



## Review

# Structure-based molecular modeling in SAR analysis and lead optimization

Veronika Temml<sup>a</sup>, Zsofia Kutil<sup>b,\*</sup><sup>a</sup> Institute of Pharmacy, Department of Pharmaceutical and Medicinal Chemistry, Paracelsus Medical University Salzburg, Strubergasse 21, 5020 Salzburg, Austria<sup>b</sup> Institute of Biotechnology of the Czech Academy of Sciences, BIOCEV, Vestec, Czech Republic

## ARTICLE INFO

## Article history:

Received 4 December 2020

Received in revised form 21 February 2021

Accepted 23 February 2021

Available online 4 March 2021

## Keywords:

Docking

Pharmacophore modeling

Lead optimization

Structure-activity relationship

Molecular modeling

## ABSTRACT

*In silico* methods like molecular docking and pharmacophore modeling are established strategies in lead identification. Their successful application for finding new active molecules for a target is reported by a plethora of studies. However, once a potential lead is identified, lead optimization, with the focus on improving potency, selectivity, or pharmacokinetic parameters of a parent compound, is a much more complex task. Even though *in silico* molecular modeling methods could contribute a lot of time and cost-saving by rationally filtering synthetic optimization options, they are employed less widely in this stage of research. In this review, we highlight studies that have successfully used computer-aided SAR analysis in lead optimization and want to showcase sound methodology and easily accessible *in silico* tools for this purpose.

© 2021 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Contents

1. Introduction	1432
1.1. The scope of the review	1432
1.2. The role of molecular modeling methods in the hit-to-lead optimization	1432
2. Structure-based computational methods	1432
2.1. Molecular docking	1432
2.1.1. Theoretical background	1432
2.1.2. Practical background – data and software selection, workflow validation, and recent advances in the field	1434
2.1.3. Examples of studies successfully integrating molecular docking to the hit to lead optimization process	1435
2.2. Pharmacophore modeling	1438
2.2.1. Theoretical background	1438
2.2.2. Practical background	1439
2.2.3. Examples of studies successfully integrating pharmacophore modeling to the hit to lead optimization process	1440
3. Conclusion and perspectives	1441
Funding	1441
CRediT authorship contribution statement	1441
Declaration of Competing Interest	1441
References	1441

\* Corresponding author.

E-mail addresses: [veronika.temml@pmu.ac.at](mailto:veronika.temml@pmu.ac.at) (V. Temml), [zsofia.kutil@ibt.cas.cz](mailto:zsofia.kutil@ibt.cas.cz) (Z. Kutil).

## 1. Introduction

### 1.1. The scope of the review

In this review, we focus on the utilization of molecular modeling techniques, primarily molecular docking but also pharmacophore modeling and molecular dynamics (MD) simulations, in the hit-to-lead optimization process.

To this end, we have surveyed the research papers deposited in Web of Science using the query phrases 1) docking “lead optimization” or “lead identification and optimization” (420 results) and 2) pharmacophore “lead optimization” or “lead identification and optimization” (156 results). Subsequently, we have excluded the publications (a) containing only *in silico* work not verified by appropriate *in vitro* methods, (b) using *in silico* methods for the explanation of observed results, not involved in the optimization process (c) lacking the optimization process. The final selection of 48 publications is presented in this review together with comments on various aspects of molecular modeling techniques that have to be considered especially when they are used in the optimization process. The objective of this review is to showcase the sound methodology and guide the readers through the fractionalized topic to the relevant sources of information like specialized reviews and verified databases. We believe that given the low ratio of publications successfully utilizing *in silico* methods in the lead optimization process such a guidepost is highly useful for multiple audiences. This review is by no means meant to be an exhaustive overview of all publications using the *in silico* methods in the optimization process. While we don't aim at a detailed analysis of the docking and pharmacophore modeling processes, we provide a basic analysis of the theory and recent developments pointing the reader at several excellent review articles that have been written on this topic.

### 1.2. The role of molecular modeling methods in the hit-to-lead optimization

The ground rules of medicinal chemistry remain mostly unchanged since the 1970s when Topliss published his famous scheme about optimizing aromatic substitution patterns for maximal bioactivity [1]. From the very beginning, this kind of systematic optimization has been accompanied by computational analysis like the Hansch and Free-Wilson analysis that aim to quantify the influence specific substitutions have on defined bioactivity [2]. Over the last decades, a vast array of computational methods for medicinal chemistry purposes have become available. Many novel activities are found in pharmacophore-based virtual screening campaigns or found by other 2D and 3D similarity-based methods [3]. Nearly every publication describing a novel activity also proposes a binding mechanism predicted in a docking simulation, often refined by follow-up MD simulations that provide more details on ligand binding.

Still, the role of computational methods becomes smaller over the course of drug development. However helpful, all of these methods are also a double-edged blade. On one hand, they can save a lot of resources and unnecessary work but they can also tempt the user to oversimplify challenges, by looking only at the scope of the simulation and not beyond it. Furthermore, we always have to consider that methods based on already known ligands and binding patterns might prevent us from finding uncommon (and often also highly innovative) ligands. It is, therefore, key to employ a clean methodology that includes as much known data as possible and does not exclude “inconvenient” data points, e.g. atypical molecules like natural products. In summary, molecular modeling methods have established a supporting role for themselves during

hit discovery [4,5], still, they also have a lot to offer during the hit to lead optimization phase when used by the educated user.

## 2. Structure-based computational methods

### 2.1. Molecular docking

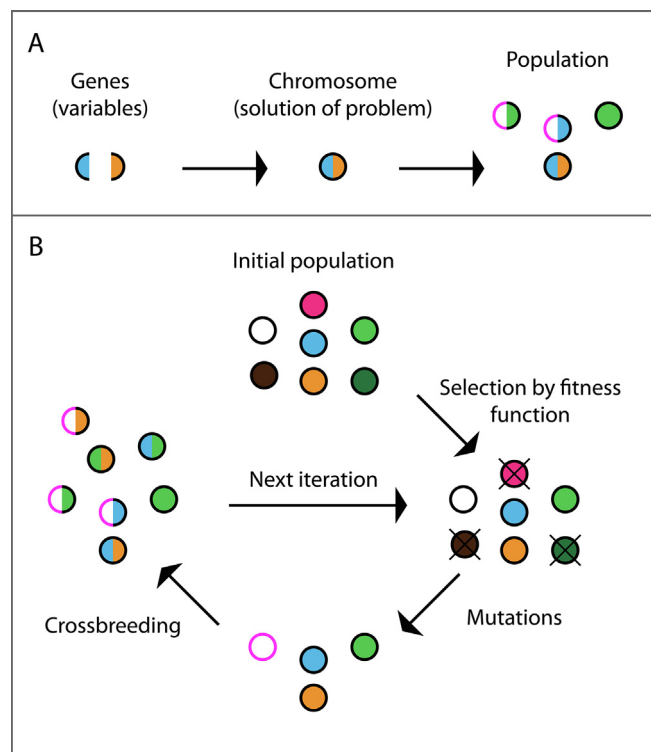
#### 2.1.1. Theoretical background

Molecular Docking has been used to predict binding conformations and interactions between ligands and protein binding sites for over 40 years, starting with Levinthal aiming to simulate the arrangement of sickle cell hemoglobin molecules (Hb-S) in the sickle cells tubular fibers [6]. Since then a plethora of molecular docking programs have been written and applied, primarily to predict binding modes of bioactive molecules. Molecular docking programs for protein–ligand docking are defined by three core functions: (1) The computational representation of protein and ligand. (2) The docking algorithm, which is used to solve the optimization problem of fitting the ligand into the protein binding pocket and generating possible binding poses. (3) The scoring function (SF) then aims to quantify the quality of said calculated binding poses. All of these factors have to balance accuracy and calculation time and therefore work with simplified representations of the physical reality of protein–ligand binding. In this section, we would like to introduce those concepts, but we also refer the readers to the specialized reviews listed in Table 1 for more information.

Due to the size and complexity of biopolymers, a full representation of a complete protein with all atomic coordinates is usually not feasible for high throughput simulations. The structural data, acquired from X-ray crystallography or lately cryogenic electron microscopy is therefore reduced and simplified for docking simulations. A segment of the protein that contains the investigated binding pocket is selected and prepared. In the most basic approaches, the protein is simply seen as a surface, which is enhanced by electrostatic properties [7]. Such approaches are still in use for protein–protein docking, where a more complex representation is

**Table 1**  
Selected reviews focused on the docking algorithms and SFs.

Algorithms	
Reference	Short description
[18]	Describes the main concepts behind biologically inspired algorithms applied to molecular docking simulations.
[19]	Reviews the implementation of genetic algorithms in drug design and quantitative structure–activity relationship.
[20]	Describes several molecular docking search algorithms, and the programs which apply such methodologies.
[21]	Reviews key aspects relevant to the field of molecular docking from forces important in molecular recognition up to the library design.
[22]	Briefly covers some of the applications of genetic algorithms in the field of drug design.
SFs	
[23]	Analyzes machine-learning SFs for drug lead optimization in the 2015–2019 period.
[24]	Summarizes the progress of traditional machine learning-based SFs in the last few years and provides insights into recently developed deep learning-based SFs.
[17]	Reviews basic types of SF based on an up-to-date classification scheme together with suitable application areas and shortcomings, challenges, and potential future study directions.
[25]	Highlight the impact of using QM methods to investigate the docking of ligands to protein targets by using QM-computed parameters to either develop or to be used directly as SFs.
[26]	Describes authors' work on developing the PDBbind database and the CASF benchmark together with relevant work done by other researchers.

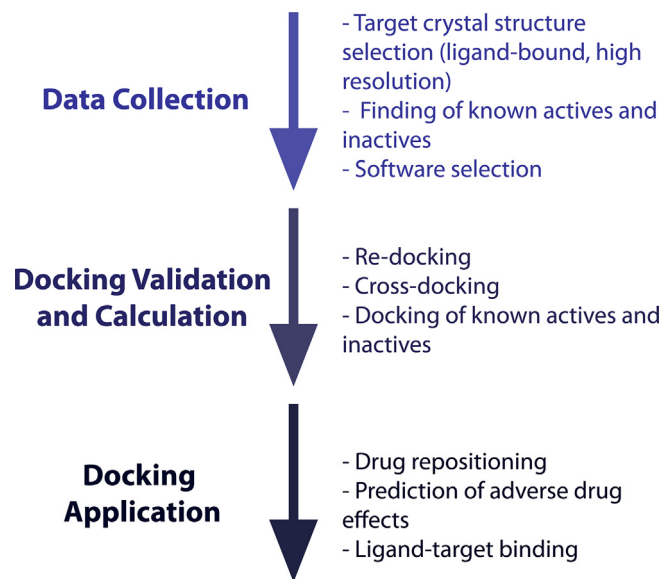


**Fig. 1.** The schematic representation of genetic algorithms. (A) Genetic algorithms are inspired by the principles of natural selection, representing sets of variables as genes, individual solutions as chromosomes, and the pool of the solutions as a population. In the case of docking, gene examples are translational, rotational, and torsional degrees of freedom. (B) The docking calculation process starts with the evaluation of the initial population, derived from a starting conformation. Then a mutation factor that ensures variation in the genes is introduced. The next step is crossbreeding – the genes are combined, creating a new (offspring) population of chromosomes. This process is iterated until (a) the fitness level is satisfactory, (b) the population has converged and does not produce offspring significantly different from the previous generation, (c) the fixed number of iterations is reached [13,14].

not feasible [8]. In protein-ligand docking, the selected binding site is represented in more detail. In the oldest docking methods both the protein and the ligands were regarded as rigid (rigid docking). These approaches were quickly replaced by flexible docking methods, which treat the protein mostly as rigid but are optimizing the conformations of the ligand to fit the binding pocket. In more advanced methods also the receptor has a degree of flexibility, allowing e.g. individual amino acids to assume different orientations [9,10]. Full receptor flexibility is mostly taken into account in docking refinement with MD simulations.

Flexible docking algorithms can be roughly divided into energy-based methods and stochastic methods. Energy-based methods aim to represent the binding free energy as a function of the binding geometry. Stochastic methods work by randomly changing the translational, rotational, and torsional degrees of freedom of a molecule [11]. The resulting poses of these changes are then evaluated by the SF and kept if they represent an improvement and discarded if they do not. To avoid the algorithm finishing in a local energy minimum, worse poses are accepted on a random basis (e.g. with a Monte Carlo algorithm [12]). One of the most widely applied group of docking algorithms are genetic algorithms (employed by software programs like GOLD and Autodock). The principle of genetic algorithms is shown in Fig. 1.

