13], [21], [147]. Even though they involve coarse-grained representations of the protein-ligand complex, these approaches represent very innovative trends for on-the-fly docking.

As a research field, molecular docking has experienced a rapid evolution and has grown to include very diverse applications. Unfortunately, no single method can address the variety of docking tasks: combining different tools and using target-specific parameters is often required to achieve best results. Future docking methods are expected to combine features of different approaches, accounting simultaneously for ligand flexibility, binding site rearrangement, large-scale ligand-induced motions in the receptor, as well as the overall dynamics of the interaction. In turn, this more realistic exploration of the binding event will also reflect the flexibility of the protein-ligand complex, probably predicting not only one “best” structure, but an ensemble of top-ranked alternative binding modes. Considering alternative binding modes will also have an impact on scoring methods, with respect to binding aspects that are often neglected, such as entropy [3], [5]. Indeed, more accurate binding energy estimations are a pressing issue for several docking applications [9], [149]. Finally, considering the diversity of methods and the difficulty to assess their performance, benchmarking efforts using stan-

standard datasets are needed to highlight the advantages of each strategy [3], [150]. Future developments in the field will greatly benefit from the popularization of such practices.

ACKNOWLEDGEMENT

Molecular graphics have been produced with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

REFERENCES

- [1] Villoutreix BO, Lagorce D, Labbé CM, et al. One hundred thousand mouse clicks down the road: Selected online resources supporting drug discovery collected over a decade. *Drug Discov Today*. 2013;18:1081–9.
- [2] Dhanik A, Kavraki LE. Protein-ligand interactions: Computational docking. In: eLS. John Wiley & Sons, Ltd; 2012. .
- [3] Sousa SF, Ribeiro AJ, Coimbra JT, et al. Protein-ligand docking in the new millennium - A retrospective of 10 years in the field. *Curr Med Chem*. 2013;20:2296–314.
- [4] Yuriev E, Holien J, Ramsland PA. Improvements, trends, and new ideas in molecular docking: 2012-2013 in review. *J Mol Recognit*. 2015; Published online 21 March 2015, doi:10.1002/jmr.2471.
- [5] Guedes IA, de Magalhães CS, Dardenne LE. Receptor-ligand molecular docking. *Biophys Rev*. 2014;6:75–87.
- [6] Grinter SZ, Zou X. Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. *Molecules*. 2014;19:10150–76.
- [7] Weill N, Therrien E, Campagna-Slater V, Moitessier N. Methods for docking small molecules to macromolecules: A user's perspective. 1. The theory. *Curr Pharm Des*. 2014;20:3338–59.
- [8] Campagna-Slater V, Therrien E, Weill N, Moitessier N. Methods for docking small molecules to macromolecules: A user's perspective. 2. Applications. *Curr Pharm Des*. 2014;20:3360–72.
- [9] Lill MA. Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. *Biochemistry*. 2011;50:6157–69.
- [10] Bello M, Martínez-Archundia M, Correa-Basurto J. Automated docking for novel drug discovery. *Exp Op Drug Discov*. 2013;8:821–34.
- [11] Audie J, Swanson J. Recent work in the development and application of protein-peptide docking. *Future Med Chem*. 2012;4:1619–44.
- [12] Dhanik A, McMurray JS, Kavraki LE. DINC: A new AutoDock-based protocol for docking large ligands. *BMC Struct Biol*. 2013;13:S11.
- [13] Li H, Lu L, Chen R, et al. PaFlexPepDock: Parallel *ab-initio* docking of peptides onto their receptors with full flexibility based on Rosetta. *PLoS ONE*. 2014;9:e94769.
- [14] Rentzsch R, Renard BY. Docking small peptides remains a great challenge: An assessment using AutoDock Vina. *Brief Bioinform*. 2015; Published online 21 April 21 2015, doi:10.1093/bib/bbv008.
