

Virtual screening strategies in drug design – methods and applications

EWA BIELSKA^{1*}, XAVIER LUCAS^{1,2*}, ANNA CZERWONIEC¹, JOANNA M. KASPRZAK¹,
KATARZYNA H. KAMINSKA², JANUSZ M. BUJNICKI^{2,1**}

¹Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Poznań, Poland

²International Institute of Molecular and Cell Biology, Warszawa, Poland

(* These two authors contributed equally)

** Corresponding author: iamb@genesilico.pl

Abstract

Virtual screening (VS) overcomes the limitations of traditional high-throughput screening (HTS) by applying computer-based methods in drug discovery. VS takes advantage of fast algorithms to filter chemical space and successfully select potential drug candidates. A key aspect in structure-based VS is the sampling of ligand-receptor conformations and the evaluation of these poses to predict near-native binding modes. The development of fast and accurate algorithms during the last few years has allowed VS to become an important tool in drug discovery and design. Herein, an overview of widely used ligand-based (e.g., similarity, pharmacophore searches) and structure-based (protein-ligand docking) VS methods is discussed. Their strengths and limitations are described, along with many successful stories. This review not only serves as an introductory guide for inexperienced VS users but also presents a general overview of the current state and scope of these powerful tools.

Key words: high-throughput virtual screening, drug design, drug discovery, ligand-based, similarity searches, pharmacophore, receptor-based, protein-ligand docking

Introduction to drug development and design

The process of drug development aims towards the identification of compounds with pharmacological interest to assist in the treatment of diseases and ultimately to improve the quality of life. The compounds used in pharmacology are mainly small organic molecules (ligands) which interact with specific biomolecules (receptors/targets). Usually, compounds with some shared physico-chemical property or obtained according to a particular protocol are compiled into large collections that are termed as libraries. Experimental identification of small molecules with the desired activity can be achieved by a high-throughput screening (HTS). Since the drug-like libraries often contain millions of compounds, their acquisition (e.g., by chemical synthesis) and efficient testing by sophisticated robots is very expensive (Plewczynski et al., 2010). A limitation that also reduces the usability of HTS is its dependency on experimental factors. For example, compounds selected for screening should have high stability for long-term storage and high solubility in the testing media.

An alternative to experimental HTS is high-throughput virtual screening (HTVS or VS), which has become a standard tool in medicinal chemistry (Leach et al., 2006). Virtual screening (VS) uses computational power to test large sets of chemical compounds in a few days at low costs. Moreover, not only real compounds but also purely theoretical ones can be included in the virtual library and screened *in silico*. This is a very powerful feature that allows purchasing or synthesizing only a reduced set of selected compounds thus reducing the cost and time of the entire study significantly.

Several screening campaigns have been launched to compare results of virtual and experimental approaches (Doman et al., 2002; Evensen et al., 2003; Jenkins et al., 2003; Paiva et al., 2001). These results suggest that VS may complement HTS and successfully identify false-negatives produced by experimental testing in order to increase the hit rate (Doman et al., 2002). Even though HTS has proven to be quite beneficial in identifying compounds with greater structural diversity compared to VS (Paiva et al., 2001), in some cases virtual hits are

Table 1. Classification of virtual screening methods based on the amount and type of information available about the system under inspection

	Known ligand(s)	Unknown ligands
Known structure of target or close homologue	structure-based virtual screening: protein-ligand docking	<i>de novo</i> structure-based virtual screening: protein-ligand docking
Unknown target structure	ligand-based virtual screening: few ligands: similarity searches several ligands: pharmacophore searches	virtual screening cannot be applied

found to be better drug candidates than the experimental ones (Brenk et al., 2005).

In contrast to high-throughput screening, VS is a knowledge-driven approach. The quality and the amount of information regarding the system under inspection is a critical factor when designing a computer-assisted drug design experiment (Klebe, 2006) – see Table 1. Knowledge about the substrates may be the starting point to retrieve similar compounds by 2D/3D similarity or pharmacophore searches (see section Pre-filtering). In cases where the target structure is available, docking methods that sample ligand poses with respect to the receptor binding site can be used (see section Protein-ligand docking).

This paper presents an overview of VS methods and applications, starting with the discussion on strategies for the preparation of ligand libraries. The selection of appropriate sets of small molecules for screening is a demanding task requiring adequate filtering strategies, for example, similarity searches by fingerprints or pharmacophores. The review also points toward the strengths and weaknesses of target-based VS methods. Finally, successful cases of VS campaigns are introduced to highlight the ability of the presented technique to assist in the drug discovery process.

Pre-filtering

The virtual hunt for bioactive compounds is a sequential process (Bajorath, 2004), which is preceded by a suitable compound library design. Thanks to the widely available molecule collections, chemical information is accessible for almost every scientist. Despite this wide accessibility, the most striking question remains: which chemical molecules are of interest and should be chosen for a VS procedure?

The chemical space is vast, but most of it remains biologically uninteresting (Shoichet, 2004). It is neces-

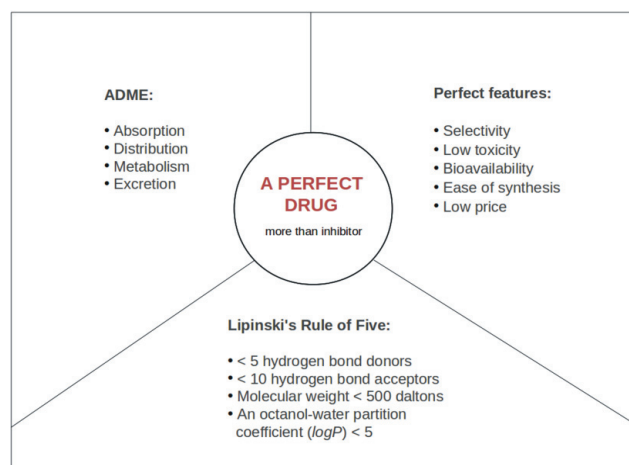


Fig. 1. Presentation of “perfect drug” properties of pharmaceutical compounds, and their relation to the bioavailability in humans (ADME properties) and to the Lipinski’s Rule of Five

sary to filter the compounds to obtain those with a high likelihood of bio-medical relevance. It is still unknown how to exhaustively screen the vastness of the chemical space and create diverse and balanced drug-like libraries (Klebe, 2006). However, a number of limitations exist in achieving the same. A wide range of filters may be applied to discredit compounds with unfavorable pharmacodynamic or pharmacokinetic properties (Varnek and Tropsha, 2008). Taking into account the knowledge-driven nature of VS, all data concerning liganda and target structure should be included in the filtering step, e.g. information regarding active compounds (substrates, known inhibitors), receptor structure and receptor-ligand interactions (Walters et al., 1998). The VS method takes advantage of all available biological information, which when annotated in chemical libraries may help to define nontoxic and biologically interesting molecules. Experimental data about toxicity, affinity or ADME (Absorption, Distribution, Metabolism and Excretion) properties (see Figure 1) are typically taken into account

when designing libraries of small molecules. However, even though the filtering methods ensure fast identification of molecules that are likely to be biologically relevant (Gasteiger, 2003), they also risk the omission of potentially interesting compounds with diverse and novel structure.

In the following section, commonly used methods and applications of ligand-based strategies are presented. First, a focus is put on the commonly applied drug-like filters and their performance in compound selection, and then similarity searches and the corresponding tools are introduced as an efficient way to retrieve bioactive molecules. The validity of the hereafter presented strategies has been proven by the successful application of VS workflows.

Filtering by drug-like properties

There is a series of requirements for organic molecules for meeting drug-likeness criteria. Compounds are usually considered as drug-like if they fulfill the Lipinski's Rule of Five and the ADME restrictions (Figure 1). This "rule of thumb" was formulated after the observation that most drugs are small, lipophilic molecules. This specifies drug-like requirements for molecular weight, hydrogen bonding capabilities, solubility and ADME properties. Other validation methods take into account the number of rotatable bonds, aggregation, polar surface area (Mestres and Knegtel, 2000) or a composition of, exclusively, bioelements (e.g. H, C, N, O, P, S, Cl, Br) (Bohacek et al., 1996). It is also necessary to consider information about ligand flexibility, which is the key factor in finding the ligand's spatial orientation within the receptor cavity. Thus, a "perfect drug" structure should be carefully balanced between flexibility and rigidity. The compound may lose binding specificity because of numerous rotatable bonds, while rigid rings or constrained scaffolds can lead to binding inability. Moreover, while designing potential chemicals, it is also important to exclude all structural features known to have unfavorable influence on an organism (i.e. known toxic chemical groups). There are many other filtration criteria (Gleeson, 2008), which may be applied in pre-filtering of compound libraries, but the main conclusion remains the same: when creating a data set of chemicals for VS purposes, it is advised to take into consideration at least basic drug-like criteria to guard against screening of chemical spaces that are too large.

