Archives of Biochemistry and Biophysics 583 (2015) 105–119

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/ locate/yabbi

Review article

Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description

Francesca Spyrakis [a], Claudio N. Cavasotto [b, *]

[a] Department of Life Sciences, University of Modena and Reggio Emilia, Via Campi 183, 41125, Modena, Italy
[b] Instituto de Investigación en Biomedicina de Buenos Aires (IBioBA) — CONICET — Partner Institute of the Max Planck Society (IBioBA-MPSP), Godoy Cruz 2390, C1425FQD, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 30 March 2015
Received in revised form
3 August 2015
Accepted 3 August 2015
Available online 10 August 2015

Keywords:
Structure-based drug discovery
Homology modeling
Ligand docking
Virtual screening
Protein flexibility
Active site water molecules

ABSTRACT

Structure-based virtual screening is currently an established tool in drug lead discovery projects. Although in the last years the field saw an impressive progress in terms of algorithm development, computational performance, and retrospective and prospective applications in ligand identification, there are still long-standing challenges where further improvement is needed. In this review, we consider the conceptual frame, state-of-the-art and recent developments of three critical "structural" issues in structure-based drug lead discovery: the use of homology modeling to accurately model the binding site when no experimental structures are available, the necessity of accounting for the dynamics of intrinsically flexible systems as proteins, and the importance of considering active site water molecules in lead identification and optimization campaigns.

1. Introduction

The formation of non-covalent complexes between macromolecules and small-molecules is essential for the proper functioning of cellular processes, such as enzyme catalysis and signal transduction. With the aim of designing chemical modulators of therapeutic relevant targets, the pharmaceutical industry basically relied on high-throughput screening, a costly strategy involving the experimental screening of chemical libraries against a specific target. Since about 40 years ago, three-dimensional (3D) structures of protein-ligand complexes have been used to guide the optimization of drug leads in terms of potency and selectivity [1], thus incorporating structural knowledge into the drug discovery process. Later on, computational methods became also available to model protein-ligand interaction, and more recently, to *in silico* screen large chemical libraries against a biomolecular target. Structure-based virtual screening (SBVS) strategies thus became the *in silico* counterpart of older high-throughput screening approaches [2–7]. Since then, SBVS has experienced a continuous improvement in terms of algorithm development, computational performance, and retrospective and prospective applications in drug lead identification [8,9]. The dramatic surge of CPU power, and the advent of GPUs, also significantly increased the feasibility of computational

Abbreviations: GPCR, G Protein-coupled receptor; HM, homology modeling; HTD, high-throughput docking; LSHM, Ligand-steered homology modeling; MD, molecular dynamics; MRC, multiple receptor conformations; PDB, Protein Data Bank; PMF, potential of mean force; REMD, replica-exchange molecular dynamics; SAR, structure-activity relationships; SBVS, structure-based virtual screening; SDM, site-directed mutagenesis; VS, virtual screening.

* Corresponding author.
E-mail addresses: cnc@cavasotto-lab.net, ccavasotto@ibioba-mpsp-conicet.gov.ar (C.N. Cavasotto).

http://dx.doi.org/10.1016/j.abb.2015.08.002
0003-9861/

simulations: screening million-compound libraries towards pharmaceutically relevant receptors is now quite feasible even with standard computational resources.

There are four main components in an SBVS protocol (Fig. 1):

i) *The target structure*. A crucial element in any SBVS campaign is the availability of the 3D structure of the macromolecular target of interest. The rate of X-ray structures solved and deposited in the Protein Data Bank (PDB) is constantly increasing with thousands of proteins being added each year. At the moment this review is written (August 2015) ~110000 biological macromolecular structures are present in the PDB [10]. The probability of successfully crystallizing a new target depends on many factors, such as the sequence length, the inherent flexibility, the presence of transmembrane helices, and the net charge, among others [11]. Thus, not all pharmaceutically relevant targets can be easily crystallized. NMR also represents a great source of protein structural information, even though somehow limited by protein solubility and molecular weight. The inclusion of water molecules in the structure should be also decided at this stage (cf. section "Modeling waters within the active site"). Target structures should be carefully prepared to ensure structural integrity, assign the correct residue protonation and tautomer states, and inspect Asn and Gln flips. When no experimental structure is available, or easily obtainable, comparative or homology modeling (HM) can furnish reliable target models to SBVS (cf. section "*In silico* target models in structure-based virtual screening").

ii) *The compound library*. Compound libraries to be screened against the aforementioned target also need to be carefully prepared, taking into account the most representative tautomers, protomers and stereoisomers at the pH of interest (usually 7.4) [12]. In cases of retrospective protein-ligand docking, the availability of ligand activity information for the considered target, and an un-biased set of inactive molecules, may also contribute to obtain meaningful results from docking-based screening [13,14].

iii) *The docking strategy*. Each molecule from the chemical library has to be placed within the target binding site optimizing protein-ligand interactions, and retaining the most favorable poses; each pose has to be scored by a native or external scoring function, according to which the library of screened molecules is ranked. Many docking algorithms and scoring functions have been developed and implemented along these years, in order to properly estimate the free energy of binding, or to maximize the separation of potential ligands and decoys, and thus to place the potential binders at the top of the hit list [15–18]. In the choice of the docking or virtual screening strategy, different aspects should be considered. The level of flexibility for both the target and the molecule library has to be specified. During docking, although both entities could be considered rigid (as in the original "lock and key" model),

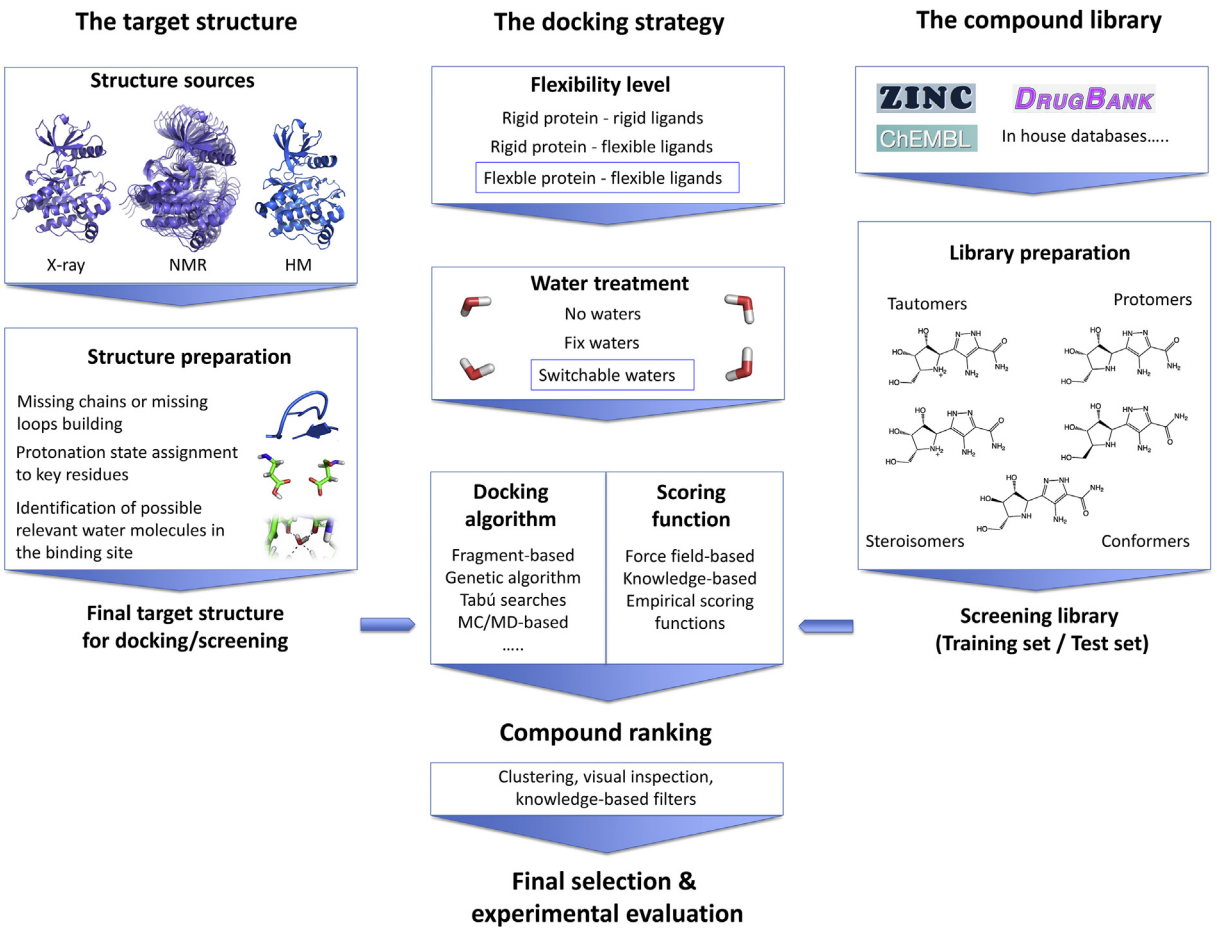

Fig. 1. Structure-based virtual screening workflow. The principal steps to be followed to perform virtual screening experiments and the available strategies are shown.

more often small-molecules are modeled as fully flexible while proteins are kept rigid, or only with a few moving relevant side-chains. Recently, approaches considering whole receptor flexibility or multiple receptor conformations strategies have become more common and feasible (cf. section "Accounting for protein flexibility in drug discovery"). The treatment of water molecules requires particular attention when water molecules mediating protein-ligand interaction are present within the binding site (see the section "Modeling waters within the active site").

iv) *The post-screening step*: After ranking by docking score, molecules can be chemically clustered, examined for key interaction patterns, and visually inspected. Compounds from this prioritized final hit-list are then purchased or synthetized, and finally evaluated in biochemical and functional assays.

Accounting for the intrinsic dynamical behavior of the system would certainly improve the simulations outcome. Proteins undergo different level of conformational adjustments, ranging from side-chains motions, up to large domain movements. Screening tools considering these effects should be preferred with respect to methodologies handling proteins as rigid objects [19,20] (cf. section "Accounting for protein flexibility in drug discovery"). The plasticity of proteins and, in particular, of binding sites is even increased by the possible presence of water molecules mediating catalytic reactions and/or protein-ligand recognition processes. In this perspective, algorithms able to rationalize water contribution, and to simulate multi-part interactions could provide better and more reliable results in a structure-based drug design scenario [21–23] (cf. section "Modeling waters within the active site").

Virtual screening offers a number of benefits with respect to traditional high-throughput screening approaches [7], such as low set-up cost, computational speed, and flexibility in changing the simulation conditions. Virtual procedures are also not affected by ligand purity, protein stability, or by the different samples and assays conditions [24]. However, VS suffers from a number of limitations stemming from the representation of the molecular system, the potential energy chosen to describe molecular interactions, and the inherent approximations. These factors may strongly affect docking poses, and the ranking and selection of the screened library. In spite of the plethora of techniques and tools currently available, a magic recipe to select the right structure, the most promising compound library, the best performing program, and the optimal post-screening strategies is not available yet [16]. Molecular simulations are everything but simple, and require a proper balance of the most appropriate tools, the adequate expertise and, most of all, an appropriate level of knowledge of the system under investigation.

In this review, we present the conceptual frame, methodological state-of-the-art, recent developments, and standing challenges of three critical "structural" issues in structure-based drug lead discovery:

i. The accurate modeling of the binding site and the use of homology models in retrospective and prospective SBVS;
ii. The strict necessity of accounting for the dynamics of intrinsically flexible systems, i.e. proteins when interacting with equally flexible ligands;
iii. The importance of classifying water molecules and of deciphering their possible contribution in ligand identification and lead optimization campaigns.

2. *In silico* target models in structure-based virtual screening

Even though the number of structures in the PDB has been increasing over the last ten years, partially thanks to structural genomics initiatives aimed to accelerate the characterization of unknown representative structures [25,26], the gap between annotated sequences and 3D protein structures is still very large, mainly due to the progress in DNA sequencing technology [27]. This gap is actually even larger when considering the level of redundancy within the PDB, and the underrepresentation of certain families such as membrane proteins, in spite of the recent impressive achievements to solve structures for ~30 different G protein-coupled receptors (GPCRs) [28,29]. It is thus clear that, in the years to come, a complete coverage of the druggable target space will not be available for structure-based drug lead discovery and optimization endeavors. In this scenario, accurate *in silico* homology modeling appears as a reliable alternative to fill the gap and expand by even two orders of magnitude the number of available structures [30,31].

The objective of homology modeling is to predict the unknown structure of a protein ("target") using the solved 3D structure(s) of homologous protein(s) ("templates") [32,33]. Modeling of a target protein consists of the following sequential steps: 1. selection of one (or more) known 3D structure(s) from sequence-based related template(s); 2. pairwise (or multiple) target-template sequence alignment; 3. construction of an initial model (crude model) based on the template, where the correspondence between amino acids in the target and template(s) is extracted from the sequence alignment; 4. refinement of the crude model incorporating biochemical and biophysical data whenever available; 5. model validation. A detailed description of each of these steps is beyond the scope of this work, and can be found elsewhere (cf. Refs. [34–37]).

