



# Protein structure-based drug design: from docking to molecular dynamics

Paweł Śledź and Amedeo Caflisch



Recent years have witnessed rapid developments of computer-aided drug design methods, which have reached accuracy that allows their routine practical applications in drug discovery campaigns. Protein structure-based methods are useful for the prediction of binding modes of small molecules and their relative affinity. The high-throughput docking of up to  $10^6$  small molecules followed by scoring based on implicit-solvent force field can robustly identify micromolar binders using a rigid protein target. Molecular dynamics with explicit solvent is a low-throughput technique for the characterization of flexible binding sites and accurate evaluation of binding pathways, kinetics, and thermodynamics. In this review we highlight recent advancements in applications of ligand docking tools and molecular dynamics simulations to ligand identification and optimization.

## Address

Department of Biochemistry, University of Zurich, Winterthurerstr. 190, 8057 Zürich, Switzerland

Corresponding authors: Śledź, Paweł ([p.sledz@bioc.uzh.ch](mailto:p.sledz@bioc.uzh.ch)), Caflisch, Amedeo ([caflisch@bioc.uzh.ch](mailto:caflisch@bioc.uzh.ch))

Current Opinion in Structural Biology 2018, 48:93–102

This review comes from a themed issue on **Folding and binding**

Edited by **Amedeo Caflisch** and **Stefano Gianni**

<https://doi.org/10.1016/j.sbi.2017.10.010>

0959-440X/© 2017 Published by Elsevier Ltd.

## Introduction

Computational methods have played pivotal role in drug discovery efforts for many years [1]. Development of several approved drugs including early examples of captopril [2], saquinavir, ritonavir, indinavir [3], and tirofiban [4], has benefited substantially from the use of computer-aided drug design (CADD), which nowadays constitutes an essential part of the discovery pipeline at pharmaceutical companies [5,6]. The CADD tools are commonly classified into ligand-based (two-dimensional, 2D) and protein structure-based (3D). In this review we will focus on the 3D methods and discuss their potential and limitations. Their principles and implementations have evolved together with the concepts of molecular recognition on protein surface. In particular, the historic ‘lock and

key’ mechanism that served as a textbook explanation of substrate recognition at the enzyme active site has gradually developed into ‘hand and glove’ concept to account for protein flexibility and mutual adaptability of receptor and ligand.

Structure-based CADD supports hit identification and medicinal chemistry optimization by addressing two major tasks: predicting how small molecules bind to the protein target, and estimating (relative) binding affinity. We first review docking, originally inspired by the lock and key concept, which is used for both tasks. We then present a fragment-based method for high-throughput docking based on molecular mechanics and transferable force field. Finally, we discuss molecular dynamics (MD) protocols, which provide atomistic details of hand and glove-like association events. The use of MD simulation-based methods is increasing steadily as they are most adequate for the analysis of thermodynamics and kinetics of ligand binding and unbinding. The section on fragment docking focuses on the methods and programs developed in the group of the last author, while the review of MD simulations of binding is more general.

## Docking of small molecules to proteins

Automatic docking is concerned with the determination of the optimal position(s) and orientation(s) of a small molecule in a protein target. It has been reported that while the success of the approach is target-dependent and software suite-dependent, it poorly correlates with the binding affinity but rather depends on the quality of interactions that the ligand makes to the protein [7]. Quality of protein–ligand interactions can be to some extent expressed by the ligand efficiency (LE), the average binding energy per non-hydrogen (or heavy) atom of the ligand. However, it should be noted that most studies of the predictive ability of docking are biased toward the molecules that bind the protein target with detectable affinity and available crystal structure. A study of about 300 kinase inhibitors has shown that a simple scoring function (van der Waals energy only) outperforms total energy (i.e. van der Waals and electrostatics) in fitting binding affinity values but has poor predictive power (i.e. lower enrichment than ranking by total energy) for *in silico* screening by high-throughput docking [8]. The real challenge of *in silico* screening is the calculation of relative binding energies with sufficient accuracy such that there are as many true positives as possible among the final selection of compounds for *in vitro* testing. In turn, the successful evaluation of binding energy relies on the

accurate prediction of the binding mode. Recent studies have reported high success rate of fragment screening by docking using transferable force fields with implicit solvent treatment of electrostatics desolvation effects [9,10].

### Virtual screening by high-throughput docking

The principle of virtual screening is to evaluate the library of molecules for possibility of binding to the protein, and to shortlist the ones that are most likely to bind with the highest affinity. As mentioned above, the main challenge is not to identify the few nanomolar binders in the small-molecule library (if any at all) but rather to reduce the number of false negatives in the subset of compounds that are selected for validation by *in vitro* assays. There are few studies that systematically analyze the success ratio of docking campaigns (also called the hit rate), that is, the percentage of compounds correctly predicted to bind the protein target. While many papers report very good hit rates [11–13,14\*,15], the criteria defining a hit are always subjective and study-dependent. The most stringent criterion is to consider as validated only those hits confirmed by the crystal structure of target-ligand complex. In this context it has to be noted that even for millimolar binders it is possible to obtain the crystal structure of the complex with the target protein. On the other hand, it can be very difficult and sometimes impossible to solve the crystal structure of complex with a potent ligand (e.g. nanomolar affinity) because the binding site can be either occluded by crystal contacts or not accessible to the ligand due to the tight packing of the protein molecules in the crystal (which mainly affects soaking experiments). Most commonly used criteria for the hit rate are based on affinity as measured in biochemical assays or biophysical experiments *in vitro* (typically  $K_D$  or  $IC_{50}$  below 100  $\mu M$ ) or semi-quantitative data, for example, from ligand-based NMR spectroscopy [16]. Such success stories have to be approached with caution, as it is clear that the selection process often involves visual inspection and examination through users with significant expertise and can be biased toward scaffolds disclosed previously in the literature. To properly benchmark the performance of different software suites, common criteria should be introduced and human intervention should be minimized which is not simple because of the complexity of the analysis of binding poses [17] and/or costs related to the *in vitro* validation.

### Computer programs for flexible ligand docking

There is a plethora of software suites developed for the automatic docking of flexible small molecules into (mainly rigid) protein structures. On the other hand, only very few docking programs have gained broad recognition and are used by a large community [18]. These include Dock [19], GOLD [20], and AutoDock [21]. These solutions have gained high popularity due to their pioneering role in the field and thanks to extensive developments, which have turned them into user-friendly

