

Recent Development and Application of Virtual Screening in Drug Discovery: An Overview

Tingjun Hou and Xiaojie Xu*

College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China



Abstract: Virtual screening, especially the structure-based virtual screening, has emerged as a reliable, cost-effective and time-saving technique for the discovery of lead compounds. Here, the basic ideas and computational tools for virtual screening have been briefly introduced, and emphasis is placed on aspects of recent development of docking-based virtual screening, scoring functions in molecular docking and ADME/Tox-based virtual screening in the past three years (2000 to 2003). Moreover, successful examples are provided to further demonstrate the effectiveness of virtual screening in drug discovery.

Key Words: Virtual screening, molecular docking, database screening, pharmacophore modeling, ADME, structure-based drug design, scoring function, *de novo* drug design.

1. INTRODUCTION

The discovery of innovative leads is the key element and starting point for any new drug discovery project. More recent advances in molecular biology, robotics and microelectronics, especially, the complete analysis of the human genome, modern drug discovery will make everlasting impact on human diseases. The human genomics expects to provide us with the sequences of all genes coding for proteins that make up the biology of a particular organism. As a consequence, the number of potential targets for drug therapy may be increased significantly. It is estimated that the number of potential drug targets may increase from about 500 at present to about 5000-10000 in the next few years [1].

Meanwhile, the development of combinatorial chemistry gives us many opportunities to synthesize thousands upon thousands of compounds in a very short period. In order to give rapid and effective assay to the huge pool of molecules, enormous efforts have been put into the large-scale automation of high-throughput screening (HTS) [2]. It does not mean that the combination of combinatorial chemistry and HTS will spell the end of any rational and knowledge-based approach, this is because initial euphoria that designated this technique as a universal lead generator has subsided as a result of the considerable costs involved and disappointingly low hit rates. The low hit rates are frequently due to inadequacies in quality and quantity of the compound libraries used for testing. Collections of synthesized compounds or natural products often contain far less chemical diversity than is desired, are not bottomless resources, and are very time-consuming to replenish. Moreover, not every assay can be automated for an HTS system.

The application of computational screening is also called "virtual" or *in silico* screening. Virtual screening (VS) is the use of high-performance computing to analyze large databases of chemical compounds to identify possible drug candidates, which is seen as a complementary approach to experimental HTS [3-5]. Virtual screening is knowledge-driven, which means that some information is available regarding either the nature of receptor binding pocket or the type of ligand that is expect to bind productively, or both. It should be noted that VS encompasses a variety of computational screens, from the simplistic to the sophisticated, and hence, can effectively exploit different types of information describing the receptor. The type of method/methods used in VS depends upon information available as input and the type of results required for output. For example, if a 3-D structure of target protein is available, molecular docking or combinatorial drug design can be used to perform receptor-based, fine-grained compounds sieving. If a 3-D receptor structure is unavailable, then pharmacophore model derived from bioactive ligands or molecular property profiles, such as molecular weight, lipophilicity, ADME properties or drug-like properties, can be used as filters in VS. In the following sections, recent developed techniques used in VS and the advances made in VS are reviewed.

2. PHARMACOPHORE-BASED VIRTUAL SCREENING

A pharmacophore is the spatial arrangement of key structural features of a set of known ligands or of the target receptor. Gund was probably the first who described that functional groups (pharmacophores) could be used for searching databases to identify molecules that may share the same structural features [6]. This has led to the successful development and application of 3D-database pharmacophore searching for discovering novel lead compounds in drug discovery [7-8]. Pharmacophore-based virtual screening is involved in two steps: identification of pharmacophore model and 3-D search based on the specific constraints. The

*Address correspondence to this author at the College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, P. R. China; E-mail: xiaojxu@chem.pku.edu.cn

attraction of VS based on pharmacophore model is that a reasonably focused query on 3-D structural grounds can be quickly applied to a large database.

