



Mini Review

Computational approaches in target identification and drug discovery

Theodora Katsila^{a,1}, Georgios A. Spyroulias^a, George P. Patrinos^{a,b}, Minos-Timotheos Matsoukas^{a,*,1}^a University of Patras, School of Health Sciences, Department of Pharmacy, University Campus, Rion, Patras, Greece^b Department of Pathology, College of Medicine and Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates

ARTICLE INFO

Article history:

Received 31 January 2016

Received in revised form 21 April 2016

Accepted 25 April 2016

Available online 7 May 2016

Keywords:

Data integration

Information technologies

Target identification

Computer-aided drug discovery

ABSTRACT

In the big data era, voluminous datasets are routinely acquired, stored and analyzed with the aim to inform biomedical discoveries and validate hypotheses. No doubt, data volume and diversity have dramatically increased by the advent of new technologies and open data initiatives. Big data are used across the whole drug discovery pipeline from target identification and mechanism of action to identification of novel leads and drug candidates. Such methods are depicted and discussed, with the aim to provide a general view of computational tools and databases available. We feel that big data leveraging needs to be cost-effective and focus on personalized medicine. For this, we propose the interplay of information technologies and (chemo)informatic tools on the basis of their synergy.

© 2016 Katsila et al. Published by Elsevier B.V. on behalf of the Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

1. Introduction	177
2. Target identification	178
3. Target validation	179
4. Computer-aided drug design	179
4.1. Ligand-based CADD	179
4.2. Structure-based CADD	180
5. Summary and outlook	182
Acknowledgements	182
References	182

1. Introduction

Current trends in drug discovery focus on disease mechanisms and their understanding, followed by target identification and lead compound discovery. In the era of personalized medicine and better-informed cost-effective public health outcomes, a system of personalized medicine that is based on molecular states (and changes, from DNA to RNA to protein) have become fundamental in drug discovery [1,2]. To build such a system, the molecular characterization of disease is necessary, while environmental influences and the gut microbiome

needs to be also considered [3,4]. At the same time, regulatory requirements of safety are increasing [5].

To address the above-mentioned interplay in high-throughput formats, we feel that information technologies and chemoinformatic tools need to be employed on the basis of a synergy that even extends to artificial and human intelligence interplay - humans can detect patterns, which computer algorithms may fail to do so, whereas data-intensive and cognitively complex settings and processes limit human ability [6]. We propose that this synergy will (i) facilitate collaborative data analysis and (ii) guide sense- and decision-making towards rapid and efficient data output. Big and diverse data demand strict filtering and thorough analysis and interpretation. At the same time, biomedicine scientists need to efficiently and effectively collaborate and make decisions. For this, large-scale volumes of complex multi-faceted data need to be meaningfully assembled, mined and analyzed [7]. In such a context, reliable target identification and validation in cooperation

* Corresponding author at: University of Patras, School of Health Sciences, Department of Pharmacy, University Campus, Rion GR-265 04, Patras, Greece. Tel.: +30 2610 962351; fax: +30 2610 997693.

E-mail address: minosmatsoukas@upatras.gr (M.-T. Matsoukas).

¹ These authors contributed equally to this work.

Table 1
Drug targets and computational methods used for compound identification and interaction prediction.

Drug target	Computational approach	Reference
p38 α MAP kinase	Ligand-based interaction fingerprint (LIFt)	[14]
GPR17	Protein ligand interaction fingerprints (PLIF) method	[15]
Transforming growth factor- β 1 receptor kinase (TGF β)	Shape-based screening (CatShape, Catalyst)	[49]
T-type calcium channel (CaV)	Bidimensional pharmacophoric fingerprints (ChemAxon and CCG's GpiDAPH3 fingerprints)	[68]
Metabotropic glutamate receptor 5 (mGlu5)	Artificial neural network (ANN) quantitative structure–activity relationship (QSAR)	[72]
prostaglandin D2 receptor 2 (CRTH2)	Proteochemometrics modeling (PCM)	[76]
HUMAN immunodeficiency virus 1 reverse transcriptase (HIV-1 RT)	Molecular mechanics energies combined with the Poisson-Boltzmann surface area (MM-PBS)	[84]
Biotin	Molecular dynamics/free energy perturbation (FEP)	[85]
β -Secretase (BACE)	Linear interaction energy (LIE)	[86]
Chemo-attractant receptor (OXE-R)	Docking virtual screening (PyP \times and AutoDock Vina)	[90]
Angiotensin II receptor type 1 (AT1)	Ligand based pharmacophore modeling (Catalyst)	[91]
Pim-1 kinase	Docking virtual screening (Glide)	[104]
Epidermal growth factor receptor (EGFR)/Bromodomain-containing protein 4 (BRD4)	Docking virtual screening (Glide)	[105]
Calcineurin	Structure based pharmacophore virtual screening (Discovery Studio)	[106]

with drug discovery methods will pave the way to more efficient computer aided drug discovery. Moreover, new network-based computational models and systems biology integrate omics databases and optimize combinational regimens of drug development.

2. Target identification

Chemoinformatic tools present a tremendous potential to advance *in silico* drug design and discovery, as they serve the integration of information in several levels to enhance the reliability of data outcomes. To name a few, chemical structure similarity searching [8], data mining/machine learning [9], panel docking [10], and bioactivity spectra based algorithms [11] have been routinely and successfully implemented [12,13]. Some examples are the ligand-based interaction fingerprint (LIFt) approach [14] in predicting potential targets for small-molecule drugs using physics-based docking and sampling methods and the protein ligand interaction fingerprints (PLIF) method [15] for summarizing interactions between ligands and proteins using a fingerprint scheme. In both cases, compounds were identified for the p38 α MAP kinase and GPR17, respectively (Table 1).

Target identification can also be studied through network-based drug discovery, a field integrating different levels of information in drug-protein and protein-disease networks. This approach involves a highly collaborative scheme between databases and correlations across genomics, transcriptomics, proteomics, metabolomics, microbiome, pharmacogenomics, which highly depends on the development of relevant computational and systems biology tools for such data interpretation [16,17]. Such approaches, for example relating pharmacological and genomic spaces can be used to develop computational frameworks for drug target identification [18]. Another recent network-based application was the integration of large-scale structural genomics and disease association studies, to generate three-dimensional human interactome, that resulted in the identification of candidate genes for unknown disease-to-gene associations with proposed molecular mechanisms [19].

