

Virtual Screening: A Fast Tool for Drug Design

Vivek VYAS *, Anurekha JAIN, Avijeet JAIN, Arun GUPTA

B. R. Nahata College of Pharmacy, Department of Medicinal Chemistry, Rajiv Gandhi Technical University, Mhow-Neemuch Road 458001 Mandsaur, India

Abstract

Computational screening of databases has become increasingly popular in the pharmaceutical research. Virtual screening uses computer based methods to discover new ligands on the basis of biological structures. Virtual screening is divided into structural based screening (docking) and screening using active compounds as templates (ligand based virtual screening). Ligand based screening techniques mainly focus on comparing molecular similarity analyses of compounds with known and unknown moiety, regardless of the methods of the used algorithm. Docking is a computational tool of structure based drug design to predict protein ligand interaction geometries and binding affinities. In this review we provide an overview of the already used ligand based virtual screening and the docking with various databases, filters, scores and applications in the recent research in the pharmaceutical field.

Keywords

Drug Design • Virtual Screening • Docking • Pharmacophore

Introduction

For those engaged in drug design, such as medicinal and computational chemists, the research phase can be broken down into two main tasks: identification of new compounds showing some activity against a target biological receptor, and the progressive optimization of these leads to yield a compound with improved potency and physicochemical properties *in-vitro*, and, eventually,

* Corresponding author:
E-mail: vicky_1744@yahoo.com (V. Vyas)

improved efficacy, pharmacokinetic, and toxicological profiles *in-vivo*. Identification of leads is driven either by random screening or a directed design approach, and traditionally both strategies have been of equal importance, depending on the problem in hand. The directed approach needs a rational starting point for medicinal chemists and molecular modeling scientist to exploit. Examples include the design of analogs of a drug known to be active against a target receptor and mimics of the natural substrate of an enzyme. Increasingly, the three-dimensional structure of many biological targets is being revealed by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, opening the way to the design of novel molecules that directly exploit the structural characteristics of the receptor-binding site. In recent years, this approach of *structure-based design* has had a major impact on the rational design and optimization of new lead compounds in those cases where the receptor structure is well characterized [1–3]. The practice of testing of large number of molecules for the activity in the model system that is representative of the human disease, known as screening, is a well-established fact in the pharmaceutical industry. High throughput screening technology allows for the testing of thousands to million of the molecules for activity against a new target system as a part of new drug discovery process [4–5]. Virtual screening, sometime also called *in-silico* screening, is a new branch of medicinal chemistry that represents a fast & cost effective tool for computationally screening database in search for the novel drug leads. The routes for the virtual screening go back to the structure-based drug design & molecular modeling [6].

Concept of Virtual Screening

Virtual screening uses computer based methods discover new ligand on the bases of biological structure [7]. The basic goal of the virtual screening is the reduction of the enormous virtual chemical space of small organic molecules, to synthesize and/or screen against a specific target protein, to a manageable number of the compound that inhibit a highest chance to lead to a drug candidate [8]. In theory, the applicability of virtual screening is limited only by what properties of a

compound can be calculated computationally and the perceived relevance of those properties to the problem in hand. On a practical level, further considerations include the timescale for calculation of the properties, which may be considerable for a database of, say, one million compounds. The software and hardware required yielding a timely answer. Many drug candidates fail in the clinical trials because of the reasons unrelated to the potency against intended drug target. Pharmacokinetic & toxicity issues are blamed for more than half of the failure in the clinical trials. Therefore first part of the visual screening evaluates the drug likeness of the same molecules most independent of their intended drug target [1]. The term *virtual screening* has been used to describe a process of computationally analyzing large compound collections in order to prioritize compounds for synthesis or assay [7]. A broad range of computational techniques can be applied to the problem. In our work we have focused on explicit receptor–ligand molecular docking as a means of yielding the most detailed model of the way in which a given ligand will bind to a receptor, and hence the most informative basis on which to assess which ligands are useful candidates for synthesis or assay. Although the underlying methods of virtual screening have been in use in various guises for several years, it is worth noting the recent impact on molecular modeling made by the increased availability of high-performance computing platforms. Affordable multiprocessor workstations and PC (personal computer) clusters have enabled the modeler to employ computationally demanding algorithms on a routine basis. This change is particularly relevant in the case of virtual screening, where as in the work described in this paper, computationally intensive methods such as molecular docking must be applied to very large databases of chemical structures [6].

Drug likeness screening [18–23]

Many drug candidate fail in the clinical trials reasons is unrelated in the potency against the intended drug target. Pharmacokinetic and toxicity issues are blamed for more than half of all failure in the clinical trials.

Therefore first part of the virtual screening evaluates druglikeness of small molecules, drug like molecules exhibit favorable absorption, distribution, metabolism, excretion, toxicological (ADMET) parameters. Using following types of method currently assesses Druglikeness

- Simple counting method
- Functional group filter
- Topological filter
- Pharmacophore filter

Counting scheme [17–22]

Database collections of known drug are typically used to extract knowledge about structure properties of potential drug molecules. Molecular weight, lipophilicity, charge are profiled to extract simple counting rules for relevant description of ADMET- related parameter.

Functional group filters [23]

Reactive, toxics, or otherwise unsuitable compounds, such as natural product derivatives, are removed using specific filters. Typical reactive functional groups include, for the example, reactive alkyl halide peroxide, and carbazide, Unsuitable leads may include crown ethers, disulfide, and aliphatic methylene chain seven or more long and Unsuitable natural product may include quinones, polyenes, and cycloheximidine derivative removed by using filters. Screening out the compound that contain certain atom groups are associated with toxicity provide a practical and fast way to reduce large database. Better description of toxicity may provide structure-based method to assess toxicity of the compound.

Topological drug classification [23]

It is generally assumed that compound those having the structure similarity with known drug may exhibit drug like properties themselves, such as oral bioavailability, low toxicity membrane permeability and metabolic stability. Its first part is artificial neural networks and decision trees very fast filter tool in virtual

screening approaches. Data's also collected to find structural motifs and pharmacophore features of small molecules that characterize drugs. For the analysis of virtual libraries according to the presence or absence of drug like framework, side chain or structural motifs can be used for virtual screening.