Chemical similarity searching

Ligand-based similarity methods rely on the basic paradigm that structural likeness enhances the chances to share a common bioactive profile. Thus, selecting compounds similar to known drugs increases the possibility of choosing a potential lead (Willet, 1998). It is common to use similarity searches in the identification of compounds based on their similarity to active ones. Moreover, this concept may be used when creating "targeted" libraries containing compounds similar to at least one true binder to its specific receptor (e.g., a substrate, inhibitor or cofactor). Based on a structural similarity search among small molecules, it is possible to retrieve compounds containing identical substructures that share affinity to the same receptors. This tendency is reflected by the term of *privileged structures* (Evans et al., 1988), which explains the role of the substructures/scaffolds that may exhibit dependencies for a particular target space (e.g., ligands with high affinity for certain target family representatives). A recent analysis have demonstrated (Schnur et al., 2006) that such dependencies are not necessarily translated into the level of chemical functionalization of the scaffold, and only a narrow range of substructures are selective for a specific target family/group. Still, it is important to keep in mind that certain substructures are repeatedly found in databases of known drugs and seem to be "well-defined islands" in the lead-like chemical space (Ertl et al., 2006). In the chemical similarity search, three-dimensional properties of compounds may be also included. When using 3D descriptors, it is possible to apply information about a ligand's shape and obtain useful information regarding possible interactions (Zauhar et al., 2003). Data describing the shape of a small molecule can be equally retrieved from known binder structures or features in the binding pocket of the receptor (Bender and Glen, 2004; Willet, 1998; Zauhar et al., 2003).

Generally, when analyzing biological systems, neither active compounds nor the binding mode of the natural cofactors are known. If there is no structural information to be used in the similarity searches, it is possible to define preferable ligand binding modes, by screening vast and structurally diverse libraries of compounds. In accordance to the funnel strategy (Beautrait et al., 2008), vast amount of chemical data are first studied to extract a smaller set of compounds. Thereafter, this set is analyzed in more detail. Using structurally diverse

libraries, more varied results may be obtained, showing novel and innovative chemical structures. Structurally diverse scaffolds selected by this method need to be further analyzed and optimized for the identification of interesting chemical structures.

How to filter?

We have already described filtering concepts based on drug-like properties, similarity and diversity measures. But what are the exact tools and strategies to be taken into consideration in order to achieve high efficiency of similarity searches?

Introduction to fingerprints

A key aspect for ligand-based filtering is the selection of appropriate molecular “fingerprints”, which may be defined as a mathematical representation of the molecules. In other words, the chemical information of the compound is translated into data that can be easily understood and processed by the computational methods. The main strength of this approach lies in its ability to compare multiple fingerprints and compute their similarity by using e.g. the Tanimoto coefficient (Hamersa et al., 2002), which greatly facilitates similarity searches. In VS, fingerprints can be used to screen the database identifying compounds that match the features of a single bioactive ligand.

Fingerprints may be classified according to their dimensionality ranging from one dimensional (1D) to 3D (Bender and Glen, 2004) – see Table 2 and Figure 2. Among all commonly used, the most popular and efficient are 2D fingerprints. They range in the family of simple 2D substructure sketches, graph-based representations and more important, bit strings. In the latter type, molecules are represented by a sequence of zeros and ones that reflect the presence or absence of certain structural features (e.g. substructure, bond, torsion angle) and transformed into a hashed number. Moreover, 2D fingerprints are considered more successful in retrieving active compounds than 3D shape or docking methods (Zhang and Muegge, 2006).

Pharmacophores

When disposing three-dimensional information about the ligand and/or receptor atomic structure, it is possible to take advantage of pharmacophore-based methods, which use geometric and topological constraints. This approach is based on ligand-receptor 3D dependen-

cies and stereo-chemical rules that determine the preferred binding mode. Ligand-based pharmacophore is usually defined as a relational model of features that are common for ligands interacting with the same receptor. The above-mentioned features usually include hydrogen bonding donors and acceptors, charged groups and aromatic rings. Spatial relationships of pharmacophoric features can be then easily transformed into 3D coordinates and distances, which may form a basis for similarity searches. This approach is recommended when disposing information about a set of ligands for a given target without a solved 3D structure. In such a case, details about a target binding site should be retrieved from known binders and used in pharmacophoric screening. Pharmacophore-based strategies are widely applied in screening projects and in some cases outperform docking results (Chen et al., 2009b).

Interaction fingerprints (IFP)

Klebe introduced the concepts of “forward” and “backward” filtering (Klebe, 2006). The former one refers to applying hierarchical filtering methods prior to docking (including drug-like features and/or structure-based pharmacophore), to significantly reduce the amount of data to be screened. Conversely, “backward” filtering starts with a high-throughput docking and focuses on post-docking strategies as a subsequent step. The latter approach was recently employed in the interaction-based filtering of ligand poses generated during docking (Marcou and Rognan, 2007). Interaction fingerprint (IFP) translates 3D information about protein-ligand interaction into a bit vector that can be easily compared with other fingerprints. IFPs – which details the interaction of a known binder with a target structure – may be treated as a reference to assess other ligand-target poses (e.g. docking results). Using IFPs allows retrieval of compounds that form interactions similar to the true binding modes. Thus, including usage of IFPs into a VS procedure may significantly improve the docking result analysis and lead to identification of new interesting compounds.

Success stories

There are several successful VS stories involving 2D/3D fingerprinting methods (Mohan et al., 2005). Fingerprints can be used at the first stage of the project to create a set of chemical compounds with the desired properties. A library focused on a specific biological sy-

Table 2. Specification of small-molecule descriptors

Space	Character	Description
1D	general	global properties retrieved from chemical formula with no need of 2D/3D structure (e.g. molecular weight, number and types of atoms/bonds, number of rings, logP)
2D	topological	variety of the fragmentation strategy (e.g. type of connectivity, atom paths, appearance of fragments), or substructures (e.g. maximum common substructure)
3D	three dimensional	shape, interaction-based pharmacophores, receptor-based pharmacophore

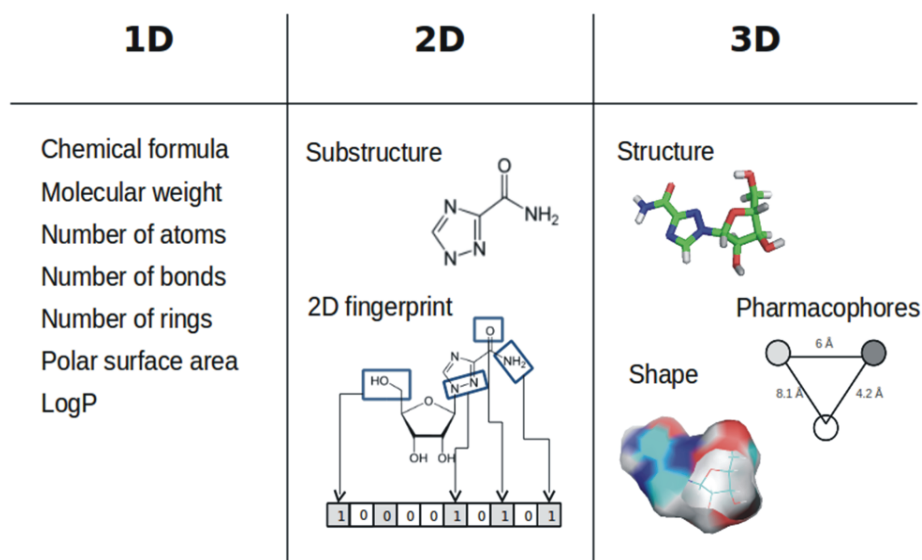


Fig. 2. Fingerprints as ligand descriptors. Depending on the information type they may range from one to three dimensional. Based on the work by Rognan, 2007

stem is obtained by selecting compounds that meet specific structural/chemical requirements (for example, ligands containing a specific substructure able to match the structural requirements of the receptor binding pocket). Using such a library allows for more accurate analyses and reduces computational demands during the docking step. The above-mentioned approach was followed in the discovery of leads for the human carbonic anhydrase II (Gruneberg et al., 2002), where 2D substructure searches and 3D pharmacophore queries combined with docking methods were successfully used in the selection of inhibitors. The activity of virtually identified compounds was then confirmed experimentally. The pharmacophore features can also be derived from the target structures analysis (receptor-based pharmacophore). More precisely, a binding site analysis may help to identify regions of favorable protein-ligand interaction. This approach, with subsequent docking procedure, was successfully applied in the discovery of novel inhibitors

of the alanine racemase (Mustata and Briggs, 2002) and *Plasmodium falciparum* hydrofolate reductase (Rastelli et al., 2003).

Final steps

Having chosen a set of organic molecules for VS, we move to the preparation stage. Considering a situation where fingerprints or chemical formula are the only available information about the ligand (for example in 2D searches) its spatial coordinates should be defined. The 3D structure can be obtained from online databases containing structural information from crystallographic/NMR experiments, like the *Cambridge Structural Database* (Allen, 2002) (CSD: 541,748 small molecule structures as of January 2011) or the *Protein Data Bank* (Berman et al., 2000) (PDB: 10,263 structures with bound ligands as of December 2010). Another option is to transform 1D/2D information into 3D spatial coordinates using commercial methods (for example, CORINA

(Sadowski et al., 1994) or open-source packages, like OpenBabel (Guha et al., 2006). The methods designed to build low-energy 3D models may usually take advantage of the already known properties of molecules deposited in the PDB or CSD (data on bond lengths, angles, torsions or molecular fragments).