To facilitate gaining in-depth knowledge of disease mechanisms and/or phenotypes information technologies are greatly needed today more than ever [20]. Indeed, the study of disease mechanisms and/or phenotypes has turned from investigating a particular gene or protein into the analysis of entire sets of biomolecules [21]. The advent of omics technologies further complicates storing, visualizing and analyzing voluminous biological data. For this, information technologies provide the means towards extensive data processing and interpretation. Tools such as the human metabolome database [22] and MetaboAnalyst [23] support integrative omics pathway analysis. The human metabolome database contains metabolite entries linked with chemical, clinical, and molecular biology data, that can assist applications in metabolomics, clinical chemistry and biomarker discovery. Metaboanalyst is a web-based analytical pipeline for high-throughput metabolomics

studies, which offers a variety of procedures for metabolomic data processing and integrates biomarker and pathway analysis. MAGENTA (<http://www.broadinstitute.org/mpg/magenta/>) and Ingenuity (<http://www.ingenuity.com/>) users can further exploit several curated biological pathways. Databases play a key role and no doubt, an extremely rich repertoire is available today (Table 2). When kinome is of interest, a computational platform ReKINect has been recently reported to identify network-attacking mutations and validated with the interpretation of exomes and quantitative proteomes of ovarian cancer cell lines and the global cancer genome repository [24]. Another useful approach helping to identify functional connections between diseases, genes and drugs is the Connectivity Map [25]. Connectivity Map is a collection of genome-wide transcriptional expression data from cultured human cells treated with bioactive small molecules and simple pattern-matching algorithms that together enable the discovery of functional connections between drugs, genes and diseases through the transitory feature of common gene-expression changes [26]. Other computational methods have been also applied to reconstruct biological networks and extract information from them, such as Bayesian [27] and Boolean networks [28] and graph based models [29].

Furthermore, applications and web services, enable sharing of data and resources for visualization and analysis purposes. The Biological General Repository for Interaction Datasets (BioGRID) [30] is an interaction repository with compiled biological data freely available in standardized formats, linked with software platforms for visualization of complex interaction networks such as Osprey [31] and Cytoscape [32]. BioMart, is a community-driven project, which call for scientists to

Table 2
Web-accessible databases for drug target identification.

Utility	Url
Human metabolome data	http://www.hmdb.ca
<i>In silico</i> target identification	http://www.dddc.ac.cn/pdtd/ http://www.genome.jp/kegg/ http://www.geneontology.org http://www.reactome.org
Pathway analysis	http://www.pantherdb.org http://www.biocarta.com http://www.ingenuity.com/ http://www.ebi.ac.uk/chembl/db
Chemogenomic data	http://pubchem.ncbi.nlm.nih.gov
Drug target database	http://www.drugbank.ca
Protein data bank	http://www.pdb.org
Disease specific target database	http://thomsonreuters.com/metacore
Pharmacogenomic data	http://www.pharmgkb.org
Multi-level drug data	http://r2d2drug.org/DMC.aspx
Comparative toxicogenomic database	http://ctdbase.org
Target-toxin database	http://www.t3db.org
Protein expression information	http://www.proteinatlas.org
Therapeutics target database	http://bidd.nus.edu.sg/group/cjttd/

share data and provides free software and data services to the scientific community in order to facilitate scientific collaborations and the scientific discovery process [33]. OncoPrint, is a cancer microarray database and web-based data-mining platform aimed at facilitating discovery from genome-wide expression analyses, providing with query and visualization tools for selected or multiple genes across all analyses [34]. The online Cancer-Related Analysis of Variants Toolkit (CRAVAT) can assist the high-throughput assessment and prioritization of genes and missense alterations important for cancer tumorigenesis, by providing predictive scores for germline variants, somatic mutations and relative gene importance [35]. The Sorting Intolerant from Tolerant (SIFT) algorithm predicts the effect of coding variants on protein function through a web server. It provides users with predictions on their variants and is widely used for characterizing missense variations [36]. PROVEAN (Protein Variation Effect Analyzer) is a software tool which predicts whether an amino acid substitution or indel has an impact on the biological function of a protein [37] and GenePattern provides with analytical tools for the analysis of gene expression, sequence variation and network analysis. MetaMapR, an open source software integrates enzymatic transformations with metabolite structural similarity, mass spectral similarity and empirical associations to generate connected metabolic networks [38]. Protein and DNA visualization software such as VMD [39] and Chimera [40] are widely used in the 3D analysis of biomolecules and drug interactions. Several networks (caBIG, <http://cabig.cancer.gov>; BIRN, <http://www.nbirn.net>) and projects (Genotype-Tissue Expression Project [41]; RD-Connect [42]) have been initiated towards data exchange for target identification.

Another important field in information technologies is the semantics field, which could give insights to associations between heterogeneous data of diseases and drug targets. Such network-based computational approaches have gained popularity recently, proposing novel therapeutic targets and deciphering disease mechanisms. However, little effort has been devoted to investigating associations among drugs, diseases, and genes in an integrative manner. In such a study, Zhang et al., constructed an association network by extracting pair-wise associations between diseases, drugs and genes in Semantic MEDLINE and applied a network-based approach to mine the local network structure [43]. This could result in the formulation of novel research hypotheses, which is critical for translational medicine research and personalized medicine.

3. Target validation

Target validation is a time-consuming and costly process that demonstrates relevance – is the identified target of relevance to a particular biological pathway, molecular process or disease? We agree that target validation efficiency can be greatly improved when combined to strict data filtering and statistics, as high throughput screening sheds light to cellular responses in disease models of interest. Network validation can be performed by comparing the network of interest to 100 random networks generated using random shuffling of the graph with degrees preserved as implemented in the Randomized network plugin in Cytoscape2.6.3 [32]. Gene function and/or gene regulatory networks can be validated *via* genome-wide approaches [44] and functional screens, such as RNAi and CRISPR-Cas9 [45]. Inter-individual variability upon drug administration/ intervention can be tracked and analyzed recently, as electronic medical records and clinical trial data become available. In addition to the molecular and clinical data, free-text data presented in literature are also useful in drug discovery *via* extensive data mining processes [46].

4. Computer-aided drug design

Once a target has been identified, there are several *in silico* tools to initiate a drug design process. The use of these methods depends on the nature of the target and the available information on the system.