To quantify the quality of binding poses, docking programs use SFs. SFs can be categorized into physics-based and empirical SFs although in the meantime a lot of hybrids have been developed. Physics-based SFs consist of a sum of energy terms from a force-



**Fig. 2.** Scheme of docking workflow. The sound docking setting validation and careful data selection must precede the docking calculation and post-processing.

field (e.g. the MERCK or AMBER force field) that represent interaction energies of the protein–ligand complex, internal ligand energy, and sometimes also solvation models [15]. Empirical SFs are based on regression analysis of structural descriptors and experimentally elucidated affinity data of known protein–ligand complexes [16]. Scoring is a hugely controversial topic in molecular docking and a wide variety of SFs and post-process rescoring functions have been developed [17].

Next to the theoretical background of molecular docking, we would like to stress that while docking is a valuable tool to find out how a ligand molecule binds to a target protein it can also lead an inexperienced user to over-interpret the results. A docking program will almost always propose one or several possible binding poses and it is the user who needs to judge the quality of the result. Unfortunately, there are a lot of studies published where docking is used with very low-quality standards. Often the biological experiments do not necessarily translate to binding affinity to the specified target, docking scores are often sold as “activities” and only a few studies fully include their workflow validation. Haphazard

**Table 2**  
Example studies focused on the evaluation of docking programs.

Special focus on	Well rated software	Other tested software	Reference
None (generally small molecules and proteins)	GOLD, AutoDock, Vina, LeDock	LigandFit, Glide, MOE Dock, Surflex-Dock, AutoDock, AutoDock Vina, rDock, UCSF DOCK	[32]
Peptides	FRODOCK, ZDOCK	Hex, pepATTRACT, PatchDock, ATTRACT	[33]
Peptides	HPEPDOCK, AutoDock, CrankPep	ZDOCK, FRODOCK, GOLD, Surflex-Dock, AutoDock Vina, GalaxyPepDock, MDockPeP, HPEPDOCK, HawkDock, CABS-dock, pepATTRACT, DINC	[34]
Metalloproteins	PLANTS, LeDock, Quick Vina, AutoDock Vina	AutoDock, UCSF DOCK	[35]
Dihydropteroate Synthase	Glide, Surflex	FlexX, GOLD, DOCK	[36]

**Table 3**

Selected publications dealing with the advances in docking, recent research, and/or specialized reviews.

Reference	Short description
[50]	Examine a representative set of currently used computational approaches to identify repurposable drugs for COVID-19.
[51]	Introduces public-private partnership that has been established worldwide and then describes the background, frame, activities, and uniqueness of this partnership.
[52]	Reviews the experience with Ebola and Zika viruses drug development, its implications to SARS-CoV-2 drug discovery, gaps in the field, and computational approaches applied.
[53]	Provides a review of the available computational methods that employ water molecules for the analysis of macromolecules' properties and structure dynamics.
[54]	Research article; Application of Gaussian Boson Samplers for prediction of accurate molecular docking configurations.
[55]	Intends to provide readers with guidance for practically applying MM/PBSA and MM/GBSA in drug design and related research fields.
[56]	An opinion of a panel of scientists from the industry who work at the interface of machine learning and pharma on the past, present, and future role of AI for ADME/Tox in drug discovery and development.
[31]	Covers the current field of in silico docking to nucleic acids, available programs, as well as challenges faced in the field.
[57]	Introduces in silico approaches and tools that have been developed to predict drug metabolism and fate, and assess their potential to facilitate the virtual discovery of promising drug candidates.
[58]	Introduces three-dimensional matched molecular pairs concept and discusses the successful applications.
[59]	Discuss the development and application of strategies for the structure-based design of cancer-targeting peptides against GRP78.
[60]	Outlines the evolution of decoys selection in benchmarking databases as well as current benchmarking databases that tend to minimize the introduction of biases, and secondly, propose recommendations for the selection and the design of benchmarking datasets.
[61]	Focuses on the specifics of docking calculations with covalent ligands.
[62]	Discuss applications and practical aspects of MD simulations with mixed solvents.
[63]	Provides an overview of protein-peptide docking methods and outlines their capabilities, limitations, and applications in structure-based drug design.
[64]	Examines the successes, limitations, and new avenues for modeling metalloenzyme inhibitors and metalloenzymes.
[65]	Gives a historical account of the development of ensemble docking and discusses some pertinent methodological advances in conformational sampling.
[66]	Presents an overview of the evolution of structure-based drug discovery techniques in the study of ligand-target recognition phenomenon, going from static molecular docking toward enhanced MD strategies.
[67]	Summarizes progress in the prediction of RNA-ligand interactions, available methods for calculating the mode of binding for various small RNA molecules, the range of SFs for ligands ranking, accommodating RNA flexibility, and example studies.
[68]	Provides an overview of the state of the art of experimental and computational approaches for investigating drug metabolism.

simulations are almost certainly meaningless for lead structure optimization. It is therefore key to rely on high-quality data and extensive validation when constructing a fine-tuned docking workflow that should be able to predict the activity of structural modifications in a quantitative manner. The optimal docking workflow is simplified in Fig. 2 and discussed in the next section.

### 2.1.2. Practical background – data and software selection, workflow validation, and recent advances in the field

The main resource for any publically available structural data is the protein databank (<https://www.rcsb.org/>), a constantly

updated database of all published protein structures [27], invaluable for molecular modeling purposes [28]. All structural data available on the target should be carefully examined since the crystallized protein differs from the wildtype and the researcher needs to judge if a modified structure is still relevant for their study (e.g. if the investigated binding site is conserved in the mutant structure). If the resolution of the crystal structure is sufficient for molecular modeling (ideally below 2.5Å), the electron density map should be checked to identify flexible, or otherwise badly resolved parts of the protein. Many flexible proteins assume different binding conformations upon binding to different small-molecule ligands. If the investigated scaffold is known beforehand it is beneficial to select a co-crystallized structure that is similar to the investigated scaffold. Next, to build an optimal docking workflow it is helpful to have a high-quality test set of known ligands and non-binding compounds, to determine if the workflow can discriminate between them. It must be stressed that aside from the docking software, the quality of the data used for the modeling and the way they are processed (e.g. protein preparation and conformer generation) is the most important factor [29].

Various studies were performed to benchmark the docking software and their SFs (see table 2; reviewed in [17,30,31]). Nevertheless, due to the number of programs and the variable nature of the targets and ligands, it is almost impossible to recommend a single software or method. As it is shown in table 2, the programs not performing well in one study, were well-rated in studies focused on other targets or ligands (e.g. AutoDock Vina). Therefore, we suggest following the procedures in the benchmarking studies and use multiple docking programs during validation, to determine the one that yields the best results for a particular research question. We are aware that the availability of the docking software often plays a role in their selection and we have specified it in Tables 4 and 6. Still, the users should keep in mind that not all commercial and open-source molecular modeling software are fully compatible.

The minimum validation of a docking workflow is the redocking of the co-crystallized ligand into the empty binding pocket of the protein. The calculated pose is then compared to the bioactive conformation from the crystal structure and the root mean square distance (RMSD) is calculated. Typically, RMSD values below 2 Å are considered acceptable, although depending on the size and flexibility of the ligand and the quality of the crystal structures, researchers might set a more ambitious benchmark [37]. If multiple structures for a protein are available, they can be used for cross-docking validation. A co-crystallized ligand is docked to another protein structure and the poses are compared to the co-crystallized complex, to see if the orientation and interactions in the binding site can be reproduced correctly [38] (Fig. 3).

Ideally, the workflow is also able to enrich active over inactive compounds in a theoretical validation with a test set of known actives and inactives from the literature or standardized benchmarking datasets [40]. In this case, the user is assuming that bioactivity correlates directly with binding affinity. While this is true for specialized methods like radioligand binding assays, other factors than direct ligand-binding may play a role in other assay types to different degrees. It is therefore critical to evaluate how many other factors influence the outcome of the used assay and if activity in the assay translates quantitatively to binding affinity.

The validity of the workflow is represented by its ability to rank active over inactive compounds, which can be represented in the receiver operating characteristic (ROC) curve. All compounds found in a test set are ranked according to their score. Each true active is counted on the vertical scale, while each false positive is counted on the horizontal scale until all compounds are included and the curve reaches the top right corner of the plot. Every ROC curve above the diagonal is an improvement over random selection. The integral of the ROC curve is the area under the curve,

**Table 4**

The selected studies using molecular docking in successful lead optimization.

Desired therapeutic use <sup>1</sup>	Target	Fold improvement <sup>2</sup>	Software used	Validation method	Reference
Allergic asthma treatment	ITK	13	Glide <sup>8</sup>	Enzymatic	[83,84]
Alzheimer's disease therapy	BACE1	DM <sup>3</sup>	Glide	Enzymatic	[85]
Antibacterial therapy	AcpS	9	FLO99 <sup>9</sup>	Enzymatic	[86]
Anti-inflammatory drug	mPGES-1, 5-LO	5; 2	Gold <sup>8</sup>	Cell-free	[48]
Anti-inflammatory, cardiovascular diseases and cancer treatment	BRD4(BD1)	DM, >60 <sup>4</sup>	SAMPLE <sup>10</sup>	Direct <sup>5</sup>	[87]
Antinociceptive, anti-inflammatory	FABP	4	AutoDock 4.2 <sup>10</sup>	Enzymatic	[88]
Antiretroviral therapy	HIV-1 RT	1600	Dock 4.0 <sup>10</sup>	Cell based	[89]
Antiretroviral therapy	HIV-1 RT	10	Sybyl <sup>8</sup>	Enzymatic	[90]
Antitubercular drug	GyrB	2	Glide XP	Enzymatic	[91]
Cancer therapy	BAX	1	Glide 5.5	Enzymatic	[73]
Cancer therapy	CSF1R	2	Glide	Enzymatic	[76]
Cancer therapy	G-quadruplex DNA	DM, 10x <sup>4</sup>	ICM-Pro 3.4 <sup>8</sup>	Direct <sup>6</sup>	[92]
Cancer therapy	MAGL	13	AutoDock 4.0	Enzymatic	[69]
Cancer therapy	MAP4K4	12,941	Glide SP	Enzymatic	[93]
Cancer therapy	MDM2	12	Glide 5.5	Cell free	[78]
Cancer therapy	mTOR	32	Glide	Enzymatic	[94]
Cancer therapy	PKC-1	24	AutoDock 3.05	Enzymatic	[95]
Cancer therapy	PI3Ka	0 <sup>7</sup>	FlexSIS <sup>9</sup>	Enzymatic	[77]
Cancer therapy	Pim-1	2	Discovery Studio <sup>8</sup>	Enzymatic	[96]
Cancer therapy	SENP1	2	Glide 4.5	Enzymatic	[97]
Complications of diabetes	ALR	20	Dock	Enzymatic	[98]
Complications of diabetes	ALR2	136	Autodock 4	Enzymatic	[99]
Diabetes mellitus therapy	PTP1B	DM	Glide 9.2 <sup>8</sup>	Enzymatic	[100]
Diabetes mellitus therapy	PTP1B	541	QXP <sup>9</sup>	Enzymatic	[101]
Huntington's disease	Jnk3	80	Glide	Enzymatic	[102]
Hyperuricemia therapy	XO	5	Glide	Enzymatic	[103]
Hyperuricemia therapy	XO	470	Glide	Enzymatic	[104]
Malaria treatment	pfENR	167	AutoDock 3.05	Enzymatic	[105]
Osteoarthritis therapy	MMP-13	5	AutoDock 4.0	Enzymatic	[106]
Pain therapy	NaV1.7	2	ZDOCK <sup>10</sup>	Cell based	[107]

<sup>1</sup> The possible drug disposition in pathophysiological conditions are taken from referenced studies.<sup>2</sup> In IC<sub>50</sub>, Ki or KD.<sup>3</sup> Data missing.<sup>4</sup> Reported improvement.<sup>5</sup> Thermal-shift assay.<sup>6</sup> DNA polymerase stop assay.<sup>7</sup> Improvement in metabolic stability.<sup>8</sup> Commercial software.<sup>9</sup> Academic.<sup>10</sup> Freeware.

which additionally contains information about the “early enrichment” in the dataset if the actives are ranked higher than inactives [41].