Before applying the docking procedure, physico-chemical properties for the selected ligands need to be assigned: protonation states, tautomers and stereoisomers. Although default charges are assigned to meet standard biological conditions (pH 7), it is possible to model lower or higher pH. Thus, focused libraries (Walters et al., 1998) can be created to meet specific requirements for certain target structures (e.g. positively or negatively charged binding pockets).

The importance of the ligand preparation step is often underestimated and may easily lead to false-positive predictions in VS campaigns (Brenk et al., 2005). It is advisable to use the already prepared and compiled small molecule collections known as “*ready-to-dock*” ligand libraries. ZINC (Irwin and Shoichet, 2005) is a widely used, freely available database that contains millions of small ligands that can be used in VS. The 3D coordinates of compiled compounds can be downloaded with their protonation states and atomic charges already assigned. Using “*ready-to-dock*” sets may significantly improve the quality of results, especially when used by inexperienced researchers.

Current state, challenges and successful stories in structure-based virtual screening

The amount and quality of available information on the system under analysis is a key factor when designing a VS project (Klebe, 2006) – see Table 1. If the 3D structure of the macromolecule of interest is known, the coordinates of the features contained in the target region are valuable data and can be used to improve the quality of results. Exemplary methods of obtaining 3D models of biomolecules include NMR spectroscopy, X-ray crystallography or homology modeling. In the last case, the solved structure from at least one closely related homologue is a *sine qua non* requirement. Currently, PDB contains more than 70,000 experimentally solved 3D structures of proteins that can be used as targets in VS and in homology modeling. NMR spectroscopy obtains several real conformations for the receptor while X-ray crystallography offers one single state of the crystallized

protein. Even though all these methods have been successfully applied in structure-based VS, X-ray crystallography remains the most powerful source of structural data.

In this section, we focus on the *in silico* approaches that profit from 3D structural features encoded in the interaction interface to perform VS. First, an overview of the current state and widely used software is presented along with several recently published success stories in computer-assisted drug design. This is followed by a detailed description of the VS process, focusing on the recent challenging aspects in the field that still remain the bottlenecks of this method.

The protein-ligand docking (PLD) problem

The aim of a PLD experiment is to predict the 3D structure formed upon the mutual binding of two or more molecules. For this purpose, two complementary components are required: 1) a method that samples the conformational space of the ligand with respect to the target (ideally: also by sampling the internal conformational space of one or both of these molecules) thereby generating target-ligand poses and 2) a mathematical expression (the *scoring function* – SF), which evaluates the proposed poses according to the potential binding strength (Leach and Gillet, 2007). Both aspects affect the quality of the results. While the generation of poses should ensure that most of the conformational space is explored, the scoring function should efficiently evaluate each of the poses and identify active compounds and near-native conformations. In practice, these two fundamental components limit the applicability of the approach, provided that for an effective high-throughput docking (HTD) a compromise between speed and accuracy is achieved.

Over the years, many small molecule docking methods have been developed and reviewed (Cross et al., 2009; Kellenberger et al., 2004; Kirchmair et al., 2008; Moitessier et al., 2009; Plewczynski et al., 2011a). They have been applied not only to protein-ligand scenarios but also in RNA-ligand docking (Chen et al., 1997; Filikov et al., 2000; Lang et al., 2009; Leclerc and Cedergren, 1998), DNA-ligand docking (Dailey et al., 2009; Ricci and Netz, 2009) and to the modulation of protein-protein interactions (Casey et al., 2009; Meireles and Mustata, 2011). They are commonly referred to as ligand-docking software. Several widely used docking implementations are

Table 3. A list of widely used protein-ligand docking software (Sousa et al., 2006). The main strengths and weaknesses are based on the work by Kellenberger et al., 2004, and the selection of targets studied by high-throughput virtual screening is followed by favorable experimental testing. GA stands for Genetic Algorithm, HF for Hierarchical Filtering, IC for Incremental Construction, MA for Matching Algorithm and MC for Monte Carlo sampling (Moitessier et al., 2009)

Program and algorithm	Strengths	Weaknesses	Examples of recent successful virtual screening experiments
AutoDock and AutoDock Vina (Morris et al., 1998; Osterberg et al., 2002; Trott and Olson, 2010) – GA	small ligands large binding sites freely available	accuracy in highly flexible ligands low speed	Glutamate Transporter 1 (GLT1) inhibitors (Luethi et al., 2010) Cdc25 phosphatase inhibitors (Park et al., 2008) D-Ala:D-Ala ligase inhibitors (Kovac et al., 2008) Cyclodextrin-based receptors (Steffen et al., 2007)
DOCK (Ewing et al., 2001; Kuntz et al., 1982; Lang et al., 2009; Moustakas et al., 2006; Oshiro et al., 1995) – IC	small binding sites opened cavities small hydrophobic ligands freely available	accuracy in highly flexible ligands highly polar ligands	Hepatitis C virus helicase inhibitors (Chen et al., 2009a) SARS-CoV 3C-like proteinase inhibitors (Liu et al., 2005) Cyclooxygenase (COX-2) inhibitors (Mozziconacci et al., 2005)
FlexX (Rarey et al., 1996) – IC	small binding sites small hydrophobic ligands	very flexible ligands	Bacterial NAD synthetase inhibitors (Moro et al., 2009) Lymphoid phosphatase inhibitors (Wu et al., 2009) RNA polymerase inhibitors (Kim et al., 2008) ATP phosphoribosyltransferase (HisG) inhibitors (Cho et al., 2008) Human histamine H4 receptor ligands (Kiss et al., 2008)
Glide (Friesner et al., 2004) – HF+MC	flexible ligands small hydrophobic ligands	ranking for very polar ligands low speed	Liver X receptor modulators (Cheng et al., 2008) HIV-1 reverse transcriptase inhibitors (Barreiro et al., 2007)
GOLD (Verdonk et al., 2003; Verdonk et al., 2005) – GA	small binding sites small hydrophobic ligands buried binding pockets	ranking for very polar ligands or large cavities	HIV-1: CD4-gp120 binding inhibitors (Lalonde et al., 2011) Serotonin 5-HT(7)R antagonists (Kurczab et al., 2010) Non-peptide β -secretase inhibitors (Xu et al., 2010) Sarco/endoplasmic reticulum calcium ATPase inhibitors (Deye et al., 2009) <i>Trypanosoma cruzi</i> trans-sialidase inhibitors (Neres et al., 2009)
Surflex (Jain, 2003; Jain, 2007) – IC+MA	large and opened cavities small binding sites very flexible ligands	low speed for large ligands	Triple helical DNA intercalators (Holt et al., 2009) ErmC methyltransferase inhibitors (Feder et al., 2008) Hepatitis C virus NS5B polymerase inhibitors (Musmuca et al., 2010)

summarized in Table 3 (Sousa et al., 2006). Although most of them require a commercial license, academic users and researchers can experiment using freely available software like DOCK (Lang et al., 2009) or AutoDock Vina (Trott and Olson, 2010). Each docking program has its own algorithms for the generation and scoring of the ligand poses, and as a result exhibits individual strengths and weaknesses for specific ligands (see Table 3) and/or targets (Cross et al., 2009). While most of the software available performs poorly in generating near native poses for highly flexible ligands, Glide software (Friesner et al., 2004) overcomes this issue by running a more accurate search along the conformational space (random search followed by structure refinement), which results in an overall slow speed.

Although the available methods are far from being perfect, their applicability in high-throughput VS has been largely demonstrated and reported. Table 3 summarizes recent VS campaigns performed using each of the implementations followed by experimental confirmation of the identified hits. It is worth noting that divergence among the targeted receptors (several protein families, RNA, DNA) indicates that each of the methods reported here represents a valuable tool in drug design that can manifest itself in different types of the system under examination. For example, the discovery and experimental confirmation of novel quinolone-based triple helical DNA intercalators (Holt et al., 2009) were based on using Surflex software (Jain, 2007). The compounds identified in that experiment were found to selectively

bind the triplex DNA poly(dA)-[poly(dT)]₂ in the *in vitro* tests. Recently, our group also used Surflex to identify potential inhibitors of the ErmC' methyltransferase which is responsible for bacterial resistance against macrolide antibiotics (Feder et al., 2008). Further experimental tests of the best scored docking ligands positively identified eight active compounds. Another interesting VS experiment revealed novel potential non-peptide inhibitors for the β -secretase (Xu et al., 2010). After running a VS experiment with GOLD (Verdonk et al., 2003), *in vitro* testing demonstrated not only high affinity for the receptor but also low cytotoxicity against neuroblastoma cells. This discovery provides new chemical entities that target an enzyme that is of great interest in the treatment of Alzheimer's disease. More examples of successful *in silico* assisted drug design stories can be found in the exhaustive research book *Success Stories of Computer-Aided Design* (Kubinyi, 2006).