In the past decade, computer aided drug design (CADD) has offered valuable tools in the identification of compounds, minimizing the risk of later rejection of lead compounds. Even though high throughput screening (HTS) usually offers several hit compounds, success rates are often very low and many of the identified compounds are later rejected due to their physicochemical properties. CADD plays a significant role in high success rates of hit compound identification [47], as well as the prioritization of HTS active compounds. One of many examples of the importance of CADD compared to HTS, was the identification of inhibitors against the transforming growth factor- β 1 receptor kinase. While the HTS for compound identification at Eli Lilly resulted in a potent lead compound [48], at the same time, a fully computational approach by Biogen Idec [49], resulted in the identification of 87 hits, the best one being Eli Lilly's initial lead compound [50]. In this case, a fully computational work was able to produce the same result as a wet lab approach, which traditionally is more costly and time consuming.

There are generally two distinct methods for computational drug design, *structure based* and *ligand based* (Fig. 1). These depend on the available information on the identified target. Most of them are analyzed in detail elsewhere [51,52], however the scope of this mini-review is to highlight and review the most commonly used. When there is no information on the structure of the target, computational methods for new molecules are based on information of known active or inactive compounds against it. This is the ligand based CADD approach, where tools such as ligand chemical similarity or pharmacophore mapping can be very useful. Generally, the most commonly used methods in hit compound identification rely on virtual screening techniques on the targets' binding site. These methods mostly rely on docking large libraries of small molecules such as ZINC [53], or chemical information on known compounds such as Pubchem [54] using docking or pharmacophore modeling tools. The use of such libraries, however, is expensive from a computational perspective. If no adequate computational resources are available, cascade virtual screening protocols are applied in a way that databases are filtered based on physicochemical or other properties of the compounds to avoid using databases as a whole [55,56].

There are studies in which ligand similarity-based virtual screening and structure based virtual screening results for the same targets and compound sets have been compared. For example, in a study comparing these two methods in several drug targets (CDK2, COX2, estrogen receptor, neuraminidase, HIV-1 protease, p38 MAP kinase and thrombin), the ligand-based virtual screening methods performed better in most of the cases [57]. In another example, the comparison between ligand-based and structure-based methods (vROCS and FRED) demonstrated that the ligand based method performed in a better predictive matter [58].

The combination of ligand and structure based molecular modeling methods, however, has become a common approach in virtual screening through sequential, parallel or hybrid approaches [59]. For example, hybrid protein–ligand pharmacophore methods have been successfully applied in virtual screening [60] as well as ligand profiling studies [61]. These methods can also be integrated in network-based approaches towards drug discovery [16,62]. Furthermore, other combinatorial tools can be implemented for the aim of multi-targeted drug design [63]. Such drugs, produced with one single chemical or with a composition of several chemicals, should be able to target the characteristic pathological network of a disease. A characteristic example, combination of dasatinib, a ABL/T315I inhibitor in combination with imatinib, a tyrosine-kinase inhibitor are proposed to treat chronic myelogenous leukemia (CML) by targeting BCR–ABL fusion proteins [64].

4.1. Ligand-based CADD

Ligand based CADD methods take advantage of information of small molecules interacting with the target in question in order to identify new, more potent compounds. This information includes binding affinities, chemical structure, physicochemical properties *etc.* These methods

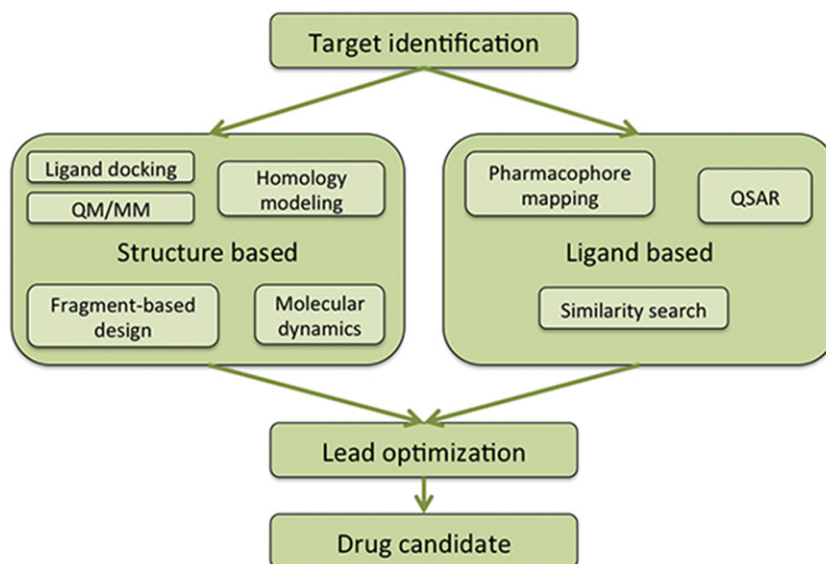


Fig. 1. General computer aided techniques for drug design. The usage largely relies on the available structural information on the drug target to be assessed. No structural information leads to ligand based drug design methods, where known active compounds are used for the discovery of similar, more potent drug candidates.

are considered in some cases more successful than structure based techniques [65].

One ligand-based approach is the selection of new compounds based on chemical similarity of known active ones. This can be done using several fingerprint methods, which allow the representation of a molecule in a way that can be effectively compared against other molecules. These methods rely on the chemical information of compounds, giving a highly qualitative approach in the search of new more potent ligands [66]. A representative example of identifying T-type calcium channel blockers, which are implicated in epilepsy and neuropathic pain [67] is the work reported by Ijjaali and co-workers [68]. In this work, a ligand based virtual screening was made on a two million compound database using ChemAxon's PF and CCG's GpiDAPH3 fingerprints to test 38 molecules for their ability to affect the functional activity of recombinant human $Ca_v3.2$ (Table 1). Sixteen out of the 38 molecules were active hits as they showed more than 50% blockade of the $Ca_v3.2$ mediated T-type current.