Pharmacophore points filter [22]

A simple pharmacophore filter has been introduced recently. It is based on the assumption that drug like molecules should contain at least two distinct pharmacophore groups four functional motifs have been identified that guarantee hydrogen bonding capability that are essential for the specific interaction of the drug molecules with its biological target. These motifs can be combined to functional groups that are also referred to here as pharmacophore points; they include: amine amide, alcohol, ketone, sulfone, sulfonamide, carboxylic acid carbamate, guanidine, amidine, urea, and ester.

Pharmacophore Based Virtual Screening [24–25]

It is the processes of matching atoms or functional group and the geometric relations between them to the pharmacophore in the query. Examples of the programs that perform pharmacophore based searches are 3D search UNITY, MACCS-3D and ROCS. Usually pharmacophore based search are done in two steps. First the software checks whether the compound has the atom types or functional groups required by the pharmacophore, than its checks whether the spatial arrangement of this element matches the query. Flexible 3D searches identified a higher number of hits than rigid searches do. However flexible searches are more time consuming than rigid ones. There are two main approaches for including conformational flexibility in to the search one is top generate a user defined number of representative conformation for each molecules when the database is to created, the other is to generate conformation during the search.

ROCS is using as shape based super position for identifying compound that have similar shaped. Pharmacophore model provide powerful filter tools for virtual

screening even in case where the protein structure is not available, pharmacophore filter are much faster than docking approaches, and there for greatly reduce the number of compound subjected to the more expensive docking application. Another interesting aspect of pharmacophore in virtual screening is 3D- pharmacophore diversity.

Introduction to pharmacophores

The first identification of the pharmacophore formulated by Paul Ehrlich was “a molecular framework that carries (phoros) the essential feature responsible for a drug’s (pharmacon) biological activity” [26]. This definition is slightly modified by Peter Gund to “a set of structural features in a molecule that is recognized at a receptor site and is responsible for that molecule’s biological activity” [27].

Note that in addition distance that describes the 3D relation among pharmacophore points, angles, dihedrals, and exclusion volumes are also used. Pharmacophore hypotheses for searching can be generated using structural information from active inhibitors, ligands, or from the protein active site itself [28–29].

2-D Pharmacophore searching

Searching of 2D database is of great importance for accelerating the drug discovery different strategies are pursued to search a 2D database to identified the compound of the interest Substructure search identified larger molecules that contain user define query irrespective of the environment in which the query substructure occur [30]. Biochemical data obtainable from these compounds can be used for generating structure-activity-relationship (SAR) even before synthetic plans are made for lead optimization [31]. In contrast, superstructure search are used to find smaller molecules that are embedded in the query. One problem that can arise from substructure search is that the number of the compound identified can reach into the thousands. One solution o this problem is raking of the compound based on similarity between compound in the database and in the query

[32–33]. Beyond structure similarity, activity similarity has also been subject of several studies. Similarity search can be combined with substructure for limiting the number of compound selected. Flexible searches are used to identify the compound that differs from the query structure in user-specified ways.

3-D Pharmacophore searching [34–5]

1. Ligand based pharmacophore generation

Ligand based pharmacophores are generally used when crystallographic; solution structure or molded structure of protein cannot be obtained. When a set of active compound is known and it is hypothesized that all the compounds bind in the similar way to the protein, then common group should interact with the same protein residue. Thus, a pharmacophore capturing this compound feature should be able to identified from a database novel compounds that binds to the same site of the protein as the known compounds do. The process of deriving pharmacophore is known as pharmacophore mapping, consist of three steps (1) identifying common binding element that are responsible for the biological activity; (2) generating potential conformations that active compound may adopt; and (3) determining the 3D relationship between pharmacophore element in each conformation generated.

2. Manual pharmacophore generation

Manual pharmacophore generation is used when there is an easy way to identify the common feature in a set of active compounds and/or there is experimental evidence that same functional groups should be present in the ligand for good activity. An example is the development of a pharmacophore model for dopamine-transporter (DAT) inhibitor. Pharmacophores should also have some flexibility built in, thus justifying the use of distance ranges.

3. Automatic pharmacophore generation [36]

Pharmacophore generation through conformational analysis and manual alignment is a very time consuming task, especially when the list of the active

ligands is large and the elements of the pharmacophore model are not obvious. There are several programs Hip Hop, Hypogen, Disco, Gaps, flo, APEX, and ROCS, that can automatically generate potential pharmacophore from a list of known inhibitors. The performance of these programs in automated pharmacophore generation varies depending on the training set.

These all program use algorithms that identified the common pharmacophore features in the training set molecules; they scoring function to rank the identified pharmacophores.

4. Receptor based pharmacophore generation

If the 3D structure of receptor is known, a pharmacophore model can be derived based on the receptor active site [37]. Biochemical data used to identify the key residue that is important for substrate and/or inhibiting binding. This information can be used for binding pharmacophores targeting the region defined by key residue or for choosing among pharmacophore generated by automated program. This can greatly improve the chance of finding small molecules that inhibit the protein because the search is focused on a region of the binding side that is crucial for binding substrate and inhibitors.

Structure Based Virtual Screening (Docking)

High throughput screening docking and scoring techniques can be applied to computationally screening a database of hundreds of thousand of compound against a target of proteins. Computational methods that predict the 3D structure of a protein ligand complex are often referred to as molecular docking approaches [38].

Virtual screening as a computational task can be trivially run using parallel computing because the protein ligand docking events are completely independent of each other. Although docking has initially been developed as a specialist modeling tool run on computer work station, now a day in expensive Linux clusters are distributed computing over networked PCs can be used for virtual screening.