Designing a high-throughput docking experiment

A high-throughput docking (HTD) experiment aims to identify active compounds from large libraries of possible ligands. The general steps are shown on the schematic Figure 3. Being a knowledge-driven approach, structure-based VS requires three-dimensional information about the binding pocket of the receptor. As mentioned earlier, this information can be collected from experimental or computational sources. When choosing a receptor structure for docking, careful inspection of the physico-chemical properties of the pocket is advised. The pocket size and shape must be suitable for docking and should consist of chemical groups able to tightly interact with the ligand (mainly charged and hydrogen-bond donors/acceptors) (Pérot et al., 2010). If true binders are known, a comparison of the docked poses and their affinities with experimental data is strongly recommended. For targets in which the location of the binding pocket is unknown, annotated homologue proteins based tools like AnnoLyze (Marti-Renom et al., 2007) can be used to predict the binding region of small ligands.

After the selection of the binding pocket, a careful preparation of the structure should be carried out; the undesired solvent and cofactors deleted, the protonation state for each amino acid determined and the hydrogen atoms must be added accordingly, and the charge for each atom should be specified. Chimera (Pettersen et al., 2004) is a freely available structure viewer (for

academic purposes) that can be used for automated preparation of molecules.

The preparation of ligand libraries has already been discussed in the previous section. Several approaches can be followed to generate ligand libraries suitable for HTD. As shown in Table 1, if there are known binders, a similarity or a pharmacophore search can be performed in order to generate a library of compounds that retain the physico-chemical properties of the ligands. Otherwise, general libraries that include diversity from the vast chemical space should be selected.

Prior to docking, the 3D coordinates, protonation states and charges of the ligands included in the generated library must be defined (see section Pre-filtering - Final Steps). If the compounds are gathered from "ready-to-dock" libraries, this step can be omitted.

High-throughput virtual screening: a useful tool for *in silico* drug discovery

After selecting and preparing the target and the ligands for screening, a PLD software and methodology should be chosen. Given that each implementation has its own specific features (see Table 3), this is a crucial step in structure-based VS. Since the first version of DOCK was made available in 1982, many changes have been introduced to PLD methods. Initially, both the target and the ligand were treated as rigid entities during the generation of poses, simplifying the space of possible protein-ligand complexes to a six-dimensional problem (three rotations and three translations). The rigidity imposed during this so-called "rigid-body" docking reduces the computational requirements dramatically. However, the conformations of both the ligand and receptor are altered upon the formation of the complex and, therefore, most of the rigidbody docking tools are not able to accurately reproduce the experimental observations. As a result, most of the current implementations take into account ligand flexibility by default.

In the DOCK and Surflex software, an "anchor-and-grow" algorithm divides the ligand into rigid fragments and docks the largest fragment into the binding pocket. The rest of the fragments are bound sequentially to the anchor in order to maximize the score of the final conformer. Another option is to generate and evaluate the poses *on-the-fly*. This is the case in the ICM (Abagyan et al., 1994) software, which uses the Metropolis/Monte-Carlo algorithm to generate random poses and

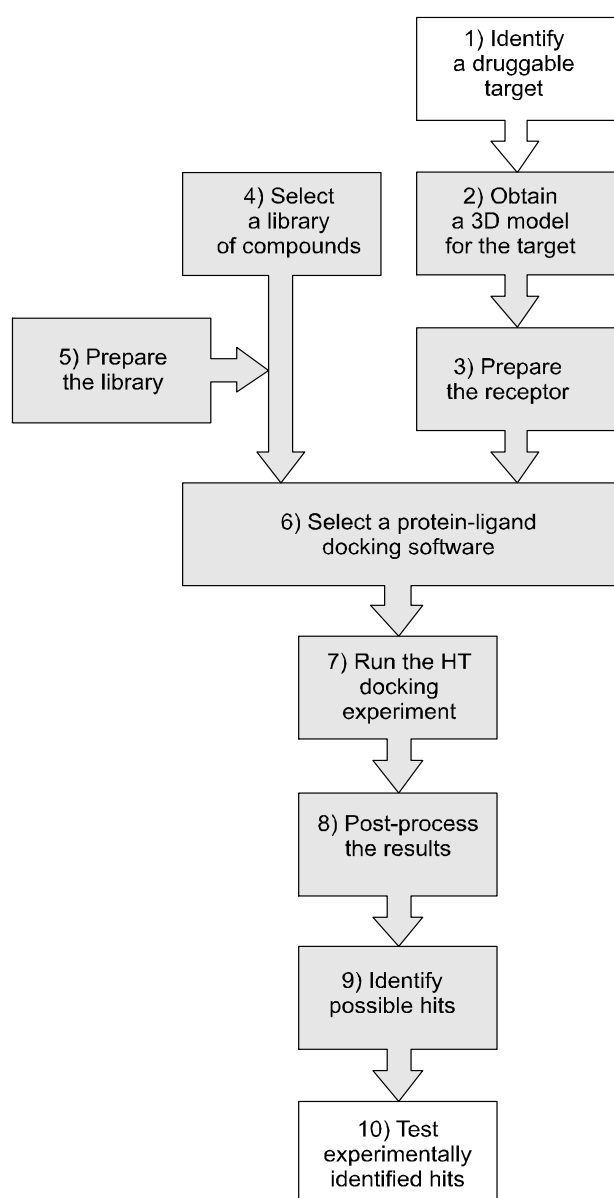


Fig. 3. Proposed workflow in a high-throughput docking experiment. Shaded steps are covered in this paper. 1) A biologically relevant target modulated by a small ligand is identified and 2) a model of its 3D structure is obtained. 3) A model for the selected receptor is prepared. 4) A library of small chemical compounds is selected and, if needed, 5) prepared for the virtual screening process. The information regarding the nature of the target and the compounds to be screened can be used to 6) select the appropriate protein-ligand docking software and method. After 7) screening the compounds, several methods have been developed to 8) post-process the results and assist in 9) the identification of possible hits. Finally, 10) the identified hits are purchased or synthesized, and their activity is validated by experimental testing

guide the exploration. There are other methods, such as AutoDock and GOLD, which rely on a genetic algorithm to explore the conformational space of the system. Ini-

tially, the coordinates of ligand fragments are randomly mixed. The new conformations are evaluated and ranked, and the highly scored ones are selected to build the next “generation” of poses. To ensure an appropriate sampling of the conformational space, this process is repeated millions of times for each ligand. In HADDOCK (Dominguez et al., 2003) software, an initial rigid-body search is performed to screen poses rapidly, and it is followed by a minimization to accommodate the ligand into the binding pocket. An interesting feature in this implementation is the possibility of using knowledge about the target-ligand binding mode to define a set of distance constraints. A complete analysis of implementations and algorithms has been recently reported (Moitesier et al., 2009) and may serve as a detailed guide for algorithms involved in VS.

A method that allows experimentation in VS without dealing with large highly flexible ligands relies on so-called fragment-based VS (Feyfant et al., 2011; Na and Hu, 2011). In this approach, instead of docking the entire structure of the ligands, libraries of chemical fragments are prepared and docked into the pocket, thus allowing for fast screen of large and diverse sets of compounds. After screening, the selected fragments can be used as scaffolds to retrieve ligands from the database. This method has been successfully applied in the *in silico* drug discovery of novel compounds (Englert et al., 2010; Hartenfeller and Schneider, 2011). For example, novel fragment inhibitors of class 1 phosphatidylinositide 3-kinases were reported to use a homology modeled receptor (Giordanetto et al., 2011). Experimental testing confirmed the potency of compounds belonging to several different structural classes.

In many real cases, rearrangements in the receptor’s binding pocket upon complex formation have been observed. Moreover, they dramatically affect the accuracy of the prediction of binding modes and affinities, and require a proper treatment of flexibility for both the ligand and receptor during the VS (Lin, 2011; Totrov and Abagyan, 2008). This is a particularly challenging aspect in HTD due to the dimensionality of the conformational space, which significantly influences the efficiency of screening. The fastest way to model target flexibility relies on the use of “softer” scoring functions. The weight of the intermolecular clashing atoms is lowered and allows a deeper penetration of the ligand into the target surface. Another method that can be used in such

cases is the so-called multiple receptor docking. In this procedure, the ligands are sequentially docked into an ensemble of conformations for the target, from which the best scored poses are selected (Popov et al., 2007). It is noteworthy that each additional conformation used in an HTD experiment increases the chance of retrieving false positives, and therefore a careful and critical analysis of the results is advisable. The conformations needed for multiple receptor docking can be obtained directly from NMR spectroscopy or multiple X-ray crystallography, or may be selected from the frames obtained in a Molecular Dynamics simulation or by means of the Normal Mode Analysis (Sperandio et al., 2010). Computational resources limit an explicit treatment of receptor flexibility in HTD. Generally, a few side-chains in the binding pocket are selected to be flexible and their conformational space is explored during docking. There are several approaches to this principle. For example, while AutoDock Vina performs an exhaustive search for each active rotatable bond, other algorithms select *on-the-fly* the conformations to be screened by taking into account the physical features of the pocket (e.g. SLIDE – see Zavodszky and Kuhn, 2005). The conformations for the side-chains can also be directly obtained from external libraries of rotamers. In the case of Surflex, after rigid receptor docking, the binding pocket is shortly optimized to accommodate the ligand, and the pose is re-scored. After pocket adaptation, significant improvements in the binding mode prediction were reported (Jain, 2009). Figure 4 shows the result of a docking experiment performed using the Surflex.