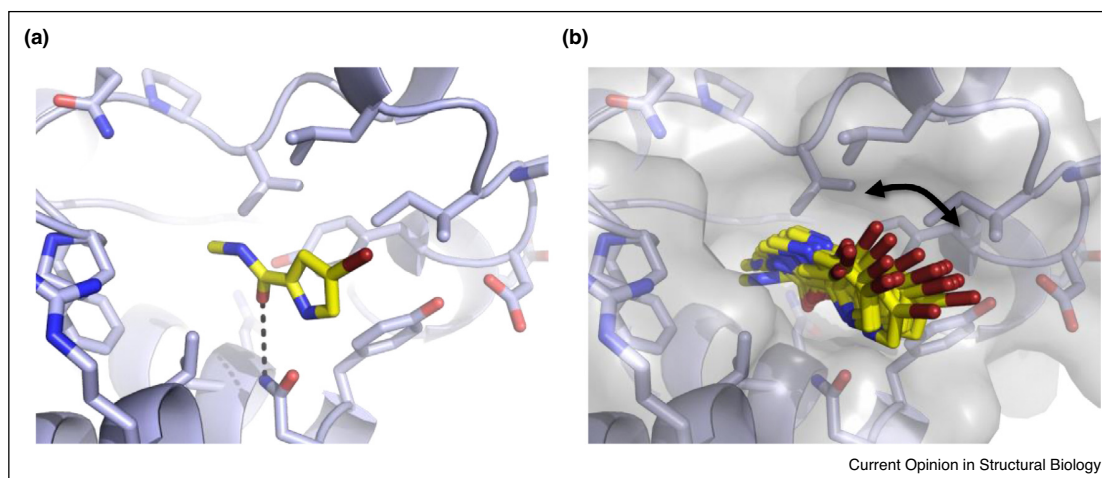
computer programs. More recently, rDock has emerged as an efficient docking tool distributed as open source code [22\*]. The most popular docking tools share similar sampling procedures (genetic algorithms-based optimization in the conformational space of the rotatable bonds or grid-based searches) and some of them use force field-based evaluation of the binding energy. A high degree of convergence toward the same pose in multiple docking runs of the same ligand (with different initial random populations of the genetic algorithm) was reported as necessary condition for successful prediction of the binding mode [23], despite being frequently neglected. Importantly, the probability of successful prediction of the binding mode decreases substantially as the intrinsic flexibility of the ligand grows [23], and depends on high-quality interactions made with the receptor [7\*]. Thus predictive ability has been validated for rigid fragments [9,24], while docking of peptides with more than a dozen rotatable bonds (backbone  $\varphi$  and  $\psi$  angles and side chain  $\chi$  angles) is considered speculative.

### Fragment docking

Nearly 20 years ago, the group of the last author developed a program for high-throughput docking of rigid fragments called SEED (Solvation Energy for Exhaustive Docking) [25]. SEED performs an exhaustive search in a discrete space defined by rotations around individual protein/fragment hydrogen bonds and/or hydrophobic contacts (Figure 1). This way, the essential feature of fragment-based drug discovery — making the high quality interactions with the protein [26] — is considered as a prerequisite and allows to reduce the complexity of search in the conformational space, and to enrich the docked poses in positives. A very efficient evaluation of bad contacts for filtering out poses with steric clashes and a two-step evaluation of the binding energy make the execution of SEED extremely rapid (about 1s per fragment). In both steps the energy evaluation is based on a transferable force field with continuum dielectric treatment of desolvation effects. The first step filters out the majority of the poses by the rapid evaluation of the van der Waals and Coulombic interactions on a 3D grid with a crude and very efficient approximation of desolvation effects [27]. In the second step the nonbonding interactions are calculated without grid-based approximation, and desolvation penalties are evaluated by the generalized Born equation with numerical calculation of the Born radii [28]. Importantly, the SEED binding energy does not require any fitting parameter and thus SEED can be used also for protein targets for which inhibitors have not been reported.

Successful high-throughput docking campaigns with SEED have been published for proteases, kinases, and bromodomains [9,24,29\*]. In a recent application SEED was used to screen for the CREBBP bromodomain a library of nearly 1500 fragments, which took less than one hour on a commodity computer, and resulted in a 50%

Figure 1



SEED [25,27] performs an exhaustive search in the discrete space defined by rotations around optimal hydrogen bond vectors (black dashed line). **(a)** Optimal pose according to calculated binding energy. **(b)** Bundle of SEED poses of 4-bromo-2-carboxymethylamide-pyrrole (C, N, O, and Br atoms in yellow, blue, red, and brown, respectively). Note that the fragment is compound **1** of Figure 2A.

success ratio (i.e. of 39 putative binders 20 were confirmed by ligand-observed NMR spectroscopy), and four crystal structures [14]. Moreover, the binding mode of the fragment hits predicted by SEED is essentially identical to the one observed in the crystal structures, which were solved *a posteriori* (Figure 2). The program SEED is available for free from the homepage of A.C.