Pharmacophore models are typically used when some active compounds have been identified but the three-dimensional (3D) structure of the target protein or receptor is still unknown. Given a set of active molecules, first of all, analyzing these molecules to identify pharmacophoric features, these are atoms probably interacting with a receptor, then aligning the active conformations of the molecules to find the best overlay of the corresponding features. The main difficulty in pharmacophore generation is in handling the conformational flexibility since the active conformations of the molecules are usually postulated. There are several commercial programs available for the automatic identification of pharmacophore models, including catalyst (Accelrys, <http://www.accelrys.com>) [9], Discotech (Tripos, <http://www.tripos.com>) [10] and GASP (Tripos, <http://www.tripos.com>) [11-12]. The main differences between these programs lie in the algorithms used for the alignment and the way in which conformational flexibility is handled. A pharmacophore model not only can be obtained from a set of active ligands, but also can be derived from the 3-D structure of receptor. Such a structure of receptor or receptor/ligand complex provides information on possible and essential points of interactions between receptor and ligand. The information provided by the 3-D structure can be analyzed to identify interaction points in the binding site as pharmacophore model, which can be used as a query in a 3-D database screening. For example, when a crystal complex is available, the atoms of ligand contributing to receptor-ligand interaction can be defined as features in a pharmacophore model. In order to avoid retrieving molecules that present the correct 3-D arrangement of chemical features, but do not fit into the active site due to a wrong shape, the protein backbone atoms around the active site are usually used to define the coordinates of centers of exclusion spheres. These spheres were merged with the feature points into a single pharmacophore. In many cases especially when the 3-D starting structure is only protein, not protein/ligand complex, one has to generate multiple queries in order to represent different possible binding modes and explore different possible pharmacophores ranging from full ligand-mimic pharmacophores to subsets. Both of UNITY (Tripos, <http://www.tripos.com>) and CATALYST can define pharmacophore model based on the structure of protein. But sometimes, due to the complexity of the receptor structure, it is very necessary to find important features for the definition of pharmacophore. The *Structure Based Focusing* module in Cerius2 (Accelrys, <http://www.accelrys.com>) uses the known or suspected active site of a protein to select compounds which are likely to bind within the defined active site [13]. In *Structure Based Focusing*, the defined active site is first analyzed to generate an interaction map for the active site consisting of a list of features (such as lipophilic, hydrogen donor, hydrogen acceptor) that a ligand is expected to satisfy for a reasonable interaction with the protein. The LUDI interaction site identification procedure was then used to generate the interaction sites [14-15], using a set of rules that are intended to cover the complete range of energetically favorable orientations for hydrogen bonds and hydrophobic

contacts. Finally, a set of 3-D queries is then derived from the interaction map and this database is searched with the set of queries. These hits are scored using LUDI.

After the pharmacophore model is identified, 3-D database search is performed to find compounds bearing these pharmacophoric features. Now, several programs are available for 3-D database search based on pharmacophore model. The most widely used programs include catalys, unity (Tripos, <http://www.tripos.com>) [16], MDL (MDL Information Systems, <http://www.mdl.com>) [17], Chem-X [18] *et al.*

2.1. Recent Advances and Successful Applications of VS Based on Pharmacophore Model

Compared with other approaches used in VS, the methods of VS based on pharmacophore model may be the most mature, so the technical advances in this area are not very exciting. To identify a pharmacophore model based on ligands seems no problem, and to generate the important interactions in receptor for ligand binding need more attentions. Fox and Haaksma represented a computational protocol to identify possible ligands from the analysis of the 3-D structure of receptor [19]. The computation protocol was divided into two stages. In the first stage, the essential interactions in the binding site with the program GRID [20]. From the multitude of probe types supplied with GRID, the authors chose a few representatives to sample the different interactions types, typically including the hydrophobic DRY₂, the alkyl hydroxy group OH, and alcoholate probe (O-), sp² oxygen (e.g., O::, O=), and ammonium probe (N2+, NM3), organic halogen (Br, Cl), water, amides and amidine *et al.* Then, the resulting regions of favorable interactions between receptor and ligand are translated into a database query. 2-D queries are specified by chemical substructures that have to be matched in the resulting GRID energies. The local minima of the interactions maps (or the geometric average if several minima are close), and the corresponding distances between these minima are simply measured to act as the mutual constraints between these features. In the second stage, a flexible 3-D database search in UNITY is performed to retrieve the possible active compounds. The feasibility of this approach is calibrated with thrombin as the target. In GRID calculations, three probes including DRY, NM3, and P4H were used, and a pharmacophore model with three-point interactions sites (Fig. 1) which consists of one positively charged center (at the bottom of the S1 pocket) and two hydrophobic site hp1 and hp2 (in the P and D pockets of thrombin). Owing to the fact that the test databases used by Fox *et al* only contain benzamidine compounds, as the