Limitations of widely used PLD routines related to the efficiency and sensibility to ligand and receptor nature have already been discussed and summarized in Table 3. The consistency of the methods can be affected by several random factors. For example, in SLIDE and QXP (McMartin and Bohacek, 1997) the input coordinates of the ligands were found to have an impact on the docking pose (as could be seen in the experiment) (Kellenberger et al., 2004). Effect of physico-chemical properties of ligands such as protonation, tautomerism and isomerism have been recently analyzed in the docking results (ten Brink and Exner, 2009). The inclusion of such modifications to extend ligand libraries remains an area of active research. Even though these features are known to play an important role in binding, current scoring functions are generally unable to handle such

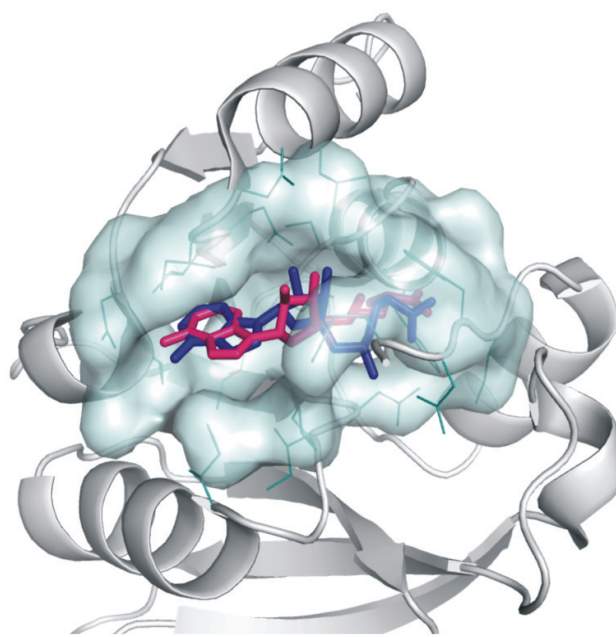


Fig. 4. Crystallographic pose and predicted binding mode of the inhibitor sinefungin (Schluckebier et al., 1999) in the binding pocket of ErmC' methyltransferase (in blue and pink, respectively). The receptor structure was obtained using the PDB database (code 1QAQ). Surflex was used to generate the poses using flexible ligand docking. Rescoring using short optimization of the binding pocket was performed afterwards and the best scored pose was selected. The figure was generated using PyMOL

structural variations dynamically, i.e. during the docking procedure. Nevertheless, Glide can automatically generate tautomeric forms for the ligands and ZINC offers the possibility of downloading ligands with several protonation and tautomeric variants.

To overcome the lack of generality and consistency of different methods (see Table 3), a consensus docking procedure has been proposed. Under this approach, the binding modes of protein-ligand complexes are obtained by comparing the results of several PLD software packages. Although significant increases in accuracy have been reported, the efficiency of running several simultaneous screens remains a holdup point for high-throughput virtual screening. Nevertheless, the combination of docking methods has been applied for the *in silico* discovery of novel chemicals (Bai et al., 2010; Choowongkamon et al., 2010; Zhang et al., 2010). Many automated consensus docking methods have been proposed and developed over the years (Okamoto et al., 2010; Paul and Rognan, 2002; Wolf et al., 2007). The Ginalski group

focuses on the critical assessment of docking methodologies and has developed VoteDock (Plewczynski et al., 2011b) – a consensus docking routine. Internal benchmarking showed that the performance of the method, in terms of both accuracy and consistency, exceeds the performance of individual protein-ligand software.

Post-processing of high-throughput docking results

The results obtained from HTD calculations should be carefully post-processed and analyzed before hits are proposed for experimental validation. For the fast generation of poses, simplified scoring functions (SF) are required and a common step after docking involves re-scoring of the poses using more accurate models. The Zou group recently published a detailed review on SF (Huang et al., 2010). Here, we present a brief overview of current state-of-the-art and the challenges in this field. There are three main goals of scoring: 1) to identify native binding modes, 2) to accurately predict binding affinities, and 3) to identify potential binders from huge libraries of compounds. For high-throughput virtual screening, the complementary prediction of affinities and selection of binders are critical for successful results. Existing SF can be classified into three categories depending on the parameterization methodology: 1) force-field based, derived from physical potentials, 2) empirical, derived from regression analysis, and 3) knowledge-based, derived from statistical analysis of experimentally determined complexes.

In a recent publication (Cheng et al., 2009), the performance of 16 widely used SF belonging to the three categories was analyzed and no single scoring function was able to consistently out-perform other functions in a set of benchmarking scenarios. Nevertheless, the knowledge-based methods GOLD::ASP and DrugScore and the empirical X-Score, were found to perform better than the other SF. In another benchmark, a set of 16 SF was tested and it was found that a knowledge-based SF (Huang and Zou, 2010), namely ITScore and DrugScore, outperformed the other methods. Thus, empirical and knowledge-based SFs are reported to have higher success rates and better correlation with the experimental affinities than force-field based methods.

A relevant measure for HTS is the enrichment factor – the ability of an SF to identify binders from huge libraries of compounds and score them accordingly, so that they are ranked in the first positions in the database.

Current SFs strongly depend on the nature of the target, irrelevant of the method used to parameterize the models. However, significant enrichment factors were observed using SF of any source (Huang et al., 2010).

In an attempt to address the lack of consistency of the methods available, multiple scoring functions can be used to evaluate the poses. Afterwards, simple statistics like linear combination of weighted methods or averaged rank among the solutions is used to determine the final score for each pose. By analogy to consensus docking, this method is known as consensus scoring (Charifson et al., 1999). The method was applied, for example, in the *in silico* screening of inhibitors of Hepatitis C virus RNA-dependent RNA polymerase (Lin et al., 2011). In that study, 11 scoring functions were used to rank the HTD results. Afterwards, 14 compounds were found to be highly scored across the rankings and selected for experimental testing, to identify several novel active compounds. Simultaneously, the Ginalski group has recently developed HarmonyDOCK (Plewczynski et al., 2010), a consensus scoring method for automated evaluation of protein-ligand results. The current state of consensus scoring and improvements in its performance were recently analyzed (Feher, 2006).

Visualization and analysis of the results

Main limitations of PLD methods result from simplified scoring functions and weak exploration of the conformational space for the ligand. As a result, some highly scored ligands may be incorrectly identified as binders. There are some general hints that assist in the identification of undesired hits. Due to limitations in scoring, compounds can exhibit poses with poor physico-chemical interactions with the binding pocket, imperfect hydrogen-bonding network or poses based purely on hydrophobic interactions and shape complementarity. Solvation effects are roughly implemented in scoring functions and, as a result, residual chains on the ligand may not interact with the target. The generation of poses outside the binding pocket is also common.

To exclude ligands with improper high scores, visual inspection of VS results is required. Several software packages are available specifically for the structural analysis of biomolecules. For example, protein-ligand complexes can be visualized and inspected using the freely available tools PyMOL, VMD (Humphrey et al., 1996), Swiss-PDB Viewer (Guex and Peitsch, 1997) and Chime-

ra (Pettersen et al., 2004). PyMOL was used to generate the PLD example shown in Figure 4.

Visual filtering of identified hits and toxicity as well as checking the physico-chemical properties of the remaining compounds is strongly advised if pre-filtering methods have not been applied prior to docking (see section Pre-filtering). Finally, the list of virtually identified hits should be tested experimentally to determine the activity of the compounds.

Conclusions

Structure-based VS techniques have been presented as a promising alternative to experimental HTS. In contrast to experimental methods, VS strategies require prior biological information to identify active compounds.

A general rule for VS analysis is to search only among biologically relevant and synthetically accessible compounds. The filtering step ensures rejection of compounds that do not meet specific drug-like criteria. However, the use of biased filters carries the risk of omitting interesting ligands. Identification of compounds with particular drug-like properties may be performed by using fingerprints or pharmacophores. These methods allow selective searches of compounds based on information retrieved from ligand structures or target-ligand interactions.

If the structure of the target is known, ligands from pre-filtered libraries can be virtually docked into the binding pocket using protein-ligand docking software. Structure-based VS introduces a set of drawbacks, among which the most significant ones are poor sampling methods and inaccurate scoring functions.

Notwithstanding filtering and docking caveats, VS is a widely used tool in drug design. The method is used to screen large libraries of compounds, complementing experimental HTS techniques in the identification of hits with pharmacological interest. These features make VS a useful and promising tool for modern computer-assisted drug discovery and design.