Although as a general rule, the degree of target-template sequence similarity is considered an early predictor of the model quality [38], further consideration is needed when using homology models in docking and high-throughput docking (HTD), since global sequence identity may mask important discrepancies in the binding site region. In case of poor sequence identity (either global or at the binding site), structural conservation should be examined whenever possible. For example, the average low sequence identity in GPCRs (around 20%, less than the 30–50% threshold generally necessary to develop good homology models [39]) is compensated by the conserved seven-transmembrane fold, which allows the generation of high-quality models usable in drug design campaigns [31,40–44]. In cases of low sequence similarity, multiple sequence alignment might help in improving the quality of the final model [45–47], though it should be stressed that a correct sequence alignment is necessary -though hardly sufficient-to develop reliable homology models. Regarding the construction of the crude model, multiple-template approaches have also been used to model the target "by pieces" [48–52].

Recent studies, including the community-wide GPCR Dock assessments [42,44,53] showed that refining the crude model complexed with a ligand, and accounting for experimental data from site-directed mutagenesis (SDM) and quantitative SAR (QSAR), may yield reliable models for further SBVS studies (for a comprehensive review on GPCR homology modeling, and its impact on drug design, the reader may refer to ref. [31]).

2.1. Integral binding-site modeling and refinement

The refinement stage is critical to obtain an accurate binding site representation due to target-template structural diversity

stemming from differences in sequence identity, chemical diversity of target ligands, and structural flexibility associated with ligand efficiency [54,55]. In the case of GPCRs, for example, it was recently demonstrated that model performance in docking correlates well with the local structural quality, the RMSD of the binding site residues, and the number of key ligand-protein interactions [56]. The ability of proteins to adjust at the binding site level, and interact with diverse ligands highlights the importance of incorporating receptor flexibility at the modeling stage within the protein-ligand complex refinement stage [34,57].

Information from SDM and SAR studies in terms of the residues and ligand groups involved in ligand recognition, and protein-ligand interaction patterns observed in experimental structures of other related family members could be used to derive pharmacophore/geometrical constraints between the receptor and the ligand, or among the ligand and receptor themselves. The use of such constraints at the refinement stage proved to be a critical step to develop high-quality binding site models. In the so-called ligand-supported homology modeling method, the use of restraints derived from manual or rigid-receptor docking was incorporated within MODELLER [58] as a way to account for ligand information in the modeling process [59]. This methodology was successfully applied to identify novel antagonists for the neurokinin-1 receptor [60].

In the ligand-steered homology modeling (LSHM) method [43,61], the six ligand rigid coordinates, the conformational space of the ligand internal degrees of freedom, and the binding site side-chains are sampled together through flexible ligand-flexible receptor Monte Carlo-based docking, in order to structurally optimize the binding site. This constitutes an extension to HM of the protocol originally designed to optimize the binding site of receptors complexed with non-native ligands [62–67]. The LSHM method was used to model the melanin concentrating hormone receptor 1 (MCH-R1), an anti-obesity GPCR target. The obtained conformations were validated and prioritized through small-scale retrospective docking of known ligands, and then used to virtually screen ~200,000 small-molecules followed by functional assays of the top-scoring compounds, where six low-micromolar novel-chemotype inhibitors were discovered [43]. Later, a benchmarking study for existing GPCRs was published [41] (see Section 2.2). The ligand steered method was also used to rationalize structure-activity relationships (SAR) data of agonists and antagonists binding to the cannabinoid CB2 receptor [49,51,68].

Moro and co-workers exploited the spatial information encoded in an ensemble of small-molecule poses obtained through rigid-receptor docking to a crude model, to adjust the orientations of binding site side-chains. Both side-chains and ligand were then optimized by local energy minimization and the ligand was finally re-docked to the lowest energy model [69]. Sherman et al. adopted the Costanzi's approach, coupling experimental information of ligand binding, with *in silico* modeling of induced fit effects [40], to obtain β_2 adrenergic receptor models from bovine rhodopsin (bRho) [70]. Then, models for dopamine D(2), D(3), and D(4), serotonin 5-HT(1B), 5-HT(2A), 5-HT(2B), and 5-HT(2C), histamine H(1), and muscarinic M(1) receptors were generated using as template the structure of the β_2-adrenergic receptor [70] and assessed in high-throughput docking campaigns [71]. This strategy was also applied in the blind prediction of the D(3) receptor structure complexed with the D(2)/D(3) selective antagonist eticlopride (GPCR DOCK 2010 challenge [44]), where a successful prediction was achieved in one out of three submitted models [72]. A detailed description of the modeling of 5-hydroxytryptamine receptors 1B and 2B in GPCR DOCK 2013 can be found elsewhere [53].

The use of the induced-fit concept was also fundamental to model the ligand and key binding-site residues of angiotensin-converting enzyme 2 (ACE2) from ACE, CDK5 (cyclin-dependent kinase 5) from CDK2, thymidine phosphorylase from a bacterial homologue, and dihydrofolate reductase from a recombinant variant. It was also proved that applying this methodology more accurate ligand-protein models were obtained, with respect to docking in already refined models [73]. Recently Thomas et al. developed homology models for the acetylcholine muscarinic receptors M_1R-M_5R using the β_2-adrenergic receptor as template, and highlighted how the optimization of the binding site incorporating ligand information has a stronger impact on the quality of the final model than target-template sequence similarity [74]. Several other works confirmed the importance of accounting for ligand information during the modeling process [75–79].

2.2. Retrospective and prospective docking using homology models

When experimentally solved structures are not available, SBVS is usually carried out using homology models [34]. It should be stressed that model quality should correlate not only with high enrichment factors, but also with accurate docking poses. Clearly, the assessment of the performance of models in docking depends on many factors which do not belong to the realm of modeling itself, such as the quality of the docking program, the choice of ligand and decoy libraries, and the particular target, as it has been observed when using experimental structures [80]. *Ad hoc* scoring functions for docking in homology models have also been developed, such as eSimDock [81], which is quite insensitive to binding site deformations and represents a valid approach to perform large-scale high-throughput docking using homology models.

A large-scale study of modeling and docking in 38 different targets using small-molecules extracted from the DUD library [14], confirmed that docking on comparative models usually outperform random selection [82]. Models were also competitive in virtual screening (assessed using the enrichment factor) when compared to crystal structures, though in this study experimental structures usually displayed a better performance. The authors interestingly reported better results for receptor-ensemble docking approaches performed by using multiple homology modeled structures, with respect to single receptor conformation screening campaigns.

Orozco and co-workers demonstrated that structures derived from homology modeling are often similar in quality to crystal structures in high-throughput docking, even in scenarios with moderate target-template sequence similarity [83]. The use of multiple structures as a way to account for protein flexibility in docking was also explored, finding that even in the case of homology modeling, using an ensemble of structures enhances the chance of success.

A benchmarking study of the performance of homology models in HTD was presented for Class A GPCRs [41]. The aforementioned ligand-steered homology modeling technique was used to cross-model bRho, β_2-adrenergic, and A_{2A} adenosine receptors, and then to investigate the performance of both crystal structures and models in docking. The ligand-steered models, obtained optimizing binding site-ligand interaction, provided an improved performance in docking in terms of enrichment factors when compared to crude homology models prior to refinement, and to random selection. Models showing low ligand RMSD exhibited a good performance not only in small-scale docking, but also in terms of selectivity prediction, with enrichment curves at least comparable to those generated by crystal structures. This study strongly highlighted the importance of performing an adequate model refinement, which

can compensate for poor target-template sequence similarity. Other studies of ligand-protein docking on homology models [84] reached very similar conclusions regarding the necessity of accurately modeling binding sites architecture, and the importance of accounting for receptor flexibility, in agreement with what previously found for GPCRs [74]. In an interesting study, the performance of the X-ray active and inactive states of β_2-adrenergic receptor was compared with that provided by historical homology models developed prior to the experimental structures became available. Although the models presented a different global packing with respect to the crystallographic structures, they exhibited a surprisingly good performance in HTD campaigns [56]. These results suggest that models with an accurate representation of the binding site could be even better suited for HTD than experimental structures.

Nguyen et al. used Rosetta to model 14 different GPCRs structures on templates with over 50% target-template sequence identity, testing their docking performances. The authors found that the top ranked models, i.e. the ones giving better results in docking campaigns, showed RMSD within 2.9 Å with respect to the experimental target structures, and that ligand sampling performance was enhanced by ~100-fold using knowledge-based and energy-based filters [85]. The use of RosettaLigand [86,87] docking into homology models was benchmarked using nine protein-ligand complexes built during CASP8 [88], and other 21 additional complexes in order to expand the chemotype space. RosettaLigand found an acceptable native-like pose among the top ten scoring binding modes [89] in 21 out of 30 docking runs. It was also observed that docking performance depends more on template selection based on ligand occupancy than on target-template sequence identity.

A further study on the accuracy of small-molecule docking into experimentally solved and homology modeled GPCR structures was recently performed by Beuming and Sherman, using both rigid receptor and induced-fit approaches [90]. The A_{2A} adenosine, β_1- and β_2-adrenergic, D_3 dopamine, H_1 histamine, M_2 and M_3 muscarine, $S1P_1$, κ-opioid, and CXC chemokine 4 receptors were modeled using multiple templates, and tested according to their capability of correctly predicting the binding mode of known ligands. In this study, docking results were significantly worse in homology models with respect to crystal structures, and it was observed that docking accuracy correlated well with the degree of target-template sequence similarity. Interestingly, the authors observed that the best results were obtained by incorporating additional information, either from experimental (SDM), or computational (pharmacophore modeling) sources.

On Table 1 we report a list of the latest applications where homology models have been used in prospective structure-based drug discovery campaigns.

3. Accounting for protein flexibility in drug discovery

As previously highlighted, modeling protein dynamics can strongly improve docking performances in terms of pose prediction and ligand specificity. The use of a single static structure and the performance of a single receptor conformation in docking and screening are not always reliable, in particular when dealing with comparative models. As proteins move to perform their functions, SBVS methodologies need to account for protein flexibility in an accurate and efficient way, in order to give reliable predictions [125–129]. Nevertheless, modeling flexibility still represents one of the most challenging and difficult issues in structure-based drug discovery and design.

3.1. The ligand binding event: old and new perspectives

The original "lock and key" model according to which proteins and ligands should be perfectly complementary to generate a complex is clearly an oversimplification. By that time, the visionary model proposed by Fisher in 1894 was able to explain enzyme specificity, but failed to explain the stabilization of the ligand transition state and, fundamentally, did not consider the possibility of a mutual conformational adjustment [130]. Enhanced models were proposed since the pioneering lock and key, including the induced fit (based on the induced structural changes occurring upon ligand binding), the conformational ensemble (based on the presence of a pre-existing ensemble of equilibrated structures) [131], and the random fluctuation of receptors and ligands [128,132,133]. In a more complex but, clearly more realistic perspective, ligand binding can be thought as a combination of a conformer selection-stage followed by minor adjustments within the binding pocket modulated by the ligand (induced-fit stage) i.e., the structural changes upon ligand binding occur within the ensemble of pre-existing receptor conformers [134].

The pre-existing or ligand-induced conformations represent the conformational space in which proteins move, and are the result of the interaction and collective motion of their constituent atoms. The accurate modeling of these collective motions is far from trivial. In spite of the unbelievable advancements experienced in the current GPU era, there is still need to balance speed and accuracy, i.e. reasonable computing time *vs* accurate simulations.

A variety of computational methods have been developed in the last years combining accuracy, reliability and computational efficiency. *In silico* methodologies able to deal implicitly or explicitly with protein flexibility in drug design include soft docking [135,136], docking with libraries of side-chain rotamers [136], algorithms accounting for induced-fit adjustments [70,137], Monte Carlo sampling [62,138] and MD simulations [139]. More recently docking into multiple conformations coming from X-ray [140–142], NMR [143,144], a combination of both [145,146] or computationally generated by MD simulations [147,148], normal-mode analysis [63,66,149], Monte Carlo [62,150], or homology models [83] have been successfully applied.

3.2. Docking with protein flexibility

Flexible ligand-flexible protein methods, simulating both ligand and protein dynamics, currently represent the first-end choice with respect to rigid-body or flexible ligand-rigid protein approaches.

The simplest approaches rely on soft potentials, rotamer libraries or local energy minimization used to simulate the movements of residues side chains lining the binding site. These approaches, usable when protein motions are limited and when ligands easily fit into a fairly static binding site [151], have been successfully applied in different scenarios [152,153]. In order to reduce the computational time a small number of degrees of freedom is usually considered and the variables driving the protein conformational changes are optimized along with the ligand translational, rotational and torsional changes.

Soft docking approaches allow a certain structural overlap between the protein and the ligand by means of soft core potentials, i.e. tolerant scoring functions. The main advantage is represented by the low computational cost, and by the possibility of identifying conformations or induced fit effects normally not experienced by more complex and demanding methodologies. Only small adjustments limited to side chains and not involving backbone rearrangements can be generally investigated [127,154]. Soft docking

Table 1
Recent applications of homology models in combination with high-throughput docking.