This increases the *in-silico* throughput into the realm of 100,000 compounds per day on a Linux clusters there by reaching the speed of today's high throughput screening. Energy function that evaluates the binding free energy between protein and ligand sometimes employs rather heuristic terms. Therefore, those functions are more broadly referred to as *scoring functions* the necessary steps include protein structure preparation, ligand database preparation, docking calculation, and post processing. The protein has to be prepared only once for the virtual screening experiment unless different protein conformation are considered. The receptor site needs to be determined and charges have to be assigned. The protein sites have to be modeled as accurately as possible. Determining protein surface atoms and site points as well as assignment of interaction data such as marking hydrogen-bond donors/acceptors and so forth, are sometimes internally include in the docking software (e.g., in FlexX) and sometimes done separately (e.g., DOCK) [39–45].

Because of the large number of the molecules, manual step in the preparation of the ligand database obviously have to be avoided. Starting typically form 2D structure, bond types have to be checked protonation states must be determine charges must be assigned and solvent molecules removed. 3D coordination can generated using a program such as CORD or CORINA. "Scoring" [48] refers to the fact that any docking procedure must evaluate and rank the configurations generated by the search process. The scoring scheme most closely related to experiment- the "*ab initio*" calculation of the binding free energies, is not easily accessible to computation. Scoring is actually composed of three different aspects relevant to the docking and design:

1. Ranking of the conformations generated by the docking search for one ligand interacting with a given protein; this aspect is essential to detect the binding mode best approximating the experimentally observed situation.

2. Ranking different ligands with respect to the binding to one protein, that is, prioritizing ligands according to their affinity; this aspect is essential in virtual screening.

3. Ranking one or different ligands with respects to their binding affinity to different; this aspect is essential for the consideration of selectivity and specificity of ligands. [43-45].

Types of docking

Following different types of docking have been reported [46]

1. Flexible protein-ligand docking

Flexible protein-ligand docking is extremely important for the discovery of new drugs. Through the exhaustive integration and fine tuning of different variables, the latest programs and algorithms are capable of predicting the behavior of chemical compounds and protein molecule in order to better help researchers find a more efficient drug leads. This method significantly reduces the necessary cost money and time consumed, as well as minimizing the non-specific interaction of drug molecule proteins; this aspect is essential for the consideration of selectivity and specificity [47].

2. Flexible protein-protein docking

Protein-protein interactions are also extremely important, since they are responsible for many necessary biological functions. Prediction of such interactions is extremely important to the complete understanding of human physiology. Association of two biological macromolecules is a fundamental biological phenomenon and an unsolved theoretical problem. In recent years, several groups have developed a variety of tools in an attempt to solve the so called protein-protein docking problem, that is, the prediction of the geometry of a complex from the atom coordinates of its uncomplexed constituents.

3. Hydrophobic docking

In view of the higher occurrence of hydrophobic groups at contact sites, their contribution results in more intermolecular atom-atom contacts per unit area for correct matches than for false positive fits. The hydrophobic groups are also

potentially less flexible at the surface. Thus, from a practical point of view, a partial representation of the molecules based on hydrophobic groups should improve the quality of the results in finding molecular recognition sites, as compared to full representation [48].

Special aspects of docking

Proteins are inherently dynamic systems [47] the average data provided by a crystal structure, may not be an adequate representation of the flexible structural characteristics of proteins, unless the system is very rigid. The actual approach could be the comparison of experimental protein structure in the ligand free and in the complexed state, this frequently exhibits conformational changes induced by or associated with ligand binding. The spectrum of phenomenon ranges from side chain rotation to the loop arrangement and movement of entire domains. The protein itself remains fixed but either through an adapted geometric representation or using a tolerant scoring function a certain amount of overlap between the protein and the ligand is allowed emulating some "Plasticity" of the receptor.

An alternative, to account in principle, for an arbitrary degree of protein flexibility is the use of protein structure ensembles. The ensembles could be assembled from multiple crystal structures of a given protein, from NMR structure determination, or from trajectories of molecular dynamics simulations. In addition, a rotamer library can be used to create a minimal set of new conformations [48]. Three different ways to use protein ensembles for docking can be distinguished:

The first and straightforward way is to carry out docking sequentially with each member of the ensemble using rigid-receptor docking [49–52].

Another way is to use a weight-average representation of the ensemble. Knegt *et al* followed this approach by generating composite grids that were used for scoring within the DOCK program [53]. Recently, it has also been tested with Autodock [55]. Broughton has developed another method by combining statistical analysis of a conformational ensemble from short MD simulations with grid –based docking protocols [55].

The third and most sophisticated approach to handle protein ensembles is implemented into FlexE, a variant of the FlexX program [56].

Assessment of docking method

Docking methods are usually assessed by their ability to reproduce the binding mode of experimentally resolved protein ligand complexes, the ligand is removed from the complex, a search area is defined around the actual binding site, the ligand is redocked into the protein, and the achieved binding mode is compared with the experimental positions usually in terms of a root mean square deviation (RMSD). If the RMSD is below 2 Å it is generally considered a successful prediction. The obvious goal is that such a "near native" solution is ranked best amongst the set of ligand poses generated. Virtually any introduction of a new docking method has been accompanied by such a test. The number of complexes used varies as much as the reported success rates, which are between 10% and 100%. Clearly, success rates of 100% are rather a consequence of the limited test set size than a reflection of the mere quality of the docking method [57–61].

Docking and QSAR

As long as the problem of accurate binding free energy prediction on the basis of a given complex geometry has not been resolved, computational methods establishing quantitative structure-activity relationships (QSARs) to estimate relative binding affinity differences within a set of ligands remain a pragmatic alternative. Both classical and 3D QSAR methods have been developed as ligand-based approaches [62]. They rely exclusively on ligand information and try to correlate experimental binding data with features described by a set of relevant descriptors. In 3D QSAR such as CoMFA (Comparative Molecular Field Analysis) these descriptors are essentially virtual interaction energies calculated using an appropriate probe atom placed at the intersection of a regularly spaced grid surrounding the molecule. They can be interpreted as a surrogate representation of the binding site.