Acknowledgements

We would like to thank Agnieszka Bielska, Andrea Hall, Mairead Mooney, and Piotr Szczechowiak for the proofreading of this manuscript. The authors' work on methods development has been supported by the Foundation for Polish Science (FNP, grant TEAM/2009-4/2) and Polish Ministry of Science and Higher Education (MNiSW, grant 0067/P01/2010/70 to

AC and JMK, grant N N301 123138 to JMK), and by the European Research Council (StG grant RNA+P=123D to JMB, contract number 261351). E.B. and K.H.K. have been supported by the MNiSW (grants Iuventus Plus 0160/H03/2010/70 and N N401 585738 to K.H.K.). JMK has been a scholarship holder of the Adam Mickiewicz University Foundation. X.L. has been supported by the Ministry of Education and Science of Spain (MEC, grants MAS2010-00070 and EDU/1781/2010). K.H.K. has been supported by the START fellowship from the FNP. JMB has been additionally supported by the "Ideas for Poland" fellowship from the FNP.

References

- Abagyan R., Totrov M., Kuznetsov D. (1994) *ICM – A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation*. J. Comput. Chem. 15: 488-506.
- Allen F.H. (2002) *The Cambridge Structural Database: a quarter of a million crystal structures and rising*. Acta Crystallogr. B. 58: 380-388.
- Bai Y., Watt B., Wahome P.G., Mantis N.J., Robertus J.D. (2010) *Identification of new classes of ricin toxin inhibitors by virtual screening*. Toxicon. 56: 526-534.
- Bajorath J. (2004) *Cheminformatics: Concepts, Methods, and Tools for Drug Discovery*. Totowa: Humana Press Inc.
- Barreiro G., Guimarães C.R.W., Tubert-Brohman I., Lyons T.M., Tirado-Rives J., Jorgensen W.L. (2007) *Search for non-nucleoside inhibitors of HIV-1 reverse transcriptase using chemical similarity, molecular docking, and MM-GB/SA scoring*. J. Chem. Inf. Model. 47: 2416-2428.
- Beautrait A., Leroux V., Chavent M., Ghemtio L., Devignes L., Smail-Tabbone M., Cai W., Shao X., Moreau G., Bladon P. (2008) *Multiple-step virtual screening using VSM-G: overview and validation of fast geometrical matching enrichment*. J. Mol. Model. 14: 135-148.
- Bender A., Glen R.C. (2004) *Molecular similarity: a key technique in molecular informatics*. Org. Biomol. Chem. 2: 3204-3218.
- Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000) *The Protein Data Bank*. Nucl. Acids Res. 28: 235-242.
- Bohacek R.S., McMartin C., Guida W.C. (1996) *The art and practice of structure-based drug design: a molecular modeling perspective*. Med. Res. Rev. 16: 3-50.
- Brenk R., Irwin J.J., Shoichet B.K. (2005) *Here be dragons: docking and screening in an uncharted region of chemical space*. J. Biomol. Screen. 10: 667-674.
- Casey F.P., Pihan E., Shields D.C. (2009) *Discovery of small molecule inhibitors of protein-protein interactions using combined ligand and target score normalization*. J. Chem. Inf. Model. 49: 2708-2717.
- Charifson P.S., Corkery J.J., Murcko M.A., Walters W.P. (1999) *Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins*. J. Med. Chem. 42: 5100-5109.

- Chen Q., Shafer R.H., Kuntz I.D. (1997) *Structure-based discovery of ligands targeted to the RNA double helix*. *Biochemistry* 36: 11402-11407.
- Chen C.S., Chiou C.T., Chen G.S., Chen S.C., Hu C.Y., Chi W.K., Chu Y.D., Hwang L.H., Chen P.J., Chen D.S., Liaw S.H., Chern J.W. (2009a) *Structure-based discovery of triphenylmethane derivatives as inhibitors of hepatitis C virus helicase*. *J. Med. Chem.* 52: 2716-2723.
- Chen Z., Li H., Zhang Q., Bao X., Yu K., Luo X., Zhu W., Jiang H. (2009b) *Pharmacophore-based virtual screening versus docking-based virtual screening: a benchmark comparison against eight targets*. *Acta Pharm. Sinic.* 30: 1694-1708.
- Cheng J.F., Zapf J., Takedomi K., Fukushima C., Ogiku T., Zhang S.H., Yang G., Sakurai N., Barbosa M., Jack R., Xu K. (2008) *Combination of virtual screening and high throughput gene profiling for identification of novel liver X receptor modulators*. *J. Med. Chem.* 51: 2057-2061.
- Cheng T., Li X., Li Y., Liu Z., Wang R. (2009) *Comparative assessment of scoring functions on a diverse test set*. *J. Chem. Inf. Model.* 49: 1079-1093.
- Cho Y., Ioerger T.R., Sacchetti J.C. (2008) *Discovery of novel nitrobenzothiazole inhibitors for Mycobacterium tuberculosis ATP phosphoribosyl transferase (HisG) through virtual screening*. *J. Med. Chem.* 51: 5984-5992.
- Choowongkamon K., Sawatdichaikul O., Songtawee N., Limtrakul J. (2010) *Receptor-based virtual screening of EGFR kinase inhibitors from the NCI diversity database*. *Molecules* 15: 4041-4054.
- Cross J.B., Thompson D.C., Rai B.K., Baber J.C., Fan K.Y., Hu Y., Humblet C. (2009) *Comparison of several molecular docking programs: pose prediction and virtual screening accuracy*. *J. Chem. Inf. Model.* 49: 1455-1474.
- Dailey M.M., Hait C., Holt P.A., Maguire J.M., Meier J.B., Miller M.C., Petraccone L., Trent J.O. (2009) *Structure-based drug design: from nucleic acid to membrane protein targets*. *Exp. Mol. Pathol.* 86: 141-150.
- Deye J., Elam C., Lape M., Ratliff R., Evans K., Paula S. (2009) *Structure-based virtual screening for novel inhibitors of the sarco/endoplasmic reticulum calcium ATPase and their experimental evaluation*. *Bioorgan. Med. Chem.* 17: 1353-1360.
- Doman T.N., McGovern S.L., Witherbee B.J., Kasten T.P., Kurumbail R., Stallings W.C., Connolly D.T., Shoichet B.K. (2002) *Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B*. *J. Med. Chem.* 45: 2213-2221.
- Dominguez C., Boelens R., Bonvin A.M.J.J. (2003) *HADDOCK: a protein-protein docking approach based on biochemical or biophysical information*. *J. Am. Chem. Soc.* 125: 1731-1737.
- Englert L., Silber K., Steuber H., Brass S., Over B., Gerber H.D., Heine A., Diederich W.E., Klebe G. (2010) *Fragment-based lead discovery: screening and optimizing fragments for thermolysin inhibition*. *ChemMedChem.* 5: 930-940.
- Ertl P., Jelfs S., Muhlbacher J., Schuffenhauer A., Selzer P. (2006) *Quest for the rings. In silico exploration of ring universe to identify novel bioactive heteroaromatic scaffolds*. *J. Med. Chem.* 49: 4568-4573.
- Evans B.E., Rittle K.E., Bock M.G., DiPardo R.M., Freidinger R.M., Whitter W.L., Lundell G.F., Veber D.F., Anderson P.S., Chang R.S., et al. (1988) *Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists*. *J. Med. Chem.* 31: 2235-2246.
- Evensen E., Eksterowicz J.E., Stanton R.V., Oshiro C., Grotenhuis P.D., Bradley E.K. (2003) *Comparing performance of computational tools for combinatorial library design*. *J. Med. Chem.* 46: 5125-5128.
- Ewing T.J., Makino S., Skillman A.G., Kuntz I.D. (2001) *DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases*. *J. Comput. Aid. Mol. Des.* 15: 411-428.
- Feder M., Purta E., Koscinski L., Cubrilo S., Maravic Vlahovick G., Bujnicki J.M. (2008) *Virtual screening and experimental verification to identify potential inhibitors of the ErmC methyltransferase responsible for bacterial resistance against macrolide antibiotics*. *ChemMedChem* 3: 316-322.
- Feher M. (2006) *Consensus scoring for protein-ligand interactions*. *Drug Discov. Today* 11: 421-428.
- Feyfant E., Cross J.B., Paris K., Tsao D.H.H. (2011) *Fragment-based drug design*. *Meth. Mol. Biol.* 685: 241-252.
- Filikov A.V., Mohan V., Vickers T.A., Griffey R.H., Cook P.D., Abagyan R.A., James T.L. (2000) *Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR*. *J. Comp. Aid. Mol. Des.* 14: 593-610.
- Friesner R.A., Banks J.L., Murphy R.B., Halgren T.A., Klicic J.J., Mainz D.T., Repasky M.P., Knoll E.H., Shelley M., Perry J.K., Shaw D.E., Francis P., Shenkin P.S. (2004) *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy*. *J. Med. Chem.* 47: 1739-1749.
- Gasteiger J. (2003) *Handbook of Chemoinformatics*, WILEY-VCH, Weinheim.
- Giordanetto F., Kull B., Dellsén A. (2011) *Discovery of novel class 1 phosphatidylinositide 3-kinases (PI3K) fragment inhibitors through structure-based virtual screening*. *Bioorg. Med. Chem. Lett.* 21: 829-835.
- Gleeson M.P. (2008) *Generation of a set of simple, interpretable ADMET rules of thumb*. *J. Med. Chem.* 51: 817-834.
- Gruneberg S., Stubbs M.T., Klebe G. (2002) *Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation*. *J. Med. Chem.* 45: 3588-3602.
- Guex N., Peitsch M.C. (1997) *SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling*. *Electrophoresis* 18: 2714-2723.
- Guha R., Howard M.T., Hutchison G.R., Murray-Rust P., Rzepa H., Steinbeck C., Wegner J.K., Willighagen E. (2006) *The Blue Obelisk - Interoperability in Chemical Informatics*. *J. Chem. Inf. Model.* 46(3): 991-998.