Another significant approach is the quantitative structure–activity relationship (QSAR), where a QSAR model is able to describe a correlation between structures of a set of molecules and their target response [69]. The general QSAR workflow consists of the gathering of a set of active and inactive molecules against a target and production of the descriptors describing their structural and physicochemical properties. The model can then be used to correlate these descriptors and the experimental activity, resulting in a predictive tool for new molecular entities [69]. QSAR algorithms are continuously evolving, involving the implementation of several 2D and 3D descriptors, which can be structural or physicochemical (e.g. molecular weight, volume, rotatable bonds, interatomic distances, atom types, molecular walk counts, electronegativity, atom distribution, aromaticity, solvation properties) and can be described on multiple levels of increasing complexity. A recent work, highlighted the success of such a method, being able to identify hit compounds that act as allosteric modulators of mGlu5. mGlu5 is a well established pharmaceutical target against anxiety, Parkinson's disease, and schizophrenia [70]. Building a QSAR model, with information from a previous HTS screen on mGlu5 [71] Mueller and co-workers were able to identify 27 active compounds that modulates the signaling of the protein [72]. The overall success of the QSAR model was a 3.6% hit rate [72] compared to the 0.2% hit rate of the HTS [71].

Another field gaining ground in the computational drug discovery setting is proteochemometrics and polypharmacology modeling [73]. Proteochemometrics (PCM) modeling combines both ligand information

and target information within a single predictive model in order to predict an output variable of interest [74,75]. Ligand information of the system is accompanied by information for its biological effect. Merging data from ligand and target sources into the frame of a single machine learning model allows the prediction of the most suitable pharmacological treatment for a given genotype (personalized medicine), which ligand-only and protein-only approaches are not able to perform [74, 75]. Using PCM methods, Frimurer et al managed to identify ~60 ligands for the prostaglandin D2 receptor 2 (CRTH2), after screening of a library of 1.2 million compounds [76].

4.2. Structure-based CADD

3D information on proteins and DNA started being used for drug design almost three decades ago. The protein databank (PDB) is the largest depository of biomolecule structure information determined mostly by X-ray crystallography and NMR techniques. In 1998, 2058 structures had been deposited in the protein data bank. Since then, each year there has been a ~ 7.5% increase in depositions, resulting in a total of 105,465 structures in 2014. The use of this abundant structural information has been the cornerstone for structure based drug design for the past years in academia, as well as the pharmaceutical industry.

Proteins, by nature are dynamic macromolecules. Thus, a structural snapshot is not enough to study a protein's interaction with a small molecule, or even identify its binding site. One of the most important aspects of studying the behavior of proteins is called molecular dynamics (MD) simulations. Based on Newtonian mechanics and using force fields such as Amber [77] or CHARMM [78], molecular dynamics simulations can calculate a trajectory of conformations as a function of time. In MD simulations, chemical bonds and atomic angles are modeled using simple virtual springs, and dihedral angles are modeled using a sinusoidal function. Non-bonded forces occur as van der Waals interactions, as the Lennard-Jones 6–12 potential and Coulomb's law are used to calculate hydrophobic and electrostatic interactions. These simulations, coupled with experimental data have a significant impact on the drug discovery field.

Traditionally these calculations have been performed in cpu clusters with software able to parallelize the processes of simulation complex systems. In the recent years, calculations required for these simulations have been developed to be performed by video-game and computer-graphics applications. Eventually the graphics-processing-units (GPUs) designed to speed up video games have began to be used to speed up

molecular dynamics simulations as well, usually by an order of magnitude [79,80].

Several MD applications are used for free energy calculations in order to correlate experimental binding affinities of small molecules to a protein with calculated, such as molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) [81], linear interaction energy (LIE) [82], and free energy perturbation methods (FEP) [83]. They can then be used for the prediction of binding affinities *in silico*. Some examples of the use of such methods are the reproduction of binding free energies TIBO derivatives to HIV-1 RT and prediction of the binding mode of efavirenz to HIV-1 RT by MM-PBSA [84], the relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using FEP approaches [85] and LIE models for predicting the binding mode of β -secretase (BACE) inhibitors [86].

In some cases, the structure of a protein target for a drug design project may not be yet solved. In this case, there are predictive tools for building comparative models. Comparative modeling is used to predict the structure of a protein based on a structural template with a similar sequence, with the general view that proteins with similar sequences have similar structures. Homology modeling is the most used computational technique towards this goal and is used to construct a protein model after identification of a structural template protein with similar sequence, alignment of their sequences, using coordinates of aligned regions, prediction of missing atom coordinates of the target, model building and refinement. Some of the most widely used software for homology modeling are MODELER [87] and SWISS-MODEL [88]. There are several examples of the use of homology modeling in drug design, one of which the model of chemo-attractant receptor OXE-R based on the crystal structure of CXCR4 [89] as a template. Consecutively, using virtual screening techniques a small-molecule modulator Gue1654 was identified, which inhibits a specific GPCR signaling pathway [90]. In another example, the binding mode of antihypertensive drugs to the angiotensin II receptor type 1 (AT1) was predicted [91]. A homology model of the receptor and analysis of binding mode of active compounds using MD and pharmacophore modeling was later validated when the crystal structure was determined (Fig. 2) [92].

Docking methods are used to predict the preferred orientation of one molecule to a protein when bound to each other to form a stable complex. Depending on the method, there are different considerations of the flexibility of either the ligand or the protein during the docking process [93]. The most commonly used method considers the ligand flexible, while the protein docking site is held rigid, usually pre-treated with molecular dynamics force fields. Several software packages

are available for docking, such as Gold [94], Autodock [95], AutoDock Vina [96], DOCK [97], GLIDE [98], SURFLEX [99] and others. The docking score is an evaluation of the energetic affinity of the complex, calculated by scoring functions. These scoring functions can be molecular mechanics-based, empirical, knowledge-based, or consensus-based functions. For example, DOCK uses the AMBER [77] force field for evaluating the energetics of binding, while SURFLEX uses an empirical function. Consensus scoring is a method whereby the binding affinities of compounds for a particular target are predicted by using more than one scoring algorithm and is a frequently studied method [100]. In such a study, Tuchinardi et al. evaluated the consensus docking and scoring of several different algorithms along 83 ligand-receptor X-ray structures [101].

Another widely used method for evaluating the probability of molecule binding to protein binding sites is pharmacophore modeling. A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure. By mapping the interaction of an active compound bound to its target protein, a pharmacophore can represent the geometrical and chemical properties using pharmacophore features, which include hydrogen bond acceptors and donors, basic and acidic groups, partial charge, aliphatic and aromatic hydrophobic moieties. This representation can be used for virtual screening projects, in order to identify potential binders based on this interaction. Several software packages have been developed towards pharmacophore modeling, Ligandscout [102] and The Pocket v.2 [103] use algorithms in protein-ligand complex data to map interactions between ligand and target.