- [15] Lee H, Heo L, Lee MS, Seok C. GalaxyPepDock: A protein-peptide docking tool based on interaction similarity and energy optimization. *Nucl Acids Res*. 2015;43:W431–5.
- [16] Najmanovich R, Kuttner J, Sobolev V, Edelman M. Side-chain flexibility in proteins upon ligand binding. *Proteins*. 2000;39:261–8.
- [17] Wei BQ, Weaver LH, Ferrari AM, et al. Testing a flexible-receptor docking algorithm in a model binding site. *J Mol Biol*. 2004;337:1161–82.
- [18] Sandak B, Wolfson HJ, Nussinov R. Flexible docking allowing induced fit in proteins: Insights from an open to closed conformational isomers. *Proteins*. 1998;32:159–74.
- [19] Seeliger D, de Groot BL. Conformational transitions upon ligand binding: Holo-structure prediction from apo conformations. *PLoS Comput Biol*. 2010;6:e1000634.
- [20] Leis S, Zacharias M. Efficient inclusion of receptor flexibility in grid-based protein-ligand docking. *J Comput Chem*. 2011;32:3433–9.
- [21] Wabik J, Kurcinski M, Kolinski A. Coarse-grained modeling of peptide docking associated with large conformation transitions of the binding protein: Troponin I fragment-Troponin C system. *Molecules*. 2015;20:10763–80.
- [22] Kohk DB, Wade RC, Wenzel W. Receptor flexibility in small-molecule docking calculations. *Wiley Interdiscip Rev Comput Mol Sci*. 2011;1:298–314.
- [23] Lexa KW, Carlson HA. Protein flexibility in docking and surface mapping. *Q Rev Biophys*. 2012;45:301–43.
- [24] Yuriev E, Ramsland PA. Latest developments in molecular docking: 2010-2011 in review. *J Mol Recognit*. 2013;26:215–39.
- [25] Buonfiglio R, Recanatini M, Masetti M. Protein flexibility in drug discovery: From theory to computation. *ChemMedChem*. 2015;10:1141–8.
- [26] Koshland Jr DE. Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci USA*. 1958;44:98–104.
- [27] Tsai CJ, Ma B, Kumar S, et al. Protein folding: Binding of conformationally fluctuating building block via population selection. *Crit Rev Biochem Mol Biol*. 2001;36:399–433.
- [28] Petersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605–12.
- [29] Changeux JP, Edelstein S. Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol Reports*. 2011;3.
- [30] Kar G, Keskin O, Gursoy A, Nussinov R. Allostery and population shift in drug discovery. *Curr Opin Pharmacol*. 2010;10:715–22.
- [31] Jiang F, Kim SH. “Soft docking”: Matching of molecular surface cubes. *J Mol Biol*. 1991;219:79–102.
- [32] Ferrari AM, Wei BQ, Costantino L, Shoichet BK. Soft docking and multiple receptor conformations in virtual screening. *J Med Chem*. 2004;47:5076–84.
- [33] Mizutani MY, Takamatsu Y, Ichinose T, et al. Effective handling of induced-fit motion in flexible docking. *Proteins*. 2006;63:878–91.
- [34] B-Rao C, Subramanian J, Sharma SD. Managing protein flexibility in docking and its applications. *Drug Discov Today*. 2009;14:394–400.
- [35] Beier C, Zacharias M. Tackling the challenges posed by target flexibility in drug design. *Exp Opin Drug Discov*. 2010;5:347–59.
- [36] Feixas F, Lindert S, Sinko W, McCammon JA. Exploring the role of receptor flexibility in structure-based drug discovery. *Biophys Chem*. 2014;186:31–45.
- [37] Sherman W, Day T, Jacobson MP, et al. Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem*. 2006;49:534–53.
- [38] Venkatraman V, Ritchie DW. Flexible protein docking refinement using pose-dependent normal mode analysis. *Proteins*. 2012;80:2262–74.