Target protein	Template (PDB ID)	Computational methods		Chemical library	Best compound discovered	Ref.
		Homology Modeling	Structure-based virtual screening			
E. tenella CDK-related kinase 2 (EtCRK2)	CDK2 (PDB 1OIR)	MODELLER [58]	Autodock [91]	~4 million from commercial vendors	$K_i = 0.17$ μM	[92]
Matriptase-2	Matriptase (PDB 1EAX)	MODELLER [58], MOE, SwissModel [94]	DOCK [95], FlexX [96]	15,000 filtered compounds from ZINC [97]	$K_i = 170$ nM	[98]
Penicillium digitatum Sterol 14α-demethylase (CYP51)	*Mycobacterium tuberculosis* CYP51 (PDB 1E9X)	CPHmodels 2.0 Server (http://www.cbs.dtu.dk/services/CPHmodels)	Gold [99], FlexX [96]	79,000 filered from SPEC (www.specs.com)	$K_d = 60$ nM	[100]
NF-κB inducing kinase (NIK)	hPAK1 (PDB 1YHW) [101]	ESyPred3D [102]	GOLD [99]	67,500 filtered compounds from ZINC [97]	$IC_{50} = 51$ μM	[103]
α4β2 Neuronal Nicotinic Receptor	Models from Ref. [104]	MODELLER [58] and MD	Glide [105], AutoDock [91]	10,000 from ChemBridge CNS-Set	$IC_{50} = 6.1$ μM	[106]
P90 ribosomal S6 kinase 2 (RSK2)	hMSK1 (1VZO hRSK1 (2Z7R mRSK2 (3G51)	Discovery Studio (http://accelrys.com)	Glide [105]	60,000 from MayBridge Database (www.maybridge.com)	$IC_{50} = 2.4$ μM	[107]
Tropomyosin-related kinase A (TrkA)	Insulin-like growth factor 1 receptor (3I81)	MODELLER [58]	AutoDock [91]	240,000 extracted from InterBioScreen Ltd. (www.ibscreen.com)	$K_d = 3$ μM	[108]
5-HT(2A)	β_2-adrenergic receptor (3D4S	MODELLER [58] and MD	[109,110]	1430 compounds extracted from DrugBank [111] and ZINC [97]	$K_d = 2$ μM	[79]
Histamine H_3 Receptor	Histamine H_1 Receptor (3RZE	MODELLER [58] and MD	Ligand- and structure-based using FLAP [112]	156090 compound extracted from ZINC [97]	$K_i = 0.5$ μM	[113]
GPR34	Based on earlier models of CCK1 [114]		Cerius2 v4.8/ LigandFit	1250000 compounds from their corporate chemical library	CRE-Luc $C_{50} = 450$ nM	[115]
CXCR7	Primary: CXCR4 receptor (3OE6 Secondary: bRho (1HZX, β_2-adrenergic (2RH1, turkey β_1-adrenergic (2VT4, adenosine A_{2A} receptor (3EML receptors	MOE [93]	CONSENSUS-DOCK [116]	Compounds from Namiki Shoji Co., Ltd., Kishida Chemical Co., Ltd., and TPRP	$IC_{50} = 1$ μM	[117]
Histamine H_4 Receptor	Histamine H_1 Receptor (3RZE	MacroModel [118]	Glide [105]	12,905 fragments from in-house collection	Hit ligand efficiency 0.31 −0.74	[119]
Dopamine D_2 receptor	Dopamine D_3 Receptor (3PBL	Schrodinger Suite [120]	Glide [105]	In-house fragment collection	$K_d = 5.4$ nM for the best linked compound	[121]
Anopheles gambiae octopamine receptor	β_2-adrenergic receptor (2RH1 3SN6	I-TASSER online server [122]	Glide [105]	12 million compounds extracted from ZINC [97]	$IC_{50} = 1$ μM $EC_{50} = 1$ μM	[123]
DUSP16 phosphatase	DUSP9 phosphatase (2HXP	MODELLER [58]	AutoDock [91]	260,000 extracted from InterBioScreen Ltd. (www.ibscreen.com)	$IC_{50} = 1$ μM	[124]

techniques are used in Glide for the preliminary placement of ligands. Induced fit effects are then modeled by replacing a given set of residues with alanine, which are then reconstructed after ligand placement. Ligand side chains are rebuilt, the whole binding site minimized, and then ligands are re-docked with no soft potentials [70].

Methods using rotamer libraries can also be applied to proteins characterized by limited motions at the binding site level, and for which reliable structural knowledge about the unbound and bound states is available. Alas, artifact complexes could be obtained and no information about the backbone adjustment is provided [155,156]. SLIDE [157], for instance, rotates both ligand and receptor side-chains to remove intermolecular overlaps. The receptor pocket is analyzed for the identification of hydrophobic and hydrogen-bonding template points, and all ligand hydrogen bond donor and acceptor groups are combinatorially matched within the binding pocket. Ligand fragmentation is performed to obtain fragments bearing triplets of matching points, positioned by a least square superposition algorithm. Also, water molecules can be retained in the binding site and their energetic contribution estimated.

RosettaLigand [158] and GalaxyDock2 [159] use Monte Carlo sampling and space annealing global energy minimization to simultaneously optimize both proteins and ligands, respectively. In particular, GalaxyDock2 is a geometry docking methods based on the beta-complex, a structure derived from the Voronoi diagram of receptor atoms, which could enhance both the computational speed and binding mode prediction accuracy.

When ligands are known to stabilize previously unknown structures or to select appreciably different conformations with respect to the experimental ones, the inclusion of pre-generated multiple states, obtained by means of experimental or computational techniques, is necessary to have reliable predictions. Molecular Dynamics simulations [160] and *ad hoc* protocols for fit [137,161] are among the most reliable methods to consider both side-chain and backbone flexibility. Geometry-based sampling methods can be also used to generate ensemble of conformations from a single structure in a faster and computationally undemanding way [162,163].

MD represents one of the most powerful available tools to describe intrinsic protein flexibility and to investigate protein motion whether free or complexed with a ligand. MD simulations, based on Newton's equations of motion [164], can describe the evolution of the conformational state and energy landscape of a protein as a function of time. The recent impressive progress in computer power and algorithm performance have made MD applications much more feasible than in the past, even if it is still more computationally demanding compared to the aforementioned methodologies. MD approaches have been used to generate ensemble of conformations for virtual screening campaigns [165,166], to replicate the structural dynamics of proteins in solution [167], or to identify cryptic or druggable allosteric sites, normally not present in rigid static structures [168,169]. Unfortunately, plain MD simulations are still time-limited and in many cases the length of simulation is not enough to sample biologically processes happening in the millisecond to second range. Low-energy conformations separated by high energy barriers are seldom crossed, unless the simulation runs for unrealistic timescales [170]. Thus, many MD applications in docking and screening are limited to the post-docking stage [171,172]. To overcome this limitation, different biased approaches for sampling the biological space in a reasonable computational time have been developed.

In accelerated MD the Hamiltonian of the system is modified to smooth the potential energy surface maintaining the relative positions of local minima. Since barriers separating low-energy states become lower, those states are more frequently visited by the system [173,174]. In metadynamics, the potential energy surface is described in a coarse-grain way using a few collective variables and a history-dependent potential defined by adding Gaussian contributions to the potential energy along the trajectory of the collective variables, thus forcing the system to avoid already sampled regions [175,176]. Despite the successful results recently reported, the outcome of metadynamics simulations heavily depends on the choice of the collective variables, what could be not trivial in cases where no knowledge about the final state conformation is available [177,178]. Replica-exchange molecular dynamics (REMD) simulations were also extensively used to enhance conformational sampling in biomolecular systems by increasing the system temperature [179,180]. Initially conceived to model protein folding, this methodology was successfully applied to the drug design field [181,182]. Extensions of the REMD include multidimensional replica-exchange methods (REM) [183], also referred as Hamiltonian REM [184], and replica-exchange umbrella sampling, which combines the conventional umbrella sampling methodology [185] with REM, and revealed to be particularly appropriate for free-energy calculations [183]. Umbrella sampling techniques are capable of sampling the conformational space also in high energy regions and, consequently, of giving reliable estimations of the free energy difference [186,187]. These methods often use harmonic biasing potentials in a series of windows along the reaction coordinate, or a single window in combination with an adaptive biasing potential aimed to match the whole free energy profile. In each window the sampling is performed through conventional MD or Hamiltonian RE. Again, the choice of the reaction coordinate is crucial for having predictive results [188]. Aside from plain MD and enhanced MD, other methodologies recently emerged for generating ensembles of protein conformations as the ligand-model concept proposed by Xu et al., developed for sampling the most suitable protein conformations able to bind structurally diverse sets of ligands [189].

3.3. Choosing the right receptor conformation in virtual screening

Docking and screening to flexible proteins generally improve the overall results. Therrien et al. reported that when screened compounds are similar one to each other, and they bind the same active site form, adopting a multiple receptor conformation (MRC) approach ameliorates the ranking of decoys. A much more significant improvement can be achieved when ligands are structurally different and/or are accommodated by diverse binding site conformations (Fig. 2) [24]. This is in line with what had been shown several years ago in protein kinases, where crystal structures and modeled receptor–ligand complexes were selected for MRC docking based on the chemical diversity of their bound ligands [62,63,66]. Thus, if a single receptor conformation approach could be suitable for lead optimization, in SBVS campaigns aimed at identifying new scaffolds, multiple receptor conformation docking, using either experimental or computationally generated structures, may represent a better choice.

The selection of the right protein conformation to be used in structure-based virtual screening is more crucial than initially thought. The “right” conformations should be able to provide the best results in terms of quantity and quality, where quantity corresponds to the percentage of active molecules identified in the early screening stage, and quality to the chemical diversity of these first ranked molecules. Both single- and multiple receptor conformation approaches [24] are significantly affected by the chosen binding site architecture, so that the key point is often represented by the capability of the conformational sampling to select the most representative states [145,190]. Some strategies were recently proposed to guide this conformational sampling. One of the first “recipes” to select the best performing conformations was proposed by Rueda et al. who analyzed 1086 X-ray conformations for 99 different therapeutically relevant protein targets, and compared the performances of the single and ensemble-conformation docking [142]. The authors proved that, when activity information is available, a ligand-guided approach represents the best strategy to identify the conformations better able to discriminate between binders and non-binders. They also observed that the use of conformational ensembles does not systematically outperform a single well-performing structure, and that it is generally not possible to outperform the best single conformation even when the others are handpicked according to their AUCs [140,191]. Significant improvements by using multiple structures can be reached when poor performing structures or no-experimental structures are available, as in the case of homology models.

Unfortunately, ligand information is generally available only for well-known pharmaceutical targets, which is not the case of newly discovered proteins possibly involved in untreated or neglected diseases. As suggested by Verdonk and co-workers, when no ligand information is available, an ensemble of randomly chosen 3–5 conformations seems to provide better results with respect to a single static structure. The score-based uncertainties related to a single receptor state can be, in fact, reduced and more consistent scores for ligand profiling generated [192].

The performances of single and multiple receptor conformation approaches were also compared by Bottegoni et al. carrying out a retrospective study on 36 X-ray targets [193] and proving, again, that the use of multiple *vs* single receptor conformations significantly increases the diversity of discovered chemical scaffolds. Further improvements could be achieved introducing more diverse and ligand-unbiased states coming, for instance, from MD simulations. Again, the authors underline that the conformational selection in absence of any ligand activity information still represents a

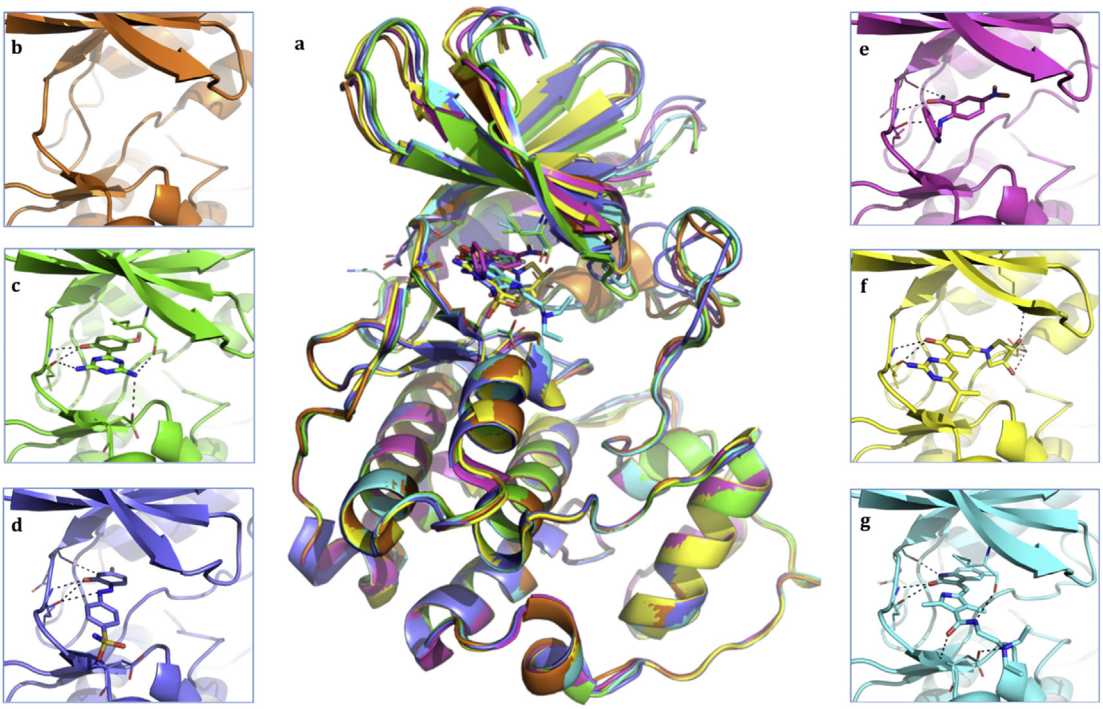

Fig. 2. Exploiting protein flexibility. a. Superposition of six different Cyclin-dependent kinase 2 structures (PDB codes 4EK3, orange; 3PXY, green; 4EZ3, slate; 3R7V, magenta; 3S2P, yellow; 3TIL, cyan). The overall superimposition highlights the conformational adjustment experienced by the P-loop covering the binding site. The following insets show how different binding site conformations can host different ligand chemotypes. b. 4EK3 (apo form). c. 3PXY. d. 4EZ3. e. 3R7V. f. 3S2P. g. 3TIL. Ligands and interacting residues are represented as capped sticks, hydrogen bonds as black dashed lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

challenging and major issue, which can greatly affect the VS results. Predicting *a priori* which conformer can provide the best results remains an unsolved issue.