- Hamersa L., Hemerycka Y., Herweyera G., Janssen M., Ketersa H., Rousseau R. (2002) *Similarity measures in scientometric research: The Jaccard index versus Salton's cosine formula*. Inform. Process Manag. 25: 315-318.
- Hartenfeller M., Schneider G. (2011) *De novo drug design*. Meth. Mol. Biol. 672: 299-323.
- Holt P.A., Ragazzon P., Strekowski L., Chaires J.B., Trent J.O. (2009) *Discovery of novel triple helical DNA intercalators by an integrated virtual and actual screening platform*. Nucl. Acids Res. 37: 1280-1287.
- Huang S.Y., Zou X. (2010) *Inclusion of solvation and entropy in the knowledge-based scoring function for protein-ligand interactions*. J. Chem. Inf. Model. 50: 262-273.
- Huang S.Y., Grinter S.Z., Zou X. (2010) *Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions*. Phys. Chem. Ch. Ph. 12: 12899-12908.
- Humphrey W., Dalke A., Schulten K. (1996) *VMD: visual molecular dynamics*. J. Mol. Graph. 14: 33-8, 27-8.
- Irwin J.J., Shoichet B.K. (2005) *ZINC – a free database of commercially available compounds for virtual screening*. J. Chem. Inf. Model. 45: 177-182.
- Jain A.N. (2003) *Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine*. J. Med. Chem. 46: 499-511.
- Jain A.N. (2007) *Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search*. J. Comp. Aid. Mol. Des. 21: 281-306.
- Jain A.N. (2009) *Effects of protein conformation in docking: improved pose prediction through protein pocket adaptation*. J. Comp. Aid. Mol. Des. 23: 355-374.
- Jenkins J.L., Kao R.Y., Shapiro R. (2003) *Virtual screening to enrich hit lists from high-throughput screening: a case study on small-molecule inhibitors of angiogenin*. Proteins 50: 81-93.
- Kellenberger E., Rodrigo J., Muller P., Rognan D. (2004) *Comparative evaluation of eight docking tools for docking and virtual screening accuracy*. Proteins 57: 225-242.
- Kim J., Kim K., Kim D.E., Chong Y. (2008) *Identification of novel HCV RNA-dependent RNA polymerase inhibitors using pharmacophore-guided virtual screening*. Chem. Biol. Drug. Des. 72: 585-591.
- Kirchmair J., Markt P., Distinto S., Wolber G., Langer T. (2008) *Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—what can we learn from earlier mistakes?* J. Comp. Aid. Mol. Des. 22: 213-228.
- Kiss R., Kiss B., Könczöl A., Szalai F., Jelinek I., László V., Noszál B., Falus A., Keseru G.M. (2008) *Discovery of novel human histamine H4 receptor ligands by large-scale structure-based virtual screening*. J. Med. Chem. 51: 3145-3153.
- Klebe G. (2006) *Virtual ligand screening: strategies, perspectives and limitations*. Drug Discov. Today 11: 580-594.
- Kovac A., Konc J., Vehar B., Bostock J.M., Chopra I., Janezic D., Gobec S. (2008) *Discovery of new inhibitors of D-alanine: D-alanine ligase by structure-based virtual screening*. J. Med. Chem. 51: 7442-7448.
- Kubinyi H. (2006) *Success Stories of Computer-Aided Design*, in: *Computer Applications in Pharmaceutical Research and Development*, ed. S. Ekins, Hoboken: John Wiley & Sons.
- Kuntz I.D., Blaney J.M., Oatley S.J., Langridge R., Ferrin T.E. (1982) *A geometric approach to macromolecule-ligand interactions*. J. Mol. Biol. 161: 269-288.
- Kurczab R., Nowak M., Chilmonczyk Z., Sylte I., Bojarski A.J. (2010) *The development and validation of a novel virtual screening cascade protocol to identify potential serotonin 5-HT(7)R antagonists*. Bioorg. Med. Chem. Lett. 20: 2465-2468.
- Lalonde J.M., Elban M.A., Courter J.R., Sugawara A., Soeta T., Madani N., Princiotta A.M., Kwon Y.D., Kwong P.D., Schön A., Freire E., Sodroski J., Smith A.B., 3rd (2011) *Design, synthesis and biological evaluation of small molecule inhibitors of CD4-gp120 binding based on virtual screening*. Bioorg. Med. Chem. 19: 91-101.
- Lang P.T., Brozell S.R., Mukherjee S., Pettersen E.F., Meng E.C., Thomas V., Rizzo R.C., Case D.A., James T.L., Kuntz I.D. (2009) *DOCK 6: combining techniques to model RNA-small molecule complexes*. RNA. 15: 1219-1230.
- Leach A.R., Gillet V.J. (2007) *An Introduction to Chemoinformatics*. Dordrecht: Kluwer Academic Publisher.
- Leach A.R., Shoichet B.K., Peishoff C.E. (2006) *Prediction of protein-ligand interactions. Docking and scoring: successes and gaps*. J. Med. Chem. 49: 5851-5855.
- Leclerc F., Cedergren R. (1998) *Modeling RNA-ligand interactions: the Rev-binding element RNA-aminoglycoside complex*. J. Med. Chem. 41: 175-182.
- Lin J.H. (2011) *Accommodating protein flexibility for structure-based drug design*. Curr. Top Med. Chem. 11: 171-178.
- Lin Y.T., Huang K.J., Tseng C.K., Chen K.J., Wang H.M., Lee J.C. (2011) *Efficient In Silico Assay of Inhibitors of Hepatitis C Virus RNA-Dependent RNA Polymerase by Structure-Based Virtual Screening and In Vitro Evaluation*. Assay Drug Dev. Tech. 3: 290-298.
- Liu Z., Huang C., Fan K., Wei P., Chen H., Liu S., Pei J., Shi L., Li B., Yang K., Liu Y., Lai L. (2005) *Virtual screening of novel noncovalent inhibitors for SARS-CoV 3C-like proteinase*. J. Chem. Inf. Model. 45: 10-17.
- Luethi E., Nguyen K.T., Bürzle M., Blum L.C., Suzuki Y., Hediger M., Reymond J.L. (2010) *Identification of Selective Norbornane-Type Aspartate Analogue Inhibitors of the Glutamate Transporter 1 (GLT-1) from the Chemical Universe Generated Database (GDB)*. J. Med. Chem. 53: 7236-7250.
- Marcou G., Rognan D. (2007) *Optimizing fragment and scaffold docking by use of molecular interaction fingerprints*. J. Chem. Inf. Model. 47: 195-207.
- Marti-Renom M.A., Rossi A., Al-Shahrour F., Davis F.P., Pieper U., Dopazo J., Sali A. (2007) *The AnnoLite and AnnoLyze programs for comparative annotation of protein structures*. BMC Bioinformatics 8(Suppl 4): S4.