When docking is used for structure optimization all experimental data that relates to the binding mode, e.g. from mutational studies or protein fishing experiments should be gathered. It should be stressed that in contrast to other *in silico* modeling applications, e.g. environmental toxicity assessments studies where *in vitro* evaluations are not desired [42–44], *in silico* lead optimization stands and falls with the combination of those methods.

If docking is used in the context of a lead structure optimization, the demands on the SF become a lot more challenging. Instead of just discriminating between active and inactive, a quantitative relationship between the docking score and measured bioactivity should be achieved. Enyedy et al showed that automated SFs have a hard time outperforming the correlation of simple physicochemical properties such as molecular weight or clogP [45]. Sinko conducted a study where 68 PDB crystal structures of complexes between acetylcholinesterase (AChE) and its ligands were evaluated by different SFs (LigScore1, LigScore2, PLP1, PLP2, Jain, PMF, and PMF04) to see if they could establish a quantitative correlation to their activity. The best results were achieved by the PLP2 function with a coefficient of determination ( $r^2$ ) of 0.591. However, also physicochemical parameters, like the number of heavy atoms or the number of sp<sup>2</sup> hybridized atoms, performed with an  $r^2$  above 0.5 [46].

Alternatively, it can be useful to identify key interactions that are vital for the activity of a specific scaffold [47]. This can lead to empirically customized SFs, especially useful in guiding the optimization process [48]. An exemplary theoretical study by Levoine et al, conducted already in 2008 aimed to construct docking-based virtual screening workflows to select high-affinity histamine H3 receptor ligands, while also excluding activity on the HERG channel and CYP2D6, prominent antitargets. For their chosen activity thresholds they achieved an accuracy of at least 70% with these methods [49].

The advances in the field and the popularity of docking led to its novel applications in specific areas of drug development. We would like to mention e. g. the docking of ligands to nucleic acids and the docking of specific ligands like peptides, but due to the broad range of the developments, it is beyond the scope of this review to discuss them in detail. Nevertheless, we have selected 20 representative specialized reviews and research publications that deal with various directions of recent developments and we provide their list in Table 3 to ease the reader the way to the information. Besides, we hope that the following “Example” section will guide the less advanced docking users in the right direction.

### 2.1.3. Examples of studies successfully integrating molecular docking to the hit to lead optimization process

Molecular docking has been successfully used in drug screening campaigns, but its ability to help in lead optimization is still gener-

**Table 5**  
Selected publications dealing with the pharmacophore modeling basics and advances.

Reference	Short description
[115]	Research article; Introduces a target-specific drug design method based on a deep learning algorithm and a water pharmacophore that can autonomously generate a series of target-favorable compounds.
[116]	Research article; Describes BioChemical Library update and shows (beside others) how the author's models can be decomposed into human-interpretable pharmacophore maps to aid in hit/lead optimization.
[117]	Provides a brief introduction to the pharmacophore modeling concept and presents examples of applications in the specific field of natural product chemistry.
[112]	Overview of the basic pharmacophore modeling concept together with recent developments in the field.
[118]	Research article; Present a new approach that incorporates flexibility based on extensive MD simulations of protein–ligand complexes into structure-based pharmacophore modeling and virtual screening.
[119]	Discuss foundations and caveats of scaffold hopping approaches and analyzes recent methodological developments
[120]	Introduces three of the fertile directions in approaching the biological activity by chemical structural causes: (i) the special computing trace of the algebraic structure–activity relationship, (ii) the minimal topological difference (MTD), and (iii) comparative molecular similarity indices analysis.
[121]	Different approaches to the generation of pharmacophore models are compared together with their strengths and weaknesses.
[122]	Reviews pharmacophore techniques that are used for modeling ADME properties (with a special focus on pharmacophore models reported for various cytochrome P450 enzymes)
[123]	Reviews pharmacophore modeling studies focused on antitargets.
[124]	Describes the specifics in the construction of structure-based pharmacophores including (i) protein structure preparation, (ii) binding site detection, (iii) pharmacophore feature definition, and (iv) pharmacophore feature selection.
[125]	Focuses on the synergistic combination of pharmacophore modeling with other molecular modeling approaches such as the hot spot analysis of protein binding sites, MD, and docking
[126]	A brief overview of structure–activity relationship methods in ligand-based drug design followed by a more detailed presentation of issues and limitations associated with empirical energy functions and conformational sampling methods
[127]	Insight into different approaches implemented by the 3D pharmacophore modeling packages like Catalyst, MOE, Phase, and LigandScout.
[128]	Describes computational models for key nuclear hormone receptor binding sites generated by ligand-based approach.

ally questioned. If we plan to use it so, we have to realize that it is not a stand-alone technique but should be embedded in a workflow of different *in silico* as well as experimental techniques. In [table 4](#), we have summarized the studies successfully integrating docking into the optimization process, that we have found. These studies share a good research methodology, where molecular docking was connected with *in vitro* verification and led to optimized compounds. The major reasons why we have excluded the remaining publications are: (1) the authors used docking just to elucidate the mechanism of action of their optimized compounds or performed retrospective molecular modeling on previously optimized compounds (~13% of the query results); (2) the authors have only proposed the optimization by *in silico* methods, but haven't performed the optimization (~3%); the authors haven't used appropriate methodology to validate their docking results (e.g. validation of molecular docking in animals or models that are too far removed from direct target ligand binding interaction; ~1%) (3); It was not clear, how the authors implemented the docking to the optimization process (~1%) (4) The authors identified the hit but did not develop it further (~1%); (5) The docking approaches did not lead to a more potent compound (~1%). Those reasons mainly point towards user-based issues in the implementation of

the docking in the optimization process and we hope that the sound methodology discussed in the following section will help to improve the reputation of the molecular docking in the hit-to-lead optimization.

When we aim to optimize a specific activity, it is vital to be familiar with the exact molecular mechanism of action, e.g. Granchi et al employed molecular docking to optimize a reversible monoacylglycerol lipase (MAGL) inhibitor [69]. Most known MAGL inhibitors act via an unfavorable irreversible mechanism of action, it was therefore vital in this case to retain the reversibility of inhibition. A docking simulation was conducted and the resulting poses were clustered to identify the dominant pose orientations and interaction patterns. This process was used to select poses that underwent an MD simulation to check their reliability. Synthesizing the structures that fared best in the computer simulations led to a compound (17b) with a tenfold improved  $K_i$  value compared to the lead compound. Additionally, aware of the fact that this class of compounds could act as artifacts and promiscuous bioactive molecules, the authors carried out a number of control and verification experiments including Pan Assay Interference Compounds (PAINS) analysis. The acronym PAINS was first used by Baell and Holloway, who described a number of substructural features that can help to identify false-positive compounds and included them into publicly available filter-it software [70]. The electronic filters formulated to recognize PAINS can process hundreds and thousands of compounds in seconds and are a very useful tool in current medicinal chemistry. Still, even the authors of this concept, later on, wrote: "It has become increasingly clear that overzealous or simplistic use of these filters may inappropriately exclude a useful compound from consideration and inappropriately tag a useless compound as worthy of development" [71]. We thus highly recommend including this type of analysis into the optimization process but keep in mind its limitations. It might be tempting to employ such filter tools at the beginning of a virtual screening campaign to exclude all potentially problematic compounds from the start. This can, however, lead to overlooking interesting hit structures.

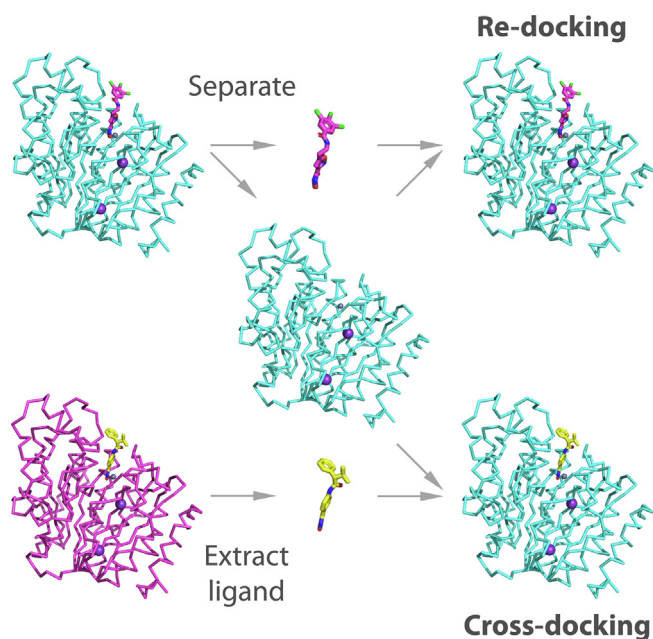
Another crucial aspect of optimization by docking is the interpretation of the interaction pattern with the protein. It should be noted that docking primarily describes reversible inhibition without covalent binding. There are however some specialized approaches (see [61]) where docking is used to predict covalent (irreversible binding). This is usually accomplished by anticipating the reaction that takes place and predicting if the reaction partners will be in close enough vicinity to perform the reaction. Non-covalent binding is usually driven by electrostatic interactions, either by ionic interactions or by weaker polar interactions such as hydrogen bonds. These bonds do not only occur directly between the ligand and the protein but are also often mediated by water molecules. Different interaction points often vary in their relevance for the overall activity. While interactions with the catalytic or substrate binding residues are most likely to play an important role, it is also key to systematically analyze all interactions, in multiple crystal structure complexes. A handy freeware tool for conducting interaction analysis is protein–ligand interaction profiler (PLIP), a web server that generates lists of interacting residues for a list of given pdb entries (available at [www.projects.biotec.tu-dresden.de/plip-web](http://www.projects.biotec.tu-dresden.de/plip-web)) [72].

If possible, also compounds that do not possess a suspected key interaction point should be synthesized and tested to verify the relevance of the interaction by loss of activity. Such proof of principle was shown in several cited studies. We would like to highlight a 2015 study by Stornaiuolo et al, where a structure-based docking approach was used to develop the first direct activator of BCL-2-associated X protein (BAX), a pro-apoptotic member of the B-cell lymphoma-2 family [73]. The authors of this study carefully prepared a design strategy by the analysis of their target and its small

**Table 6**

The selected studies using pharmacophore modeling and other approaches in successful lead optimization.

Desired therapeutic use <sup>1</sup>	Target	Fold improvement <sup>2</sup>	Modeling method	Software used	Validation method	Reference
Antimitotic agents	Tubulin	DM <sup>3</sup>	PM <sup>4</sup>	HypoGen <sup>5</sup>	Cell based	[153]
Antiretroviral therapy	HIV-1 RT	2	FEP	NAMD <sup>6</sup>	Enzymatic	[154]
Cancer	Dvl PDZ	30	PM, Docking	LigandScout, Glide <sup>5</sup>	Direct1	[146,147]
Cancer	PARP1	21	PM, MD	Discovery Studio 3.5 <sup>5</sup>	Enzymatic	[143]
Cancer	PRMT5	3	SS	Pipeline Pilot 7.5 <sup>5</sup>	Enzymatic	[155]
Cancer, Alzheimer's disease, Parkinson's disease, sleep disorders, inflammation	CK1	DM	PM	Discovery Studio 3.1 <sup>5</sup>	Enzymatic	[156]
Cancer, glaucoma, asthma	A <sub>3</sub> AR	3	FB	Chemaxon <sup>5</sup>	Cell based	[157]
Cognitive diseases, renal and cardiac failure, Alzheimer's disease	A <sub>1</sub> AR	4	PM, MD	SYBYL 6.2, AMBER 4.1 <sup>5</sup>	Cell-free	[158]
Glucocorticoid-dependent diseases	11 $\beta$ -HSD	31	PM	LigandScout	Cell based	[134,151,152]
Growth hormone therapy	GH	DM	PM, FB	DistComp <sup>6</sup>	Cell based	[159]
Hyperlipidemia	ACAT	DM	PM	HypoGen	Cell-free	[160]
Inflammatory diseases	Caspase-3	DM	MD	MacroModel, Amber 6.0 <sup>5</sup>	Enzymatic	[161]
Parkinson's disease	A <sub>2</sub> A AR	5	SS	MOE <sup>5</sup>	Cell-free	[162]

<sup>1</sup> The possible drug disposition in pathophysiological conditions are taken from referenced studies.<sup>2</sup> In IC<sub>50</sub>, Ki or KD.<sup>3</sup> Data missing.<sup>4</sup> Methods are abbreviated as follows: PM, Pharmacophore modeling; FEP, Free energy perturbation; MD, Molecular dynamics; SS, Substructure search; FB, Fragment-Based.<sup>5</sup> Commercial software.<sup>6</sup> Academic.