The same conclusions were basically drawn by Korb et al. when investigating the performance of ensemble docking in virtual screening [194]. Ensemble docking always outperformed the use of the worst single structure and, in some cases, provided even better results with respect to the best single protein conformation. Since, as previously stated, it is not possible to foresee *a priori* the performances of single conformations, MRC approaches should at least minimize the risk of bad or worst performances in virtual screening. Good results were also obtained in terms of chemotype diversity. The strong dependence of docking results on scoring functions and sampling accuracy was highlighted, since different performances were observed for the same set of structure when using different scoring functions. The pressing need of developing ensemble selection protocols for choosing the best conformations given a target and a scoring function was again underlined.

More recently Spyrakis et al. outlined an innovative pipeline to extract the best performing conformations from MD simulations when ligand information is available [165]. The authors applied an integrated MD-FLAP approach including MD simulations, and Clustering and Linear Discriminant Analysis for enhancing the accuracy and efficiency in VS campaigns. They first extracted a limited number of representative structures from tens of nanoseconds of MD trajectories by means of the k-medoids clustering algorithm. Then, instead of applying arbitrary selection criteria, they let the linear discriminant analysis implemented in FLAP (Fingerprint for Ligand and Proteins) [112] to automatically choose the best performing conformational states among medoids and X-ray structures. Retrospective virtual screenings confirmed that ensemble receptor protocols outperform single rigid receptor approaches, proved that computationally generated conformations comprise the same quantity/quality of information included in X-ray structures, and identified the MD-FLAP approach as a valuable tool for improving VS performances.

4. Modeling waters within the active site

Protein mobility is possibly enhanced by the presence of moving and displaceable water molecules inside and outside the protein matrix and, in particular, within the binding site. The cases of ligands properly designed to fit hydrated binding sites, or to displace structural waters are known and well documented in the literature [195,196]. Water molecules are, in fact, critical actors in biomolecular recognition, and their role has been extensively investigated [197–205]. They are responsible for hydrophobic effects [206,207], discrete hydrogen bonded water networks [208], and they are also involved in the stabilization of protein-ligand complexes [209,210]. Water-mediated effects are many, though far from being clearly understood, and modeling their behavior still remains extremely challenging. Many different approaches, starting from continuum solvent models [211–214] up to methods considering waters as explicit components [215–218], or at a quantum level [219,220] have been developed to predict water effects and energetics in biological systems.

Waters in proteins have been classified into different categories, i) water molecules in active sites, ii) water molecules deeply inserted within cavities, iii) buried water molecules, iv) first shell external water molecules, and v) second shell external water molecules [198]. From a drug design perspective, water molecules mediating protein-ligand interactions within active sites deserve special attention. When a ligand enters its target binding site, in

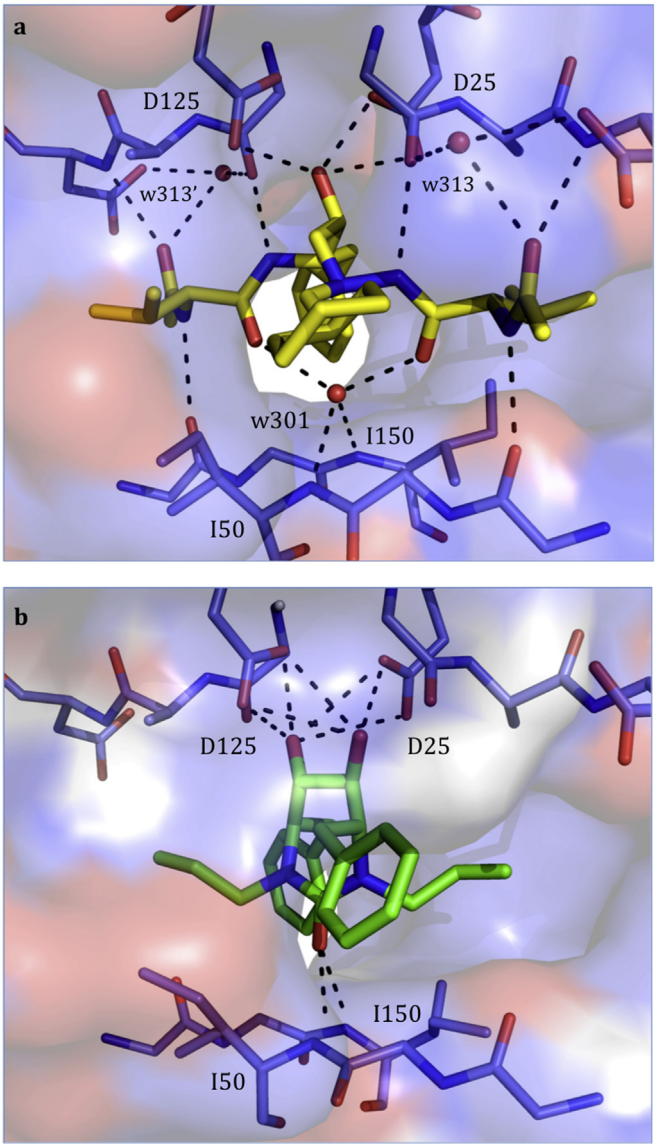

Fig. 3. Example of successful water displacement. a. HIV1 protease complexed with the pseudosymmetric inhibitor CGP53820 (PDB code 1HIH). Wat301 mediating the interaction between the protein and ligand is labeled and represented as a sphere. Two other water molecules stabilizing the interaction at a less extent are highlighted. Key residues are labeled, i.e. the catalytic aspartates and the isoleucines interacting with wat301. b. HIV1 protease complexed with the cyclic urea inhibitor XK216 (PDB code 1HWR). No water molecule is retained within the binding site. In particular wat301 is displaced by the ureic carbonyl group. Key residues are labeled. Ligands and interacting residues are represented in capped stick, hydrogen bonds as black dashed lines. The protein is shown in surface representation.

fact, the surrounding waters, and those filling the cavity need to rearrange. Most of them will be displaced or removed, likely increasing the entropic contribution to the free energy of binding [221–223]. The disruption of water networks, in particular those surrounding the ligand hydrophobic moieties, increases the solvent disorder and stabilizes protein-ligand association [224]. As well, waters filling hydrophobic clefts in the protein will be also favorably released to the solvent, resulting in a positive entropic and enthalpic gain stemming from the additional interactions of the ligand with the residues lining up the binding site [223]. In ligand design, conserved waters in binding sites can be considered either part of the site to be targeted, or displaceable by ligand functional groups, thus providing an additional entropic gain [225,226]. The case of cyclic urea-based HIV-1 inhibitors designed to displace the well-known water301 is remarkable (Fig. 3). In this system water removal positively amplified the hydrophobic interactions and increased the ligand rigidity leading to a smaller loss in ligand entropy upon binding to the receptor [196]. Another known example of successful water displacement is given by the p38a MAPK, where the displacement of a water molecule by a ligand cyano group yielded a 60-fold improvement in the inhibition constant [227]. As well, the replacement of water by a cyano group resulted in an improvement in ligand binding to scyatolone dehydratase [228].

Other examples of the advantage given by water inclusion in pharmacophore [229], docking [230], *de novo* design [225,226], and many others computational approaches were given and are reported in the literature [21,23,197–199,231–243]. Unfortunately, there is no general rule to predict whether a water molecule will be retained upon protein-complex formation, or will be removed [22]. Trying to replace bridging waters led, also, to some unfavorable consequences in binding [244,245], highlighting how challenging is binding energy prediction, and the identification of the different contributions when water molecules are present. Removing a water from a binding site and transferring it to the bulk implies an entropic gain as well as an enthalpic variation depending on the number and strength of contacts made, first, by water with the protein residues and, then, with the surrounding bulk water. Even if the net free energy of water interaction was estimated in the 0–2 kcal/mol range, the individual contribution is highly affected by the specific protein environment, and its calculation is not an easy task [246]. Attention should be also paid to waters tightly bound in crevices on the protein surface or inside the active site. In this case, the entropic gain for displacing them is unfavorably outweighed by the breaking of hydrogen bonds made with the surrounding residues, and by the related enthalpic loss [247]. The displacement of these waters could be not only difficult, but also energetically unfavorable for protein-ligand association [221,248], and their energetic contribution should be carefully considered. The thermodynamics of these phenomena is anything but simple and, up to now, accurate calculations of solvation/desolvation effects are still elusive.

4.1. *Wet or dry binding sites?*

It has been recently reported that compounds interacting with tightly bound bridging waters are as potent as those interacting only with binding site residues [22]. Ligands not stabilized by waters are generally bulkier, more hydrophobic, more complex, and have larger steric effects than those making interactions with structurally conserved water molecules within the target binding site. They also show a higher number of aromatic rings, which decrease their plasticity, their solubility and also cause frequent problems in the subsequent development stages. On the contrary, water-mediated ligands are often smaller, less hydrophobic, less flat, more flexible, and have a better ligand efficiency [249]. More hydrophilic molecules exhibit also better pharmacokinetic properties, are more specific and less promiscuous than more lipophilic compounds [250,251], which might also suffer of molecular obesity [252]. This further demonstrates that tightly bound waters can stabilize a protein-ligand complex acting as a glue or a third partner, and might contribute to improve the specificity of interaction and the ligand pharmacodynamic properties [253,254]. Trying to displace these waters to increase the binding entropy does not *a priori* represent the most promising strategy, although it has occasionally led to positive results [196]. Water displacement might, in fact, have unpredictable consequences on ligand binding affinity, as well as on ligand physicochemical properties,

pharmacokinetics, specificity and safety [255]. In many cases, as previously reported, replacing tightly bound bridging waters only resulted in a slight improvement in the binding energy [244,256–258], or even in a loss of binding affinity [259]. Overall, promoting enthalpic with respect to pre-supposed entropic gain could represent the winning strategy. In this perspective, designing or removing a bridging water should be done only if the subsequent ligand modifications would also improve ligand efficiency, enthalpy, entropy, specificity and pharmacokinetic properties, which will at least have benefits on ligand optimization and development [22].

4.2. How to simulate water in docking and virtual screening

Several attempts to predict the importance of water in structure-based drug design approaches and to evaluate its energetic contributions have been made through the last years. Huang and Shoichet demonstrated that crystallographic waters were able to improve VS enrichment for at least half of the DUD dataset targets [260]. As well, Fukunishi and Nakamura confirmed that including the free energy contribution of bridging water molecules significantly improved the scoring performances [261]. On the contrary, Therrien et al. showed that the impact of waters on virtual screening accuracy was not important as expected, pointing out the need of improving the scoring of key water molecules. The authors applied the FITTED algorithm to investigate a set of 171 proteins representing 40 unique enzymes and demonstrated that, overall, including crystallographic waters, and accounting for protein flexibility improved the enrichment in active compounds, with flexibility effects being more important than water inclusion [24]. They interestingly highlighted that the choice of the contributions to be considered cannot be easily standardized, as it is strongly protein- and system-dependent. Thus, even if the presence of waters and the inclusion of protein flexibility generally improve screening performances, this is not always true. For instance, lead optimization analyses might take advantage of using a rigid structure, while large virtual screening campaigns looking for new different scaffolds should account for protein dynamics. Santos et al. showed that including water molecules in the case of CYP2D6 did not yield significant improvements over docking without retaining any water, and concluded that the effect of water presence strongly depends on the protein conformation and on the ligand identity [262]. Similar conclusions were also drawn by Hritz et al. [144]. The only solution still relies in a deep knowledge of the dataset to be screened, and of the target or of related known proteins, and in an improvement of the currently available scoring functions [21,263,264].

Different programs able to compute the free energy of binding of water molecules and predict their presence in apo (uncomplexed) and holo (complexed with small ligands) protein binding site have been developed and are reported in Table 2.

In 1985 Goodford developed the well-known GRID, to identify favorable interaction sites for probe molecules, being water one of the possible probes [265]. GRID energy was simply calculated by summing Lennard-Jones, electrostatic and hydrogen bonding interactions. Subsequently, in 1996 Raymer and co-workers developed *Consolv* and demonstrated its capability of predicting the conservation of active site water molecules upon ligand binding with 75% accuracy. The algorithm only used the crystallographic temperature factor, the number of hydrogen bonds formed with the protein, the density and the hydrophilicity of neighboring protein atoms [266]. Similar structural properties were implemented by Garcia-Sosa et al. in WaterScore. The authors performed a multivariate logistic regression analysis to identify a statistical correlation between the structural properties of water molecules in the binding site of apo proteins, with the probability of finding them again in the corresponding holo forms [209]. The validation of WaterScore yielded a 67% accuracy.

Kellogg and co-workers combined in a statistically robust method the HINT score and the geometric descriptor Rank, to identify relevant water molecules to be possibly considered in protein-ligand docking and structure-based drug discovery approaches [267]. No other information or parameter was adopted to predict water relevance. The authors showed that conserved waters made at least two hydrogen bonds with the surrounding residues

Table 2
Available software for predicting water displacement and energetic contribution in docking and screening.