Essential for the success of all 3D QSAR approaches is an appropriate alignment of the ligands; their relative spatial superposition must reflect the differences in binding geometry experienced at the binding site of the structurally unknown protein [61]. These methods are also applied if the receptor structure is known. These results in “receptor-based 3D QSAR,” a combination of a ligand-based 3D QSAR approach with information extracted from receptor structures. In case of receptor structure is known, the ligand alignment can be obtained by docking. Variety of studies shows that model generated with docking alignment could be shown to higher relevance than traditional CoMFA model based on ligand alignment [63–64].

Another concept to combine docking with QSAR has recently been proposed by Vieth & Cummins in their DoMCoSAR approach [65]. DoMCoSAR is used for statistically determining the docking mode that is consistent with a structure-activity relationship, based on the explicit assumption that all molecules exhibit the same binding mode. In a first step, all molecule of a chemical series with common substructure are docked in an unbiased way to the protein-binding site and the results are clustered to establish the most favorable docking modes for the common substructure. Subsequently, forcing all molecules to align with the common substructure in the major docking modes performs constrained docking. In final stage, interaction-energy-based descriptors are calculated for all major docking modes. QSAR models are then derived to determine the statistically significant and most consistent with a given structure-activity relationship.

Docking and homology modeling

Often, the crystal structure of the therapeutic target is not available, but the three- dimensional structure of a homologous protein will have been determined. Depending on the degree of homology between the two proteins, it may be useful to model –build the structure of the unknown protein based on the known structure.

So, in absence of an experimental protein structure, a homology model may be used for docking and structure based design. Comparative modeling based on

homologous proteins of known structure can generate such a model. Obviously it is most reliable in the regions of highest homology between the templates and the target protein. Using this method overall skeleton of the target protein can frequently be obtained with sufficient accuracy, but the structural details of the binding site are often not clear. In fact members of a homologous protein family may show considerable differences in the binding region. An approach developed especially for the purpose of docking ligands into approximate protein models generated by homology modeling is the DragHome method [72] in which the binding site is analyzed in terms of putative ligand interaction sites and translated using Gaussian function into a functional binding-site description represented by physicochemical properties. Similarly, ligands are translated into a description based on Gaussian functions and the docking is computed by optimizing the overlap between the two functional descriptions. The use of soft Gaussian functions to describe protein- ligand interaction is one possibility to take into account the limited accuracy of modeled structures for the purpose of docking. The method for generating and optimizing ligand orientation relative to the binding-site representation was adapted from the ligand alignment program SEAL [66–67] for a set of different ligands; the generated solutions are analyzed with respect to the mutual ligand alignment. The alignment is then used to generate 3-D QSAR models, which in turn can be interpreted with respect to the surrounding protein model.

The docking calculation is typically done for one ligand at a time. Depending on optimization and sampling parameters as well as on the flexibility of the compound, typically between a few seconds and a few minute of CPU time is needed to dock a ligand. Because the individual docking events are independent of each other they can run on parallel hardware. Task schedulers that distribute ligand docking on available CPUs are used in many dockings programs.

Research in virtual screening

Maxwell D. Cummings *et al* studied [68] the performances of several commercially available docking programs and compared them with context of virtual screening. Five different protein targets are used with several known ligands. For many of the known ligand crystal structure of the relevant protein-ligand complexes were available. For a given docking method, hits rates were improved versus that would be expected for random selection for most protein targets. However, the ability to prioritize known ligand on the basis of the docking that resembles known crystal structures is both method-and target-dependant. Jean-Christophe Mozziconacci *et al* studied [69] to exploit available structural information about the cyclooxygenase enzyme for the virtual screening of large chemical libraries; a docking protocol was turned and validated. The screening accuracy was assessed using a series of known inhibitor and a set of diverse a priori inactivate compound that was seeded with known active ligand. Andeas Evers *et al*. studied [70] the homology modeling of the alpha 1A receptor based on the X-ray structure of bovine rhodopsin. They applied a hierarchical virtual screening procedure guided by 2D filter and three-dimensional pharmacophore model. Their finding suggest that rhodopsin based in homology model may be used has the structural bases for GPCR lead finding and compound optimization. Maria J. Duart *et al* studied [71] the utility of the virtual combinatorial chemistry coupled with computational screening. A library of amine and urea derivatives was designed by virtual combinatorial synthesis and eventually computationally screened by a mathematical topological model as antihistaminic compounds. The results reveal that virtual combinatorial synthesis and virtual screening together with molecular topology is a powerful tool in the design of new drugs.

Liqin Zhao *et al* studied [72] that ER β has been associated with estrogen-induced promotion of memory function and neuronal survival. Based on the optimized complex structure of human ER β LBD bound with genistein, computer-aided structure-based virtual screening against a natural source chemical database

was conducted to determine the occurrence of the plant-based ER β -selective ligands. Twelve representative hits derived from database screenings were assessed for their binding profile to both ERs, three of which displayed over 100-fold binding selectivity to ER β over ER α .

Simone Sciabola *et al* studied [73] Ligand- (GRIND) and structure (GLUE/GRIND) 3D-QSAR approach were compared for 55 (aryl-) bridged 2-aminobenzonitriles inhibiting HIV –1 reverse transcriptase (HIV-1 RT) the ligand-based model was build from conformers selected by in vacuo minimization. The available X-ray structure of 3D in complex with HIV-1 RT allowed comparative structure based calculation using the new docking software GLUE for conformer selection. Both models were validated via statistics and via virtual receptor site (VRS) considering pharmacophore regions and mutual distance, which are also compared with experimental evidence. The statistics show slight superiority of the structure-based approach in terms of fitting and prediction. By encoding relevant molecular interaction field (MIF) into pharmacophoric region, 10 such regions were derived from both models; they all fit the real receptor except HBD2. Also mutual distances highly agree between the real site and both VRS although distance from the structure based approach are closer to the real receptor; present data prove the validity of the ligand-based GRIND approach. Timega Polar and Gyorgy Keseru studied [74] Virtual screening for β -Secretase (BACE1) inhibitors revealed the important c at Asp32 and Asp228. A comparative virtual screening for β -Secretase (BACE1) inhibitors using different docking methods (FlexX and Flex-pharm), scoring function (Dock, Gold, Chem PMF, FlexX), protonation states (default and calculated), and protein conformation (apo and ligand bound) has been performed. Marco Scarsi *et al* studied [75] that sulfonylureas and glinides additionally bind to PPAR γ and exhibit PPAR γ agonistic activity. This activity was predicted in silico by virtual screening and confirmed in vitro in a binding assay. Szewczuk LM [76] *et al* used a structure-based computational approach to identify the first druglike and selective inhibitors of AANAT (arylalkylamine N-acetyltransferase). Approximately 1.2 million compounds were virtually screened by