**Fig. 3.** Re- and cross-docking. The re-docking sometimes called self-docking involves the separation of the target from the co-crystallized ligand, subsequent docking back, and comparison of the calculated pose to the conformation from the crystal structure. In the case of cross-docking, ligand and target from two different crystal complexes are used for the docking experiment. The Figure was prepared using the program PyMOL [39].

molecule activator by means of docking experiments. Conformational changes from the inactivated to the activated state were considered in the activator binding model, by superimposing structures of the active and inactive state of the protein. Based on observed interactions, the authors synthesized a series of analogs, including compounds serving as negative controls. The negative controls were prepared to evaluate the importance of particular moieties of the activator. Next, their affinity and selectivity profiles

of the synthesized compounds were measured and the whole SAR was rationalized. The resulting lead compound was further evaluated in cellular assays and *in vivo* on mice models showing efficiency in tumor mass reduction together with the absence of gross toxicity.

When moving further along in the drug development process the effects on the whole organism become more and more relevant. A marvelously low IC<sub>50</sub> on an isolated target is worth nothing if the drug can never reach the intended target. ADMET properties should, therefore, be considered early on in the drug development process to avoid inauspicious surprises in *in vivo* trials. There already a lot of online prediction tools from a wide variety of sources available to profile compounds for their drug-likeness and for their ability to be orally absorbed Online predictors like swissADME (<http://www.swissadme.ch/>) [74] give a good idea about oral bioavailability, ability to penetrate the blood–brain barrier and also for the likelihood of a compound to bind to metabolic enzymes and some anti-targets. Biotransformer, an open-source prediction tool for drug metabolism was presented at the beginning of 2019. It uses a knowledge and machine learning-based approach to predict putative metabolites of an input molecule under different conditions (e.g. gut microbiota and water/soil microbiota) [75]. An excellent resource to stay on top of novel tools and databases that become available is <http://www.click2drug.org/>, a comprehensive list of computer-aided drug design (CADD) software.

Nevertheless, even docking itself can be utilized in the improvement of ADMET properties. In a study, focusing on the optimization of bis-amide derivatives as CSF1R inhibitors, published in 2017, the authors employed a docking workflow in GLIDE to evaluate possible replacements for a metabolically labile and poorly permeable methyl piperazine group in the lead compound. Possible replacement candidates were evaluated in a docking workflow and compounds with good chances of retaining activity were synthesized and tested. Replacement of the methyl piperazine group and further optimization led to more stable compounds and oral bioavailability in a mouse model [76]. Another example of a successful docking application in improving the ADMET properties can be

found in [77]. The lead phosphoinositide 3-kinase (PI3K) inhibitor was metabolically unstable because of rapid glucuronidation of the phenol moiety. Based on the X-ray structure of the PI3K, the authors concentrated their SAR on the interactions with Asp836 and Lys833, but the results of this approach were not successful. The subsequent virtual docking experiments took into consideration not only Asp841 and Tyr867, but also Asp836 and Lys833. Based on this docking simulation the authors identified aminopyrimidine as a bioisostere of phenol and designed compounds showing comparable PI3K inhibitory activity to the lead compound and greatly improved metabolic stability. The final compound showed strong tumor growth inhibition against a KPL-4 breast cancer xenograft model *in vivo*.

Docking can be valuable when investigating multi-target inhibitors since it allows us to compare the binding mode within two different proteins that often share binding site similarities. Here, we would like to discuss one successful and one unsuccessful example for this application. Giustiniano et al used a docking-based virtual screening to find dual inhibitors of a murine double minute (MDM) 2 and 4 homologs [78]. Inhibitors that target MDM2 selectively induce the upregulation of MDM4. Targeting both proteins, therefore, increases the chance for an effective clinical response. The visual inspection of docking results to MDM2 performed by Glide docking software led the authors to the synthesis of compound effective against both the MDM2 and MDM4. The lead structure was then optimized in an *in silico* guided process to retain both activities and led to a nanomolar inhibitor of both targets. A binding mode predicted for the series was then confirmed with NMR spectroscopy.

However, docking showed the limitations in a study performed by Cheung et al. The authors performed docking experiments for the binding sites of the two target enzymes 5-Lipoxygenase (5-LO) and Microsomal prostaglandin E synthase (mPGES-1) to gain more insights into their SAR. The docking was performed with GOLD and the docking poses were analyzed by taking into account three factors: the presence of key hydrogen bonds, the docking score, and the similarity of the outcome poses. Nevertheless, the detailed analysis of the results revealed that the docking simulations underscore the similarities between the two binding modes, explaining why so many of the derivatives also showed activities against both, 5-LO and mPGES-1 [48].

In depth *in silico* analysis was used by a research team led by Artem Cherkasov, to explore various binding sites on the androgen receptor (AR). The AR contains a ligand-binding domain, which serves as a binding site for both androgens and other small molecule ligands. A particularly dangerous mutation in prostate cancer, however, produces AR-V7, a mutant that lacks the ligand-binding domain and thereby cannot be controlled by the common androgen deprivation therapy [79]. Inhibitors of the AR targeting alternative binding sites are therefore of great interest to treat prostate cancer. The group used *in-silico* methods to identify two alternative binding sites on the DNA binding domain of the AR [80]. Docking was used to identify inhibitors of the D-BOX dimerization interface of the AR DNA binding domain [81].

The last study that we would like to showcase is a textbook example for docking applications in lead discovery and lead optimization. Lyo et al used a large-scale docking approach to evaluate 170 million of make-on-demand compounds that can be formed by 130 popular chemical reactions against the AmpC  $\beta$ -lactamase (AmpC) and the D4 dopamine receptor [82]. They have identified phenolate inhibitor of AmpC which revealed a group of inhibitors with unknown precedent and optimized this compound lowering the Ki 29 times. Against the D4 dopamine receptor, 81 new chemotypes were discovered, 30 showed submicromolar activity, including a 180-pM subtype-selective agonist of the D4 dopamine receptor. This study was executed under rare conditions that also

allowed the researchers to address several key questions of the molecular docking. The authors not only verified their hits and leads *in vitro* (over 500 molecules), but they also co-crystallized selected compounds with AmpC and confirmed their fidelity to the docking predictions. This large and verified compound data set allowed them to evaluate SFs and take a closer look at the performance of the human over the computer. The compounds were spread to the bins covering the highest-ranking, mid-ranking, and low-ranking scores. The hit rates followed the score after a plateau defined by the highest-ranking molecules. The hit rates of sets selected by docking score alone and by human visual evaluation were comparable (at around 24%). Nevertheless, the molecules prioritized by human inspection typically had better affinities and a disproportionate number of the most potent compounds. This study shows that it is possible to achieve a quantitative correlation between score and experimental activity (at least for the D4 receptor) and molecular docking is extremely helpful in drug discovery, but still confirms the need for human expert inroad.

## 2.2. Pharmacophore modeling

### 2.2.1. Theoretical background

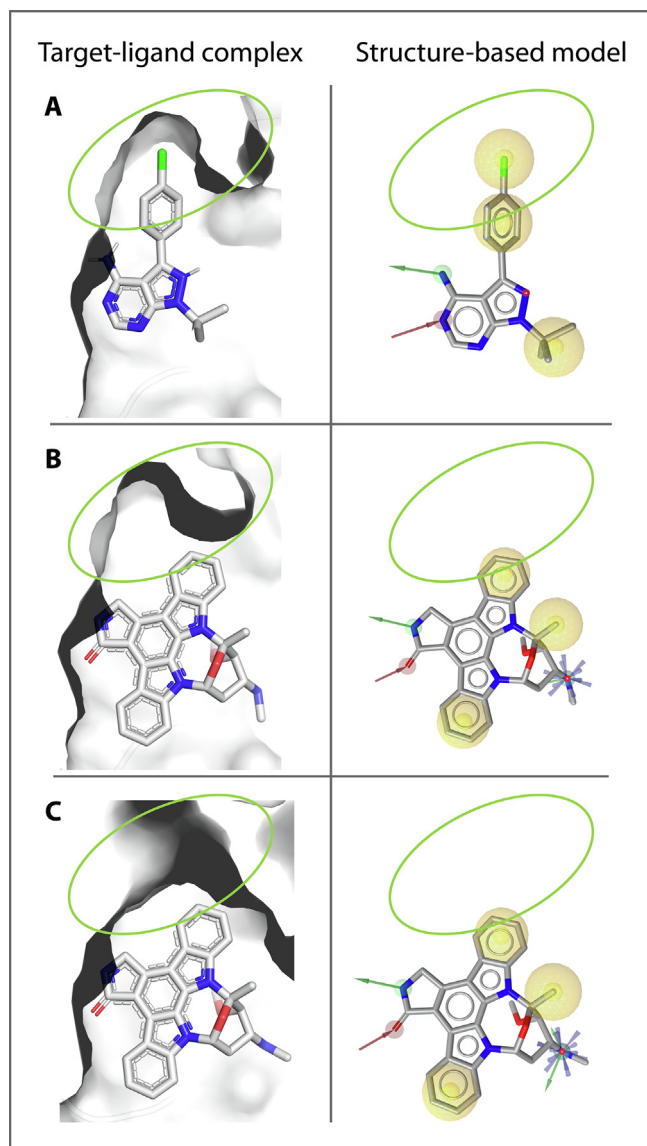
Pharmacophore modeling has found a place in drug discovery, especially for fast and effective screening for new bioactive molecules [108]. A pharmacophore is defined by IUPAC as “an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response” [109]. These abstracted three-dimensional interaction patterns can be used to examine key interaction points that define ligand binding and to determine which parts of a molecule can be modified while retaining or even improving ligand binding. Established pharmacophore models for a target binding site can then be compared to the pharmacophores of conformational libraries in a virtual screening that requires comparatively little calculation time [110]. Since a pharmacophore only defines interaction as a general interaction type, e.g. as a hydrophobic area or a hydrogen bond acceptor, a pharmacophore screening can also perform scaffold hopping and propose different functional groups for bioisosteric replacement [111].

There are two general approaches to generate pharmacophore models: (i) the structure-based, built on an experimentally elucidated protein–ligand complex, and (ii) the ligand-based approach, where known ligands are superimposed, to elucidate common structural elements and pharmacophore features.

There are a variety of different molecular modeling software packages available for pharmacophore modeling. They are using different approaches to four key aspects of pharmacophore modeling: Calculation of the pharmacophore and feature definitions, conformational sampling, screening algorithm, and SF. Several pharmacophore modeling software programs are reviewed in-depth under [112]. They are mainly distinguished by the available types of features and how they are placed. Some features like ionic interactions or hydrogen bonds are available in all programs and their definition is rather straightforward based on the physical properties of these bonds. Many programs also contain more specialized features e.g. metal binding features for different metals or even allow the user to define their own features in accordance with their requirements [110,113].