Docking and pharmacophore modeling have been widely used in virtual screening studies to identify novel compounds against drug targets. There are several successful examples of hit compounds that later on proceeded to a hit to lead process. Human Pim-1 kinase, a highly conserved serine-threonine kinase is a valuable anticancer drug target. In a virtual screening study using docking methods and a database of 700,000 commercially available compounds, four compounds were identified, having affinity in the micromolar range [104]. In another study, using a docking screening approach to identify novel dual kinase (EGFR)/bromodomain (BRD4) inhibitors from six million commercially available small molecules, Allen and co-workers selected and tested 24 compounds [105]. The result was the identification of several novel BRD4 binders and one novel dual EGFR-BRD4 inhibitor (2870), a first in class compound that could target multiple cancer promoting pathways. The use of pharmacophore modeling in high throughput virtual

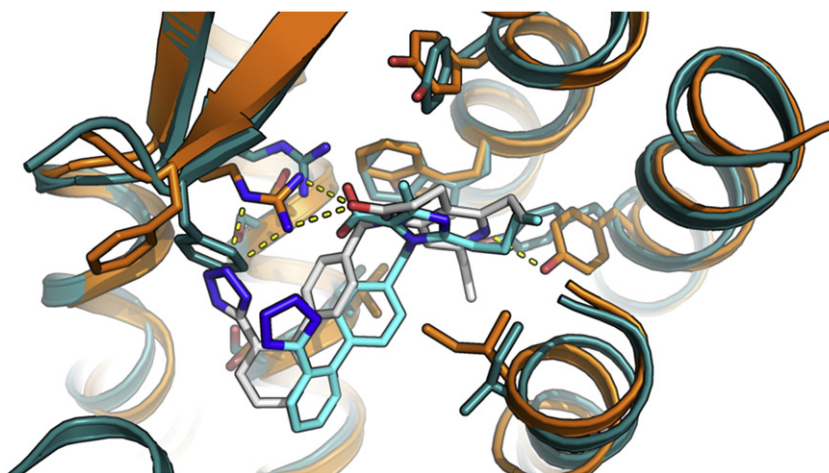


Fig. 2. Predicted homology model of the angiotensin II type 1 receptor (deep blue) with compound EXP3174 docked (cyan) superimposed with the crystal structure of AT1 (orange) co-crystallized with ZD7155 (white). Hydrogen bonds are depicted in yellow dashed lines.

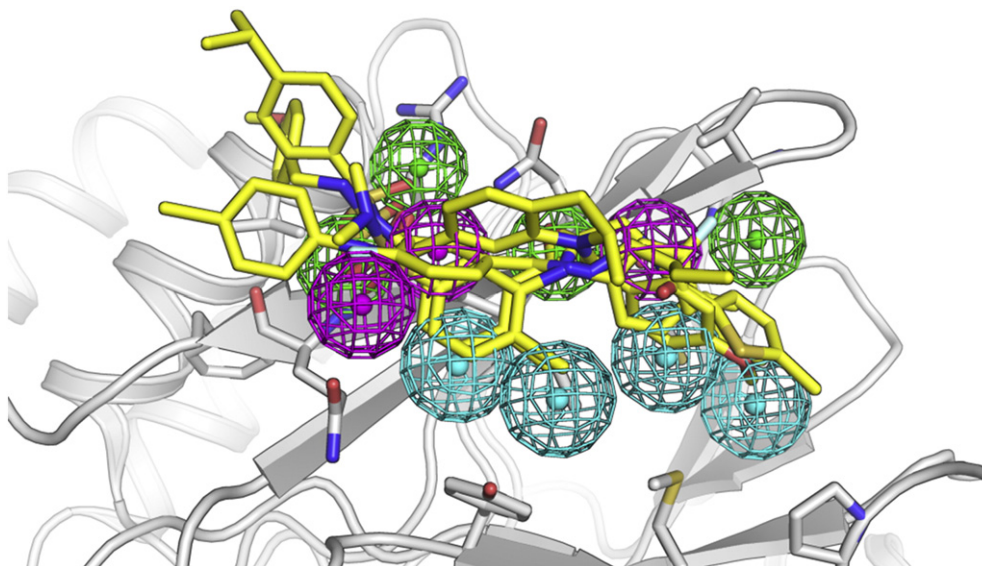


Fig. 3. Four pharmacophore model hit compounds (shown in yellow stick representation), which were identified through a virtual screening lead compound identification process to target Calcineurin (in white ribbon and stick representation). The pharmacophore model is shown in colored spheres (cyan for hydrophobic, magenta for hydrogen bond donor and green for hydrogen bond acceptor features).

screening has also proven valuable in identifying novel active compounds. In a study to identify Calcineurin inhibitors that could be further developed into novel immunosuppressant agents, a fifteen million database of purchasable compounds from ZINC was screened on a pharmacophore model mapping a protein–protein interaction. Out of the 32 compounds tested, seven showed micromolar affinity, with four of them having the ability to inhibit the expression of nuclear factor of activated T cells (NFATc) dependent genes, cytokine production, and cell proliferation, suggesting that these may have therapeutic potential as immunosuppressive agents (Fig. 3) [106].

Once a computational hit identification method has produced compounds of usually low affinity, fragment-based drug discovery (FBDD) is utilized for finding lead compounds as part of the drug discovery process. Commercial databases of compounds have the limitation of representing only a small amount of the available chemical space. It has been estimated that the Lipinski virtual chemical space might contain as many as 10^{60} compounds [107]. This is where FBDD comes to enrich hit compounds and produce more potent lead compounds [108]. The two commonly used approaches for the optimization of fragment hits into lead-like compounds are Fragment growing and Fragment linking. The first is the addition of functional groups to the active fragment core in order to optimize interactions with the binding site, while the second is a less commonly used method, which links fragments that bind in adjacent sites of a target protein to turn low affinity fragments into high affinity leads. Successful examples of hit optimization are well documented, such as the discovery of Beta-site amyloid precursor protein cleaving enzyme 1 (BACE1) by Amgen towards inhibitors against Alzheimer's disease [109]. Another example is the fragment based discovery of inhibitors against the phosphatidylinositol-3 kinases (PI3Ks) [110] which are involved in cancer, rheumatoid arthritis, cardiovascular disease and respiratory disease.

5. Summary and outlook

Computational methods have provided a powerful toolbox for target identification, discovery and optimization of drug candidate molecules. Information technologies coupled to statistics and chemoinformatic tools shed light to disease mechanisms and phenotypes revealing potential drug targets to be further validated by high throughput screening technologies.