- [39] Desmet J, Wilson IA, Joniau M, et al. Computation of the binding of fully flexible peptides to proteins with flexible side chains. *FASEB J*. 1997;11:164–72.
- [40] Schaffer L, Verkhivker GM. Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side-chain optimization. *Proteins*. 1998;33:295–310.
- [41] Leach AR. Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol*. 1994;235:345–56.
- [42] Ding F, Yin S, Dokholyan NV. Rapid flexible docking using a stochastic rotamer library of ligands. *J Chem Inf Model*. 2010;50:1623–32.
- [43] Shin WH, Seok C. GalaxyDock: Protein-ligand docking with flexible protein side-chains. *J Chem Inf Model*. 2012;52:3225–32.
- [44] Schumann M, Armen RS. Systematic and efficient side chain opti-

- mization for molecular docking using a cheapest-path procedure. *J Comput Chem.* 2013;34:1258–69.
- [45] Abagyan R, Totrov M, Kuznetsov D. ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comput Chem.* 1994;15:488–506.
- [46] Schnecke V, Kuhn LA. Virtual screening with solvation and ligand-induced complementarity. *Perspect Drug Discov Des.* 2000;20:171–90.
- [47] Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol.* 1995;245:43–53.
- [48] Smieško M. DOLINA – Docking based on a local induced-fit algorithm: Application toward small-molecule binding to nuclear receptors. *J Chem Inf Model.* 2013;53:1415–23.
- [49] Schnecke V, Swanson CA, Getzoff ED, et al. Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. *Proteins.* 1998;33:74–87.
- [50] Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem.* 2009;30:2785–91.
- [51] Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010;31:455–61.
- [52] Abreu RM, Froufe HJ, Queiroz MJ, Ferreira IC. Selective flexibility of side-chain residues improves VEGFR-2 docking score using AutoDock Vina. *Chem Biol Drug Des.* 2012;79:530–4.
- [53] Morris GM, Green LG, Radić Z, et al. Automated docking with protein flexibility in the design of femtomolar “click chemistry” inhibitors of acetylcholinesterase. *J Chem Inf Model.* 2013;53:898–906.
- [54] Correa-Basurto J, Ramos-Morales FR, Matus MH, et al. Docking and DFT studies to explore the Topoisomerase II ATP pocket employing 3-substituted 2,6-piperazindiones for drug design. *Mol Simul.* 2012;38:1072–84.
- [55] Davis IW, Baker D. RosettaLigand docking with full ligand and receptor flexibility. *J Mol Biol.* 2009;385:381–92.
- [56] Flick J, Tristram F, Wenzel W. Modeling loop backbone flexibility in receptor-ligand docking simulations. *J Comput Chem.* 2012;33:2504–15.
- [57] Pang YP, Kozikowski AP. Prediction of the binding site of 1-benzyl-4-[(5,6-dimethoxy-1-indanon-2-yl)methyl]piperidine in acetylcholinesterase by docking studies with the SYSDOC program. *J Comput Aided Mol Des.* 1994;8:683–93.
- [58] Knegtel RM, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. *J Mol Biol.* 1997;266:424–40.
- [59] Barril X, Fradera X. Incorporating protein flexibility into docking and structure-based drug design. *Expert Opin Drug Discov.* 2006;1:335–49.
- [60] Corbeil CR, Therrien E, Moitessier N. Modeling reality for optimal docking of small molecules to biological targets. *Curr Computer-Aided Drug Des.* 2009;5:241–63.
- [61] Dietzen M, Zotenko E, Hildebrandt A, Lengauer T. On the applicability of elastic network normal modes in small-molecule docking. *J Chem Inf Model.* 2012;52:844–56.
- [62] Martinez-Ramos F, Fonseca-Sabater Y, Soriano-Ursúa MA, et al. o-Alkylselenenylated benzoic acid accesses several sites in serum albumin according to fluorescence studies, Raman spectroscopy and theoretical simulations. *Protein Pept Lett.* 2013;20:705–14.