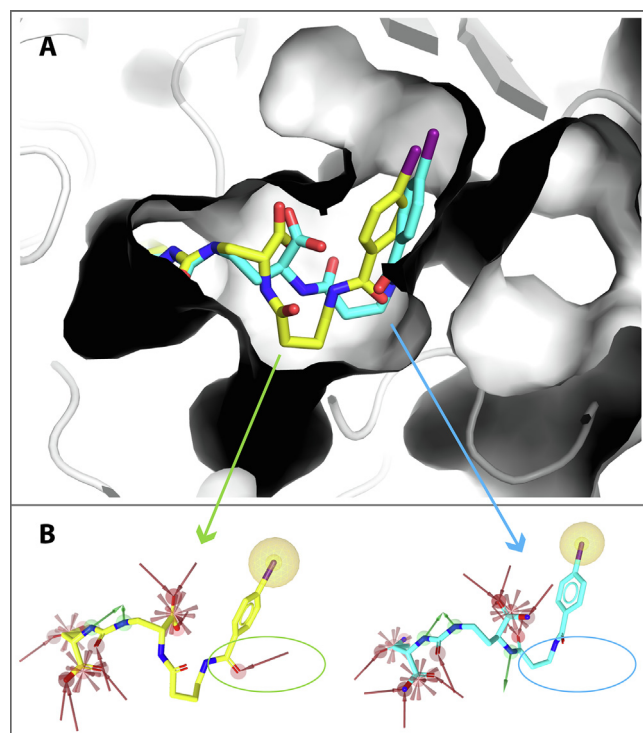
The screening algorithm then compares the query pharmacophore to the pharmacophores of molecules within a conformational library. Similar models in different software programs can find different hit molecules within the same database, due to differences in the algorithm [114]. Finally, the fit of a compound to the pharmacophore model is quantified by a SF (often called pharmacophore fit value). Since the fit into these simplified models is a more clearly defined task than the fit to an actual binding pocket,





**Fig. 4.** The comparison of non-selective and Src kinase family selective inhibitors with the help of pharmacophore models. (A) PP2, Src kinase family selective inhibitor access a deep, hydrophobic pocket of the lymphocyte-specific kinase Lck (pdb: 1qpe), not occupied by staurosporine, a potent but non-selective inhibitor (B; pdb: 1qpd). (C) Staurosporine co-crystallized with kinase from a different family (cyclin-dependent kinase 2; pdb: 1aq1). This kinase lacks the binding pocket near the ATP-binding cleft seen in A and B. The inhibitors share a common binding mode in the ATP-binding cleft, but PP2 binds selectively by making additional contacts in a deep, hydrophobic pocket present in Src kinase family [129,130]. This phenomenon is visualized by the software PyMOL (left side of the panel; [39]), but even better reflected by pharmacophore models created by the software LigandScout (presented at the right site; [110]) and showing the feature important for the selective binding. The yellow spheres represent hydrophobic features, the blue star is a positive ionizable feature, the green arrow is a hydrogen bond donor feature and the red arrow marks a hydrogen bond acceptor feature. These pharmacophore models can serve for the search of new Src kinase family selective inhibitors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

SFs are not as controversial as they are in molecular docking. Recent developments are aiming to combine pharmacophore modeling with MD simulations, creating dynamic pharmacophores that account for the dynamic process of protein/ligand binding [112]. For more information about the basics of pharmacophore modeling and the recent developments in the field, please see the publications listed in Table 5.



**Fig. 5.** The comparison of the binding modes of two similar inhibitors with the help of pharmacophore models. (A) Two similar inhibitors co-crystallized with human glutamate carboxypeptidase II (pdb: 6RBC and 6S1X) visualized by PyMol software [39]. (B) Their pharmacophore models prepared by software LigandScout show that the yellow inhibitor lacks one interaction which is probably responsible for its lower potency measured *in vitro* [135]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 2.2.2. Practical background

Constructing a pharmacophore model requires an in-depth knowledge of the target protein. Many proteins have a variety of different binding sites and even for a single binding site there are usually multiple different binding modes since different ligands interact with different parts of the pocket (Fig. 4 A and B). Subtypes of proteins often contain very small differences that can determine the selectivity of a pharmacophore, e.g. in the case of the Src kinase family: A hydrophobic pocket is specific for lymphocyte-specific kinase Lck. If a pharmacophore model is constructed to reflect that it contains a hydrophobic feature within that pocket (Fig. 4A). A pharmacophore based on a non-selective inhibitor does not contain a feature there but has the ability to also find molecules that bind to other kinases that do not have the side pocket (Fig. 4B and C) [129,130].

Since one pharmacophore model can only represent a single binding mode it is often necessary to create multiple models for a target to cover its active space [131]. Usually automatically generated models from a single crystallographic complex need to be further refined by using a test set of known binding compounds. Same as for docking it is vital to carefully curate the dataset and use only experimental data that shows direct ligand/target interactions, undisturbed by interactions with other targets or assay components, to arrive at a high-quality model [132]. The test set allows us to calculate several typical quality parameters that can be used to compare individual models and select the best one. The enrichment factor shows how a model increases the yield of actives compared to random selection. The Ef is the ratio between the number of true positives (tp) divided by all found virtual hits (vh) and the number of actives (A) divided by the total number of compounds in the database (N) [133].

$$E_f = (tp/vh)/(A/N)$$

Model quality can also be represented in the ROC curve. In this case, the compounds are ranked by the fit value and all compounds that are not found by the model are counted with a fit value of zero and represented by a line to the upper right corner of the plot. In a final experimental validation step, the model is sometimes set to screen a large database of available compounds. The best hits are then biologically evaluated and the gained activity data can in turn be used to even further refine the model [134].

The application of pharmacophore modeling in lead optimization is often about identifying new interaction points that might increase the potency of an inhibitor. For example two pharmacophore models of similar inhibitors with the same binding site of glutamate carboxypeptidase II [135] help to understand why a single modification leading to a new hydrogen bond acceptor feature leads to an increase of potency (Fig. 5). In an optimization process, this new feature can then be used to select molecules with a better chance for high binding affinity.

Sometimes also the *de novo* design is guided by pharmacophore models, to evaluate synthesis proposals. To this end, computational structure generators aim to cover the chemically possible space and to generate structures that are synthetically feasible with a reasonable effort [136]. Instead of using databases with known and available molecules, databases generated by structure generators can be used to discover novel chemical entities. The program PhDD combines these methods by generating structures that fit a specific pharmacophore [137].

Pharmacophores have also been used to design multi-target ligands, by combining known pharmacophores of two targets, sometimes with computational methods [138], sometimes simply by creating chimeric molecules [138,139]. More examples of the pharmacophore modeling application in the hit to lead optimization could be found in the following section.

### 2.2.3. Examples of studies successfully integrating pharmacophore modeling to the hit to lead optimization process

Pharmacophore-based thinking is a common strategy in lead optimization, even if it is not always supported by computational methods (see Figs. 4 and 5). Many medicinal chemists simply divide a molecule into different areas and functionalities, to then modify them by replacing functional groups with moieties with similar functionality. Even though pharmacophore-based *in silico* modeling is not used in lead optimization as often as docking, several high-quality publications using pharmacophore models could be found. The reasons, why we have excluded most of the found studies are quite similar like in the case of the docking, i.e.: mostly they (1) used pharmacophore modeling just to elucidate the mechanism of action of already optimized compounds (2) the authors have only proposed the optimization, but haven't performed it (3) haven't used appropriate methodology to validate their modeling results, etc. Those reasons are again pointing towards user-based issues and not the limitations of the pharmacophore modeling. The following section is showing the successful application of pharmacophore modeling in the hit-to-lead optimization process.

This strategy was for example used to design a ligand for botulinum neurotoxin serotype A light chain (BoNT/A LC), a zinc metalloprotease that cleaves components of the SNARE (soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor). Proteolysis of SNARE proteins inhibits the exocytosis of acetylcholine into neuromuscular junctions and results in life-threatening flaccid paralysis [140]. Burnett et al. have focused on the pharmacophore-based identification and development of non-Zn(II)-coordinating small molecule, non-peptidic, inhibitors of BoNT/A LC [141]. The authors decided to use a pharmacophore modeling strategy since they have been unable to generate an X-ray co-crystal of their lead candidate in complex with the BoNT/

A LC and the synthetic chemistry guided by structure-based molecular docking studies has failed to provide a derivative possessing increased potency [142]. Due to the absence of reliable binding site models, they used three-dimensional search queries derived from their gas-phase pharmacophore for BoNT/A LC inhibition. This model defined separation between the overlaps of several different, non-zinc(II)-coordinating small molecule chemotypes and resulted in a new more potent structural hybrid possessing a  $K_i = 600$  nM [141].

In contrast, Fu et al. successfully integrated both *in silico* pharmacophore modeling, MD simulations, and co-crystallization studies [143]. They focused on the development of inhibitors targeting poly (ADP-ribose) polymerase-1 (PARP1). PARP1 is an enzyme involved in the self-repair of cellular DNA damage and its inhibitors are used for breast cancer therapy. In this study, the authors screened the DrugBank and ZINC databases (in total 40 215 compounds) via the docking and successfully co-crystallized one of the hit compounds. To explore how to modify and further optimize their hit compound, they constructed the structure-based pharmacophore model. This model was based on the co-crystal structures of their hit compound as well as ten other co-crystal structures of PARP1 inhibitors from the literature. Using structure-based pharmacophore features, they designed and synthesized new derivatives of their hit and tested their PARP1 inhibition activities. The first series of compounds had negligible effects on PARP1 activity and led the authors to the re-evaluation of the X-ray crystal structure and the pharmacophore model. The second round of the following synthesis has led to several compounds with significantly enhanced activity and finally to a compound with a novel chemical scaffold, unique binding interaction with PARP1 protein, and almost 20 fold lower  $IC_{50}$  in comparison to the parent compound. The binding mode of this compound was investigated with 10-ns MD simulations.

Crystal structures provide a “frozen” picture of a ligand–protein complex, a system that is, in fact, dynamic in its biological environment. This raises the question of how ligand–protein interactions change in the dynamic state. MD simulations aim to answer this question and are often combined with other methods to analyze changes in the binding pattern. MD simulations are frequently used to refine docking poses and elucidate binding affinity patterns [144], as also shown by Fu et al for PARP1 [143], but we can also find examples in their combination with pharmacophore modeling. Wieder et al investigated the differences between pharmacophore models built from the original protein–ligand complexes obtained from the PDB and pharmacophore models built on the final structures of an MD simulation [118,145]. They showed that the resulting models differ in feature number and type from those built on the original complexes. Furthermore, some of them displayed a better enrichment of active compounds than the original models. This shows that a combination of these techniques could lead to more successful models in the future.

Another study successfully integrating several *in silico* methods is a project by Shan and Zheng conducted in 2009 [146]. The authors aimed at the optimization of an inhibitor of a dishevelled PDZ domain, a potential cancer therapeutic target [147]. To this end, they have prepared the pharmacophore model of their previously identified hit compound and two non-binders. The pharmacophore derived from the hit complex structure was examined and essential components were selected based on the differences between their hit and two similar compounds that do not bind to the PDZ domain. Using this model, they screened the ChemDiv database with an algorithm that combines similarity search and docking procedures. The virtual hits were analyzed with pharmacophore modeling and the 15 selected virtual hits were examined with NMR spectroscopy for their binding affinity towards the dishevelled PDZ domain. All tested compounds showed improve-

ment over the original compound, the best of them with a 30-fold improvement in KD.

In a series of studies Schuster and Vuorinen built and then refined pharmacophore models of 11b-hydroxysteroid dehydrogenases (11b-HSD) 1 and 2. The 11b-hydroxysteroid dehydrogenases are enzymes regulating the intracellular availability of glucocorticoids and activation of glucocorticoid receptors and their inhibition has considerable therapeutic potential for glucocorticoid-associated diseases [148–150]. While the aim of these studies was not the lead optimization, they illustrate how a pharmacophore-based virtual screening approach can be refined with experimental data to optimally predict activity and selectivity for a specific target. At the beginning of Schusters and Vuorinens modeling studies, no X-ray crystal structure of 11b-HSD 1 was available. Accordingly, they employed ligand-based pharmacophore models as virtual screening tools for the identification of novel classes of 11b-HSD inhibitors [151]. First models with the ability to identify 11b-HSD inhibitors were generated and experimentally validated. The results of the validation run and more experimental data from virtual hits of the models were then used to evaluate them and to refine them. The final models had a higher hit rate and better selectivity among 11b-HSD subtypes [134,151,152].

### 3. Conclusion and perspectives

As shown by the variety of success stories published in the scientific literature molecular modeling techniques have the ability to support the drug development process at various stages. It is, however vital not to oversimplify the complex problems faced in drug development. The meaningful application of computational models requires careful curating of datasets, extensive model validation, and thorough analysis and contextualization of experimental results.

Even more, than in other CADD applications, SAR optimization requires high-quality standards that we would like to sum up in the following questions as a take-home message:

What is my target? Even though it seems redundant to stress this, it is vital to thoroughly familiarize oneself with the available literature on the target structure. Many studies suffer from poor background reading which often leads to mistakes that could have easily been avoided or the repetition of experiments that have already been conducted by other groups. Different binding sites, binding modes, and structural changes undergone by the protein upon binding to different ligands are usually well described in the publications related to the available structural data and can help to select the right workflow for a lead optimization problem.

How do the biological test systems work? The test systems that were and are used to gather activity data should be thoroughly scrutinized. It is very common that the applied assay systems do not directly reflect ligand/protein binding making it difficult to use the generated data for the optimization of an in silico workflow. Especially when multiple different binding modes are possible it becomes key to determine how they can be experimentally distinguished. It is also important to be aware of putative assay interference, often caused by PAINS and other well-known promiscuous motifs.