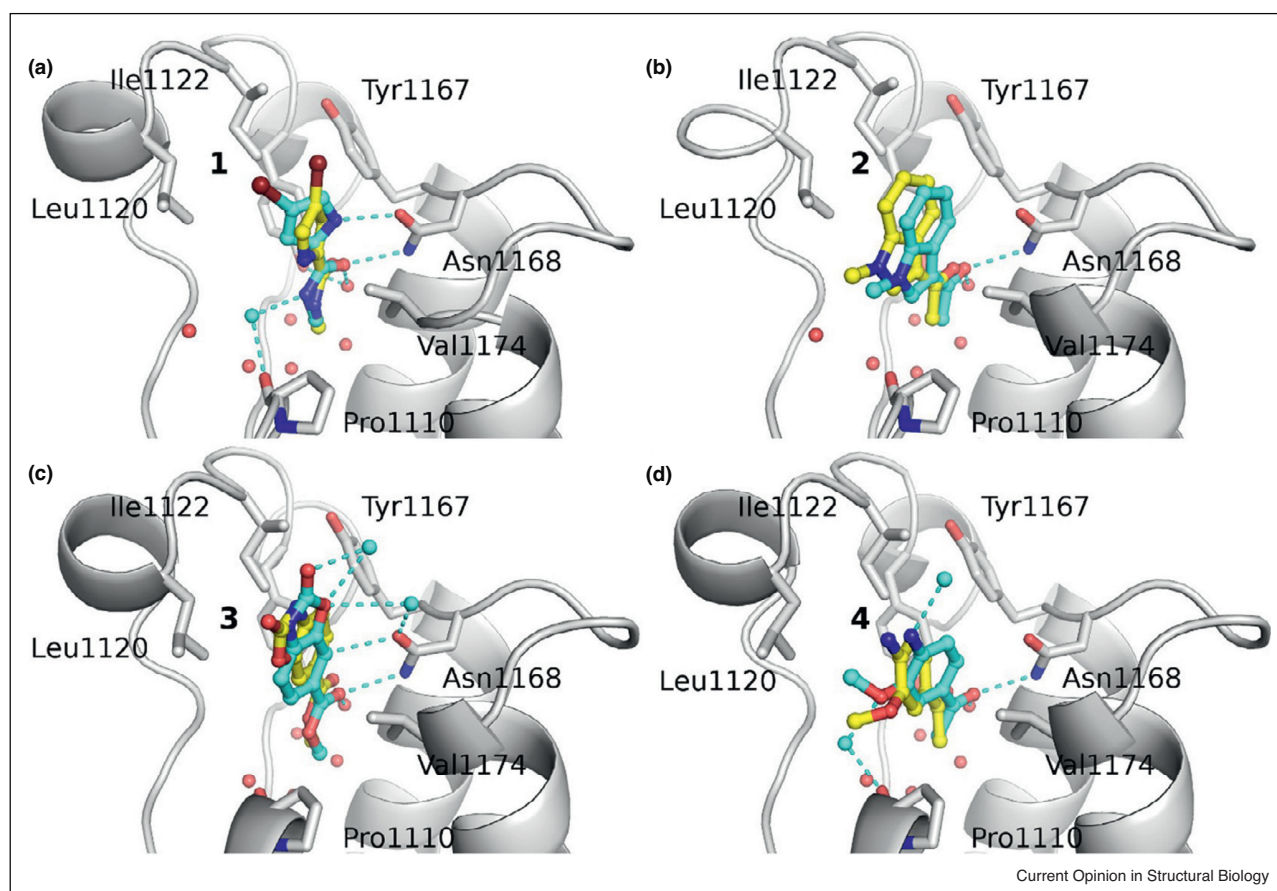
### Fragment growing

Once the hits are identified, they are subjected to chemical elaboration for the optimization of their binding affinity and/or drug-like properties. In one approach called fragment growing, further chemical moieties are added to the binding fragment to pick up additional interactions in the binding site. Such modifications can also benefit from docking of potential derivatives and other CADD approaches. As the accuracy of docking is reduced with the conformational complexity of the ligand, restraining the scaffold with confidently-determined binding pose of an anchor head group allows for more accurate binding mode and energy predictions of flexible molecules. Tethered docking relies on maintaining the binding mode of the fragment hit during the determination of the binding pose of the derived molecule (Figure 3a). Thus, such approach can be considered a knowledge-based implementation of the essence of fragment growing, which is realized by means of restraints to reduce the computational complexity.

A computational protocol called ALTA-VS (Anchor-based Library Tailoring Virtual Screening) that combines the advantages of high-throughput docking with those of fragment-based hit identification (Figure 3b) was first applied to  $\beta$ -secretase, a protease implicated in

Alzheimer's disease [30,31]. It consists of four steps (Figure 3b, [15,32]): (1) decomposition of each compound of the initial library to its rigid fragments by cutting at rotatable bonds; (2) docking of the fragments by SEED with evaluation of van der Waals interactions, Coulombic energy, and desolvation penalties (using the generalized Born equation); (3) flexible docking of the parent molecules that contain the top ranking fragments which are used as non-covalent binding anchors during docking; and (4) energy minimization with final evaluation of binding energy including desolvation effects by numerical solution of the finite-difference Poisson equation. Note that both steps (2) and (4) make use of the continuum dielectric approximation but the final step is based on the Poisson equation which is more accurate than the generalized Born approximation. Two essential advantages characterize the ALTA-VS approach with respect to the brute-force docking of a large library of flexible molecules. The first key element is the substantially smaller computational requirements for the fragment-anchored docking of a set of  $10^3$ – $10^4$  parent molecules than the non-anchored docking of a multi-million library of compounds (Figure 3b). The other advantage is much higher accuracy of docked poses of rigid fragments [14,29,33] than flexible molecules [23]. The successful identification of hit compounds by the ALTA-VS approach has been reported for several protein targets [8,15,30–32,34]. When starting from a known fragment, it constitutes an attractive implementation of computationally driven 'SAR by catalogue'. More recently, the group of the last author has also been implementing a fragment-growing algorithm that goes beyond the commercially available known chemical space. Importantly, it takes into account synthetic feasibility of the predicted molecules (Batiste *et al.*, submitted).

Figure 2



Structural validation of the fragment-based *in silico* screening campaign for the CREBBP bromodomain [14<sup>\*</sup>]. (a–d) The binding mode in the crystal structures (carbon atoms of ligands in cyan) are compared to the binding pose predicted by docking with SEED (carbon atoms in yellow) for compounds (a) **1**, (b) **2**, (c) **3**, and (d) **4** which correspond to the PDB structures 5MQE, 5MQK, 5MPZ, and 5MQG, respectively. Conserved water molecules and water molecules present in the crystal structure but not used for docking are shown as red and cyan spheres, respectively. Compounds **1** and **3** have a different relative orientation of the substituents, which could not be predicted by SEED since the compounds were docked as rigid molecules. Reprinted from [14<sup>\*</sup>] with permission from Elsevier.

### Contemporary challenges in docking

Even though software solutions for docking based on end point calculations (i.e. evaluation of binding energy using a single, rigid structure of the protein) have achieved performance that guarantees their standing involvement in discovery pipelines, several issues are to be addressed.

#### Improved description of binding energetics: polarization effects

The continuous and vigorous development of additive force fields for proteins [35–37] and small molecules [38–40] plays a fundamental role in the correct evaluation of relative binding affinities upon docking or MD-based (un)binding protocols. Furthermore, efficient tools for automatic atomtyping and parametrization of large libraries of compounds have been developed [41<sup>\*</sup>]. Despite the robustness and accuracy of classical force fields, a number of energetic effects of binding remain challenging to be

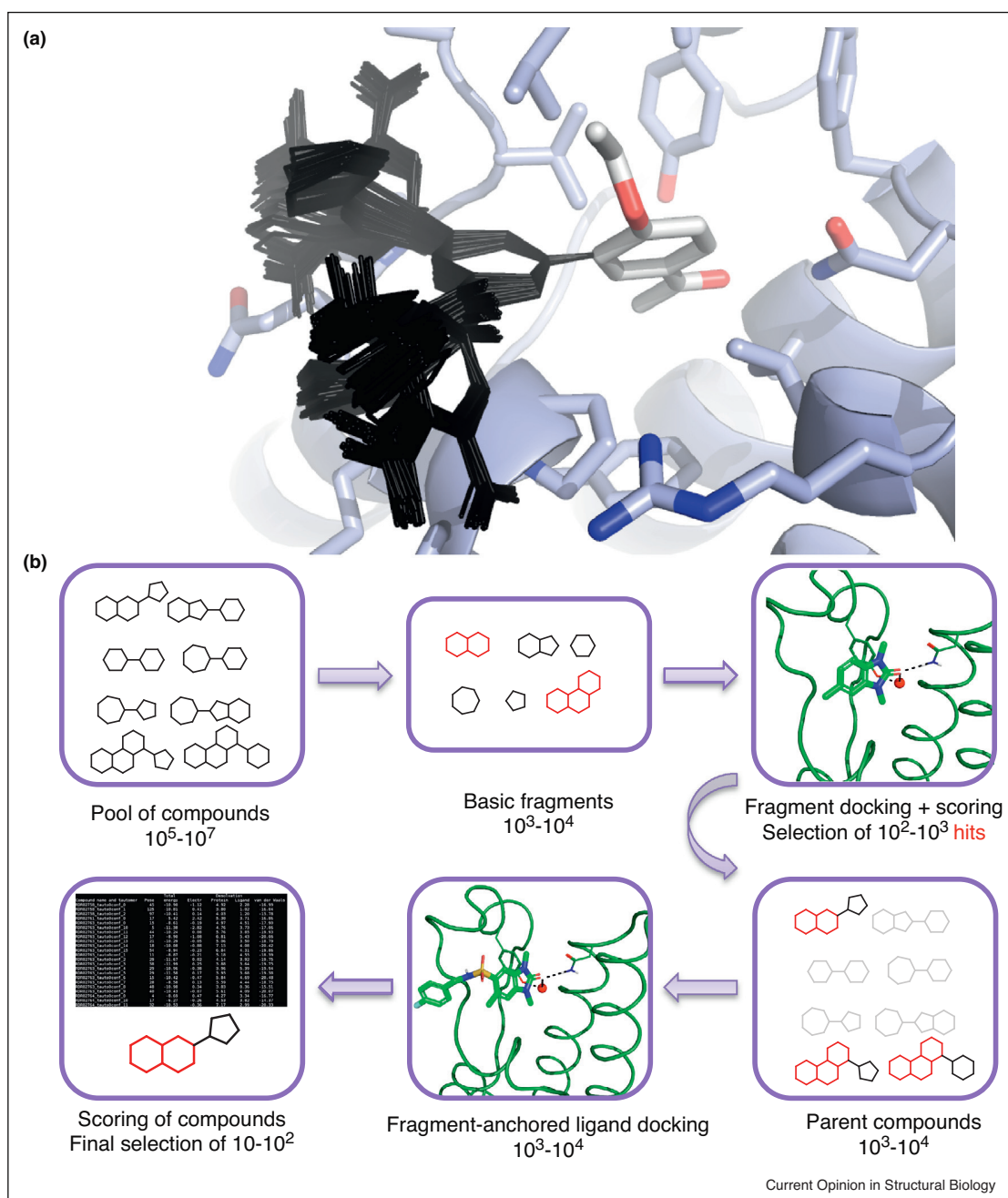
described by energy functions based on fixed partial charges. Key backbone and/or side chain groups can be treated as quantum mechanical (QM) probes to accurately approximate the local electronic structures in the binding site [42,43]. Briefly, QM probes use semi-empirical Hamiltonian for the evaluation of the interaction energy between ligand and individual polar groups of the protein to reproduce polarization and charge-transfer effects. One advantage is that partial charges are not needed. The main drawback is that dispersion effects are not captured by efficient semi-empirical methods. Alternatively, application of polarizable force fields is expected to improve the description of binding energetics in the coming years [44].