- [63] Österberg F, Morris GM, Sanner MF, et al. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins.* 2002;46:34–40.
- [64] Carlson HA, Masukawa KM, Rubins K, et al. Developing a dynamic pharmacophore model for HIV-1 integrase. *J Med Chem.* 2000;43:2100–14.
- [65] Joseph-McCarthy D, Thomas IV BE, Belmarsh M, et al. Pharmacophore-based molecular docking to account for ligand flexibility. *Proteins.* 2003;51:172–88.
- [66] Barril X, Morley SD. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J Med Chem.* 2005;48:4432–43.
- [67] Craig IR, Essex JW, Spiegel K. Ensemble docking into multiple crystallographically derived protein structures: An evaluation based on the statistical analysis of enrichments. *J Chem Inf Model.* 2010;50:511–24.
- [68] Brooijmans N, Humblet C. Chemical space sampling by different scoring functions and crystal structures. *J Comput Aided Mol Des.* 2010;24:433–47.
- [69] Huang SY, Zou X. Efficient molecular docking of NMR structures: Application to HIV-1 protease. *Protein Sci.* 2007;16:43–51.
- [70] Damm KL, Carlson HA. Exploring experimental sources of multiple protein conformations in structure-based drug design. *J Am Chem Soc.* 2007;129:8225–35.
- [71] Novoa EM, de Poupiana LR, Barril X, Orozco M. Ensemble docking from homology models. *J Chem Theory Comput.* 2010;6:2547–57.
- [72] Carlson HA, McCammon JA. Accommodating protein flexibility in computational drug design. *Mol Pharmacol.* 2000;57:213–8.
- [73] Nichols SE, Baron R, Ivetac A, McCammon JA. Predictive power of molecular dynamics receptor structures in virtual screening. *J Chem Inf Model.* 2011;51:1439–46.
- [74] Rueda M, Bottegoni G, Abagyan R. Consistent improvement of cross docking results using binding site ensembles generated with Elastic Network Normal Modes. *J Chem Inf Model.* 2009;49:716–25.
- [75] Rueda M, Totrov M, Abagyan R. ALiBERO: Evolving a team of complementary pocket conformations rather than a single leader. *J Chem Inf Model.* 2012;52:2705–14.
- [76] Sperandio O, Mouawad L, Pinto E, et al. How to choose relevant multiple receptor conformations for virtual screening: A test case of Cdk2 and normal mode analysis. *Eur Biophys J.* 2010;39:1365–72.
- [77] Cavasotto CN, Kovacs JA, Abagyan RA. Representing receptor flexibility in ligand docking through relevant normal modes. *J Am Chem Soc.* 2005;127:9632–40.
- [78] Kovacs JA, Cavasotto CN, Abagyan RA. Conformational sampling of protein flexibility in generalized coordinates: Application to ligand docking. *J Comput Theor Nanosci.* 2005;2:354–61.
- [79] Park SJ, Kufareva I, Abagyan R. Improved docking, screening and selectivity prediction for small molecule nuclear receptor modulators using conformational ensembles. *J Comput Aided Mol Des.* 2010;24:459–71.
- [80] Li Y, Kim DJ, Ma W, et al. Discovery of novel checkpoint kinase 1 inhibitors by virtual screening based on multiple crystal structures. *J Chem Inf Model.* 2011;51:2904–14.
- [81] Sinko W, Lindert S, McCammon JA. Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design. *Chem Biol Drug Des.* 2013;81:41–9.
- [82] Rueda M, Bottegoni G, Abagyan R. Recipes for the selection of experimental protein conformations for virtual screening. *J Chem Inf Model.* 2010;50:186–93.
- [83] Korb O, Olsson TS, Bowden SJ, et al. Potential and limitations of ensemble docking. *J Chem Inf Model.* 2012;52:1262–74.