3D high-throughput docking into the active site of X-ray structures for AANAT, and in total 241 compounds were tested as inhibitors. Cavasotto CN [77] *et al* studied A ligand-steered homology modeling approach has been developed (where information about existing ligands is used explicitly to shape and optimize the binding site) followed by docking-based virtual screening. Top scoring compounds identified virtually were tested experimentally in an MCH-R1 (Melanin-concentrating hormone receptor 1) competitive binding assay, and six novel chemotypes as low micromolar affinity antagonist "hits" were identified. Jie-Fei Cheng [78] *et al* conducted virtual docking studies using GLIDE with modified LXR β ligand-binding domain (LBD) on internal compound collection followed by the gene profiling with ArrayPlate mRNA assay. A total of 69 compounds were found to upregulate LXR α and certain LXR regulated genes from 1308 compounds selected by virtual screen (hit rate: 5.3%).

Future prospective

Virtual screening, especially the structure-based virtual screening, has emerged as a reliable, cost-effective and timesaving technique for the discovery of lead compounds. This review focuses on the generation and use of virtual compound libraries, and also on studies in which chemical feature-based pharmacophore models and docking are used in combination with in silico screening. These procedures are generally used to obtain hits (or leads) that are more likely to give successful clinical candidates. Virtual screening of virtual libraries (VSVL) is a rapidly changing area of research. Great efforts are being made to produce better algorithms, selection methods and infrastructure. It is a fact that these tools remain quite daunting for the majority of scientists working at the bench. The routine use of these methods is not simply a matter of education and training. The authors are confident that the synergy of these technologies will bring great benefit to the industry, with more efficient production of higher quality clinical candidates. The future is bright. The future is virtual focused assessment of

corporate databases by virtual screening using well validated pharmacophore models will yield to a significant improvement in lead structure determination.

Tab.1. Available Software for Virtual Screening

Target	Software Package	Ligand Data Set
Thymidylate synthase	DOCK	153000ACD compounds [8]
FK506-binding protein	SANDOCK	ACD and Cambridge Crystallographic Database[9]
Retinoic acid receptor	ICM	153000ACD compounds [10]
HIV-1 RNA transactivation response element	ICM	153000 ACD compounds[11]
Farnesyl transferase	EUDOC	67928 ACD compounds[12]
DNA gyrase	LUDI, CATALYST	350000 ACD + in-house compounds[13]
Kinesin	DOCK	110000 ACD compounds [14]
Hypoxanthine-guanine-xanthine Phosphoribosyl transferase	DOCK	599 compound virtual library [15]
Thrombin; factor Xa; estrogen receptor	PRO_LEADS	10000 ChemBridge compounds [16]

References

- [1] Greer J, Erickson JW, Baldwin JJ, Varney MD. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J Med Chem.* 1994; 37:1035–1054. [doi:10.1021/jm00034a001]

- [2] Bohacek RS, McMartin C, Guida WC.
The art and practice of structure-based drug design: A molecular modeling perspective.
Med Res Rev. 1996; 16: 3–50.
[doi:10.1002/(SICI)1098-1128(199601)16:1<3::AID-MED1>3.0.CO;2-6]
- [3] Shoichet BK, Bussiere DE.
The Role of Macromolecular Crystallography and Structure for Drug Discovery: Advances and Caveats.
Curr Opin Drug Discov Dev. 2000; 3: 408–422.
- [4] Wodak SJ, Janin J.
Computer analysis of protein-protein interaction.
J Mol Biol. 1978; 124:323–342.
[doi:10.1016/0022-2836(78)90302-9]
- [5] Kuntz ID, Blany JM, Oatley SJ, Langridge R, Ferrin TE.
A geometric approach to macromolecule-ligand interactions.
J Mol Biol. 1982; 161: 269–288.
[doi:10.1016/0022-2836(82)90153-X]
- [6] Mugge I, Enyedy I.
Virtual Screening.
In Burger's Medicinal chemistry and Drug discovery.
6 th ed Volume 1; Verginia : Wiley Interscience , 2003: 243–279.
[doi:10.1002/0471266949.bmc006]
- [7] Shoichet BK.
Virtual screening of chemical libraries.
Nature. 2004; 432:862–865.
[doi:10.1038/nature03197]
- [8] Tondi D, Slomczynska U, Costi MP, Watterson DM, Ghelli S, Shoichet BK.
Structure based discovery and in-parallel optimization of novel competitive inhibitors of thymidylate synthase.
Chem Biol. 1999; 6: 319–331.
[doi:10.1016/S1074-5521(99)80077-5]
- [9] Burkhard P, Hommel U, Sanner, M Walkinshaw, MD.
The discovery of steroids and other novel FKBP inhibitors using a molecular docking program.
J Mol Biol. 1999; 287: 853–858.
[doi:10.1006/jmbi.1999.2621]
- [10] Schapira M, Raaka BM, Samuels HH, Abagyan R.
Rational discovery of novel nuclear hormone receptor antagonists.
Proc Natl Acad Sci U S A. 2000; 97: 1008–1013.
[doi:10.1073/pnas.97.3.1008]