#### Solvent treatment: conserved water molecules

The treatment of water molecules remains a challenge, as it is difficult to predict which solvent molecules are



Figure 3



(a) The principle of tethered ligand docking — only the movement of the tethered black part of the molecule is allowed, the fragment-derived anchor shown in white is restrained to its position. (b) Anchor-based Library Tailoring Approach for Virtual Screening (ALTA-VS) [32]. A chemical library, with up to tens of millions of compounds, is decomposed into non-rotatable fragments, which are docked and scored. Parent compounds containing the top ranking fragments (red) are retrieved and docked with tethering of the fragment head-group. Those docked molecules are then further energy minimized with a force field and evaluation of electrostatic desolvation effects by the finite-difference Poisson approach. Thus, the ALTA-VS protocol selects  $10-10^2$  compounds for *in vitro* validation (bottom, left panel) from libraries of  $10^5-10^7$  molecules (top, left) by docking only  $10^3-10^4$  fragments (top, middle) and  $10^3-10^4$  compounds (bottom, middle).

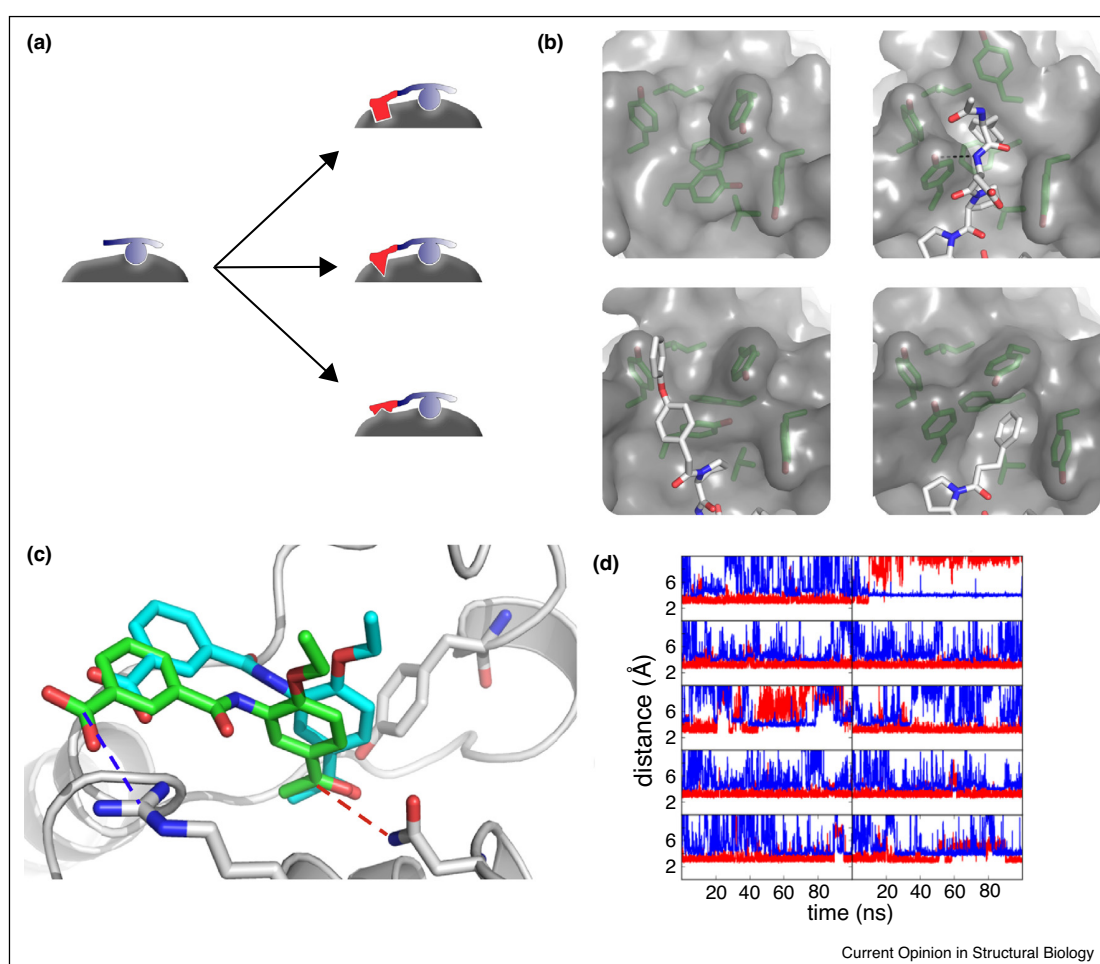
obligate in the binding site, and which can be displaced by incoming ligand. The water stability in the binding site can be determined based on analysis of multiple crystal structures, or more thoroughly by running MD simulations with explicit solvent [45]. The treatment of solvent during high-throughput docking can be mixed [12,14<sup>•</sup>,29<sup>•</sup>,33,46<sup>•</sup>,47,48], i.e. some water molecules that are structurally conserved in the binding site (i.e. those that act as bridges in intermolecular hydrogen bonds) are usually considered explicitly, while bulk water can be approximated efficiently by implicit solvent models to account for desolvation effects [25,27].