- [84] Bolstad ES, Anderson AC. In pursuit of virtual lead optimization: Pruning ensembles of receptor structures for increased efficiency and accuracy during docking. *Proteins.* 2009;75:62–74.
- [85] Bottegoni G, Rocchia W, Rueda M, et al. Systematic exploitation of multiple receptor conformations for virtual ligand screening. *PLoS ONE.* 2011;6:e18845.
- [86] Korb O, McCabe P, Cole J. The ensemble performance index: An improved measure for assessing ensemble pose prediction performance. *J Chem Inf Model.* 2011;51:2915–19.

- [87] Lin JH, Perryman AL, Schames JR, McCammon JA. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J Am Chem Soc.* 2002;124:5632–3.
- [88] 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.
- [89] Machado KS, Winck AT, Ruiz DD, de Souza ON. Mining flexible-receptor docking experiments to select promising protein receptor snapshots. *BMC Genomics.* 2010;11:S6.
- [90] Rocchia W, Masetti M, Cavalli A. Enhanced sampling methods in drug design. In: Luque J, Barril X, editors. *Physico-Chemical and Computational Approaches to Drug Discovery.* The Royal Society of Chemistry; 2012. p. 273–301.
- [91] 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;8:2204–14.
- [92] Nowosielski M, Hoffmann M, Kuron A, et al. The MM2QM tool for combining docking, molecular dynamics, molecular mechanics, and quantum mechanics. *J Comput Chem.* 2013;34:750–6.
- [93] Armen RS, Chen J, Brooks III CL. An evaluation of explicit receptor flexibility in molecular docking using molecular dynamics and torsion angle molecular dynamics. *J Chem Theory Comput.* 2009;5:2909–23.
- [94] Xu M, Lill MA. Significant enhancement of docking sensitivity using implicit ligand sampling. *J Chem Inf Model.* 2011;51:693–706.
- [95] Xu M, Lill MA. Utilizing experimental data for reducing ensemble size in flexible-protein docking. *J Chem Inf Model.* 2012;52:187–98.
- [96] Cavasotto CN, Abagyan RA. Protein flexibility in ligand docking and virtual screening to protein kinases. *J Mol Biol.* 2004;337:209–25.
- [97] Yoon S, Welsh WJ. Identification of a minimal subset of receptor conformations for improved multiple conformation docking and two-step scoring. *J Chem Inf Comput Sci.* 2004;44:88–96.
- [98] McRobb FM, Capuano B, Crosby IT, et al. Homology modeling and docking evaluation of aminergic G protein-coupled receptors. *J Chem Inf Model.* 2010;50:626–37.
- [99] Vilar S, Ferino G, Phatak SS, et al. Docking-based virtual screening for ligands of G protein-coupled receptors: Not only crystal structures but also in silico models. *J Mol Graph Model.* 2011;29:614–23.
- [100] Das R, Baker D. Macromolecular modeling with Rosetta. *Annu Rev Biochem.* 2008;77:363–82.
- [101] Meiler J, Baker D. ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility. *Proteins.* 2006;65:538–48.
- [102] Lemmon G, Meiler J. Rosetta Ligand docking with flexible XML protocols. In: Baron R, editor. *Computational Drug Discovery and Design.* vol. 819 of *Methods in Molecular Biology.* Springer; 2012. p. 143–55.
- [103] Apostolakis J, Plückthun A, Caflisch A. Docking small ligands in flexible binding sites. *J Comput Chem.* 1998;19:21–37.
- [104] Borrelli KW, Cossins B, Guallar V. Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J Comput Chem.* 2010;31:1224–35.
- [105] Tsfadia Y, Friedman R, Kadmon J, et al. Molecular dynamics simulations of palmitate entry into the hydrophobic pocket of the fatty acid binding protein. *FEBS Lett.* 2007;581:1243–7.