Method	Strategy	Website	Ref.
Consolv 1.0	Hybrid k-nearest-neighbors classifier/genetic algorithm	http://www.kuhnlab.bmb.msu.edu/software/consolv/	[266]
WaterScore	Multivariate logistic regression analysis	http://www.cus.cam.ac.uk/	[209]
HINT	HINT (Hydropathic Interactions) score and geometric Rank descriptor	http://www.edusoft-lc.com/hint/	[198]
Water free energy calculation	Double decoupling method, with replica exchange thermodynamic integration in Monte Carlo simulations		[268]
Superstar	Propensity maps	https://www.ccdc.cam.ac.uk/	[279]
Just Add Water Molecules (JAWM)	Statistical thermodynamic analysis based on Monte Carlo simulations		[269]
WaterMap	Statistical thermodynamic analysis of water molecules from explicit solvent MD simulation	http://www.schrodinger.com/WaterMap.php	[280]
Water PMF	Potential of Mean Forces		[271]
WaterFLAP	GRID Molecular Interaction Fields [265]	http://www.moldiscovery.com/software/flap/	[112]
GOLD	Estimation of the free-energy change associated with transferring a water molecule from the bulk solvent to its binding site in a protein-ligand complex	http://www.ccdc.cam.ac.uk/Solutions/GoldSuite/Pages/GOLD.aspx	[241]
DOCK	Flexible-receptor docking method considering displaced water and retained water states with variable water position	http://dock.compbio.ucsf.edu/	[260]
FlexX	Particle concept	http://www.biosolveit.de/FlexX/	[23]
AutoDock	Hydration force field accounting for the entropic and enthalpic contributions of discrete waters to ligand binding	http://autodock.scripps.edu/	[233]
SLIDE	Consolv	http://www.kuhnlab.bmb.msu.edu/software/slide/	[281]
FITTED	Potential energy function to estimate water displacement	http://www.fitted.ca/	[282]
Glide	Statistics about the number of hydrogen bonds formed by polar and apolar groups	http://www.schrodinger.com/Glide/	[278]

and gain 0.6–2.0 kcal/mol binding energy than non-conserved waters [198,199].

Thermodynamic integration and Bayesian statistics were applied by Barillari et al. to calculate the propensity of a water molecule to be displaced. Two classes of waters were identified: those conserved and not displaced by any of the ligands, and those removed upon some ligand entrance, observing that conserved waters are generally more tightly bound than displaced ones. The proposed approach allowed the quantitative identification of waters to be considered and possibly displaced in rational drug design strategies [268].

Michel et al. developed the Just Add Water Molecules (JAWM) procedure, using a double decoupling approach to compare the energetic cost of removing a water molecule from the bulk and from a binding site. The energetic cost is calculated using a scaling parameter Θ, which represents a degree of freedom during Monte Carlo simulation of the solvated protein-ligand complex and modulates intermolecular interactions for a given water molecule [269].

MD-based approaches were exploited to develop WaterMap, based on explicit solvent MD simulations followed by statistical thermodynamic analyses of water clusters based on inhomogeneous solvation theory [270]. Abel et al. demonstrated that water displacement can significantly affect ligand potency and calculated the related binding free energy by combining a water displacement term with complementary terms derived from implicit solvent molecular mechanics calculations [263,270].

Recently Zheng et al. developed the Water PMF (wPMF) approach, using the potential of mean force on 3946 non-redundant high resolution crystal structures. Water PMF was applied to predict the potential hydration sites in protein structures, showing an accuracy of ~80%. Also the algorithm was able to assess whether or not a given water molecule should be targeted for displacement in ligand design. When compared to 3D-RISM, wPMF gave comparable performances with much less computational cost [271].

As the aforementioned methodologies aim to predict *a priori* which water molecules have to be retained in subsequent VS and docking simulations, different softwares also allow an “on the fly” displacement of waters during docking. FLAP (Fingerprint or Ligands and Proteins), based on the GRID Molecular Interaction Fields [112,272,273], allows the user to perform docking simulations automatically calculating the probability of crystallographic or predicted water molecules to be retained upon ligand binding to the protein target [274]. The possibility of including or removing waters during docking was also implemented in DOCK 3.5.54 [260] and in GOLD [241], where a 2 kJ/mol penalty is associated to each water displaced from the binding site. Autodock combines different protein structures, thus considering both flexibility and water contribution at a time, into a single grid where, for each point of the grid, a weighed average potential is calculated. This approach offers the advantage of avoiding steric clashes and allows docking of ligands into cavities occupied by waters without causing any steric clash [233,275]. Consolv has been implemented in SLIDE in order to predict which waters have to be retained during docking simulations. Similarly to Autodock, to each displaced water molecule an energy penalty is associated and deducted from the corresponding ligand score [157]. FlexX applies a particle concept defining *a priori* the possible hydration sites by placing water particles in the pocket. During the fragment-based docking, whenever the fragment clashes with any possible water, that water is removed and the building process continues. Otherwise the water is retained until the next iteration [23]. FITTED predicts the retaining or displacement of waters switching them on and off during the docking process. The choice is guided by scaling factors ranging from 0 to 1 according to the distance between the ligand and the waters. The authors observed a slight improvement in docking performances when waters were retained with respect of using completely unhydrated binding sites [276,277]. Differently, Glide places waters into the binding site after ligands have been docked, filling the holes left between the residues lining the cavity and the ligands. The position of water molecules is predicted by using statistics about the number of hydrogen bonds formed by polar and apolar groups [278].

Even if many different methodologies have been developed and a number of investigations performed, understanding and quantifying water behavior, and the associated energy contributions still remain elusive. From a drug design perspective, there are still two main issues to be solved, i) if waters have to be retained in lead discovery, i.e. docking-based virtual screening, ii) if the displacement of protein-ligand bridging waters in lead optimization could improve binding affinity [243]. Actually, when there is no previous knowledge about the conservation of water molecules within the protein binding site upon ligand binding, waters are normally disregarded in modeling simulations. We also have to consider that retaining waters in a binding site changes the size and the chemical properties of the pocket, thus reducing the chemical diversity of the possible scaffolds. As also suggested by Wong and Lightstone [243], docking simulations could include water contribution by explicitly modeling them or by using implicit models, or a combination of explicit water molecules with implicit solvent rescoring. Lead optimization requires, instead, a more accurate evaluation of the energetic contribution for displacing or retaining waters upon ligand optimization, since displacement of bridging waters does not guarantee stronger binding. Unfortunately, no *a priori* indication is reliable, and only an accurate analysis of the literature and of the system itself can help in making a successful choice.

5. Conclusions

Many steps forward have been made in structure-based drug discovery since the first simulations were performed and reported. A number of algorithms, computer programs and developments have been and are currently developed with the aim of representing biological interactions in the most realistic way. We are perfectly aware that such an objective will probably never be reached. Nevertheless, experimental difficulties, energetic barriers and hardware limits are overcome again and again and, by now, computational simulations accompany almost any experimental research.

In this review we focused on three “structural” critical issues to structure-based virtual screening, demonstrating how *in silico* approaches can help in the accurate representation of biological systems whenever experimental data are missing. First we discussed homology modeling of the binding site, able to provide essential information when experimental structures of the system of interest are missing. In the future, the development of more accurate modeling methods incorporating the wealth of experimental knowledge from biochemical and biophysical data should translate into the discovery of an increasing number of potent, selective and efficient drug leads with high therapeutic value. Either using experimental or modeled structures, a correct representation of the protein dynamics is fundamental to provide reliable results in drug design. We reported the significant improvements made during the last years in considering and predicting target flexibility in computational simulations. As well, a great effort has been poured to consider and predict the role of water molecules within the binding site, as waters considerably increase protein flexibility and versatility. We thus centered on

solvent effects and on the importance of estimating water contribution in lead identification and development to give, again, a proper representation of all the actors playing in biological interactions.

Future developments in these areas, coupled with an enhanced description of the physics underlying macromolecular interaction [232,283–296], will be key for more predictive computational tools in structure-based drug discovery and design. Building the most reliable models is still based, and will probably always be, on the deep knowledge of the system or event under study. We are still far from the final goal: as of today, not any structure can be modeled, and not all binding events can be simulated or energetically quantified. Nevertheless we do believe that there is plenty of space for further improvements and progresses.

Acknowledgments

This work has been supported by the Agencia Nacional de Promoción Científica y Tecnológica, Argentina (PICT-2011-2778 to CNC) and FOCEM-Mercosur (COF 03/11).