- McMartin C., Bohacek R.S. (1997) *QXP: powerful, rapid computer algorithms for structure-based drug design*. J. Comp. Aid. Mol. Des. 11: 333-344.
- Meireles L.M.C., Mustata G. (2011) *Discovery of modulators of protein-protein interactions: current approaches and limitations*. Curr. Top Med. Chem. 11: 248-257.
- Mestres J., Knegtel R. (2000) *Similarity versus docking in 3D virtual screening*. Persp. Drug Disc. Des. 20: 191-207.
- Mohan V., Gibbs A.C., Cummings M.D., Jaeger E.P., DesJarlais R.L. (2005) *Docking: successes and challenges*. Curr. Pharm. Des. 11: 323-333.
- Moitessier N., Englebienne P., Lee D., Lawandi J., Corbeil C.R. (2009) *Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go*. Br. J. Pharmacol. 153: S7-S26.
- Moro W.B., Yang Z., Kane T.A., Brouillette C.G., Brouillette W.J. (2009) *Virtual screening to identify lead inhibitors for bacterial NAD synthetase (NADs)*. Bioorg. Med. Chem. Lett. 19: 2001-2005.
- Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., Belew R.K., Olson A.J. (1998) *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function*. J. Comput. Chem. 19: 1639-1662.
- Moustakas D.T., Lang P.T., Pegg S., Pettersen E., Kuntz I.D., Brooijmans N., Rizzo R.C. (2006) *Development and validation of a modular, extensible docking program: DOCK 5*. J. Comp. Aid. Mol. Des. 20: 601-619.
- Mozziconacci J.C., Arnoult E., Bernard P., Do Q.T., Marrot C., Morin-Allory L. (2005) *Optimization and validation of a docking-scoring protocol; application to virtual screening for COX-2 inhibitors*. J. Med. Chem. 48: 1055-1068.
- Musmuca I., Caroli A., Mai A., Kaushik-Basu N., Arora P., Ragno R. (2010) *Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors*. J. Chem. Inf. Model. 50: 662-676.
- Mustata G.I., Briggs J.M. (2002) *A structure-based design approach for the identification of novel inhibitors: application to an alanine racemase*. J. Comp. Aid. Mol. Des. 16: 935-953.
- Na J., Hu Q. (2011) *Design of screening collections for successful fragment-based lead discovery*. Meth. Mol. Biol. 685: 219-240.
- Neres J., Brewer M.L., Ratier L., Botti H., Buschiazzi A., Edwards P.N., Mortenson P.N., Charlton M.H., Alzari P.M., Frasci A.C., Bryce R.A., Douglas K.T. (2009) *Discovery of novel inhibitors of Trypanosoma cruzi trans-sialidase from in silico screening*. Bioorg. Med. Chem. Lett. 19: 589-596.
- Okamoto M., Masuda Y., Muroya A., Yasuno K., Takahashi O., Furuya T. (2010) *Evaluation of docking calculations on X-ray structures using CONSENSUS-DOCK*. Chem. Pharm. Bull. 58: 1655-1657.
- Oshiro C.M., Kuntz I.D., Dixon J.S. (1995) *Flexible ligand docking using a genetic algorithm*. J. Comp. Aid. Mol. Des. 9: 113-130.
- Osterberg F., Morris G.M., Sanner M.F., Olson A.J., Goodsell D.S. (2002) *Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock*. Proteins 46: 34-40.
- Paiva A.M., Vanderwall D.E., Blanchard J.S., Kozarich J.W., Williamson J.M., Kelly T.M. (2001) *Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of Mycobacterium tuberculosis*. Biochim. Biophys. Acta 1545: 67-77.
- Park H., Bahn Y.J., Jung S.K., Jeong D.G., Lee S.H., Seo I., Yoon T.S., Kim S.J., Ryu S.E. (2008) *Discovery of novel Cdc25 phosphatase inhibitors with micromolar activity based on the structure-based virtual screening*. J. Med. Chem. 51: 5533-5541.
- Paul N., Rognan D. (2002) *ConsDock: A new program for the consensus analysis of protein-ligand interactions*. Proteins 47: 521-533.
- Pérot S., Sperandio O., Miteva M.A., Camproux A.C., Villoutreix B.O. (2010) *Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery*. Drug Discov. Today 15: 656-667.
- Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E. (2004) *UCSF Chimera—a visualization system for exploratory research and analysis*. J. Comput. Chem. 25: 1605-1612.
- Plewczynski D., Philips A., von Grotthuss M., Rychlewski L., Ginalski K. (2010) *HarmonyDOCK: the structural analysis of poses in protein-ligand docking*. J. Comp. Biol. 18: 1-10.
- Plewczynski D., Łażniewski M., Augustyniak R., Ginalski K. (2011a) *Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database*. J. Comput. Chem. 32: 742-755.
- Plewczynski D., Łażniewski M., von Grotthuss M., Rychlewski L., Ginalski K. (2011b) *VoteDock: consensus docking method for prediction of protein-ligand interactions*. J. Comput. Chem. 32: 568-581.
- Popov V.M., Yee W.A., Anderson A.C. (2007) *Towards in silico lead optimization: scores from ensembles of protein/ligand conformations reliably correlate with biological activity*. Proteins 66: 375-387.
- Rarey M., Kramer B., Lengauer T., Klebe G. (1996) *A fast flexible docking method using an incremental construction algorithm*. J. Mol. Biol. 261: 470-489.
- Rastelli G., Pacchioni S., Sirawaraporn W., Sirawaraporn R., Parenti M.D., Ferrari A.M. (2003) *Docking and database screening reveal new classes of Plasmodium falciparum dihydrofolate reductase inhibitors*. J. Med. Chem. 46: 2834-2845.
- Ricci C.G., Netz P.A. (2009) *Docking studies on DNA-ligand interactions: building and application of a protocol to identify the binding mode*. J. Chem. Inf. Model. 49: 1925-1935.
- Rognan D. (2007) *Chemogenomic approaches to rational drug design*. Br. J. Pharmacol. 152: 38-52.

- Sadowski J., Gasteiger J., Klebe G. (1994) *Comparison of automatic three-dimensional model builders using 639 X-ray structures*. J. Chem. Inf. Comp. Sci. 34: 1000-1008.
- Schluckebier G., Zhong P., Stewart K.D., Kavanaugh T.J., Abad-Zapatero C. (1999) *The 2.2 Å structure of the rRNA methyltransferase ErmC' and its complexes with cofactor and cofactor analogs: implications for the reaction mechanism*. J. Mol. Biol. 289: 277-291.
- Schnur D.M., Hermsmeier M.A., Tebben A.J. (2006) *Are target-family-privileged substructures truly privileged?* J. Med. Chem. 49: 2000-2009.
- Sherman W., Day T., Jacobson M.P., Friesner R.A., Farid R. (2006) *Novel procedure for modeling ligand/receptor induced fit effects*. J. Med. Chem. 49: 534-553.
- Shoichet B.K. (2004) *Virtual screening of chemical libraries*. Nature 432: 862-865.
- Sousa S.F., Fernandes P.A., Ramos M.J. (2006) *Protein-ligand docking: current status and future challenges*. Proteins 65: 15-26.
- Sperandio O., Mouawad L., Pinto E., Villoutreix B.O., Perahia D., Miteva M.A. (2010) *How to choose relevant multiple receptor conformations for virtual screening: a test case of Cdk2 and normal mode analysis*. Eur. Biophys. J. 39: 1365-1372.
- Steffen A., Thiele C., Tietze S., Strassnig C., Kämper A., Lengauer T., Wenz G., Apostolakis J. (2007) *Improved cyclodextrin-based receptors for camptothecin by inverse virtual screening*. Chemistry 13: 6801-6809.
- Ten Brink T., Exner T.E. (2009) *Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results*. J. Chem. Inf. Model. 49: 1535-1546.
- Totrov M., Abagyan R. (2008) *Flexible ligand docking to multiple receptor conformations: a practical alternative*. Curr. Opin. Struct. Biol. 18: 178-184.
- Trott O., Olson A.J. (2010) *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading*. J. Comput. Chem. 31: 455-461.
- Varnek A., Tropsha A. (2008) *Cheminformatics Approaches to Virtual Screening*. Royal Society of Chemistry, Cambridge.
- Verdonk M.L., Cole J.C., Hartshorn M.J., Murray C.W., Taylor R.D. (2003) *Improved protein-ligand docking using GOLD*. Proteins 52: 609-623.
- Verdonk M.L., Chessari G., Cole J.C., Hartshorn M.J., Murray C.W., Nissink J.W.M., Taylor R.D., Taylor R. (2005) *Modeling water molecules in protein-ligand docking using GOLD*. J. Med. Chem. 48: 6504-6515.
- Walters W.P., Stahl M., Murcko M.A. (1998) *Virtual Screening – an overview*. Drug Discov. Today. 3: 160-178.
- Willet P. (1998) *Chemical Similarity Searching*. J. Chem. Inf. Comp. Sci. 38: 983-996.
- Wolf A., Zimmermann M., Hofmann-Apitius M. (2007) *Alternative to consensus scoring – a new approach toward the qualitative combination of docking algorithms*. J. Chem. Inf. Model. 47: 1036-1044.
- Wu S., Bottini M., Rickert R.C., Mustelin T., Tautz L. (2009) *In silico screening for PTPN22 inhibitors: active hits from an inactive phosphatase conformation*. ChemMedChem 4: 440-444.
- Xu W., Chen G., Zhu W., Zuo Z. (2010) *Identification of a sub-micromolar, non-peptide inhibitor of β -secretase with low neural cytotoxicity through in silico screening*. Bioorg. Med. Chem. Lett. 20: 5763-5766.
- Zauhar R.J., Moyna G., Tian L., Li Z., Welsh W.J. (2003) *Shape signatures: a new approach to computer-aided ligand- and receptor-based drug design*. J. Med. Chem. 46: 5674-5690.
- Zavodszky M.I., Kuhn L.A. (2005) *Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis*. Protein Science: A Publication of the Protein Society 14: 1104-1114.
- Zhang Q., Muegge I. (2006) *Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring*. J. Med. Chem. 49: 1536-1548.
- Zhang Q., Li D., Wei P., Zhang J., Wan J., Ren Y., Chen Z., Liu D., Yu Z., Feng L. (2010) *Structure-based rational screening of novel hit compounds with structural diversity for cytochrome P450 sterol 14 α -demethylase from *Penicillium digitatum**. J. Chem. Inf. Model. 50: 317-325.