Consecutively, multiple methods allow for the prediction and characterization of binding sites, studying the dynamic nature of drug targets, identifying new active molecular entities and their optimization. Nowadays, large databases of readily commercially available compounds and ligand chemical space exploration offer drug discovery scientists with enormous data to handle. Different methods, based on readily available information on the biological system under study are evolving to assist the manipulation and handling of this data. Moreover, integration of ‘-omics’ technologies and databases may facilitate the identification of novel drug targets or the design of network-based multi-target drugs. Structure and ligand based methods are the most commonly used in the drug discovery field, however, emerging combinatorial techniques such as proteochemometrics are emerging.

All the computational methods mentioned in this review, either towards target identification, either towards novel ligand discovery continue to evolve and their synergy is what we envisage that will facilitate cost-effective and reliable outcomes in an era of big data demands.

Acknowledgements

This work was supported by a grant to M.-T.M. from the Greek State Scholarships Foundation.

The authors have no conflict of interests to disclose.

References

- [1] Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011;12:56–68.
- [2] Chen B, Butte A. Network medicine in disease analysis and therapeutics. *Clin Pharmacol Ther* 2013;94:627–9.
- [3] Culligan EP, Sleator RD, Marchesi JR, Hill C. Metagenomics and novel gene discovery: promise and potential for novel therapeutics. *Virulence* 2014;5:399–412.
- [4] Garber K. Drugging the gut microbiome. *Nat Biotechnol* 2015;33:228–31.
- [5] Wang L, Xie X-Q. Computational target fishing: what should chemogenomics researchers expect for the future of in silico drug design and discovery? *Future Med Chem* 2014;6:247–9.
- [6] Agrawal D, Bernstein P, Bertino E, Davidson S, Dayal U, et al. Challenges and opportunities with big data. *Proc VLDB Endowment* 2012;5:2032–3.
- [7] Tsiliki G, Karacapilidis N, Christodoulou S, Tzagarakis M. Collaborative mining and interpretation of large-scale data for biomedical research insights. *PLoS One* 2014;9, e108600.
- [8] Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, et al. Predicting new molecular targets for known drugs. *Nature* 2009;462:175–81.
- [9] Nidhi GM, Davies JW, Jenkins JL. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J Chem Inf Model* 2006;46:1124–33.