- [11] Filikov AV, Mohan V, Vickers TA.
Identification of Ligands for RNA Targets via Structure-Based Virtual Screening.
J Comp Aid Mol Des. 2000; 14: 561–593.
[doi:10.1023/A:1008121029716]
- [12] Perola E, Xu K, Kollmeyer TM, Kaufmann SH, Prendergast FG.
Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads.
J Med Chem. 2000; 43: 401–408.
[doi:10.1021/jm990408a]
- [13] Boehm HJ, Boehringer M, Bur D, Gmuender H, Huber W, Klaus W, Kostrewa D, Kuehne H, Luebbers T, Meunier-Keller N, Mueller F.
Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening.
J Med Chem. 2000; 43: 2664–2674
[doi:10.1021/jm000017s]
- [14] Hopkins SC, Vale RD, Kuntz ID.
Inhibitors of kinesin activity from structure-based computer screening.
Biochemistry. 2000; 39: 2805–2814.
[doi:10.1021/bi992474k]
- [15] Aronov AM, Munagala NR, Ortiz de Montellano PR, Kuntz ID, Wang CC.
Rational Design of Selective Submicromolar Inhibitors of *Tritrichomonas foetus* Hypoxanthine-Guanine-Xanthine Phosphoribosyltransferase.
Biochemistry. 2000; 39: 4684–4689.
[doi:10.1021/bi992555g]
- [16] Baxter CA, Murray CW, Waszkowycz B.
New Approach to Molecular Docking and Its Application to Virtual Screening of Chemical Databases.
J Chem Inf Comput Sci. 2000; 40: 254–262.
[doi:10.1021/ci990440d]
- [17] Walters WP, Ajay A, Murcko MA.
Recognizing molecules with drug-like properties.
Curr Opin Chem Biol. 1999; 3: 384–387.
[doi:10.1016/S1367-5931(99)80058-1]
- [18] Walters WP, Murcko MA.
Library filtering systems and prediction of drug-like properties.
In *Virtual Screening for Bioactive Molecules.*
Wiley-VCH, 2000: 10: 15–32.
[doi:10.1002/9783527613083.ch2]

- [19] Clark DE, Pickett SD.
Computational Methods for the Prediction of Drug-Likeness.
Drug Discov Today. 2000; 5: 49–58.
[doi:10.1016/S1359-6446(99)01451-8]
- [20] Podlogar BL, Muegge I, Brice LJ.
Computational methods to estimate drug development parameters.
Curr Opin Drug Discov Dev. 2001; 4: 102–109.
- [21] Hann M, Hudson B, Lewell X, Lively R, Miller L, Ramsden N.
Strategic pooling of compounds for high-throughput screening.
J Chem Inf Comput Sci. 1999; 39: 897–903.
[doi:10.1021/ci990423o]
- [22] Muegge I, Heald, SL, Brittelli D.
Simple Selection Criteria for Drug-like Chemical Matter.
J Med Chem. 2001; 44: 1841–1846.
[doi:10.1021/jm015507e]
- [23] Ajay A, Walters WP, Murcko MA.
Can we learn to distinguish between "drug-like" and "nondrug-like" molecules?
J Med Chem. 1998; 27: 3314–3324.
[doi:10.1021/jm970666c]
- [24] Van Drie JH, Weininger D, Martin YC.
ALADDIN: an integrated tool for computer-assisted molecular design and
pharmacophore recognition from geometric, steric, and substructure searching
of three-dimensional molecular structures.
J Comput-Aided Mol Des. 1989; 3: 225–251.
[doi:10.1007/BF01533070]
- [25] Walters WP, Stahl MT, Murcko MA.
Virtual screening: an overview.
Drug Discov Today. 1998; 3: 160–178.
[doi:10.1016/S1359-6446(97)01163-X]
- [26] Ehrlich P.
Present status of chemotherapy.
Ber Dtsch Chem Ges. 1909; 42: 17–47
[doi:10.1002/cber.19090420105]
- [27] Gund P.
Three-dimensional Pharmacophoric Pattern Searching.
Prog Mol Subcell Biol. 1977; 11: 117–143.
- [28] Marshall GR, Barry CD, Bosshard HE, Dammkoehler RA, Dunn DA.
The Conformational Parameter in Drug Design: The Active Analog Approach.
*Computer-Assisted Drug Design American Chemical Society: Washington D.
C., 1979; Vol. ACS Symposium II2.*

- [29] Sufrin JR, Dunn DA, Marshall GR.
Steric mapping of the L-methionine binding site of ATP:L-methionine S-adenosyltransferase.
Mol Pharmacol. 1981; 19: 307–313.
- [30] Barnard JM.
Substructure searching methods: Old and new.
J Chem Infor Compt Sci. 1993; 33: 532–538.
[doi:10.1021/ci00014a001]
- [31] Enyedy IJ, Wang J, Zaman WA, Johnson KM, Wang S.
Discovery of substituted 3, 4-Diphenyl-thiazoles as a novel class of monoamine transporter inhibitors through 3-D pharmacophore search using a new pharmacophore model derived from mazindol.
Bioorg Med Chem Lett. 2002; 12: 1775–1778.
[doi:10.1016/S0960-894X(02)00243-3]
- [32] Downs GM, Willett P.
Similarity searching in databases of chemical structures.
In Lipkowitz KB Boyd DB. *Reviews in Computational Chemistry 1 edn Volume 7* New York VCH Publishers, Inc 1996: 1–66.
- [33] Willett P, Barnard JM, Downs GM.
Chemical similarity searching.
J Chem Inf Comput Sci. 1998; 38: 983–996.
[doi:10.1021/ci9800211]
- [34] Hoffman BT, Kopajtic T, Katz JL, Newman AH.
2D QSAR Modeling and Preliminary Database Searching for Dopamine Transporter Inhibitors Using Genetic Algorithm Variable Selection of Molconn Z Descriptors.
J Med Chem. 2000; 43: 4151–4159.
[doi:10.1021/jm990472s]
- [35] Wang S, Sakamuri S, Enyedy IJ, Kozikowski AP, Deschaux O, Bandyopadhyay BC, Tella SR, Zaman WA, Johnson KM.
Discovery of a novel dopamine transporter inhibitor, 4-hydroxy-1-methyl-4-(4-methylphenyl)-3-piperidyl 4-methylphenyl ketone, as a potential cocaine antagonist through 3D-database pharmacophore searching. Molecular modeling, structure-activity relationships, and behavioral pharmacological studies.
J Med Chem. 2000; 43: 351–360.
[doi:10.1021/jm990516x]
- [36] Guner OF.
Pharmacophore Perception, Development, and Use in Drug Design.
International University Line, La Jolla, CA, 2000.