References

[1] M. Congreve, C.W. Murray, T.L. Blundell, Drug Discov. Today 10 (2005) 895–907.
[2] D.E. Clark, Expert Opin. Drug Discov. 3 (2008) 841–851.
[3] R.S. Ferreira, A. Simeonov, A. Jadhav, O. Eidam, B.T. Mott, M.J. Keiser, J.H. McKerrow, D.J. Maloney, J.J. Irwin, B.K. Shoichet, J. Med. Chem. 53 (2010) 4891–4905.
[4] H. Koppen, Curr. Opin. Drug Discov. Dev. 12 (2009) 397–407.
[5] C. McInnes, Curr. Opin. Chem. Biol. 11 (2007) 494–502.
[6] V. Zoete, A. Grosdidier, O. Michielin, J. Cell. Mol. Med. 13 (2009) 238–248.
[7] S.S. Phatak, C.C. Stephan, C.N. Cavasotto, Exp. Opin. Drug Discov. 4 (2009) 947–959.
[8] B. Waszkowycz, D.E. Clark, E. Gancia, Wiley Interdiscip. Rev. Comput. Mol. Sci. (2011) 229–259.
[9] E. Yuriev, J. Holien, P.A. Ramsland, J. Mol. Recognit. (2015), http://dx.doi.org/10.1002/jmr.2471.
[10] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucl. Acids Res. 28 (2000) 235–242.
[11] L. Slabinski, L. Jaroszewski, A.P. Rodrigues, L. Rychlewski, I.A. Wilson, S.A. Lesley, A. Godzik, Protein Sci. 16 (2007) 2472–2482.
[12] D. Horvath, Methods Mol. Biol. 672 (2011) 261–298.
[13] E.A. Gatica, C.N. Cavasotto, J. Chem. Inf. Model. 52 (2012) 1–6.
[14] N. Huang, B.K. Shoichet, J.J. Irwin, J. Med. Chem. 49 (2006) 6789–6801.
[15] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, C.R. Corbeil, Br. J. Pharmacol. 153 (Suppl. 1) (2008) S7–S26.
[16] F. Spyrakis, P. Cozzini, G.E. Kellogg, Burger's Medicinal Chemistry, seventh ed., Wiley-VCH, New York, 2010, pp. 601–684.
[17] F. Spyrakis, G.E. Kellogg, A. Amadasi, P. Cozzini, Scoring functions for virtual screening, in: "Frontiers in Drug Design and Discovery", vol. 3, Bentham Science Publishers, Ltd, 2007, pp. 317–379.
[18] C.N. Cavasotto, in: F.J. Luque, X. Barril (Eds.), Physico-Chemical and Computational Approaches to Drug Discovery, Royal Society of Chemistry, London, 2012, pp. 195–222.
[19] A.N. Jain, J. Comput. Aided Mol. Des. 23 (2009) 355–374.
[20] A. May, M. Zacharias, J. Med. Chem. 51 (2008) 3499–3506.
[21] S.B. de Beer, N.P. Vermeulen, C. Oostenbrink, Curr. Top. Med. Chem. 10 (2010) 55–66.
[22] A.T. Garcia-Sosa, J. Chem. Inf. Model. 53 (2013) 1388–1405.
[23] M. Rarey, B. Kramer, T. Lengauer, Proteins Struct. Funct. Genet. 34 (1999) 17–28.
[24] E. Therrien, N. Weill, A. Tomberg, C.R. Corbeil, D. Lee, N. Moitessier, J. Chem. Inf. Model. 54 (2014) 3198–3210.
[25] M. Levitt, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 3183–3188.
[26] K. Lundstrom, J. Cell. Mol. Med. 11 (2007) 224–238.
[27] Nucleic Acids Res. 42 (2014) D191–D198.
[28] S. Costanzi, K. Wang, Adv. Exp. Med. Biol. 796 (2014) 3–13.
[29] J.S. Mason, A. Bortolato, M. Congreve, F.H. Marshall, Trends Pharmacol. Sci. 33 (2012) 249–260.
[30] U. Pieper, N. Eswar, B.M. Webb, D. Eramian, L. Kelly, D.T. Barkan, H. Carter, P. Mankoo, R. Karchin, M.A. Marti-Renom, F.P. Davis, A. Sali, Nucleic Acids Res. 37 (2009) D347–D354.
[31] C.N. Cavasotto, D. Palomba, Chem. Commun. (2015), http://dx.doi.org/10.1039/c5cc05050b.
[32] A. Fiser, Expert Rev. Proteom. 1 (2004) 97–110.
[33] C.N. Cavasotto, S.S. Phatak, Drug Discov. Today 14 (2009) 676–683.
[34] C.N. Cavasotto, Curr. Top. Med. Chem. 11 (2011) 1528–1534.
[35] G.H. Lushington, Methods Mol. Biol. 1215 (2015) 309–330.
[36] S. Costanzi, Methods Mol. Biol. 857 (2012) 259–279.
[37] D. Palomba, C.N. Cavasotto, in: C.N. Cavasotto (Ed.), Silico Drug Discovery and Design: Theory, Methods, Challenges, and Applications, CRC Press, Taylor & Francis Group, 2015, pp. 215–248.
[38] C. Chothia, A.M. Lesk, EMBO J. 5 (1986) 823–826.
[39] A. Hillisch, L.F. Pineda, R. Hilgenfeld, Drug Discov. Today 9 (2004) 659–669.
[40] S. Costanzi, J. Med. Chem. 51 (2008) 2907–2914.
[41] S.S. Phatak, E.A. Gatica, C.N. Cavasotto, J. Chem. Inf. Model. 50 (2010) 2119–2128.
[42] M. Michino, E. Abola, C.L. Brooks 3rd, J.S. Dixon, J. Moult, R.C. Stevens, Nat. Rev. Drug Discov. 8 (2009) 455–463.
[43] C.N. Cavasotto, A.J. Orry, N.J. Murgolo, M.F. Czarniecki, S.A. Kocsi, B.E. Hawes, K.A. O'Neill, H. Hine, M.S. Burton, J.H. Voigt, R.A. Abagyan, M.L. Bayne, F.J. Monsma Jr., J. Med. Chem. 51 (2008) 581–588.
[44] I. Kufareva, M. Rueda, V. Katritch, R.C. Stevens, R. Abagyan, Structure 19 (2011) 1108–1126.
[45] R.C. Edgar, Nucleic Acids Res. 32 (2004) 1792–1797.
[46] K. Katoh, D.M. Standley, Mol. Biol. Evol. 30 (2013) 772–780.
[47] A. Krogh, M. Brown, I.S. Mian, K. Sjolander, D. Haussler, J. Mol. Biol. 235 (1994) 1501–1531.
[48] J. Cheng, BMC Struct. Biol. 8 (2008) 18.
[49] P. Diaz, S.S. Phatak, J. Xu, F.R. Fronczek, F. Astruc-Diaz, C.M. Thompson, C.N. Cavasotto, M. Naguib, ChemMedChem 4 (2009) 1615–1629.
[50] N. Fernandez-Fuentes, B.K. Rai, C.J. Madrid-Aliste, J.E. Fajardo, A. Fiser, Bioinformatics 23 (2007) 2558–2565.
[51] R.R. Petrov, L. Knight, S.R. Chen, J. Wager-Miller, S.W. McDaniel, F. Diaz, F. Barth, H.L. Pan, K. Mackie, C.N. Cavasotto, P. Diaz, Eur. J. Med. Chem. 69 (2013) 881–907.
[52] J.C. Mobarec, R. Sanchez, M. Filizola, J. Med. Chem. 52 (2009) 5207–5216.
[53] I. Kufareva, V. Katritch, R.C. Stevens, R. Abagyan, Structure 22 (2014) 1120–1139.
[54] T. Kenakin, Annu. Rev. Pharmacol. Toxicol. 42 (2002) 349–379.
[55] B.K. Kobilka, X. Deupi, Trends Pharmacol. Sci. 28 (2007) 397–406.
[56] H. Tang, X.S. Wang, J.H. Hsieh, A. Tropsha, Proteins 80 (2012) 1503–1521.
[57] C.N. Cavasotto, N. Singh, Curr. Comput. Aided Drug Des. 4 (2008) 221–234.
[58] A. Sali, T.L. Blundell, J. Mol. Biol. 234 (1993) 779–815.
[59] A. Evers, G. Klebe, Angew. Chem. Int. Ed. Engl. 43 (2004) 248–251.
[60] A. Evers, G. Klebe, J. Med. Chem. 47 (2004) 5381–5392.
[61] A.J.W. Orry, C.N. Cavasotto, Ligand-docking-based homology model of the Melanin-Concentrating Hormone 1 receptor, in: 231st Meeting of the American Chemical Society, Atlanta, GA, 2006.
[62] C.N. Cavasotto, R.A. Abagyan, J. Mol. Biol. 337 (2004) 209–225.
[63] C.N. Cavasotto, J.A. Kovacs, R.A. Abagyan, J. Am. Chem. Soc. 127 (2005) 9632–9640.
[64] M.C. Monti, A. Casapullo, C.N. Cavasotto, A. Napolitano, R. Riccio, ChemBioChem 8 (2007) 1585–1591.
[65] M.C. Monti, A. Casapullo, C.N. Cavasotto, A. Tosco, F. Dal Piaz, A. Ziemys, L. Margarucci, R. Riccio, Chem. Eur. J. 15 (2009) 1155–1163.
[66] J.A. Kovacs, C.N. Cavasotto, R.A. Abagyan, J. Comp. Theor. Nanosci. 2 (2005) 354–361.
[67] M. Rossi, B. Rotblat, K. Ansell, I. Amelio, M. Caraglia, G. Misso, F. Bernassola, C.N. Cavasotto, R.A. Knight, A. Ciechanover, G. Melino, Cell Death Dis. 5 (2014) e1203.
[68] P. Diaz, S.S. Phatak, J. Xu, F. Astruc-Diaz, C.N. Cavasotto, M. Naguib, J. Med. Chem. 52 (2009) 433–444.
[69] S. Moro, F. Deflorian, M. Bacilieri, G. Spalluto, Curr. Pharm. Des. 12 (2006) 2175–2185.
[70] W. Sherman, T. Day, M.P. Jacobson, R.A. Friesner, R. Farid, J. Med. Chem. 49 (2006) 534–553.
[71] F.M. McRobb, B. Capuano, I.T. Crosby, D.K. Chalmers, E. Yuriev, J. Chem. Inf. Model. 50 (2010) 626–637.
[72] C. Obiol-Pardo, L. Lopez, M. Pastor, J. Selent, Proteins 79 (2011) 1695–1703.
[73] J.A. Dalton, R.M. Jackson, J. Mol. Biol. 399 (2010) 645–661.
[74] T. Thomas, K.C. McLean, F.M. McRobb, D.T. Manallack, D.K. Chalmers, E. Yuriev, J. Chem. Inf. Model. 54 (2014) 243–253.
[75] S.P. Chin, M.J. Buckle, D.K. Chalmers, E. Yuriev, S.W. Doughty, J. Mol. Graph. Model. 49 (2014) 91–98.
[76] D. Pala, T. Beuming, W. Sherman, A. Lodola, S. Rivara, M. Mor, J. Chem. Inf. Model. 53 (2013) 821–835.
[77] M. Kolaczkowski, A. Bucki, M. Feder, M. Pawlowski, J. Chem. Inf. Model. 53 (2013) 638–648.
[78] R. Singh, M.E. Sobhia, J. Biomol. Struct. Dyn. 31 (2013) 694–715.
[79] X. Lin, X.P. Huang, G. Chen, R. Whaley, S. Peng, Y. Wang, G. Zhang, S.X. Wang, S. Wang, B.L. Roth, N. Huang, J. Med. Chem. 55 (2012) 5749–5759.
[80] G.L. Warren, C.W. Andrews, A.M. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head, J. Med. Chem. 49 (2006) 5912–5931.
[81] M. Brylinski, J. Chem. Inf. Model. 53 (2013) 3097–3112.
[82] H. Fan, J.J. Irwin, B.M. Webb, G. Klebe, B.K. Shoichet, A. Sali, J. Chem. Inf. Model. 49 (2009) 2512–2527.
[83] E.M. Novoa, L. Ribas de Pouplana, X. Barril, M. Orozco, J. Chem. Theory Comput. 6 (2010) 2547–2557.
[84] A. Bordogna, A. Pandini, L. Bonati, J. Comput. Chem. 32 (2011) 81–98.
[85] E.D. Nguyen, C. Norn, T.M. Frimurer, J. Meiler, PLoS One 8 (2013) e67302.