- [10] Li H, Gao Z, Kang L, Zhang H, Yang K, et al. TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res* 2006;34:W219–24.
- [11] Cheng T, Li Q, Wang Y, Bryant SH. Identifying compound-target associations by combining bioactivity profile similarity search and public databases mining. *J Chem Inf Model* 2011;51:2440–8.
- [12] Rognan D. *Theory and applications in computational chemistry: the first decade of the second millennium: International Congress TACC-2012*, vol. 1456. AIP Publishing; 2012 157–64.
- [13] Rognan D. Structure-based approaches to target fishing and ligand profiling. *Mol Inf* 2010;29:176–87.
- [14] Cao R, Wang Y. Predicting molecular targets for small-molecule drugs with a ligand-based interaction fingerprint approach. *ChemMedChem* 2015 [ePub ahead of print].
- [15] Eberini I, Daniele S, Parravicini C, Sensi C, Trincavelli ML, et al. In silico identification of new ligands for GPR17: a promising therapeutic target for neurodegenerative diseases. *J Comput Aided Mol Des* 2011;25:743–52.
- [16] Barabasi A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011;12:56–68.
- [17] Billur Engin H, Gursoy A, Nussinov R, Keskin O. Network-based strategies can help mono- and poly-pharmacology drug discovery: a systems biology view. *Curr Pharm Des* 2014;20:1201–7.
- [18] Zhao S, Li S. Network-based relating pharmacological and genomic spaces for drug target identification. *PLoS One* 2010;5, e11764.
- [19] Wang X, Wei X, Thijssen B, Das J, Lipkin SM, et al. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat Biotechnol* 2012;30:159–64.
- [20] Buchan NS, Rajpal DK, Webster Y, Alatorre C, Gudivada RC, et al. The role of translational bioinformatics in drug discovery. *Drug Discov Today* 2011;16:426–34.
- [21] Gentles AJ, Gallahan D. Systems biology: confronting the complexity of cancer. *Cancer Res* 2011;71:5961–4.
- [22] Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, et al. HMDB 3.0—the human metabolome database in 2013. *Nucleic Acids Res* 2013;41:D801–7.
- [23] Xia J, Sinelnikov IV, Han B, Wishart DS. MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res* 2015;43:W251–7.
- [24] Creixell P, Schoof EM, Simpson CD, Longden J, Miller CJ, et al. Kinome-wide decoding of network-attacking mutations rewiring cancer signaling. *Cell* 2015; 163:202–17.
- [25] Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;313:1929–35.
- [26] Lamb J. The connectivity map: a new tool for biomedical research. *Nat Rev Cancer* 2007;7:54–60.
- [27] Alakwaa FM, Solouma NH, Kadah YM. Construction of gene regulatory networks using biclustering and bayesian networks. *Theor Biol Med Model* 2011;8:39.
- [28] Heiser LM, Wang NJ, Talcott CL, Laderoute KR, Knapp M, et al. Integrated analysis of breast cancer cell lines reveals unique signaling pathways. *Genome Biol* 2009;10: R31.
- [29] Yeger-Lotem E, Riva L, Su LJ, Gitler AD, Cashikar AG, et al. Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. *Nat Genet* 2009;41:316–23.
- [30] Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, et al. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* 2006;34:D535–9.
- [31] Breitkreutz B-J, Stark C, Tyers M. Osprey: a network visualization system. *Genome Biol* 2003;4:R22.
- [32] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.
- [33] Guberman JM, Ai J, Arnaiz O, Baran J, Blake A, et al. BioMart central portal: an open database network for the biological community. *Database* 2011; bar041; 2011.
- [34] Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* 2004;6:1–6.
- [35] Douville C, Carter H, Kim R, Niknafs N, Diekhans M, et al. CRAVAT: cancer-related analysis of variants toolkit. *Bioinformatics* 2013;29:647–8.
- [36] Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, et al. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 2012;40:W452–7.
- [37] Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 2015;31:2745–7.
- [38] Grapov D, Wanichthanarak K, Fiehn O. MetaMapR: pathway independent metabolomic network analysis incorporating unknowns. *Bioinformatics* 2015;31: 2757–60.
- [39] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996;14:33–8.
- [40] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004;25:1605–12.
- [41] Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, et al. The genotype-tissue expression (GTEx) project. *Nat Genet* 2013;45:580–5.
- [42] Thompson R, Johnston L, Taruscio D, Monaco L, Bérout C, et al. RD-connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. *J Gen Intern Med* 2014;29:780–7.
- [43] Zhang Y, Tao C, Jiang G, Nair AA, Su J, et al. Network-based analysis reveals distinct association patterns in a semantic MEDLINE-based drug-disease-gene network. *J Biomed Semant* 2014;5:1–13.
- [44] Peter A. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat Commun* 2015;6.
- [45] Barrangou R, Birmingham A, Wiemann S, Beijersbergen RL, Hornung V, et al. Advances in CRISPR-Cas9 genome engineering: lessons learned from RNA interference. *Nucleic Acids Res* 2015;43:3407–19.
- [46] Chen B, Butte AJ. Leveraging big data to transform target selection and drug discovery. *Clin Pharmacol Ther* 2015;99:285–97.
- [47] Doman TN, McGovern SL, Witherbee BJ, Kasten TP, Kurumbail R, et al. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem* 2002;45:2213–21.
- [48] Sawyer JS, Anderson BD, Beight DW, Campbell RM, Jones ML, et al. Synthesis and activity of new aryl- and heteroaryl-substituted pyrazole inhibitors of the transforming growth factor- β type I receptor kinase domain. *J Med Chem* 2003; 46:3953–6.
- [49] Singh J, Chuaqui CE, Boriack-Sjodin PA, Lee W-C, Pontz T, et al. Successful shape-based virtual screening: the discovery of a potent inhibitor of the type I TGF β receptor kinase (T β RI). *Bioorg Med Chem Lett* 2003;13:4355–9.
- [50] Shekhar C. In silico pharmacology: computer-aided methods could transform drug development. *Chem Biol* 2008;15:413–4.
- [51] S-s O-Y, Lu J-y, Kong X-q, Liang Z-j, Luo C, et al. Computational drug discovery. *Acta Pharmacol Sin* 2012;33:1131–40.
- [52] Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational methods in drug discovery. *Pharmacol Rev* 2014;66:334–95.
- [53] Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model* 2012;52:1757–68.
- [54] Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, et al. Pubchem substance and compound databases. *Nucleic Acids Res* 2015;44:D1202–13.
- [55] Kurczab R, Nowak M, Chilmonczyk Z, Sylte I, Bojarski AJ. The development and validation of a novel virtual screening cascade protocol to identify potential serotonin 5-HT $_{7A}$ antagonists. *Bioorg Med Chem Lett* 2010;20:2465–8.
- [56] de Graaf C, Rein C, Piwnica D, Giordanetto F, Rognan D. Structure-based discovery of allosteric modulators of two related class B G-protein-coupled receptors. *ChemMedChem* 2011;6:2159–69.
- [57] Zhang Q, Muegge I. Scaffold hopping through virtual screening using 2d and 3d similarity descriptors: ranking, voting, and consensus scoring. *J Med Chem* 2006; 49:1536–48.
- [58] Ramasamy T, Selvam C. Performance evaluation of structure based and ligand based virtual screening methods on ten selected anti-cancer targets. *Bioorg Med Chem Lett* 2015;25:4632–6.
- [59] Drwal MN, Griffith R. Combination of ligand- and structure-based methods in virtual screening. *Drug Discov Today: Technol* 2013;10:e395–401.
- [60] Chen Z, Tian G, Wang Z, Jiang H, Shen J, et al. Multiple pharmacophore models combined with molecular docking: a reliable way for efficiently identifying novel PDE4 inhibitors with high structural diversity. *J Chem Inf Model* 2010; 50:615–25.
- [61] Meslamani J, Li J, Sutter J, Stevens A, Bertrand H-O, et al. Protein-ligand-based pharmacophores: generation and utility assessment in computational ligand profiling. *J Chem Inf Model* 2012;52:943–55.
- [62] Leung EL, Cao Z-W, Jiang Z-H, Zhou H, Liu L. Network-based drug discovery by integrating systems biology and computational technologies. *Brief Bioinform* 2013; 14:491–505.
- [63] Liu X, Zhu F, H. Ma X, Shi Z, Y. Yang S, et al. Predicting targeted polypharmacology for drug repositioning and multi-target drug discovery. *Curr Med Chem* 2013;20: 1646–61.
- [64] Kaiser J. Combining targeted drugs to stop resistant tumors. *Science* 2011;331: 1542–5.
- [65] Stumpfe D, Ripphausen P, Bajorath J. Virtual compound screening in drug discovery. *Future Med Chem* 2012;4:593–602.
- [66] Auer J, Bajorath J. In: Keith J, editor. *Bioinformatics*. Humana Press; 2008. p. 327–47.
- [67] Bourinnet E, Zamponi GW. Voltage gated calcium channels as targets for analgesics. *Curr Top Med Chem* 2005;5:539–46.
- [68] Ijjaali I, Barrere C, Nargeot J, Petitet F, Bourinnet E. Ligand-based virtual screening to identify new T-type calcium channel blockers. *Channels* 2007;1:300–4.
- [69] Zhang S. In: Satyanarayanan SD, editor. *Drug design and discovery*. Humana Press; 2011. p. 23–38.
- [70] Gasparini F, Bilbe G, Gomez-Mancilla B, Spooren W. mGluR5 antagonists: discovery, characterization and drug development. *Curr Opin Drug Discov Devel* 2008; 11:655–65.
- [71] Rodriguez AL, Grier MD, Jones CK, Herman EJ, Kane AS, et al. Discovery of novel allosteric modulators of metabotropic glutamate receptor subtype 5 reveals chemical and functional diversity and in vivo activity in rat behavioral models of anxiolytic and antipsychotic activity. *Mol Pharmacol* 2010;78:1105–23.
- [72] Mueller R, Rodriguez AL, Dawson ES, Butkiewicz M, Nguyen TT, et al. Identification of metabotropic glutamate receptor subtype 5 potentiators using virtual high-throughput screening. *ACS Chem Neurosci* 2010;1:288–305.
- [73] Cortes-Ciriano I, Ain QU, Subramanian V, Lenselink EB, Mendez-Lucio O, et al. Polypharmacology modelling using proteochemometrics (PCM): recent methodological developments, applications to target families, and future prospects. *Med Chem Commun* 2015;6:24–50.
- [74] Lapinsh M, Prusis P, Gutcaits A, Lundstedt T, Wikberg JES. Development of proteochemometrics: a novel technology for the analysis of drug-receptor interactions. *Biochim Biophys Acta* 2001;1525:180–90.
- [75] van Westen GJP, Wegner JK, Ijzerman AP, van Vlijmen HWT, Bender A. Proteochemometric modeling as a tool to design selective compounds and for extrapolating to novel targets. *Med Chem Commun* 2011;2:16–30.
- [76] Frimurer TM, Ulven T, Elling CE, Gerlach L-O, Kostenis E, et al. A phylogenetic method to assign ligand-binding relationships between 7TM receptors. *Bioorg Med Chem Lett* 2005;15:3707–12.