- [37] Wang R, Gao Y, Lai L.
LigBuilder: A Multiple-Purpose Program for Structure-Based Drug Design.
J Mol Model. 2000; 6: 498–516.
[doi:10.1007/s0089400060498]
- [38] Haraki KS, Sheridan RP, Venkataraghavan R, Dunn, DA, McCulloch R.
Looking for Pharmacophores in 3-D Databases: Does Conformational
Searching Improve the Yield of Actives?
Tetrahedron Comput Methodol. 1990; 3: 565–573.
[doi:10.1016/0898-5529(90)90159-6]
- [39] Clark DE, Willett P, Kenny PW.
Pharmacophoric Pattern Matching in Files of Three-Dimensional Chemical
Structures: Implementation of Flexible Searching.
J Mol Graphics. 1993; 11: 146–156.
[doi:10.1016/0263-7855(93)80066-Z]
- [40] Blaney JM, Dixon JS.
A good ligand is hard to find: automated docking methods.
Perspect Drug Discov Des. 1993; 1: 301–319.
[doi:10.1007/BF02174531]
- [41] Jones G, Willett P, Glen R C, Leach A R, Taylor R.
Development and Validation of a Genetic Algorithm for Flexible Docking.
J Mol Biol. 1997; 267: 727–748.
[doi:10.1006/jmbi.1996.0897]
- [42] Wang R, Liu L, Lai L, Tang Y.
SCORE: A New Empirical Method for Estimating the Binding Affinity of a
Protein-Ligand Complex.
J Mol Model. 1998; 4: 379–394.
[doi:10.1007/s008940050096]
- [43] Walters WP, Stahl MT, Murcko MA.
Virtual screening: an overview.
Drug Discov Today. 1998; 3: 160–178.
[doi:10.1016/S1359-6446(97)01163-X]
- [44] Leach AR, Hann MM.
The in silico world of virtual libraries.
Drug Discov Today 2000; 5: 326–336.
[doi:10.1016/S1359-6446(00)01516-6]
- [45] Lewis RA, Pickett SD, Clark DE.
Computer-aided molecular diversity analysis and combinatorial library design.
Rev Comput Chem. 2000; 16: 1–51.
[doi:10.1002/9780470125939.ch1]

- [46] Goshe AK, Crippen GM.
Quantitative Structure-Activity Relationship by Distance Geometry:
Quinazolines as Dihydrofolate Reductase Inhibitors.
J Med Chem. 1982; 25: 892–899.
[doi:10.1021/jm00350a003]
- [47] Karplus M, McCammon JA.
Dynamics of proteins: elements and function.
Annu Rev Biochem. 1983; 52: 263–300.
[doi:10.1146/annurev.bi.52.070183.001403]
- [48] Abagyan R, Totrov M.
High-throughput docking for lead generation
Curr Opin Chem Biol. 2001; 4: 375–382.
[doi:10.1016/S1367-5931(00)00217-9]
- [49] Knegtel RM, Kuntz ID, Oshiro CM.
Molecular docking to ensembles of protein structures.
J Mol Biol. 1997; 266: 424–440.
[doi:10.1006/jmbi.1996.0776]
- [50] Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS.
Automated docking to multiple target structures: incorporation of protein
mobility and structural water heterogeneity in AutoDock.
Proteins. 2001; 46: 34–40.
[doi:10.1002/prot.10028]
- [51] Pang YP, Kozikowski AP.
Prediction of the binding sites of Huperzine A in acetylcholinesterase by
docking studies.
J Comput-Aided Mol Des. 1994; 8: 662–669.
[doi:10.1007/BF00124014]
- [52] Bold G, Altmann KH, Bruggen J, Frei J, Lang M, Manley PW, Traxler P,
Wietfeld B, Buchdunger E, Cozens R, Ferrari S, Furet P, Hofmann F, Martiny-
Baron, G, Mestan J, Rosel J, Sills M, Stover D, Acemoglu F, Boss E,
Emmenegger R, Lasser L, Masso E, Roth R, Schlachter C, Vetterli W, Wyss
D, Wood JM.
New Anilinophthalazines as Potent and Orally Well Absorbed Inhibitors of the
VEGF Receptor Tyrosine Kinases Useful as Antagonists of Tumor-Driven
Angiogenesis.
J Med Chem. 2000; 43: 2310–2323.
[doi:10.1021/jm9909443]
- [53] Carlson HA, Masukawa KM, McCammon JA.
Method for Including the Dynamic Fluctuations of a Protein in Computer-aided
Drug Design.
J Phys Chem. 1999; 110: 1856–1872.
[doi:10.1063/1.477873]

- [54] Bouzida D, Rejto PA, Arthurs S, Colson AB, Freer ST, Gehlhaar DK, Larson V, Luty BA, Rose PW, Verkhivker GM.
Computer simulations of ligand-protein binding with ensembles of protein conformations: A Monte Carlo study of HIV-1 protease binding energy landscapes.
Int J Quantum Chem. 1999; 72: 73–84.
[doi:10.1002/(SICI)1097-461X(1999)72:1<73::AID-QUA7>3.0.CO;2-O]
- [55] Broughton HB.
A method for including protein flexibility in protein-ligand docking: improving tools for database mining and virtual screening.
J Mol Graph Model. 2000; 18: 247–257.
[doi:10.1016/S1093-3263(00)00036-X]
- [56] Carlson HA, McCammon JA.
Accommodating protein flexibility in computational drug design.
Mol Pharmacol. 2000; 57: 213–218.
- [57] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS.
Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy.
J Med Chem. 47; 2004: 1739–1749.
[doi:10.1021/jm0306430]
- [58] Kroemer RT, Vulpetti A, McDonald JJ, Rohrer DC, Trosset JY, Giordanetto F, Cotesta S, McMartin C, Kihlen M, Pieter FW, Stouten PF.
Interactions-Based Accuracy Classification (IBAC) versus Crystal Structure Deviations.
J Chem Inf Comput Sci. 2004; 44: 871–881.
[doi:10.1021/ci049970m]
- [59] McMartin C, Bohacek RS.
QXP: powerful, rapid computer algorithms for structure-based drug design.
J Comput-Aided Mol Des. 1997; 11: 333–334.
[doi:10.1023/A:1007907728892]
- [60] Kubinyi H.
Molecular field analysis.
In 3D QSAR in drug design: Theory, methods and Applications.
ESCOM, Leiden, the Netherlands 1993: 443–485.
- [61] Sippl W.
Receptor-based 3DQSAR analysis of estrogen receptor ligands merging the accuracy of receptor-based alignments with the computational efficiency of ligand-based methods.
J Comput-Aided Mol Des. 2000; 14: 559–572.
[doi:10.1023/A:1008115913787]