[86] S.A. Combs, S.L. Deluca, S.H. Deluca, G.H. Lemmon, D.P. Nannemann, E.D. Nguyen, J.R. Willis, J.H. Sheehan, J. Meiler, Nat. Protoc. 8 (2013) 1277–1298.
[87] G. Lemmon, J. Meiler, Methods Mol. Biol. 819 (2012) 143–155.
[88] J. Moult, K. Fidelis, A. Kryshtafovych, B. Rost, A. Tramontano, Proteins 77 (Suppl. 9) (2009) 1–4.
[89] K.W. Kaufmann, J. Meiler, PLoS One 7 (2012) e50769.
[90] T. Beuming, W. Sherman, J. Chem. Inf. Model. 52 (2012) 3263–3277.
[91] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, J. Comp. Chem. 19 (1998) 1639–1662.
[92] K. Engels, C. Beyer, M.L. Suarez Fernandez, F. Bender, M. Gassel, G. Unden, R.J. Marhofer, J.C. Mottram, P.M. Selzer, ChemMedChem 5 (2010) 1259–1271.
[93] MOE, Chemical Computing Group, Inc, Canada, Montreal, 2012.
[94] T. Schwede, J. Kopp, N. Guex, M.C. Peitsch, Nucleic Acids Res. 31 (2003) 3381–3385.
[95] E.C. Meng, B.K. Shoichet, I.D. Kuntz, J. Comput. Chem. 13 (1992) 505–524.
[96] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, J. Mol. Biol. 261 (1996) 470–489.
[97] J.J. Irwin, B.K. Shoichet, J. Chem. Inf. Model. 45 (2005) 177–182.
[98] M.T. Sisay, T. Steinmetzer, M. Stirnberg, E. Maurer, M. Hammami, J. Bajorath, M. Gutschow, J. Med. Chem. 53 (2010) 5523–5535.
[99] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, J. Mol. Biol. 267 (1997) 727–748.
[100] Q. Zhang, D. Li, P. Wei, J. Zhang, J. Wan, Y. Ren, Z. Chen, D. Liu, Z. Yu, L. Feng, J. Chem. Inf. Model. 50 (2010) 317–325.
[101] J. Mortier, R. Frederick, C. Ganeff, C. Remouchamps, P. Talaga, L. Pochet, J. Wouters, J. Piette, E. Dejardin, B. Masereel, Biochem. Pharmacol. 79 (2010) 1462–1472.
[102] C. Lambert, N. Leonard, X. De Bolle, E. Depiereux, Bioinformatics 18 (2002) 1250–1256.
[103] J. Mortier, B. Masereel, C. Remouchamps, C. Ganeff, J. Piette, R. Frederick, Bioorg. Med. Chem. Lett. 20 (2010) 4515–4520.
[104] T.F. Gonzalez-Cestari, B.J. Henderson, R.E. Pavlovicz, S.B. McKay, R.A. El-Hajj, A.B. Pulipaka, C.M. Orac, D.D. Reed, R.T. Boyd, M.X. Zhu, C. Li, S.C. Bergmeier, D.B. McKay, J. Pharmacol. Exp. Ther. 328 (2009) 504–515.
[105] Glide, Schrödinger, LLC, New York, NY, 2009.
[106] K.V. Mahasenan, R.E. Pavlovicz, B.J. Henderson, T.F. Gonzalez-Cestari, B. Yi, D.B. McKay, C. Li, ACS Med. Chem. Lett. 2 (2011) 855–860.
[107] S. Li, Y. Zhou, W. Lu, Y. Zhong, W. Song, K. Liu, J. Huang, Z. Zhao, Y. Xu, X. Liu, H. Li, J. Chem. Inf. Model. 51 (2011) 2939–2947.
[108] H. Park, O. Chi, J. Kim, S. Hong, J. Chem. Inf. Model. 51 (2011) 2986–2993.
[109] N. Huang, C. Kalyanaraman, K. Bernacki, M.P. Jacobson, Phys. Chem. Chem. Phys. 8 (2006) 5166–5177.
[110] N. Huang, C. Kalyanaraman, J.J. Irwin, M.P. Jacobson, J. Chem. Inf. Model. 46 (2006) 243–253.
[111] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, Nucleic Acids Res. 34 (2006) D668–D672.
[112] M. Baroni, G. Cruciani, S. Sciabola, F. Perruccio, J.S. Mason, J. Chem. Inf. Model. 47 (2007) 279–294.
[113] F. Sirci, E.P. Istyastono, H.F. Vischer, A.J. Kooistra, S. Nijmeijer, M. Kuijer, M. Wijtmans, R. Mannhold, R. Leurs, I.J. de Esch, C. de Graaf, J. Chem. Inf. Model. 52 (2012) 3308–3324.
[114] C. Diaz, P. Leplatois, P. Angelloz-Nicoud, M. Lecomte, A. Josse, M. Delpech, F. Pecceu, G. Loison, D. Shire, M. Pascal, P. Ferrara, E. Ferran, Mol. Inf. 30 (2011) 345–358.
[115] C. Diaz, C. Labit-Le Bouteiller, S. Yvon, A. Cambon-Kernëis, A. Roasio, M.-F. Jamme, A. Aries, C. Feuillerat, E. Perret, F. Guette, P. Dieu, B. Miloux, D. Albène, N. Hasel, M. Kaghad, E. Ferran, J. Lupker, P. Ferrara, Mol. Inf. 32 (2013) 213–229.
[116] M. Okamoto, K. Takayama, T. Shimizu, K. Ishida, O. Takahashi, T. Furuya, J. Med. Chem. 52 (2009) 7323–7327.
[117] Y. Yoshikawa, S. Oishi, T. Kubo, N. Tanahara, N. Fujii, T. Furuya, J. Med. Chem. 56 (2013) 4236–4251.
[118] MacroModel, Schrödinger, LLC, New York, NY, 2012.
[119] M. Vass, E. Schmidt, F. Horti, G.M. Keseru, Eur. J. Med. Chem. 77 (2014) 38–46.
[120] Schrödinger Suite, Protein Preparation Wizard, Epik Version 2.3, Impact Version 5.8, Prime Version 3.1, Schrödinger, LLC, New York, NY, 2013.
[121] M. Vass, E. Agai-Csongor, F. Horti, G.M. Keseru, ACS Med. Chem. Lett. 5 (2014) 1010–1014.
[122] I-TASSER Suite, http://zhanglab.ccmb.med.umich.edu/I-TASSER/, 2012.
[123] K.W. Kastner, D.A. Shoue, G.L. Estiu, J. Wolford, M.F. Fuerst, L.D. Markley, J.A. Izaguirre, M.A. McDowell, Malar. J. 13 (2014) 434.
[124] H. Park, S.Y. Park, S.W. Nam, S.E. Ryu, J. Biomol. Screen 19 (2014) 1383–1390.
[125] P. Cozzini, G.E. Kellogg, F. Spyrakis, D.J. Abraham, G. Costantino, A. Emerson, F. Fanelli, H. Gohlke, L.A. Kuhn, G.M. Morris, M. Orozco, T.A. Pertinhez, M. Rizzi, C.A. Sotriffer, J. Med. Chem. 51 (2008) 6237–6255.
[126] A. Henzler, M. Rarey, Mol. Inf. 29 (2010) 164–173.
[127] C. Rao, J. Subramanian, S.D. Sharma, Drug Discov. Today 14 (2009) 394–400.
[128] S.J. Teague, Nat. Rev. Drug Discov. 2 (2003) 527–541.
[129] M.L. Teodoro, L.E. Kavraki, Curr. Pharm. Des. 9 (2003) 1635–1648.
[130] E. Fisher, Berichte Dtsch. Chem. Ges. Berl. 27 (1894) 2985–2993.
[131] D.D. Boehr, R. Nussinov, P.E. Wright, Nat. Chem. Biol. 5 (2009) 789–796.
[132] B. Ma, M. Shatsky, H.J. Wolfson, R. Nussinov, Protein Sci. 11 (2002) 184–197.
[133] C.J. Tsai, S. Kumar, B. Ma, R. Nussinov, Protein Sci. 8 (1999) 1181–1190.
[134] H.R. Bosshard, News Physiol. Sci. 16 (2001) 171–173.
[135] A.M. Ferrari, B.Q. Wei, L. Costantino, B.K. Shoichet, J. Med. Chem. 47 (2004) 5076–5084.
[136] V. Kairys, M.K. Gilson, J. Comput. Chem. 23 (2002) 1656–1670.
[137] S.B. Nabuurs, M. Wagener, J. de Vlieg, J. Med. Chem. 50 (2007) 6507–6518.
[138] C.N. Cavasotto, A.J. Orry, R.A. Abagyan, Proteins 51 (2003) 423–433.
[139] N. Moitessier, C. Henry, B. Maigret, Y. Chapleur, J. Med. Chem. 47 (2004) 4178–4187.
[140] X. Barril, S.D. Morley, J. Med. Chem. 48 (2005) 4432–4443.
[141] I.R. Craig, J.W. Essex, K. Spiegel, J. Chem. Inf. Model. 50 (2010) 511–524.
[142] M. Rueda, G. Bottegoni, R. Abagyan, J. Chem. Inf. Model. 50 (2010) 186–193.
[143] K.L. Damm, H.A. Carlson, J. Am. Chem. Soc. 129 (2007) 8225–8235.
[144] J. Hritz, A. de Ruiter, C. Oostenbrink, J. Med. Chem. 51 (2008) 7469–7477.
[145] E.S. Bolstad, A.C. Anderson, Proteins 73 (2008) 566–580.
[146] J.H. Lin, A.L. Perryman, J.R. Schames, J.A. McCammon, J. Am. Chem. Soc. 124 (2002) 5632–5633.
[147] L.S. Cheng, R.E. Amaro, D. Xu, W.W. Li, P.W. Arzberger, J.A. McCammon, J. Med. Chem. 51 (2008) 3878–3894.
[148] I.H. Park, C. Li, J. Phys. Chem. B 114 (2010) 5144–5153.
[149] C.N. Cavasotto, Methods Mol. Biol. 819 (2012) 157–168.
[150] S. Vilar, G. Ferino, S.S. Phatak, B. Berk, C.N. Cavasotto, S. Costanzi, J. Mol. Graph. Model. 29 (2011) 614–623.
[151] P. Zavodszky, J. Kardos, G.A. Petsko Svingor, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 7406–7411.
[152] F. Jiang, S.H. Kim, J. Mol. Biol. 219 (1991) 79–102.
[153] M.I. Zavodszky, L.A. Kuhn, Protein Sci. 14 (2005) 1104–1114.
[154] C. Beier, M. Zacharias, Expert Opin. Drug Discov. 5 (2010) 347–359.
[155] A.R. Leach, J. Mol. Biol. 235 (1994) 345–356.
[156] S.F. Sousa, A.J. Ribeiro, J.T. Coimbra, R.P. Neves, S.A. Martins, N.S. Moorthy, P.A. Fernandes, M.J. Ramos, Curr. Med. Chem. 20 (2013) 2296–2314.
[157] V. Schnecke, L.A. Kuhn, Perspect. Drug Discov. Des. 20 (2000) 171–190.
[158] I.W. Davis, D. Baker, J. Mol. Biol. 385 (2009) 381–392.
[159] W.H. Shin, J.K. Kim, D.S. Kim, C. Seok, J. Comput. Chem. 34 (2013) 2647–2656.
[160] E.M. Cohen, K.S. Machado, M. Cohen, O.N. de Souza, BMC Genomics 12 (Suppl 4) (2011) S7.
[161] J. Koska, V.Z. Spassov, A.J. Maynard, L. Yan, N. Austin, P.K. Flook, C.M. Venkatachalam, J. Chem. Inf. Model. 48 (2008) 1965–1973.
[162] B.L. de Groot, D.M. van Aalten, R.M. Scheek, A. Amadei, G. Vriend, H.J. Berendsen, Proteins 29 (1997) 240–251.
[163] D. Seeliger, J. Haas, B.L. de Groot, Structure 15 (2007) 1482–1492.
[164] D. Spoel, E. Lindaul, B. Hess, G. Groenhof, A. Mark, H. Berendsen, J. Comput. Chem. 26 (2005) 1701–1718.
[165] F. Spyrakis, P. Benedetti, S. Decherchi, W. Rocchia, A. Cavalli, S. Alcaro, F. Ortuso, M. Baroni, G. Cruciani, submitted (2015).
[166] L.A. Ramirez-Duran, M.C. Rosales-Hernandez, M. Hernandez-Rodriguez, J.E. Mendieta-Wejebe, J. Trujillo-Ferrara, J. Correa-Basurto, Curr. Pharm. Des. 19 (2013) 2204–2215.
[167] L. Salmon, G. Bouvignies, P. Markwick, N. Lakomek, S. Showalter, D.W. Li, K. Walter, C. Griesinger, R. Bruschweiler, M. Blackledge, Angew. Chem. Int. Ed. Engl. 48 (2009) 4154–4157.
[168] B.J. Grant, S. Lukman, H.J. Hocker, J. Sayyah, J.H. Brown, J.A. McCammon, A.A. Gorfe, PLoS One 6 (2011) e25711.
[169] A. Ivetac, J.A. McCammon, Chem. Biol. Drug Des. 76 (2010) 201–217.
[170] N. Brooijmans, I.D. Kuntz, Annu. Rev. Biophys. Biomol. Struct. 32 (2003) 335–373.
[171] S. Alcaro, A. Artese, F. Ceccherini-Silberstein, F. Ortuso, C.F. Perno, T. Sing, V. Svicher, J. Chem. Inf. Model. 49 (2009) 1751–1761.
[172] B.L. Dong, Q.H. Liao, J. Wei, J. Mol. Model. 17 (2011) 1727–1734.
[173] D. Hamelberg, J. Mongan, J.A. McCammon, J. Chem. Phys. 120 (2004) 11919–11929.
[174] L.C. Pierce, P.R. Markwick, J.A. McCammon, N.L. Doltsinis, J. Chem. Phys. 134 (2011) 174107.
[175] A. Laio, F.L. Gervasio, Rep. Prog. Phys. 71 (2008) 126601.
[176] A. Laio, M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 12562–12566.
[177] A. Barducci, M. Bonomi, M. Parrinello, Wiley Interdiscip. Rev. Comput. Mol. Sci. 1 (2011) 826–843.
[178] V. Leone, F. Marinelli, P. Carloni, M. Parrinello, Curr. Opin. Struct. Biol. 20 (2010) 148–154.
[179] X. Cheng, G. Cui, V. Hornak, C. Simmerling, J. Phys. Chem. B 109 (2005) 8220–8230.
[180] P. Liu, B. Kim, R.A. Friesner, B.J. Berne, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 13749–13754.
[181] H. Kokubo, T. Tanaka, Y. Okamoto, J. Comput. Chem. 34 (2013) 2601–2614.
[182] Y. Okamoto, H. Kokubo, T. Tanaka, Adv. Protein Chem. Struct. Biol. 92 (2013) 63–91.
[183] Y. Sugita, A. Kitao, Y. Okamoto, J. Chem. Phys. 113 (2000) 6042–6051.
[184] F. Fukunishi, O. Watanabe, S. Takada, J. Chem. Phys. 116 (2002) 9058–9067.
[185] G. Torrie, J. Valleau, J. Comput. Phys. 23 (1997) 187–199.
[186] G. Patey, J. Valleau, J. Chem. Phys. 64 (1975) 2334–2339.
[187] G. Torrie, J. Valleau, Chem. Phys. Lett. 28 (1974) 578–581.
[188] W. Sinko, S. Lindert, J.A. McCammon, Chem. Biol. Drug Des. 81 (2013) 41–49.
[189] F. Xu, H. Wu, V. Katritch, G.W. Han, K.A. Jacobson, Z.G. Gao, V. Cherezov, R.C. Stevens, Science 332 (2011) 322–327.
[190] O. Sperandio, L. Mouawad, E. Pinto, B.O. Villoutreix, D. Perahia, M.A. Miteva, Eur. Biophys. J. 39 (2010) 1365–1372.