- [106] Sokkar P, Sathis V, Ramachandran M. Computational modeling on the recognition of the HRE motif by HIF-1: Molecular docking and molecular dynamics studies. *J Mol Model.* 2012;18:1691–700.
- [107] Schaffer L, Verkhivker GM. Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side-chain optimization. *Proteins.* 1998;33:295–310.
- [108] Haspel N, Ricklin D, Geisbrecht BV, et al. Electrostatic contributions drive the interaction between *Staphylococcus aureus* protein Efb-C and its complement target C3d. *Protein Sci.* 2008;17:1894–906.
- [109] Haspel N, Geisbrecht BV, Lambris JD, Kavraki LE. Multi-scale characterization of the energy landscape of proteins with application to the C3d/Efb-C complex. *Proteins.* 2009;78:1004–14.
- [110] da Silva ML, Gonçalves Ad, Batista PR, et al. Design, docking studies and molecular dynamics of new potential selective inhibitors of *Plasmodium falciparum* serine hydroxymethyltransferase. *Mol Simul.* 2010;36:5–14.
- [111] Yang Z, Nie Y, Yang G, et al. Synergistic effects in the designs of neuraminidase ligands: Analysis from docking and molecular dynamics studies. *J Theor Biol.* 2010;267:363–74.
- [112] Wang X, Yang W, Xu X, et al. Studies of benzothiadiazine derivatives as hepatitis C virus NS5B polymerase inhibitors using 3D-QSAR, molecular docking and molecular dynamics. *Curr Med Chem.* 2010;17:2788–803.
- [113] Di Nola A, Roccatano D, Berendsen HJ. Molecular dynamics simulation of the docking of substrates to proteins. *Proteins.* 1994;19:174–82.
- [114] Mangoni M, Roccatano D, Di Nola A. Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation. *Proteins.* 1999;35:153–62.
- [115] Luty BA, Wasserman ZR, Stouten PF, et al. A molecular mechanics / grid method for evaluation of ligand-receptor interactions. *J Comput Chem.* 1995;16:454–64.
- [116] Huang Z, Wong CF, Wheeler RA. Flexible protein-flexible ligand docking with disrupted velocity simulated annealing. *Proteins.* 2008;71:440–54.
- [117] Antes I. DynaDock: A new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility. *Proteins.* 2010;78:1084–104.
- [118] Whalen KL, Chang KM, Spies MA. Hybrid steered molecular dynamics-docking: An efficient solution to the problem of ranking inhibitor affinities against a flexible drug target. *Mol Inform.* 2011;30:459–71.
- [119] Teodoro ML, Kavraki LE. Conformational flexibility models for the receptor in structure based drug design. *Curr Pharm Des.* 2003;9:1635–48.
- [120] Zacharias M, Sklenar H. Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: Application to DNA minor groove ligand complex. *J Comput Chem.* 1999;20:287–300.
- [121] Kolossvary I, Guida WC. Low-mode conformational search elucidated: Application to $C_{39}H_{80}$ and flexible docking of 9-deazaguanine inhibitors into PNP. *J Comput Chem.* 1999;20:1671–84.
- [122] Keseru GM, Kolossvary I. Fully flexible low-mode docking: Application to induced fit in HIV integrase. *J Am Chem Soc.* 2001;123:12708–9.
- [123] Teodoro ML, Phillips Jr GN, Kavraki LE. Understanding protein flexibility through dimensionality reduction. *J Comput Biol.* 2003;10:617–34.
- [124] Zacharias M. Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: Binding of FK506 to FKBP. *Proteins.* 2004;54:759–67.
- [125] May A, Zacharias M. Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: Evaluation on kinase inhibitor cross docking. *J Med Chem.* 2008;51:3499–506.
- [126] Huang SY, Zou X. Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins.* 2007;66:399–421.
- [127] 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.
- [128] Claußen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol.* 2001;308:377–95.