- [62] Sippl W and Holtje HD.
Structure-based 3D-QSAR—merging the accuracy of structure-based alignments with the computational efficiency of ligand-based methods.
J Mol Struct (Theochem). 2000; 503: 31–50.
[doi:10.1016/S0166-1280(99)00361-9]
- [63] Vieth M, Cummins DJ.
DoMCoSAR: a novel approach for establishing the docking mode that is consistent with the structure-activity relationship. Application to HIV-1 protease inhibitors and VEGF receptor tyrosine kinase inhibitors.
J Med Chem. 2000; 43: 3020–3032.
[doi:10.1021/jm990609e]
- [64] Wojciechowski M, Skolnick J.
Docking of small ligands to low-resolution and. theoretically predicted receptor structures.
J Comp Chem. 2002; 23: 189–197.
[doi:10.1002/jcc.1165]
- [65] Moro S, Li AH, Jacobson KA.
Molecular Modeling Studies of Human A3 Adenosine Antagonists: Structural Homology and Receptor Docking.
J Chem Inf Comput Sci. 1998; 38: 1239–1248.
[doi:10.1021/ci980080e]
- [66] Kearsley SK, Smith GM.
An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap.
Tetrahedron Comput Methodol. 1990; 3: 615–633.
[doi:10.1016/0898-5529(90)90162-2]
- [67] Evers A, Klabunde T.
Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor.
J Med Chem. 2005; 48: 1088–1097.
[doi:10.1021/jm0491804]
- [68] Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP.
Docking to five proteins of a compound set seeded with known actives using DOCK, DOCKVISION, GLIDE, and GOLD suggests that GLIDE and GOLD in general identify the most reasonable binding modes for known active compounds.
J Med Chem. 2005; 48: 962–976.
[doi:10.1021/jm049798d]

- [69] Mozziconacci JC, Arnoult E, Bernard P, Do QT, Marot C, Morin-Allory L. Optimization and validation of a docking-scoring protocol; application to virtual screening for COX-2 inhibitors. *J Med Chem.* 2005; 48: 1055–1068. [doi:10.1021/jm049332v]
- [70] Evers A, Klabunde T. Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J Med Chem.* 2005; 48: 1088–1097. [doi:10.1021/jm0491804]
- [71] Duart MJ, Anton-Fos G M, Aleman PA, Gay-Roig JB, Gonzalez-Rosende ME, Galvez J, Garcia-Domenech R. New Potential Antihistaminic Compounds. Virtual Combinatorial Chemistry, Computational Screening, Real Synthesis, and Pharmacological Evaluation. *J Med Chem.* 2005; 48: 1260–1264. [doi:10.1021/jm040877z]
- [72] Zhao L, Brinton RD. Structure-Based Virtual Screening for Plant-Based ER β -Selective Ligands as Potential Preventative Therapy against Age-Related Neurodegenerative Diseases. *J Med Chem.* 2005; 48: 3463–3466. [doi:10.1021/jm0490538]
- [73] Sciabola S, Carosati E, Baroni M, Mannhold R. Comparison of Ligand-Based and Structure-Based 3D-QSAR Approaches: A Case Study on (Aryl-) Bridged 2-Aminobenzonitriles Inhibiting HIV-1 Reverse Transcriptase. *J Med Chem.* 2005; 48: 3756–3767. [doi:10.1021/jm049162m]
- [74] Polgar T, Keseru GM. Virtual Screening for β -Secretase (BACE1) Inhibitors Reveals the Importance of Protonation States at Asp32 and Asp228. *J Med Chem.* 2005; 48: 3749–3755. [doi:10.1021/jm049133b]
- [75] Scarsi M, Podvinec M, A Roth, Hug H, Kersten S, Albrecht H, Schwede T, Meyer UA, Biozentrum CR. Sulfonylureas and Glinides Exhibit PPAR $\{\gamma\}$ Activity: A Combined Virtual Screening and Biological Assay Approach. *Mol Pharmacol.* 2006; 69: 170–178. [doi:10.1124/mol.106.024596]

- [76] Szewczuk LM, Saldanha SA, Ganguly S, Bowers EM, Javoroncov M, Karanam B, Culhane JC, Holbert MA, Klein DC, Abagyan R, Cole PA.
De novo discovery of serotonin N-acetyltransferase inhibitors:
J Med Chem. 2007; 50: 5330–5338.
[doi:10.1021/jm0706463]
- [77] Cavasotto CN, Orry AJ, Murgolo NJ, Czarniecki MF, Kocsi SA, Hawes BE, O'Neill KA, Hine H, Burton MS, Voigt JH, Abagyan RA, Bayne ML, Monsma FJ Jr.
Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening.
J Med Chem. 2008; 51: 581–588.
[doi:10.1021/jm070759m]
- [78] Cheng JF, Zapf J, Takedomi K, Fukushima C, Ogiku T, Zhang SH, Yang G, Sakurai N, Barbosa M, Jack R, Xu K.
Combination of Virtual Screening and High Throughput Gene Profiling for Identification of Novel Liver X Receptor Modulators.
J Med Chem. 2008; 51: 2057–2061.
[doi:10.1021/jm7011326]

Received March 13th, 2008

Accepted (after revision) September 4th, 2008

Available online at www.scipharm.at September 6th, 2008