How good is my data? This again is a less self-explanatory point than one might think. The large amount of activity data collected in public databases like the ChEMBL (ebi.ac.uk) tempts the user to use automated scripts to create activity datasets for validation on the fast track. However, this often leads to low data quality in the set. Different assay types are mixed up and they contain data that is not representative for direct target ligand binding. Different binding sites might not always be investigated separately and

the intermixing of all these different types of data leads to contradictions and in the end poor models. A smaller high-quality dataset can lead to better results.

What are my software tools? There is a reason why non of the many available software solutions for docking and pharmacophore modeling has become dominant in the field. The success of algorithms and SFs hinges largely on their suitability for a specific project or target. Empirical SFs often outperform energy-based functions, when they are applied to structures that are also represented in the dataset they are trained with they might, however, fail if presented with a novel binding mode that is not sampled in the training data. If available, use multiple different programs and approaches and select the most suitable one in a thorough validation process. While this takes time, it leads to more high-quality simulations and results.

Is my workflow valid? Employ all available data to test and validate your workflow. If simple validation strategies like re-docking and cross-docking are performed, they vastly improve the reliability of any predictions made by the workflow.

Can my workflow be refined? Adapt and refine your computational models as new data becomes available. Build a system to learn from every compound that is tested. Build specialized SFs for your specific compound class. Find relevant key interactions and aim to elucidate the dominant binding orientation. In many cases also automated approaches can be used to analyze vast amounts of data and extract the relevant findings.

What else? Structure-based molecular modeling can be a very powerful and valuable tool in lead optimization if it is used with care and sound methodology. We hope the readers of this review can use it to find the best way to employ molecular modeling techniques in their own optimization projects.

### Funding

This research was funded by Czech Science Foundation, grant number 19-22269Y, the CAS (RVO: 86652036), and Austrian Science Fund (FWF) project “New ways to counter inflammation”, grant number T-942.

### CRediT authorship contribution statement

**Veronika Temml:** Writing - original draft, Investigation. **Zsofia Kutil:** Investigation, Visualization, Writing - review & editing.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### References

- [1] Topliss JG. Utilization of operational schemes for analog synthesis in drug design. *J Med Chem* 1972;15:1006–11.
- [2] Kubinyi H. Free Wilson Analysis. Theory, applications and its relationship to Hansch analysis. *Quant Struct-Act Rel* 1988;7:121–33.
- [3] Issa NT, Wathieu H, Ojo A, Byers SW, Dakshanamurthy S. Drug metabolism in preclinical drug development: a survey of the discovery process, toxicology, and computational tools. *Curr Drug Metab* 2017;18(6).
- [4] Schnecke V, Boström J. Computational chemistry-driven decision making in lead generation. *Drug Discov Today* 2006;11:43–50.
- [5] Jorgensen WL. Efficient drug lead discovery and optimization. *Acc Chem Res* 2009;42:724–33.
- [6] Levinthal C, Wodak SJ, Kahn P, Dadivanian AK. Hemoglobin interaction in sickle cell fibers. I: Theoretical approaches to the molecular contacts. *Proc Natl Acad Sci USA* 1975;72:1330–4.
- [7] Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 1982;161:269–88.

- [8] Rudden LSP, Degiacomi MT. Protein docking using a single representation for protein surface, electrostatics, and local dynamics. *J Chem Theory Comput* 2019;15(9):5135–43.
- [9] Leach AR. Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol* 1994;235:345–56.
- [10] Knegtel RM, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. *J Mol Biol* 1997;266:424–40.
- [11] Sousa SF, Fernandes PA, Ramos MJ. Protein–ligand docking: current status and future challenges. *Proteins: Struct., Funct Bioinf* 2006;65:15–26.
- [12] Liu M, Wang S. MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. *J Comput Aided Mol Des* 1999;13:435–51.
- [13] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 1998;19:1639–62.
- [14] Gardiner EJ, Willett P, Artymiuk PJ. Protein docking using a genetic algorithm. *Proteins: Struct., Funct Bioinf* 2001;44:44–56.
- [15] Liu J, Wang R. Classification of current scoring functions. *J Chem Inf Model* 2015;55:475–82.
- [16] Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. *Molecules* 2015;20:13384–421.
- [17] Li J, Fu A, Zhang L. An overview of scoring functions used for protein–ligand interactions in molecular docking. *Interdiscip Sci* 2019;11:320–8.
- [18] Heberlé G, de Azevedo Jr WF. Bio-inspired algorithms applied to molecular docking simulations. *Curr. Med. Chem.* 2011;18:1339–52.
- [19] Fernandez M, Caballero J, Fernandez L, Sarai A. Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM). *Mol Divers* 2011;15(1):269–89.
- [20] Dias R, de Azevedo Jr WF. Molecular docking algorithms. *Curr. Drug Targets.* 2008;9:1040–7.
- [21] Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* 2003;32:335–73.
- [22] Willett P. Genetic algorithms in molecular recognition and design. *Trends Biotechnol* 1995;13:516–21.
- [23] Li H, Sze K-H, Lu G, Ballester PJ. Machine-learning scoring functions for structure-based drug lead optimization. *WIREs Comput Mol Sci* 2020;10:e1465.
- [24] Shen C, Ding J, Wang Z, Cao D, Ding X, Hou T. From machine learning to deep learning: Advances in scoring functions for protein–ligand docking. *WIREs Comput Mol Sci* 2020;10(1). <https://doi.org/10.1002/wcms.v10.110.1002/wcms.1429>.
- [25] Adeniyi AA, Soliman MES. Implementing QM in docking calculations: is it a waste of computational time?. *Drug Discov Today* 2017;22:1216–23.
- [26] Liu Z, Su M, Han L, Liu J, Yang Q, Li Y, et al. Forging the basis for developing protein–ligand interaction scoring functions. *Acc Chem Res* 2017;50:302–9.
- [27] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res* 2000;28:235–42.
- [28] Kirchmair J, Markt P, Distinto S, Schuster D, Spitzer GM, Liedl KR, et al. The protein data bank (PDB), its related services and software tools as key components for in silico guided drug discovery. *J Med Chem* 2008;51:7021–40.
- [29] Friedrich N-O, Meyder A, de Bruyn Kops C, Sommer K, Flachsenberg F, Rarey M, et al. High-quality dataset of protein-bound ligand conformations and its application to benchmarking conformer ensemble generators. *J Chem Inf Model* 2017;57:529–39.
- [30] Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. *Biophys Rev* 2017;9:91–102.
- [31] Luo J, Wei W, Waldispühl J, Moitessier N. Challenges and current status of computational methods for docking small molecules to nucleic acids. *Eur J Med Chem* 2019;168:414–25.
- [32] Wang Z, Sun H, Yao X, Li D, Xu L, Li Y, et al. Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys Chem Chem Phys* 2016;18(18):12964–75.
- [33] Agrawal P, Singh H, Srivastava HK, Singh S, Kishore G, Raghava GPS. Benchmarking of different molecular docking methods for protein–peptide docking. *BMC Bioinf* 2019;19:426.
- [34] Weng G, Gao J, Wang Z, Wang E, Hu X, Yao X, et al. Comprehensive evaluation of fourteen docking programs on protein–peptide complexes. *J Chem Theory Comput* 2020;16(6):3959–69.
- [35] Çınaroğlu SS, Timuçin E. Comparative assessment of seven docking programs on a nonredundant metalloprotein subset of the PDBbind refined. *J Chem Inf Model* 2019;59:3846–59.
- [36] Hevener KE, Zhao W, Ball DM, Babaoglu K, Qi J, White SW, et al. Validation of molecular docking programs for virtual screening against dihydropteroate synthase. *J Chem Inf Model* 2009;49:444–60.
- [37] Castro-Alvarez A, Costa AM, Vilarrasa J. The performance of several docking programs at reproducing protein–macrolide-like crystal structures. *Molecules* 2017;22:136.
- [38] Mukherjee S, Balius TE, Rizzo RC. Docking validation resources: protein family and ligand flexibility experiments. *J Chem Inf Model* 2010;50(11):1986–2000.
- [39] The PyMOL Molecular Graphics System, Version 2.4.1, Schrödinger, LLC.
- [40] Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. *J Med Chem* 2006;49:6789–801.
- [41] Triballeau N, Acher F, Brabet I, Pin J-P, Bertrand H-O. Virtual screening workflow development guided by the “receiver operating characteristic” curve approach. application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem* 2005;48:2534–47.
- [42] Zhao Y, Li Y. Design of environmentally friendly neonicotinoid insecticides with bioconcentration tuning and Bi-directional selective toxic effects. *J Clean Prod* 2019;221:113–21.
- [43] Gu W, Li Q, Li Y. Environment-friendly PCN derivatives design and environmental behavior simulation based on a multi-activity 3D-QSAR model and molecular dynamics. *J Hazard Mater* 2020;393:122339. <https://doi.org/10.1016/j.jhazmat.2020.122339>.
- [44] Qiu Y, Zhang S, Li Y. High ultraviolet sensitivity of phthalic acid esters with environmental friendliness after modification through pharmacophore modeling associated with the solvation effect. *Pol J Environ Stud* 2020;29(3):2303–16.
- [45] Enyedy IJ, Egan WJ. Can we use docking and scoring for hit-to-lead optimization?. *J Comput Aided Mol Des* 2008;22:161–8.
- [46] Šinko G. Assessment of scoring functions and in silico parameters for AChE–ligand interactions as a tool for predicting inhibition potency. *Chem-Biol Interact* 2019;308:216–23.
- [47] Pein H, Ville A, Pace S, Temml V, Garscha U, Raasch M, et al. Endogenous metabolites of vitamin E limit inflammation by targeting 5-lipoxygenase. *Nat Commun* 2018;9:3834.
- [48] Cheung S-Y, Werner M, Esposito L, Troisi F, Cantone V, Liening S, et al. Discovery of a benzenesulfonamide-based dual inhibitor of microsomal prostaglandin E2 synthase-1 and 5-lipoxygenase that favorably modulates lipid mediator biosynthesis in inflammation. *Eur J Med Chem* 2018;156:815–30.
- [49] Levoín N, Calmels T, Poupardin-Olivier O, Labeuw O, Danvy D, Robert P, et al. Refined docking as a valuable tool for lead optimization: application to histamine h3 receptor antagonists. *Arch Pharm (Weinheim)* 2008;341(10):610–23.
- [50] Galindez G, Matschinske J, Rose TD, Sadegh S, Salgado-Albarrán M, Späth J, et al. Lessons from the COVID-19 pandemic for advancing computational drug repurposing strategies. *Nat Comput Sci* 2021;1:33–41.
- [51] Komura H, Watanabe R, Kawashima H, Ohashi R, Kuroda M, Sato T, Honma T, Mizuguchi K. (2021) A public–private partnership to enrich the development of in silico predictive models for pharmacokinetic and cardiotoxic properties, *Drug Discov. Today*. In Press.
- [52] Ekins S, Mottin M, Ramos PRPS, Sousa BKP, Neves BJ, Foil DH, et al. Déjà vu: stimulating open drug discovery for SARS-CoV-2. *Drug Discov Today* 2020;25:928–41.
- [53] Mitusińska K, Raczyńska A, Bzówka M, Bagrowska W, Góra A. Applications of water molecules for analysis of macromolecule properties. *Comput Struct Biotechnol J* 2020;18:355–65.
- [54] Banchi L, Fingerhuth M, Babej T, Ing C, Arrazola JM. Molecular docking with Gaussian Boson Sampling. *Science Advances* 2020;6: eaax1950.
- [55] Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, et al. End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem Rev* 2019;119(16):9478–508.
- [56] Bhatarai B, Walters WP, Hop C, Lanza G, Ekins S. Opportunities and challenges using artificial intelligence in ADME/Tox. *Nat Mater* 2019;18:418–22.
- [57] Kazmi SR, Jun R, Yu M-S, Jung C, Na D. In silico approaches and tools for the prediction of drug metabolism and fate: a review. *Comput Biol Med* 2019;106:54–64.
- [58] Ehmki ESR, Rarey M. Exploring structure–activity relationships with three-dimensional matched molecular pairs—a review. *Chem Med Chem* 2018;13:482–9.
- [59] Wang SH, Yu J. Structure-based design for binding peptides in anti-cancer therapy. *Biomaterials* 2018;156:1–15.
- [60] Réau M, Langenfeld F, Zagury J-F, Lagarde N, Montes M. Decoys selection in benchmarking datasets: overview and perspectives. *Front Pharmacol* 2018;9:11.
- [61] Sottriffer C. Docking of covalent ligands: challenges and approaches. *Mol Inform* 2018;37(9-10):1800062. <https://doi.org/10.1002/minf.v37.9-1010.1002/minf.201800062>.
- [62] Defelipe L, Arcon J, Modenutti C, Marti M, Turjanski A, Barril X. Solvents to fragments to drugs: MD applications in drug design. *Molecules* 2018;23(12):3269. <https://doi.org/10.3390/molecules23123269>.
- [63] Ciemny M, Kurcinski M, Kamel K, Koliński A, Alam N, Schueler-Furman O, et al. Protein–peptide docking: opportunities and challenges. *Drug Discov Today* 2018;23:1530–7.
- [64] Riccardi L, Genna V, De Vivo M. Metal–ligand interactions in drug design. *Nat Rev Chem* 2018;2:100–12.
- [65] Amaro RE, Baudry J, Chodera J, Demir Ö, McCammon JA, Miao Y, et al. Ensemble docking in drug discovery. *Biophys J* 2018;114:2271–8.
- [66] Salmaso V, Moro S. Bridging molecular docking to molecular dynamics in exploring ligand–protein recognition process: an overview. *Front Pharmacol* 2018:9.
- [67] Stefaniak F, Chudyk EI, Bodkin M, Dawson WK, Bujnicki JM. Modeling of ribonucleic acid–ligand interactions, *WIREs Comput. Mol Sci* 2015;5(6):425–39.
- [68] Kirchmair J, Göller AH, Lang D, Kunze J, Testa B, Wilson ID, et al. Predicting drug metabolism: experiment and/or computation?. *Nat Rev Drug Discov* 2015;14(6):387–404.