### Binding site flexibility

Protein flexibility constitutes another challenge of computer-aided drug discovery. Since the advent of structural biology methods and their application in structure-based drug discovery [49], it has become apparent that many

protein binding sites cannot be represented as a single snapshot, as significant structural rearrangements take place to accommodate diverse ligands. The pioneering work of Wells and coworkers targeting the interleukin-8 established protein-protein interactions as challenging targets to comprise flat and featureless surfaces that are inherently flexible and adapt to different binders [50,51]. Since then, several other examples of protein flexibility upon ligand binding have been shown. Cryptic pockets could be discovered by accident or through detailed analysis, often by more than one independent approach, as in the case of the polo-box domain of polo-like kinase 1, a mediator of phosphorylation-dependent protein-protein interaction and cell-cycle dependent cancer target [52–54] (Figure 4a,b). It has been also shown that structural adaptations are not limited to protein-protein interfaces — the enzymes are also more plastic than originally predicted

Figure 4



(a) The principle of observed ligand-dependent conformational changes of the protein surface. (b) Four crystallographic conformational states of the surface of polo-like kinase 1 exposed to different ligands ([52,61<sup>•</sup>] and unpublished data). (c) Comparison of docked and crystal structure (carbon atoms in cyan and green, respectively) of the hit compound targeting the CREBBP bromodomain [62]. (d) MD time series revealing stable hydrogen bonds between the key moieties of the compound and the protein (shown in blue and red, as in the panel c). Adapted with permission from Ref. [46<sup>•</sup>]. Copyright 2016 American Chemical Society.

by lock and key model, as shown in the case of mollusk acetylcholine binding protein, a structural model for the campaigns targeting the corresponding human enzyme [55]. Information on binding site flexibility, available from experimental and/or simulation studies, can be taken into account to prepare one or more conformations of the target protein for (high-throughput) docking [46\*].

### Molecular dynamics (MD) and receptor flexibility

MD simulations have been long proposed to provide insight into protein dynamics beyond that available crystallographically, and unravel novel cryptic binding sites, expanding the druggability of the targets. One of the first described approaches was the relaxed complex scheme, which combines all-atom nanosecond-long MD simulation of the protein target to describe its conformational flexibility with rapid docking of small molecules to the protein snapshots saved along the MD run [56,57\*]. This scheme has seen a number of successful applications, including one to the cancer-relevant MDM2/MDMx-p53 interaction [58] or HIV integrase [59]. Rewardingly, the cryptic pocket first proposed *in silico* in the latter case, has been then validated experimentally and is exploited by enzyme inhibitors approved for use against HIV infection in the clinic [60].

While the idea is conceptually correct, often the conformational change needed to uncover the pocket would not be observed at the 1- $\mu$ s timescale accessible to explicit solvent MD simulations with conventional protocols on commodity compute clusters, highlighting the need for more sophisticated sampling schemes. The molecular dynamics-induced fit (MD-IF) protocol makes use of atomistic simulations to increase (or reduce) the aperture of subpocket(s) in the active site [63]. The initial pose of the ligand that induces the fit can be obtained by manual docking of a known inhibitor into the target protein or structural alignment of the target with the holo structure of a cognate enzyme [64]. Another approach to the long timescale problem predicates on the fact that the cryptic sites are more likely to open (or remain open) in the presence of their cognate ligand, and several groups have proposed to perform ligand-mapping simulations. These include the MD-based protocol called SILCS (Site Identification by Ligand Competitive Saturation), which uses high concentration of the small-molecule ligands to map possible binding sites on protein surface. In this case the ligand aggregation is avoided by turning off the attractive part of the Lennard-Jones potential [65,66]. Alternatively, the aggregation can be avoided by running the simulation in the presence of lower ligand concentration (e.g. 0.2 M for benzene), which has been demonstrated to efficiently probe the conformational space of Plk1 and propose new cryptic pocket, subsequently validated experimentally [61\*]. More recently, enhanced sampling by Hamiltonian replica-exchange in the presence of ligand probes has

been shown to reproduce previous results (e.g. those for Plk1) and be less computationally expensive [67]. Its applicability to unprecedented targets remains to be seen though.

### MD simulations of ligand binding and unbinding

The application of MD to drug discovery projects is expanding, following the need for insights into binding, unbinding, and conformational change events at spatial and temporal resolution that is not available experimentally [10,68]. MD simulations can be used to map ligand binding sites and analyze (un)binding pathways [69,70\*,71–73]. Extensive MD simulations have been performed to determine binding sites and bound conformations of allosteric inhibitors of the M2 muscarinic acetylcholine receptor [74\*]. These simulations have revealed a new binding site dependent on cation- $\pi$  interactions which is placed 15 Å away from the classic recognition site, and was validated by radioligand binding experiments.

Unbiased simulations of ligand unbinding are useful, not only because they provide insight into affinity of the complex [71]. It is also possible to study the dynamics of complex formation and dissociation, and quantify complete energy landscape and kinetics for these processes [70\*]. It has been reported that the unbinding rates are strongly dependent on the state of the protein receptor, which in turn depends on the conditions of the experiment and in particular the concentration of the ligand. Therefore, as the unbinding of small ligands can be order(s) of magnitude faster than relaxation (i.e. conformational change) of the protein, it is crucial to select the right conformational state when executing high-throughput docking campaigns [75]. The ligand binding/unbinding kinetics can also have strong implications in pharmacology, as the residence time of the complex has been proposed to be more accurate predictor of drug efficacy than the affinity itself [76]. As a result, a number of simulation protocols have been designed to address this issue and estimate the kinetic parameters of drug (un)binding [77].

### MD simulations for ligand optimization

MD simulations are used frequently to guide further optimization of the molecules stemming from *in silico* discovery campaigns, particularly in the absence of a crystal structure of the complex with the target protein. Even if the structure has been solved, MD may provide insight as to which interactions are stable over time and contribute mostly to binding. This way, in a campaign to develop compounds targeting the bromodomain of CREBBP, MD simulations revealed the amide linker of an initial hit compound as not contributing directly to binding [46\*] (Figure 4c) and thus replaceable (Batiste *et al.*, submitted). In another example targeting the EphB4 tyrosine kinase MD simulations were used to prioritize the docking hits that were maintaining a stable hydrogen bond network in

the protein active site, which were then validated as true inhibitors with nanomolar affinity [78].

### Perspective

Two main techniques seem to emerge in the contemporary application of protein structure-based CADD methods. They are particularly useful in the initial phase (ligand identification) and advanced phase (ligand optimization) of drug discovery projects, respectively. For ligand identification, high-throughput docking of large libraries of small molecules (up to  $10^6$  molecules of 10–25 non-hydrogen atoms each) to a rigid protein structure is the method of choice. Importantly, classical force fields with implicit solvation and end point calculations (i.e. evaluation of binding energy using only one structure of the complex) have shown substantial predictive power [9,12,14<sup>•</sup>,15,24,29<sup>•</sup>,46<sup>•</sup>,79<sup>•</sup>].