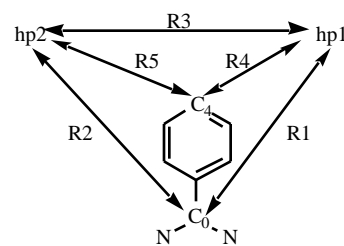


Fig. (1). Pharmacophore model as derived from the GRID results and used in the UNITY database search; hp=hydrophobic center.

positive center the amidine carbon atom of benzamidine (C_0) was used. Based on the query, 3-D flexible database search was performed, and the calculated results show that the resulting hit lists are enriched with thrombin inhibitors compared to the total database. The possible insufficiency of the Fox's approach is that it cannot be fully automated, as the success will critically depend on the identification of an anchor fragment. An anchor fragment is defined as a minimal recognition structure essential for binding. For the identification of these fragments may be accomplished using the *de novo* design program LUDI.

The possible combination of pharmacophore model with molecular docking was investigated by Hindle et al [21]. Hindel proposed an extended version of the flexible docking tool FlexX (Tripos, <http://www.tripos.com>), FlexX-PHARM, allowing the incorporation of information about important characteristics of protein-ligand binding models into a docking calculation. This information is introduced as a simple set of constraints derived from receptor-based type pharmacophore features, and all docking solutions must possess the properties prescribed by the set of constraints. In FlexX-PHARM two different types of constraint can be defined in the active site: interaction constraint and spatial constraints. For the first type, the user can specify a FlexX calculation surface in the active site that must take part in an interaction with ligand. For the second type, the user can specify inclusion volume. For the definition of interaction constraints, an interacting group and interaction type in the active site of receptor must be specified (along with an interaction surface if more than one surface exists for that interaction). FlexX-PHARM ensures that an interaction is formed between the specified interacting group in the active site and the ligand in a valid docking solution. The spatial constraint can be used to constrict ligand position in the active site and consists of a sphere plus an associated element type. In order to gain the maximum potential from FlexX-PHARM in terms of speed and optimal results, some pro-docking checks were performed to eliminate these ligands which cannot meet the pharmacophore constraints. Fig. 2 shows a set of four pharmacophore constraints in the active site of carbonic anhydrase. Examples are given where FlexX-PHARM significantly improved the results of docking in several PDB complexes where FlexX did not perform particularly well. FlexX-PHARM was also tested as a database-searching tool on a small dataset of molecules for three target proteins including thermolysin, carbonic anhydrase and dihydrofolate reductase. In two cases, FlexX-PHARM missed one or two of the active molecules due to the constraint selected. However, in general FlexX-PHARM maintained or improved the enrichment shown with FlexX. More importantly, it was able to do this using less time than required by FlexX. One challenge when using FlexX-PHARM is that the constraints must first be obtained and then introduced into FlexX-PHARM.

In the past three years, VS based on pharmacophore model has been applied to discover novel ligands over ten systems successfully (Table 1). The inhibitors discovered are novel, having little similarity to the known ligands. Most initial leads have affinities in the low-micromolar level. Among all these systems in Table 1, the pharmacophore models for 12 systems were obtained based on the ligands,

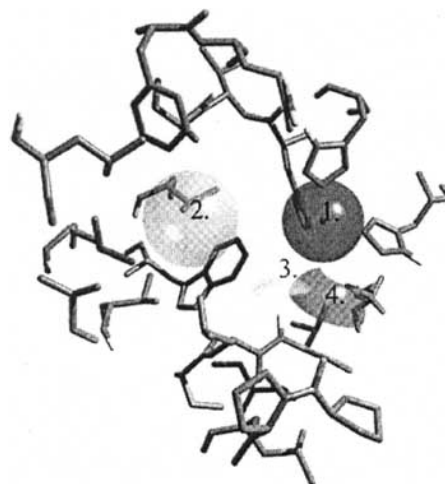


Fig. (2). A set of four pharmacophore constraints in the active site of carbonic anhydrase. Constraint 1: essential metal interaction at the zinc ion, constraint 2: essential spatial constraint for a carbon atom, constraint 3: optional hydrogen_donor interaction at the backbone nitrogen of residue Thr199, and constraint 4: optional hydrogen_acceptor interaction at the gamma oxygen of residue Thr199.