- [129] Corbeil CR, Englebienne P, Moitessier N. Docking ligands into

- flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J Chem Inf Model*. 2007;47:435–49.
- [130] Kim J, Park JG, Chong Y. FlexE ensemble docking approach to virtual screening for CDK2 inhibitors. *Mol Simulation*. 2007;33:667–76.
- [131] Corbeil CR, Moitessier N. Docking ligands into flexible and solvated macromolecules. 3. Impact of input ligand conformation, protein flexibility, and water molecules on the accuracy of docking programs. *J Chem Inf Model*. 2009;49:997–1009.
- [132] Takaya D, Yamashita A, Kamijo K, et al. A new method for induced fit docking (GENIUS) and its application to virtual screening of novel HCV NS3-4A protease inhibitors. *Bioorg Med Chem*. 2011;19:6892–905.
- [133] Nabuurs SB, Wagener M, de Vlieg J. A flexible approach to induced fit docking. *J Med Chem*. 2007;50:6507–18.
- [134] Bosshard HR. Molecular recognition by induced fit: How fit is the concept? *News Physiol Sci*. 2001;16:171–3.
- [135] Zhou HX. From induced fit to conformational selection: A continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys J*. 2010;98:L15–7.
- [136] Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: An extended view of binding events. *Trends Biochem Sci*. 2010;35:539–46.
- [137] Kuzu G, Keskin O, GURSOY A, Nussinov R. Expanding the conformational selection paradigm in protein-ligand docking. In: Baron R, editor. *Computational Drug Discovery and Design*. vol. 819 of *Methods in Molecular Biology*. Springer; 2012. p. 59–74.
- [138] Grünberg R, Leckner J, Nilges M. Complementarity of structure ensembles in protein-protein binding. *Structure*. 2004;12:2125–36.
- [139] Wlodarski T, Zagrovic B. Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc Natl Acad Sci USA*. 2009;106:19346–51.
- [140] Shehu A, Kaviraki LE, Clementi C. Multiscale characterization of protein conformational ensembles. *Proteins*. 2009;76:837–51.
- [141] Cortés J, Le D, Iehl R, Siméon T. Simulating ligand-induced conformational changes in proteins using a mechanical disassembly method. *Phys Chem Chem Phys*. 2010;12:8268–76.
- [142] Shehu A, Kaviraki LE. Modeling structures and motions of loops in protein molecules. *Entropy*. 2012;14:252–90.
- [143] Gipson B, Hsu D, Kaviraki LE, Latombe JC. Computational models of protein kinematics and dynamics: Beyond simulation. *Annu Rev Anal Chem*. 2012;5:273–91.
- [144] Gipson B, Moll M, Kaviraki LE. SIMS: A hybrid method for rapid conformational analysis. *PLoS ONE*. 2013;8:e68826.
- [145] Grant BJ, Lukman S, Hocker HJ, et al. Novel allosteric sites on Ras for lead generation. *PLoS ONE*. 2011;6:e25711.
- [146] Bowman GR, Geissler PL. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc Natl Acad Sci USA*. 2012;109:11681–6.
- [147] Raveh B, London N, Zimmerman L, Schueler-Furman O. Rosetta FlexPepDock *ab-initio*: Simultaneous folding, docking and refinement of peptides onto their receptors. *PLoS ONE*. 2011;6:e18934.
- [148] Blaszczyk M, Jamroz M, Kmiecik S, Kolinski A. CABS-fold: Server for the *de novo* and consensus-based prediction of protein structure. *Nucleic Acids Res*. 2013;41:W406–11.
- [149] Chaskar P, Zoete V, Röhrig UF. Toward on-the-fly quantum mechanical/molecular mechanical (QM/MM) docking: Development and benchmark of a scoring function. *J Chem Inf Model*. 2014;54:3137–52.
- [150] Schomburg KT, Rarey M. Benchmark data sets for structure-based computational target prediction. *J Chem Inf Model*. 2014;54:2261–74.