For ligand optimization, MD simulation-based free energy calculations can be carried out on small sets (up to a few hundreds) of related molecules. The prediction of relative binding affinity by explicit solvent MD simulations is the more accurate the higher the pairwise compound similarity as statistical convergence, that is, sufficient sampling, is easier to achieve if the binding mode does not change substantially. The MD-based protocols include free energy perturbation and thermodynamic integration [80,81], alchemical free energy calculations [82,83], and umbrella sampling (see [84] for the theory and [85] for a successful application). Furthermore, MD simulations of spontaneous (un)binding can provide atomistic information on pathways and kinetics [70<sup>•</sup>,71–73,74<sup>•</sup>,75], and have shown potential for the identification of allosteric sites [69,74<sup>•</sup>]. While the rigid-protein docking of large libraries of fragments can be carried out on a conventional desktop (or even laptop) computer, enhanced sampling protocols and the availability of a compute cluster are essential for the MD-based calculations of (relative) free energies and binding kinetics. The continuous development of graphical processing units (GPUs) is benefitting the MD-based protocols substantially.

### Note

A recent paper has disclosed four fragment hits for the bromodomain of BAZ2A which is a protein target implicated in prostate cancer [86]. These hits were identified by high-throughput docking using the SEED program (see section Fragment docking) and consensus scoring based on force field energy terms with generalized Born solvation. The predicted binding mode of the four fragment hits was validated by X-ray crystallography (PDB codes 5MGJ, 5MGK, 5MGL, 5MGM). Notably, these are the first crystal structures of the BAZ2A bromodomain in the complex with non-peptidic small molecules.

### Conflict of interest

None declared.

### Acknowledgements

The authors would like to thank the members of the Cafilisch group for their scientific input and discussions. The work described here was in part supported by the grants from Swiss National Science Foundation (grants no. 149897 and 169007 to A.C.) and Synapsis Foundation (Alzheimer Research Switzerland) and Heidi Seiler-Stiftung (to A.C. and P.S.).

### References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Jorgensen WL: **The many roles of computation in drug discovery.** *Science* 2004, **303**:1813–1818.
  2. Talele TT, Khedkar SA, Rigby AC: **Successful applications of computer aided drug discovery: moving drugs from concept to the clinic.** *Curr Top Med Chem* 2010, **10**:127–141.
  3. Van Drie JH: **Computer-aided drug design: the next 20 years.** *J Comput Aided Mol Des* 2007, **21**:591–601.
  4. Hartman GD, Egbertson MS, Halczenko W, Laswell WL, Duggan ME, Smith RL, Naylor AM, Manno PD, Lynch RJ, Zhang G: **Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors.** *J Med Chem* 1992, **35**:4640–4642.
  5. Muegge I, Bergner A, Kriegl JM: **Computer-aided drug design at Boehringer Ingelheim.** *J Comput Aided Mol Des* 2017, **31**:275–285.
- The authors review the computational chemistry workflow implemented in their pharmaceutical company.
6. Hillisch A, Heinrich N, Wild H: **Computational chemistry in the pharmaceutical industry: from childhood to adolescence.** *ChemMedChem* 2015, **10**:1958–1962.
- The authors review the computational chemistry methods used in the pharmaceutical company in which they work. They also give some statistics on compounds in clinical phase I for which computational chemistry played a significant role.
7. Verdonk ML, Giangreco I, Hall RJ, Korb O, Mortenson PN, Murray CW: **Docking performance of fragments and druglike compounds.** *J Med Chem* 2011, **54**:5422–5431.
- The authors investigate the predictive power of fragment docking.
8. Kolb P, Huang D, Dey F, Cafilisch A: **Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model.** *J Med Chem* 2008, **51**:1179–1188.
  9. Spiliotopoulos D, Cafilisch A: **Fragment-based in silico screening of bromodomain ligands.** *Drug Discov Today Technol* 2016, **19**:81–90.
  10. Zhao H, Cafilisch A: **Molecular dynamics in drug design.** *Eur J Med Chem* 2015, **91**:4–14.
  11. Chen Y, Shoichet BK: **Molecular docking and ligand specificity in fragment-based inhibitor discovery.** *Nat Chem Biol* 2009, **5**:358–364.
  12. Marchand J-R, Lolli G, Cafilisch A: **Derivatives of 3-amino-2-methylpyridine as BAZ2B bromodomain ligands: in silico discovery and in crystallo validation.** *J Med Chem* 2016, **59**:9919–9927.
  13. Perryman AL, Yu W, Wang X, Ekins S, Forli S, Li S-G, Freundlich JS, Tonge PJ, Olson AJ: **A virtual screen discovers novel, fragment-sized inhibitors of Mycobacterium tuberculosis InhA.** *J Chem Inf Model* 2015, **55**:645–659.
  14. Spiliotopoulos D, Zhu J, Wamhoff E-C, Deerrain N, Marchand J-R, Aretz J, Rademacher C, Cafilisch A: **Virtual Screen to NMR (VS2NMR): discovery of fragment hits for the CBP bromodomain.** *Bioorg Med Chem Lett* 2017, **27**:2472–2478.
- In this multidisciplinary study, fragment hits were identified by high-throughput docking and implicit solvent force field ranking, and validated by ligand-observed NMR spectroscopy and X-ray crystallography.