- [69] Granchi C, Rizzolio F, Palazzolo S, Carmignani S, Macchia M, Saccomanni G, et al. Structural optimization of 4-chlorobenzoylpiperidine derivatives for the development of potent, reversible, and selective monoacylglycerol lipase (MAGL) inhibitors. *J Med Chem* 2016;59:10299–314.
- [70] Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* 2010;53:2719–40.
- [71] Baell JB, Nissink JWM. Seven year itch: pan-assay interference compounds (PAINS) in 2017—utility and limitations. *ACS Chem Biol* 2018;13(1):36–44.
- [72] Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res* 2015;43:W443–7.
- [73] Stornaiuolo M, La Regina G, Passacantilli S, Grassia G, Coluccia A, La Pietra V, et al. Structure-based lead optimization and biological evaluation of BAX direct activators as novel potential anticancer agents. *J Med Chem* 2015;58:2135–48.
- [74] Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep* 2017;7:42717.
- [75] Djombou-Feunang Y, Fiamoncini J, Gil-de-la-Fuente A, Greiner R, Manach C, Wishart DS. BioTransformer: a comprehensive computational tool for small molecule metabolism prediction and metabolite identification. *J Cheminform* 2019;11:2.
- [76] Ramachandran SA, Jadhavar PS, Miglani SK, Singh MP, Kalane DP, Agarwal AK, et al. Design, synthesis and optimization of bis-amide derivatives as CSF1R inhibitors. *Bioorg Med Chem Lett* 2017;27:2153–60.
- [77] Kawada H, Ebiike H, Tsukazaki M, Nakamura M, Morikami K, Yoshinari K, et al. Lead optimization of a dihydropyridopyrimidine inhibitor against phosphoinositide 3-kinase (PI3K) to improve the phenol glucuronic acid conjugation. *Bioorg Med Chem Lett* 2013;23:673–8.
- [78] Giustiniano M, Daniele S, Pelliccia S, La Pietra V, Pietrobbono D, Brancaccio D, et al. Computer-aided identification and lead optimization of dual murine double minute 2 and 4 binders: structure-activity relationship studies and pharmacological activity. *J Med Chem* 2017;60:8115–30.
- [79] Snow O, Lallous N, Singh K, Lack N, Rennie P, Cherkasov A. Androgen receptor plasticity and its implications for prostate cancer therapy. *Cancer Treat Rev* 2019;81.
- [80] Ban F, Dalal K, Li H, LeBlanc E, Rennie PS, Cherkasov A. Best practices of computer-aided drug discovery: lessons learned from the development of a preclinical candidate for prostate cancer with a new mechanism of action. *J Chem Inf Model* 2017;57(5):1018–28.
- [81] Dalal K, Ban F, Li H, Morin H, Roshan-Moniri M, Tam KJ, et al. Selectively targeting the dimerization interface of human androgen receptor with small-molecules to treat castration-resistant prostate cancer. *Cancer Lett* 2018;437:35–43.
- [82] Lyu J, Wang S, Balias TE, Singh I, Levit A, Moroz YS, et al. Ultra-large library docking for discovering new chemotypes. *Nature* 2019;566:224–9.
- [83] Trani G, Barker JJ, Bromidge SM, Brookfield FA, Burch JD, Chen Y, et al. Design, synthesis and structure-activity relationships of a novel class of sulfonylpyridine inhibitors of Interleukin-2 inducible T-cell kinase (ITK). *Bioorg Med Chem Lett* 2014;24(24):5818–23.
- [84] Alder CM, Ambler M, Campbell AJ, Champigny AC, Deakin AM, Harling JD, et al. Identification of a novel and selective series of Itk inhibitors via a template-hopping strategy. *ACS Med Chem Lett* 2013;4:948–52.
- [85] Zou Y, Li L, Chen W, Chen T, Ma L, Wang X, et al. Virtual screening and structure-based discovery of indole acylguanidines as potent beta-secretase (BACE1) inhibitors. *Molecules* 2013;18:5706–22.
- [86] Joseph-McCarthy D, Parris K, Huang A, Failli A, Quagliato D, Dushin EG, et al. Use of structure-based drug design approaches to obtain novel anthranilic acid acyl carrier protein synthase inhibitors. *J Med Chem* 2005;48:7960–9.
- [87] Hoffer L, Voitovich YV, Raux B, Carrasco K, Muller C, Fedorov AY, et al. Integrated strategy for lead optimization based on fragment growing: the diversity-oriented-target-focused-synthesis approach. *J Med Chem* 2018;61:5719–32.
- [88] Yan S, Elmes MW, Tong S, Hu K, Awwa M, Teng GYH, et al. SAR studies on truxillic acid mono esters as a new class of antinociceptive agents targeting fatty acid binding proteins. *Eur J Med Chem* 2018;154:233–52.
- [89] Ranise A, Spallarossa A, Schenone S, Bruno O, Bondavalli F, Vargiu L, et al. Design, synthesis, SAR, and molecular modeling studies of acylthiocarbamates: a novel series of potent non-nucleoside HIV-1 reverse transcriptase inhibitors structurally related to phenethylthiazolylthiourea derivatives. *J Med Chem* 2003;46:768–81.
- [90] Mai A, Sbardella G, Artico M, Ragno R, Massa S, Novellino E, et al. Structure-based design, synthesis, and biological evaluation of conformationally restricted novel 2-alkylthio-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as non-nucleoside inhibitors of HIV-1 reverse transcriptase. *J Med Chem* 2001;44:2544–54.
- [91] Jeankumar VU, Renuka J, Kotagiri S, Saxena S, Kakan SS, Sridevi JP, et al. Gyrase ATPase domain as an antitubercular drug discovery platform: structure-based design and lead optimization of nitrothiazolyl carboxamide analogues. *Chem Med Chem* 2014;9:1850–9.
- [92] Wu P, Ma D-L, Leung C-H, Yan S-C, Zhu N, Abagyan R, et al. Stabilization of G-quadruplex DNA with platinum(II) Schiff base complexes: luminescent probe and down-regulation of c-myc oncogene expression. *Chemistry* 2009;15(47):13008–21.
- [93] Crawford TD, Ndubaku CO, Chen H, Boggs JW, Bravo BJ, Delatorre K, et al. Discovery of selective 4-Amino-pyridopyrimidine inhibitors of MAP4K4 using fragment-based lead identification and optimization. *J Med Chem* 2014;57:3484–93.
- [94] Poulsen A, Williams M, Nagaraj HM, William AD, Wang H, Soh CK, et al. Structure-based optimization of morpholino-triazines as PI3K and mTOR inhibitors. *Bioorg Med Chem Lett* 2012;22:1009–13.
- [95] Zhu J, Huang JW, Tseng PH, Yang YT, Fowble J, Shiau CW, et al. From the cyclooxygenase-2 inhibitor celecoxib to a novel class of 3-phosphoinositide-dependent protein kinase-1 inhibitors. *Cancer Res* 2004;64:4309–18.
- [96] Xia Z, Knaak C, Ma J, Beharry ZM, McInnes C, Wang W, et al. Synthesis and evaluation of novel inhibitors of Pim-1 and Pim-2 protein kinases. *J Med Chem* 2009;52:74–86.
- [97] Chen Y, Wen D, Huang Z, Huang M, Luo Y, Liu B, et al. 2-(4-Chlorophenyl)-2-oxoethyl 4-benzamidobenzoate derivatives, a novel class of SENP1 inhibitors: virtual screening, synthesis and biological evaluation. *Bioorg Med Chem Lett* 2012;22(22):6867–70.
- [98] Iwata Y, Arisawa M, Hamada R, Kita Y, Mizutani MY, Tomioka N, et al. Discovery of novel aldose reductase inhibitors using a protein structure-based approach: 3D-database search followed by design and synthesis. *J Med Chem* 2001;44:1718–28.
- [99] Ramunno A, Cosconati S, Sartini S, Maglio V, Angiuoli S, La Pietra V, et al. Progresses in the pursuit of aldose reductase inhibitors: the structure-based lead optimization step. *Eur J Med Chem* 2012;51:216–26.
- [100] Rakse M, Karthikeyan C, Deora GS, Moorthy NS, Rathore V, Rawat AK, et al. Design, synthesis and molecular modelling studies of novel 3-acetamido-4-methyl benzoic acid derivatives as inhibitors of protein tyrosine phosphatase 1B. *Eur J Med Chem* 2013;70:469–76.
- [101] Moretto AF, Kirincich SJ, Xu WX, Smith MJ, Wan ZK, Wilson DP, et al. Bicyclic and tricyclic thiophenes as protein tyrosine phosphatase 1B inhibitors. *Bioorg Med Chem* 2006;14:2162–77.
- [102] Wityak J, McGee KF, Conlon MP, Song RH, Duffy BC, Clayton B, et al. Lead optimization toward proof-of-concept tools for huntington's disease within a 4-(1h-pyrazol-4-yl)pyrimidine class of Pan-JNK inhibitors. *J Med Chem* 2015;58:2967–87.
- [103] Bajaj K, Burudkar S, Shah P, Keche A, Ghosh U, Tannu P, et al. Lead optimization of isocytosine-derived xanthine oxidase inhibitors. *Bioorg Med Chem Lett* 2013;23:834–8.
- [104] Khanna S, Burudkar S, Bajaj K, Shah P, Keche A, Ghosh U, et al. Isocytosine-based inhibitors of xanthine oxidase: design, synthesis, SAR, PK and in vivo efficacy in rat model of hyperuricemia. *Bioorg Med Chem Lett* 2012;22(24):7543–6.
- [105] Kumar G, Parasuraman P, Sharma SK, Banerjee T, Karmodiya K, Suroliya N, et al. Discovery of a rhodanine class of compounds as inhibitors of Plasmodium falciparum enoyl-acyl carrier protein reductase. *J Med Chem* 2007;50:2665–75.
- [106] La Pietra V, Marinelli L, Cosconati S, Di Leva FS, Nuti E, Santamaria S, et al. Identification of novel molecular scaffolds for the design of MMP-13 inhibitors: a first round of lead optimization. *Eur J Med Chem* 2012;47:143–52.
- [107] Wu B, Murray JK, Andrews KL, Sham K, Long J, Aral J, et al. Discovery of tarantula venom-derived NaV1.7-inhibitory jzTx-V peptide 5-Br-Trp24 analogue AM-6120 with systemic block of histamine-induced pruritis. *J Med Chem* 2018;61:9500–12.
- [108] Yang S-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov. Today* 2010;15(11-12):444–50.
- [109] Wermuth CG, Ganellin R, Lindberg P, Mitscher LA. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998); 1998, 70, 1129.
- [110] Wolber G, Langer T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model* 2005;45:160–9.
- [111] Schneider G, Neidhart W, Giller T, Schmid G. "Scaffold-hopping" by topological pharmacophore search: a contribution to virtual screening. *Angew Chem Int* 1999;38:2894–6.
- [112] Schaller D, Šribar D, Noonan T, Deng L, Nguyen TN, Pach S, et al. Next generation 3D pharmacophore modeling. *WIREs Comput Mol Sci* 2020;10:e1468.
- [113] Dixon SL, Smondryev AM, Rao SN. PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des* 2006;67(5):370–2.
- [114] Temml V, Kaserer T, Kutil Z, Landa P, Vanek T, Schuster D. Pharmacophore modeling for COX-1 and -2 inhibitors with LigandScout in comparison to Discovery Studio. *Future Med Chem* 2014;6:1869–81.
- [115] Kim M, Park K, Kim W, Jung S, Cho AE. Target-specific drug design method combining deep learning and water pharmacophore. *J Chem Inf Model* 2021;61:36–45.
- [116] Brown BP, Mendenhall J, Geanes AR, Meiler J. General Purpose Structure-Based drug discovery neural network score functions with human-interpretable pharmacophore maps. *J. Chem. Inf. Model*; 2021.
- [117] Seidel T, Wieder O, Garon A, Langer T. Applications of the pharmacophore concept in natural product inspired drug design. *Mol Inform* 2020;39:2000059.
- [118] Wieder M, Garon A, Perricone U, Boresch S, Seidel T, Almerico AM, et al. Common hits approach: combining pharmacophore modeling and molecular dynamics simulations. *J Chem Inf Model* 2017;57(2):365–85.