[191] L. Birch, C.W. Murray, M.J. Hartshorn, I.J. Tickle, M.L. Verdonk, J. Comput. Aided Mol. Des. 16 (2002) 855–869.
[192] M.L. Verdonk, P.N. Mortenson, R.J. Hall, M.J. Hartshorn, C.W. Murray, J. Chem. Inf. Model. 48 (2008) 2214–2225.
[193] G. Bottegoni, W. Rocchia, M. Rueda, R. Abagyan, A. Cavalli, PLoS One 6 (2011) e18845.
[194] O. Korb, T.S. Olsson, S.J. Bowden, R.J. Hall, M.L. Verdonk, J.W. Liebeschuetz, J.C. Cole, J. Chem. Inf. Model. 52 (2012) 1262–1274.
[195] A. Bortolato, B.G. Tehan, M.S. Bodnarchuk, J.W. Essex, J.S. Mason, J. Chem. Inf. Model. 53 (2013) 1700–1713.
[196] P.Y. Lam, P.K. Jadhav, C.J. Eyermann, C.N. Hodge, Y. Ru, L.T. Bacheler, J.L. Meek, M.J. Otto, M.M. Rayner, Y.N. Wong, et al., Science 263 (1994) 380–384.
[197] M.H. Ahmed, M. Habtemariam, M.K. Safo, J.N. Scarsdale, F. Spyrakis, P. Cozzini, A. Mozzarelli, G.E. Kellogg, Comput. Biol. Chem. 47 (2013) 126–141.
[198] A. Amadasi, F. Spyrakis, P. Cozzini, D.J. Abraham, G.E. Kellogg, A. Mozzarelli, J. Mol. Biol. 358 (2006) 289–309.
[199] A. Amadasi, J.A. Surface, F. Spyrakis, P. Cozzini, A. Mozzarelli, G.E. Kellogg, J. Med. Chem. 51 (2008) 1063–1067.
[200] U. Bren, d. Janezic, J. Chem. Phys. 137 (2012) 024108.
[201] S. Genheden, P. Mikulskis, L. Hu, J. Kongsted, P. Soderhjelm, U. Ryde, J. Am. Chem. Soc. 133 (2011) 13081–13092.
[202] P. Setny, R. Baron, J.A. McCammon, J. Chem. Theory Comput. 6 (2010) 2866–2871.
[203] C.M. Stegmann, D. Seeliger, G.M. Sheldrick, B.L. de Groot, M.C. Wahl, Angew. Chem. Int. Ed. Engl. 48 (2009) 5207–5210.
[204] J.R. Tame, S.H. Sleigh, A.J. Wilkinson, J.E. Ladbury, Nat. Struct. Biol. 3 (1996) 998–1001.
[205] L. Wang, B.J. Berne, R.A. Friesner, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 1326–1330.
[206] A. Biela, N.N. Nasief, M. Betz, A. Heine, D. Hangauer, G. Klebe, Angew. Chem. Int. Ed. Engl. 52 (2013) 1822–1828.
[207] P.W. Snyder, J. Mecinovic, D.T. Moustakas, S.W. Thomas 3rd, M. Harder, E.T. Mack, M.R. Lockett, A. Heroux, W. Sherman, G.M. Whitesides, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 17889–17894.
[208] R. Lemieux, Acc. Chem. Res. 29 (1996) 373–380.
[209] A.T. Garcia-Sosa, R.L. Mancera, P.M. Dean, J. Mol. Model. 9 (2003) 172–182.
[210] Y. Lu, R. Wang, C.Y. Yang, S. Wang, J. Chem. Inf. Model. 47 (2007) 668–675.
[211] D. Bashford, D.A. Case, Annu. Rev. Phys. Chem. 51 (2000) 129–152.
[212] S. Edinger, C. Cortis, P. Shenkin, R.A. Friesner, J. Phys. Chem. B 101 (1997) 1190–1197.
[213] A. Ghosh, C. Rapp, R.A. Friesner, J. Phys. Chem. B 102 (1998) 10983–10990.
[214] V. Tsui, D.A. Case, Biopolymers 56 (2000) 275–291.
[215] S. Amira, D. Spangberg, K. Hermansson, Chem. Phys. 303 (2004) 327–334.
[216] H.W. Horn, W.C. Swope, J.W. Pitera, J.D. Madura, T.J. Dick, G.L. Hura, T. Head-Gordon, J. Chem. Phys. 120 (2004) 9665–9678.
[217] M. Mahoney, W.L. Jorgensen, J. Chem. Phys. 112 (2000) 8910–8922.
[218] P. Mark, L. Nilsson, J. Phys. Chem. A 105 (2001) 9954–9960.
[219] D. Borgis, A. Staib, Chem. Phys. Lett. 238 (1995) 187–192.
[220] J. Lobaugh, G. Voth, J. Chem. Phys. 106 (1997) 2400–2410.
[221] Z. Li, T. Lazaridis, J. Phys. Chem. B 109 (2005) 662–670.
[222] C.N. Nguyen, T.K. Young, M.K. Gilson, J. Chem. Phys. 137 (2012) 044101.
[223] S. Roy, B. Bagchi, J. Phys. Chem. B 116 (2012) 2958–2968.
[224] T. Head Gordon, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 8308–8312.
[225] A.T. Garcia-Sosa, S. Firth-Clark, R.L. Mancera, J. Chem. Inf. Model 45 (2005) 624–633.
[226] A.T. Garcia-Sosa, R.L. Mancera, J. Mol. Model. 12 (2006) 422–431.
[227] C. Liu, S.T. Wrobleski, J. Lin, G. Ahmed, A. Metzger, J. Wityak, K.M. Gillooly, D.J. Shuster, K.W. McIntyre, S. Pitt, D.R. Shen, R.F. Zhang, H. Zhang, A.M. Doweyko, D. Diller, I. Henderson, J.C. Barrish, J.H. Dodd, G.L. Schieven, K. Leftheris, J. Med. Chem. 48 (2005) 6261–6270.
[228] J.M. Chen, S.L. Xu, Z. Wawrzak, G.S. Basarab, D.B. Jordan, Biochemistry 37 (1998) 17735–17744.
[229] D.G. Lloyd, A.T. Garcia-Sosa, I.L. Alberts, N.P. Todorov, R.L. Manceral, J. Comput. Aided Mol. Des. 18 (2004) 89–100.
[230] A.T. Garcia-Sosa, S. Sild, K. Takkis, U. Maran, J. Chem. Inf. Model. 51 (2011) 2595–2611.
[231] M.H. Ahmed, F. Spyrakis, P. Cozzini, P.K. Tripathi, A. Mozzarelli, J.N. Scarsdale, M.A. Safo, G.E. Kellogg, PLoS One 6 (2011) e24712.
[232] B. de Courcy, J.P. Piquemal, C. Garbay, N. Gresh, J. Am. Chem. Soc. 132 (2010) 3312–3320.
[233] S. Forli, A.J. Olson, J. Med. Chem. 55 (2012) 623–638.
[234] D. Hamelberg, J.A. McCammon, J. Am. Chem. Soc. 126 (2004) 7683–7689.
[235] D.J. Huggins, B. Tidor, Protein Eng. Des. Sel. 24 (2011) 777–789.
[236] G.E. Kellogg, M. Fornabaio, F. Spyrakis, A. Lodola, P. Cozzini, A. Mozzarelli, D.J. Abraham, J. Mol. Graph. Model. 22 (2004) 479–486.
[237] J. Michel, J. Tirado-Rives, W.L. Jorgensen, J. Am. Chem. Soc. 131 (2009) 15403–15411.
[238] P. Monecke, T. Borosch, J. Brickmann, S.M. Kast, Biophys. J. 90 (2006) 841–850.
[239] G. Rossato, B. Ernst, A. Vedani, M. Smiesko, J. Chem. Inf. Model. 51 (2011) 1867–1881.
[240] F. Spyrakis, P. Cozzini, C. Bertoli, A. Marabotti, G.E. Kellogg, A. Mozzarelli, BMC Struct. Biol. 7 (2007) 4.
[241] M.L. Verdonk, G. Chessari, J.C. Cole, M.J. Hartshorn, C.W. Murray, J.W. Nissink, R.D. Taylor, R. Taylor, J. Med. Chem. 48 (2005) 6504–6515.
[242] O. Villacanas, S. Madurga, E. Giralt, I. Belda, Curr. Comput. Aided Drug Des. 5 (2009) 145–154.
[243] S.E. Wong, F.C. Lightstone, Expert Opin. Drug Discov. 6 (2011) 65–74.
[244] V. Mikol, C. Papageorgiou, X. Borer, J. Med. Chem. 38 (1995) 3361–3367.
[245] A. Wissner, D.M. Berger, D.H. Boschelli, M.B. Floyd Jr., L.M. Greenberger, B.C. Gruber, B.D. Johnson, N. Mamuya, R. Nilakantan, M.F. Reich, R. Shen, H.R. Tsou, E. Upeslacis, Y.F. Wang, B. Wu, F. Ye, N. Zhang, J. Med. Chem. 43 (2000) 3244–3256.
[246] J.D. Dunitz, Science 264 (1994) 670.
[247] V. Lounnas, T. Ritschel, J. Kelder, R. McGuire, R.P. Bywater, N. Foloppe, Comput. Struct. Biotechnol. J. 5 (2013) e201302011.
[248] Z. Li, T. Lazaridis, J. Phys. Chem. B 110 (2006) 1464–1475.
[249] F. Lovering, J. Bikker, C. Humblet, J. Med. Chem. 52 (2009) 6752–6756.
[250] Y. Chen, B.K. Shoichet, Nat. Chem. Biol. 5 (2009) 358–364.
[251] P.D. Leeson, B. Springthorpe, Nat. Rev. Drug Discov. 6 (2007) 881–890.
[252] M.M. Hann, G.M. Keseru, Nat. Rev. Drug Discov. 11 (2012) 355–365.
[253] B.A. Katz, C. Luong, J.D. Ho, J.R. Somoza, E. Gjerstad, J. Tang, S.R. Williams, E. Verner, R.L. Mackman, W.B. Young, P.A. Sprengeler, H. Chan, K. Mortara, J.W. Janc, M.E. McGrath, J. Mol. Biol. 344 (2004) 527–547.
[254] C.P. Swaminathan, A. Nandi, S.S. Visweswariah, A. Surolia, J. Biol. Chem. 274 (1999) 31272–31278.
[255] A.T. Garcia-Sosa, U. Maran, Sar. QSAR Environ. Res. 24 (2013) 319–331.
[256] M. Andaloussi, L.M. Henriksson, A. Wieckowska, M. Lindh, C. Bjorkelid, A.M. Larsson, S. Suresh, H. Iyer, B.R. Srinivasa, T. Bergfors, T. Unge, S.L. Mowbray, M. Larhed, T.A. Jones, A. Karlen, J. Med. Chem. 54 (2011) 4964–4976.
[257] J.L. Ekstrom, T.A. Pauly, M.D. Carty, W.C. Soeller, J. Culp, D.E. Danley, D.J. Hoover, J.L. Treadway, E.M. Gibbs, R.J. Fletterick, Y.S. Day, D.G. Myszka, V.L. Rath, Chem. Biol. 9 (2002) 915–924.
[258] J. Seo, J. Igarashi, H. Li, P. Martasek, L.J. Roman, T.L. Poulos, R.B. Silverman, J. Med. Chem. 50 (2007) 2089–2099.
[259] J.I. Trujillo, J.R. Kiefer, W. Huang, J.E. Day, J. Moon, G.M. Jerome, C.P. Bono, C.M. Kornmeier, M.L. Williams, C. Kuhn, G.R. Rennie, T.A. Wynn, C.P. Carron, A. Thorarensen, Bioorg. Med. Chem. Lett. 22 (2012) 3795–3799.
[260] N. Huang, B.K. Shoichet, J. Med. Chem. 51 (2008) 4862–4865.
[261] Y. Fukunishi, H. Nakamura, Pharm. (Basel) 6 (2013) 604–622.
[262] R. Santos, J. Hritz, C. Oostenbrink, J. Chem. Inf. Model. 50 (2010) 146–154.
[263] C.R. Guimaraes, A.M. Mathiowetz, J. Chem. Inf. Model. 50 (2010) 547–559.
[264] D. Nikolíc, N. Blinov, D. Wishart, A. Kovalenko, J. Chem. Theory Comput. 8 (2012) 3356–3372.
[265] P.J. Goodford, J. Med. Chem. 28 (1985) 849–857.
[266] M.L. Raymer, P.C. Sanschagrin, W.F. Punch, S. Venkataraman, E.D. Goodman, L.A. Kuhn, J. Mol. Biol. 265 (1997) 445–464.
[267] P. Cozzini, M. Fornabaio, A. Mozzarelli, F. Spyrakis, G.E. Kellogg, D.J. Abraham, J. Quantum Chem. 106 (2006) 647–651.
[268] C. Barillari, J. Taylor, R. Viner, J.W. Essex, J. Am. Chem. Soc. 129 (2007) 2577–2587.
[269] J. Michel, J. Tirado-Rives, W.L. Jorgensen, J. Phys. Chem. B 113 (2009) 13337–13346.
[270] R. Abel, N.K. Salam, J. Shelley, R. Farid, R.A. Friesner, W. Sherman, ChemMedChem 6 (2011) 1049–1066.
[271] M. Zheng, Y. Li, B. Xiong, H. Jiang, J. Shen, J. Comput. Chem. 34 (2013) 583–592.
[272] F. Spyrakis, B. Cellini, S. Bruno, P. Benedetti, E. Carosati, G. Cruciani, F. Micheli, A. Felici, P. Cozzini, G.E. Kellogg, C.B. Voltattorni, A. Mozzarelli, ChemMedChem 9 (2014) 1501–1511.
[273] F. Spyrakis, R. Singh, P. Cozzini, B. Campanini, E. Salsi, P. Felici, S. Raboni, P. Benedetti, G. Cruciani, G.E. Kellogg, P.F. Cook, A. Mozzarelli, PLoS One 8 (2013) e77558.
[274] www.moldiscovery.com
[275] F. Osterberg, G.M. Morris, M.F. Sanner, A.J. Olson, D.S. Goodsell, Proteins Struct. Funct. Genet. 46 (2002) 34–40.
[276] C.R. Corbeil, P. Englebienne, N. Moitessier, J. Chem. Inf. Model 47 (2007) 435–449.
[277] C.R. Corbeil, N. Moitessier, J. Chem. Inf. Model. 49 (2009) 997–1009.
[278] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, J. Med. Chem. 47 (2004) 1739–1749.
[279] M.L. Verdonk, J.C. Cole, R. Taylor, J. Mol. Biol. 289 (1999) 1093–1108.
[280] T. Young, R. Abel, B. Kim, B.J. Berne, R.A. Friesner, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 808–813.
[281] V. Schnecke, L. Kuhn, Perspect. Drug Discov. Des. 20 (2000) 171–190.
[282] C.R. Corbeil, P. Englebienne, C.G. Yannopoulos, L. Chan, S.K. Das, D. Bilimoria, L. L'Heureux, N. Moitessier, J. Chem. Inf. Model. 48 (2008) 902–909.
[283] R. Sedlak, T. Janowski, M. Pitonak, J. Rezac, P. Pulay, P. Hobza, J. Chem. Theory Comput. 9 (2013) 3364–3374.
[284] H.S. Muddana, M.K. Gilson, J. Chem. Theory Comput. 8 (2012) 2023–2033.
[285] P. Dobes, J. Fanfrlik, J. Rezac, M. Otyepka, P. Hobza, J. Comput. Aided Mol. Des. 25 (2011) 223–235.
[286] V.M. Anisimov, A. Ziemys, S. Kizhake, Z. Yuan, A. Natarajan, C.N. Cavasotto, J. Comput. Aided Mol. Des. 25 (2011) 1071–1084.
[287] P. Soderhjelm, J. Kongsted, S. Genheden, U. Ryde, Interdiscip. Sci. 2 (2010) 21–37.
[288] S.A. Hayik, R. Dunbrack, K.M. Merz, J. Chem. Theory Comput. 6 (2010) 3079–3091.

[289] L. Hu, J. Eliasson, J. Heimdal, U. Ryde, J. Phys. Chem. A 113 (2009) 11793–11800.
[290] V.M. Anisimov, V.L. Bugaenko, C.N. Cavasotto, ChemPhysChem 10 (2009) 3194–3196.
[291] C.J.R. Illingworth, G.M. Morris, K.E.B. Parkes, C.R. Snell, C.A. Reynolds, J. Phys. Chem. A 112 (2008) 12157–12163.
[292] K. Raha, M.B. Peters, B. Wang, N. Yu, A.M. Wollacott, L.M. Westerhoff, K.M. Merz Jr., Drug Discov. Today 12 (2007) 725–731.
[293] V.M. Anisimov, C.N. Cavasotto, J. Comput. Chem. 32 (2011) 2254–2263.
[294] P. Soderhjelm, F. Aquilante, U. Ryde, J. Phys. Chem. B 113 (2009) 11085–11094.
[295] P.E.M. Lopes, B. Roux, A.D. MacKerell, Theor. Chem. Acc. 124 (2009) 11–28.
[296] F. Forti, C.N. Cavasotto, M. Orozco, X. Barril, F.J. Luque, J. Chem. Theory Comput. 8 (2012) 1808–1819.