15. Marchand J-R, Dalle Vedove A, Lolli G, Caflich A: **Discovery of inhibitors of four bromodomains by fragment-anchored ligand docking.** *J Chem Inf Model* 2017, **57**:2584-2597.
16. Śledź P, Abell C, Ciulli A: **Ligand-observed NMR in fragment-based approaches.** *NMR of Biomolecules.* Wiley-VCH Verlag GmbH & Co. KGaA; 2012:264-280.
17. Gathiaka S, Liu S, Chiu M, Yang H, Stuckey JA, Kang YN, Delproposito J, Kubish G, Dunbar JB, Carlson HA *et al.*: **D3R grand challenge 2015: evaluation of protein-ligand pose and affinity predictions.** *J Comput Aided Mol Des* 2016, **30**:651-668.
18. Chen Y-C: **Beware of docking!** *Trends Pharmacol Sci* 2015, **36**:78-95.
19. Allen WJ, Balius TE, Mukherjee S, Brozell SR, Moustakas DT, Lang PT, Case DA, Kuntz ID, Rizzo RC: **DOCK 6: impact of new features and current docking performance.** *J Comput Chem* 2015, **36**:1132-1156.
20. Jones G, Willett P, Glen RC, Leach AR, Taylor R: **Development and validation of a genetic algorithm for flexible docking.** *J Mol Biol* 1997, **267**:727-748.
21. 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.
22. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, Barril X, Hubbard RE, Morley SD: **rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids.** *PLoS Comput Biol* 2014, **10**:e1003571.
- This is the original publication of rDock, an open-source software tool for automatic docking.
23. Cecchini M, Kolb P, Majeux N, Caflich A: **Automated docking of highly flexible ligands by genetic algorithms: a critical assessment.** *J Comput Chem* 2004, **25**:412-422.
24. Huang D, Caflich A: **Library screening by fragment-based docking.** *J Mol Recognit* 2010, **23**:183-193.
25. Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflich A: **Exhaustive docking of molecular fragments with electrostatic solvation.** *Proteins* 1999, **37**:88-105.
26. Murray CW, Rees DC: **The rise of fragment-based drug discovery.** *Nat Chem* 2009, **1**:187-192.
27. Majeux N, Scarsi M, Caflich A: **Efficient electrostatic solvation model for protein-fragment docking.** *Proteins* 2001, **42**:256-268.
28. Scarsi M, Apostolakis J, Caflich A: **Continuum electrostatic energies of macromolecules in aqueous solutions.** *J Phys Chem A* 1997, **101**:8098-8106.
29. Zhu J, Caflich A: **Twenty crystal structures of bromodomain and PHD finger containing protein 1 (BRPF1)/ligand complexes reveal conserved binding motifs and rare interactions.** *J Med Chem* 2016, **59**:5555-5561.
- A very high hit rate was reported for high-throughput fragment docking with validation by X-ray crystallography.
30. Huang D, Lüthi U, Kolb P, Edler K, Cecchini M, Audetat S, Barberis A, Caflich A: **Discovery of cell-permeable non-peptide inhibitors of beta-secretase by high-throughput docking and continuum electrostatics calculations.** *J Med Chem* 2005, **48**:5108-5111.
31. Huang D, Lüthi U, Kolb P, Cecchini M, Barberis A, Caflich A: **In silico discovery of beta-secretase inhibitors.** *J Am Chem Soc* 2006, **128**:5436-5443.
32. Kolb P, Kipouros CB, Huang D, Caflich A: **Structure-based tailoring of compound libraries for high-throughput screening: discovery of novel EphB4 kinase inhibitors.** *Proteins* 2008, **73**:11-18.
33. Lolli G, Caflich A: **High-throughput fragment docking into the BAZ2B bromodomain: efficient in silico screening for X-ray crystallography.** *ACS Chem Biol* 2016, **11**:800-807.
34. Schenker P, Alfarano P, Kolb P, Caflich A, Baici A: **A double-headed cathepsin B inhibitor devoid of warhead.** *Protein Sci* 2008, **17**:2145-2155.
35. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C: **ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB.** *J Chem Theory Comput* 2015, **11**:3696-3713.
36. Jorgensen WL, Tirado-Rives J: **The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin.** *J Am Chem Soc* 1988, **110**:1657-1666.
37. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD: **Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\varphi$ ,  $\psi$  and side-chain  $\chi(1)$  and  $\chi(2)$  dihedral angles.** *J Chem Theory Comput* 2012, **8**:3257-3273.
38. Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, Wang L, Lupyán D, Dahlgren MK, Knight JL *et al.*: **OPLS3: a force field providing broad coverage of drug-like small molecules and proteins.** *J Chem Theory Comput* 2016, **12**:281-296.
39. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA: **Development and testing of a general amber force field.** *J Comput Chem* 2004, **25**:1157-1174.
40. Soteras Gutiérrez I, Lin F-Y, Vanommeslaeghe K, Lemkul JA, Armacost KA, Brooks CL, MacKerell AD: **Parametrization of halogen bonds in the CHARMM general force field: improved treatment of ligand-protein interactions.** *Bioorg Med Chem* 2016, **24**:4812-4825.
41. Vanommeslaeghe K, MacKerell AD: **Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing.** *J Chem Inf Model* 2012, **52**:3144-3154.
- The authors disclose a software technology for the automatic parametrization of organic compounds which is fully consistent with the CHARMM all-atom force field for proteins.
42. Zhou T, Caflich A: **High-throughput virtual screening using quantum mechanical probes: discovery of selective kinase inhibitors.** *ChemMedChem* 2010, **5**:1007-1014.
43. Zhou T, Huang D, Caflich A: **Quantum mechanical methods for drug design.** *Curr Top Med Chem* 2010, **10**:33-45.
44. Lemkul JA, Huang J, Roux B, MacKerell AD: **An empirical polarizable force field based on the classical drude oscillator model: development history and recent applications.** *Chem Rev* 2016, **116**:4983-5013.
45. Huang D, Rossini E, Steiner S, Caflich A: **Structured water molecules in the binding site of bromodomains can be displaced by cosolvent.** *ChemMedChem* 2014, **9**:573-579.
46. Xu M, Unzue A, Dong J, Spiliotopoulos D, Nevado C, Caflich A: **Discovery of CREBBP bromodomain inhibitors by high-throughput docking and hit optimization guided by molecular dynamics.** *J Med Chem* 2016, **59**:1340-1349.
- Ligands were identified by high-throughput docking and implicit solvent force field ranking. Chemical synthesis of derivatives for hit optimization was guided by the analysis of explicit solvent molecular dynamics simulations.
47. Lemmon G, Meiler J: **Rosetta ligand docking with flexible XML protocols.** *Methods Mol Biol* 2012, **819**:143-155.
48. Li L, Xu W, Lü Q: **Improving protein-ligand docking with flexible interfacial water molecules using SWRosettaLigand.** *J Mol Model* 2015, **21**:294.
49. Blundell TL: **Structure-based drug design.** *Nature* 1996, **384**:23-26.
50. Arkin MR, Randal M, DeLano WL, Hyde J, Luong TN, Oslob JD, Raphael DR, Taylor L, Wang J, McDowell RS *et al.*: **Binding of small molecules to an adaptive protein-protein interface.** *Proc Natl Acad Sci U S A* 2003, **100**:1603-1608.
51. Wells JA, McClendon CL: **Reaching for high-hanging fruit in drug discovery at protein-protein interfaces.** *Nature* 2007, **450**:1001-1009.