- [77] Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem* 2004;25:1157–74.
- [78] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, et al. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 2010;31:671–90.
- [79] Liu W, Schmidt B, Voss G, Müller-Wittig W. Accelerating molecular dynamics simulations using graphics processing units with CUDA. *Comput Phys Commun* 2008;179:634–41.
- [80] Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA, et al. How does a drug molecule find its target binding site? *J Am Chem Soc* 2011;133:9181–3.
- [81] Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* 2010;51:69–82.
- [82] Hansson T, Marelus J, Åqvist J. Ligand binding affinity prediction by linear interaction energy methods. *J Comput Aided Mol Des* 1998;12:27–35.
- [83] Singh UC, Brown FK, Bash PA, Kollman PA. An approach to the application of free energy perturbation methods using molecular dynamics: applications to the transformations of $\text{CH}_3\text{OH} \rightarrow \text{CH}_3\text{CH}_3$, $\text{H}_3\text{O}^+ \rightarrow \text{NH}_4^+$, glycine \rightarrow alanine, and alanine-phenylalanine in aqueous solution and to $\text{H}_3\text{O}^+(\text{H}_2\text{O})_3 \rightarrow \text{NH}_4^+(\text{H}_2\text{O})_3$ in the gas phase. *J Am Chem Soc* 1987;109:1607–14.
- [84] Wang J, Morin P, Wang W, Kollman PA. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J Am Chem Soc* 2001;123:5221–30.
- [85] Miyamoto S, Kollman PA. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. *Proteins Struct Funct Genet* 1993;16:226–45.
- [86] Tounge BA, Rajamani R, Baxter EW, Reitz AB, Reynolds CH. Linear interaction energy models for β -secretase (BACE) inhibitors: role of van der Waals, electrostatic, and continuum-solvation terms. *J Mol Graph Model* 2006;24:475–84.
- [87] Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, et al. Current protocols in protein science; 2006 5.6. 1–5.6. 30.
- [88] Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 2003;31:3381–5.
- [89] Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, et al. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 2010;330:1066–71.
- [90] Blättermann S, Peters L, Ottersbach PA, Bock A, Konya V, et al. A biased ligand for OXE-R uncouples $\text{G}\alpha$ and $\text{G}\beta\gamma$ signaling within a heterotrimer. *Nat Chem Biol* 2012;8:631–8.
- [91] Matsoukas M-T, Cordoní A, Ríos S, Pardo L, Tselios T. Ligand binding determinants for angiotensin II type 1 receptor from computer simulations. *J Chem Inf Model* 2013;53:2874–83.
- [92] Zhang H, Unal H, Gati C, Han Gye W, Liu W, et al. Structure of the angiotensin receptor revealed by serial femtosecond crystallography. *Cell* 2015;161:833–44.
- [93] Yuriev E, Agostino M, Ramsland PA. Challenges and advances in computational docking: 2009 in review. *J Mol Recognit* 2011;24:149–64.
- [94] Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein–ligand docking using GOLD. *Proteins Struct Funct Bioinf* 2003;52:609–23.
- [95] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 2009;30:2785–91.
- [96] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010;31:455–61.
- [97] Moustakas D, Lang PT, Pegg S, Pettersen E, Kuntz I, et al. Development and validation of a modular, extensible docking program: DOCK 5. *J Comput Aided Mol Des* 2006;20:601–19.
- [98] Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J Med Chem* 2006;49:6177–96.
- [99] Jain AN. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem* 2003;46:499–511.
- [100] Houston DR, Walkinshaw MD. Consensus docking: improving the reliability of docking in a virtual screening context. *J Chem Inf Model* 2013;53:384–90.
- [101] Tuccinardi T, Poli G, Romboli V, Giordano A, Martinelli A. Extensive consensus docking evaluation for ligand pose prediction and virtual screening studies. *J Chem Inf Model* 2014;54:2980–6.
- [102] 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.
- [103] Chen J, Lai L. Pocket v.2: further developments on receptor-based pharmacophore modeling. *J Chem Inf Model* 2006;46:2684–91.
- [104] Pierce AC, Jacobs M, Stuver-Moody C. Docking study yields four novel inhibitors of the protooncogene pim-1 kinase. *J Med Chem* 2008;51:1972–5.
- [105] Allen BK, Mehta S, Ember SWJ, Schonbrunn E, Ayad N, et al. Large-scale computational screening identifies first in class multitarget inhibitor of EGFR kinase and BRD4. *Sci Rep* 2015;5:16924.
- [106] Matsoukas M-T, Aranguren-Ibáñez Á, Lozano T, Nunes V, Lasarte JJ, et al. Identification of small-molecule inhibitors of calcineurin-NFATc signaling that mimic the PxlIT motif of calcineurin binding partners. *Sci Signal* 2015;8:ra63.
- [107] Reymond J-L, Awale M. Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem Neurosci* 2012;3:649–57.
- [108] Kumar A, Voet A, Zhang KYJ. Fragment based drug design: from experimental to computational approaches. *Curr Med Chem* 2012;19:5128–47.
- [109] Cheng Y, Judd TC, Bartberger MD, Brown J, Chen K, et al. From fragment screening to in vivo efficacy: optimization of a series of 2-aminoquinolines as potent inhibitors of beta-site amyloid precursor protein cleaving enzyme 1 (BACE1). *J Med Chem* 2011;54:5836–57.
- [110] Hughes SJ, Millan DS, Kilty IC, Lewthwaite RA, Mathias JP, et al. Fragment based discovery of a novel and selective PI3 kinase inhibitor. *Bioorg Med Chem Lett* 2011;21:6586–90.