- [119] Hu Y, Stumpfe D, Bajorath J. Recent advances in scaffold hopping. *J Med Chem* 2017;60:1238–46.
- [120] Putz MV, Duda-Seiman C, Duda-Seiman D, Putz A-M, Alexandrescu I, Mernea M, et al. Chemical structure-biological activity models for pharmacophores' 3D-interactions. *Int J Mol Sci* 2016;17:1087.
- [121] Van Drie JH. Generation of three-dimensional pharmacophore models. *WIREs Comput Mol Sci* 2013;3(5):449–64.
- [122] Guner OF, Bowen JP. Pharmacophore modeling for ADME. *Curr Top Med Chem* 2013;13:1327–42.
- [123] Thai KM, Ngo TD, Tran TD, Le MT. Pharmacophore modeling for antitargets. *Curr Top Med Chem* 2013;13:1002–14.
- [124] Sanders MPA, McGuire R, Roumen L, de Esch JJP, de Vlieg J, Klomp JPG, et al. From the protein's perspective: the benefits and challenges of protein structure-based pharmacophore modeling. *Med Chem Comm* 2012;3(1):28–38.
- [125] Caporuscio F, Tafi A. Pharmacophore modelling: a forty year old approach and its modern synergies. *Curr Med Chem* 2011;18:2543–53.
- [126] Shim J, Mackerell Jr AD. Computational ligand-based rational design: role of conformational sampling and force fields in model development. *Med Chem Comm* 2011;2:356–70.
- [127] Wolber G, Seidel T, Bendix F, Langer T. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov Today* 2008;13:23–9.
- [128] Ekins S, Mirny L, Schuetz EG. A ligand-based approach to understanding selectivity of nuclear hormone receptors PXR, CAR, FXR, LXR $\alpha$ , and LXR $\beta$ . *Pharm Res* 2002;19:1788–800.
- [129] Zhu X, Kim JL, Newcomb JR, Rose PE, Stover DR, Toledo LM, et al. Structural analysis of the lymphocyte-specific kinase Lck in complex with non-selective and Src family selective kinase inhibitors. *Structure* 1999;7:651–61.
- [130] Lawrie AM, Noble MEM, Tunnah P, Brown NR, Johnson LN, Endicott JA. Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with CDK2. *Nat Struct Biol* 1997;4:796–801.
- [131] Vuorinen A, Schuster D. Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods* 2015;71:113–34.
- [132] Kirchmair J, Markt P, Distinto S, Wolber G, Langer T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—what can we learn from earlier mistakes?. *J Comput Aided Mol Des* 2008;22(3–4):213–28.
- [133] Jacobsson M, Lidén P, Stjernschantz E, Boström H, Norinder U. Improving structure-based virtual screening by multivariate analysis of scoring data. *J Med Chem* 2003;46:5781–9.
- [134] Vuorinen A, Nashev LG, Odermatt A, Rollinger JM, Schuster D. Pharmacophore model refinement for 11 $\beta$ -hydroxysteroid dehydrogenase inhibitors: search for modulators of intracellular glucocorticoid concentrations. *Mol Inform* 2014;33(1):15–25.
- [135] Kim K, Kwon H, Barinka C, Motlova L, Nam S, Choi D, et al. Novel  $\beta$ - and  $\gamma$ -amino acid-derived inhibitors of prostate-specific membrane antigen. *J Med Chem* 2020;63:3261–73.
- [136] Schneider G, Fechner U. Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov* 2005;4(8):649–63.
- [137] Huang Q, Li L-L, Yang S-Y. PhDD: A new pharmacophore-based de novo design method of drug-like molecules combined with assessment of synthetic accessibility. *J Mol Graph Model* 2010;28(8):775–87.
- [138] Temml V, Garscha U, Romp E, Schubert G, Gerstmeier J, Kutil Z, et al. Discovery of the first dual inhibitor of the 5-lipoxygenase-activating protein and soluble epoxide hydrolase using pharmacophore-based virtual screening. *Sci Rep* 2017;7(1). <https://doi.org/10.1038/srep42751>.
- [139] Meirer K, Steinhilber D, Proschak E. Inhibitors of the Arachidonic Acid Cascade: Interfering with Multiple Pathways; 2014. 114, 83–91.
- [140] Brunger AT, Jin R, Breidenbach MA. Highly specific interactions between botulinum neurotoxins and synaptic vesicle proteins. *Cell Mol Life Sci* 2008;65:2296–306.
- [141] Burnett JC, Wang C, Nuss JE, Nguyen TL, Hermone AR, Schmidt JJ, et al. Pharmacophore-guided lead optimization: the rational design of a non-zinc coordinating, sub-micromolar inhibitor of the botulinum neurotoxin serotype A metalloprotease. *Bioorg Med Chem Lett* 2009;19(19):5811–3.
- [142] Hermone AR, Burnett JC, Nuss JE, Tressler LE, Nguyen TL, Solaja BA, et al. Three-dimensional database mining identifies a unique chemotype that unites structurally diverse botulinum neurotoxin serotype A inhibitors in a three-zone pharmacophore. *Chem Med Chem* 2008;3:1905–12.
- [143] Fu L, Wang S, Wang X, Wang P, Zheng Y, Yao D, et al. Crystal structure-based discovery of a novel synthesized PARP1 inhibitor (OL-1) with apoptosis-inducing mechanisms in triple-negative breast cancer. *Sci Rep* 2016;6(1). <https://doi.org/10.1038/s41598-016-0007-2>.
- [144] Stjernschantz E, Marelus J, Medina C, Jacobsson M, Vermeulen NPE, Oostenbrink C. Are automated molecular dynamics simulations and binding free energy calculations realistic tools in lead optimization? an evaluation of the linear interaction energy (LIE) method. *J Chem Inf Model* 2006;46(5):1972–83.
- [145] Wieder M, Perricone U, Seidel T, Boresch S, Langer T. Comparing pharmacophore models derived from crystal structures and from molecular dynamics simulations. *Monatsh Chem* 2016;147:553–63.
- [146] Shan J, Zheng JJ. Optimizing Dvl PDZ domain inhibitor by exploring chemical space. *J Comput Aided Mol Des* 2009;23:37–47.
- [147] Shan J, Shi D-L, Wang J, Zheng J. Identification of a specific inhibitor of the dishevelled PDZ domain. *Biochemistry* 2005;44:15495–503.
- [148] Masuzaki H, Yamamoto H, Kenyon CJ, Elmquist JK, Morton NM, Paterson JM, et al. Transgenic amplification of glucocorticoid action in adipose tissue causes high blood pressure in mice. *J Clin Invest* 2003;112(1):83–90.
- [149] Masuzaki H, Paterson J, Shinyama H, Morton NM, Mullins JJ, Seckl JR, et al. A transgenic model of visceral obesity and the metabolic syndrome. *Science* 2001;294:2166–70.
- [150] Paterson JM, Morton NM, Fievet C, Kenyon CJ, Holmes MC, Staels B, et al. Metabolic syndrome without obesity: Hepatic overexpression of 11 $\beta$ -hydroxysteroid dehydrogenase type 1 in transgenic mice. *Proc Natl Acad Sci USA* 2004;101:7088–93.
- [151] Schuster D, Maurer EM, Laggner C, Nashev LG, Wilckens T, Langer T, et al. The discovery of new 11 $\beta$ -hydroxysteroid dehydrogenase type 1 inhibitors by common feature pharmacophore modeling and virtual screening. *J Med Chem* 2006;49:3454–66.
- [152] Kratschmar DV, Vuorinen A, Da Cunha T, Wolber G, Classen-Houben D, Doblhoff O, et al. Characterization of activity and binding mode of glycyrrhetic acid derivatives inhibiting 11 $\beta$ -hydroxysteroid dehydrogenase type 2. *J Steroid Biochem Mol Biol* 2011;125:129–42.
- [153] Chiang YK, Kuo CC, Wu YS, Chen CT, Coumar MS, Wu JS, et al. Generation of ligand-based pharmacophore model and virtual screening for identification of novel tubulin inhibitors with potent anticancer activity. *J Med Chem* 2009;52:4221–33.
- [154] Stanton RA, Lu X, Detorio M, Montero C, Hammond ET, Ehteshami M, et al. Discovery, characterization, and lead optimization of 7-azaindole non-nucleoside HIV-1 reverse transcriptase inhibitors. *Bioorg Med Chem Lett* 2016;26:4101–5.
- [155] Zhu K, Jiang C, Tao H, Liu J, Zhang H, Luo C. Identification of a novel selective small-molecule inhibitor of protein arginine methyltransferase 5 (PRMT5) by virtual screening, resynthesis and biological evaluations. *Bioorg Med Chem Lett* 2018;28(9):1476–83.
- [156] Yang LL, Li GB, Yan HX, Sun QZ, Ma S, Ji P, et al. Discovery of N6-phenyl-1H-pyrazolo[3,4-d]pyrimidine-3,6-diamine derivatives as novel CK1 inhibitors using common-feature pharmacophore model based virtual screening and hit-to-lead optimization. *Eur J Med Chem* 2012;56:30–8.
- [157] Ranganathan A, Stoddart LA, Hill SJ, Carlsson J. Fragment-based discovery of subtype-selective adenosine receptor ligands from homology models. *J Med Chem* 2015;58:9578–90.
- [158] Da Settimo F, Primofiore G, Taliani S, Marini AM, La Motta C, Novellino E, et al. 3-Aryl[1,2,4]triazino[4,3-a]benzimidazol-4(10H)-ones: a new class of selective A1 adenosine receptor antagonists. *J Med Chem* 2001;44:316–27.
- [159] Huang P, Loew GH, Funamizu H, Mimura M, Ishiyama N, Hayashida M, et al. Rational design, discovery, and synthesis of a novel series of potent growth hormone secretagogues. *J Med Chem* 2001;44:4082–91.
- [160] Chhabria MT, Brahmshatriya PS, Mahajan BM, Darji UB, Shah GB. Discovery of novel acyl coenzyme a: cholesterol acyltransferase inhibitors: pharmacophore-based virtual screening, synthesis and pharmacology. *Chem Biol Drug Des* 2012;80:106–13.
- [161] Choong IC, Lew W, Lee D, Pham P, Burdett MT, Lam JW, et al. Identification of potent and selective small-molecule inhibitors of caspase-3 through the use of extended tethering and structure-based drug design. *J Med Chem* 2002;45:5005–22.
- [162] Tian S, Wang X, Li L, Zhang X, Li Y, Zhu F, et al. Discovery of novel and selective adenosine A2A receptor antagonists for treating Parkinson's disease through comparative structure-based virtual screening. *J Chem Inf Model* 2017;57:1474–87.