52. Śledź P, Stubbs CJ, Lang S, Yang Y-Q, McKenzie GJ, Venkataraman AR, Hyvönen M, Abell C: **From crystal packing to molecular recognition: prediction and discovery of a binding site on the surface of polo-like kinase 1.** *Angew Chem Int Ed Engl* 2011, **50**:4003-4006.
53. Śledź P, Lang S, Stubbs CJ, Abell C: **High-throughput interrogation of ligand binding mode using a fluorescence-based assay.** *Angew Chem Int Ed Engl* 2012, **51**:7680-7683.
54. Liu F, Park J-E, Qian W-J, Lim D, Gräber M, Berg T, Yaffe MB, Lee KS, Burke TR: **Serendipitous alkylation of a Plk1 ligand uncovers a new binding channel.** *Nat Chem Biol* 2011, **7**:595-601.
55. Bouzat C, Gumilar F, Spitzmaul G, Wang H-L, Rayes D, Hansen SB, Taylor P, Sine SM: **Coupling of agonist binding to channel gating in an ACh-binding protein linked to an ion channel.** *Nature* 2004, **430**:896-900.
56. Lin J-H, Perryman AL, Schames JR, McCammon JA: **The relaxed complex method: accommodating receptor flexibility for drug design with an improved scoring scheme.** *Biopolymers* 2003, **68**:47-62.
57. 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.
- The authors review the relaxed complex method which combines the advantages of molecular dynamics and docking.
58. Barakat K, Mane J, Friesen D, Tuszyński J: **Ensemble-based virtual screening reveals dual-inhibitors for the p53-MDM2/MDMX interactions.** *J Mol Graph Model* 2010, **28**:555-568.
59. Schames JR, Henchman RH, Siegel JS, Sottriffer CA, Ni H, McCammon JA: **Discovery of a novel binding trench in HIV integrase.** *J Med Chem* 2004, **47**:1879-1881.
60. Hazuda DJ, Anthony NJ, Gomez RP, Jolly SM, Wai JS, Zhuang L, Fisher TE, Embrey M, Guare JP, Egbertson MS *et al.*: **A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase.** *Proc Natl Acad Sci U S A* 2004, **101**:11233-11238.
61. Tan YS, Śledź P, Lang S, Stubbs CJ, Spring DR, Abell C, Best RB: **Using ligand-mapping simulations to design a ligand selectively targeting a cryptic surface pocket of polo-like kinase 1.** *Angew Chem Int Ed Engl* 2012, **51**:10078-10081.
- The authors report an efficient MD protocol to predict new cryptic binding sites. The predictions are validated through the design and synthesis of ligand targeting the new pocket.
62. Unzue A, Xu M, Dong J, Wiedmer L, Spiliotopoulos D, Caffisch A, Nevado C: **Fragment-based design of selective nanomolar ligands of the CREBBP bromodomain.** *J Med Chem* 2016, **59**:1350-1356.
63. Zhao H, Huang D, Caffisch A: **Discovery of tyrosine kinase inhibitors by docking into an inactive kinase conformation generated by molecular dynamics.** *ChemMedChem* 2012, **7**:1983-1990.
64. Zhao H, Caffisch A: **Discovery of ZAP70 inhibitors by high-throughput docking into a conformation of its kinase domain generated by molecular dynamics.** *Bioorg Med Chem Lett* 2013, **23**:5721-5726.
65. Raman EP, Yu W, Guvench O, MacKerell AD: **Reproducing crystal binding modes of ligand functional groups using Site-Identification by Ligand Competitive Saturation (SILCS) simulations.** *J Chem Inf Model* 2011, **51**:877-896.
66. Guvench O, MacKerell AD: **Computational fragment-based binding site identification by ligand competitive saturation.** *PLoS Comput Biol* 2009, **5**:e1000435.
67. Oleinikovas V, Saladino G, Cossins BP, Gervasio FL: **Understanding cryptic pocket formation in protein targets by enhanced sampling simulations.** *J Am Chem Soc* 2016, **138**:14257-14263.
68. Durrant JD, McCammon JA: **Molecular dynamics simulations and drug discovery.** *BMC Biol* 2011, **9**:71.
69. Huang D, Caffisch A: **How does darunavir prevent HIV-1 protease dimerization?** *J Chem Theory Comput* 2012, **8**:1786-1794.
70. Huang D, Caffisch A: **The free energy landscape of small molecule unbinding.** *PLoS Comput Biol* 2011, **7**:e1002002.
- Complex network analysis reveals heterogeneous unbinding pathways of small molecules from FKBP and a bound state that consists of several subbasins.
71. Huang D, Caffisch A: **Small molecule binding to proteins: affinity and binding/unbinding dynamics from atomistic simulations.** *ChemMedChem* 2011, **6**:1578-1580.
72. Buch I, Giorgino T, De Fabritiis G: **Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations.** *Proc Natl Acad Sci U S A* 2011, **108**:10184-10189.
73. Dror RO, Pan AC, Arlow DH, Borhani DW, Maragakis P, Shan Y, Xu H, Shaw DE: **Pathway and mechanism of drug binding to G-protein-coupled receptors.** *Proc Natl Acad Sci U S A* 2011, **108**:13118-13123.
74. Dror RO, Green HF, Valant C, Borhani DW, Valcourt JR, Pan AC, Arlow DH, Canals M, Lane JR, Rahmani R *et al.*: **Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs.** *Nature* 2013, **503**:295-299.
- The authors report molecular dynamics simulations of the spontaneous binding of allosteric inhibitors of the M2 muscarinic acetylcholine receptor. Validation was carried out by radioligand binding experiments.
75. Xu M, Caffisch A, Hamm P: **Protein structural memory influences ligand binding mode(s) and unbinding rates.** *J Chem Theory Comput* 2016, **12**:1393-1399.
76. Copeland RA, Pompliano DL, Meek TD: **Drug-target residence time and its implications for lead optimization.** *Nat Rev Drug Discov* 2006, **5**:730-739.
77. Mollica L, Decherchi S, Zia SR, Gaspari R, Cavalli A, Rocchia W: **Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations.** *Sci Rep* 2015, **5**:11539.
78. Zhao H, Dong J, Lafleur K, Nevado C, Caffisch A: **Discovery of a novel chemotype of tyrosine kinase inhibitors by fragment-based docking and molecular dynamics.** *ACS Med Chem Lett* 2012, **3**:834-838.
79. Röhrig UF, Majjigapu SR, Grosdidier A, Bron S, Stroobant V, Pilotte L, Colau D, Vogel P, Van den Eynde BJ, Zoete V *et al.*: **Rational design of 4-aryl-1,2,3-triazoles for indoleamine 2,3-dioxygenase 1 inhibition.** *J Med Chem* 2012, **55**:5270-5290.
- The authors report the design of very potent indoleamine 2,3-dioxygenase 1 inhibitors by means of docking followed by implicit solvent force field scoring.
80. Steinbrecher TB, Dahlgren M, Cappel D, Lin T, Wang L, Krilov G, Abel R, Friesner R, Sherman W: **Accurate binding free energy predictions in fragment optimization.** *J Chem Inf Model* 2015, **55**:2411-2420.
81. Aldeghi M, Heifetz A, Bodkin MJ, Knapp S, Biggin PC: **Predictions of ligand selectivity from absolute binding free energy calculations.** *J Am Chem Soc* 2017, **139**:946-957.
82. Armacost KA, Goh GB, Brooks CL: **Biasing potential replica exchange multisite  $\lambda$ -dynamics for efficient free energy calculations.** *J Chem Theory Comput* 2015, **11**:1267-1277.
83. Ding X, Vilseck JZ, Hayes RL, Brooks CL: **Gibbs sampler-based  $\lambda$ -dynamics and Rao-Blackwell estimator for alchemical free energy calculation.** *J Chem Theory Comput* 2017, **13**:2501-2510.
84. Roux B: **The calculation of the potential of mean force using computer simulations.** *Comput Phys Commun* 1995, **91**:275-282.
85. Herrmann US, Schütz AK, Shirani H, Huang D, Saban D, Nuvolone M, Li B, Ballmer B, Aslund AKO, Mason JJ *et al.*: **Structure-based drug design identifies polythiophenes as antiprion compounds.** *Sci Transl Med* 2015, **7**:299ra123.
86. Spiliotopoulos D, Wamhoff EC, Lolli G, Rademacher C, Caffisch A: **Discovery of BAZ2A bromodomain ligands.** *Eur J Med Chem* 2017, **139**:564-572.