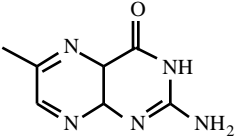
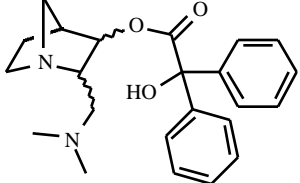
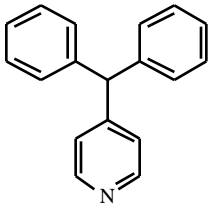
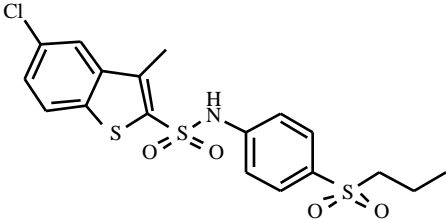
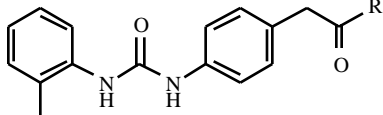
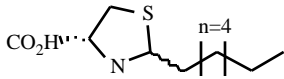
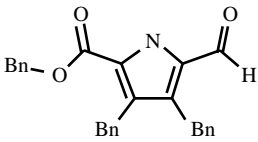
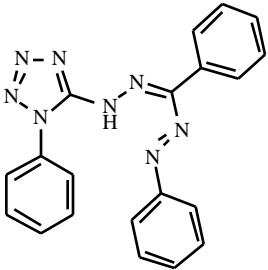
and only one reported by Brenk *et al.* was obtained based on the structural information of receptor. Brenk *et al.* superimposed to crystal structures of two ligands complexed with TGT by least squares fit. The hydrogen donor and acceptor features were defined using UNITY by picking the appropriate atoms of ligands. In order to consider the directionality of the hydrogen bonds, corresponding sites are attributed to the neighboring atoms of the protein and the interstitial water molecules. The hydrophobic moiety has been spatially characterized as the centroid of the benzoic ring of a ligand (Fig. 3). The proposed pharmacophore hypothesis was used in database search of UNITY.

3. COMBINATORIAL DRUG DESIGN

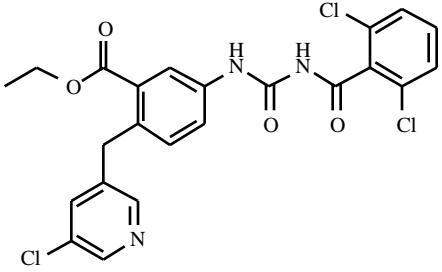
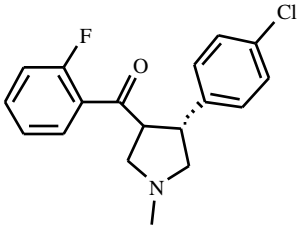
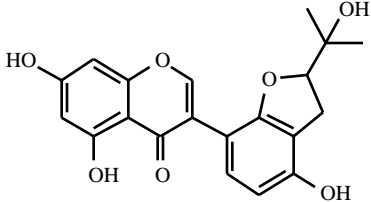
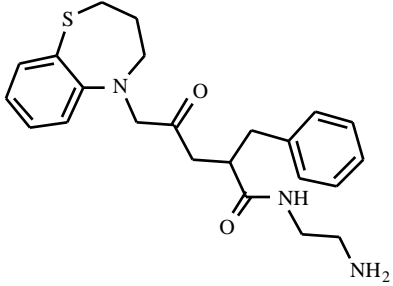
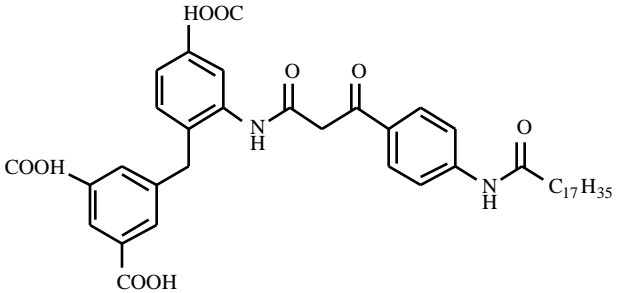
Rational or *de novo* drug design may be considered as an optimization process to find the best combination of molecular fragments which can be constituted into a whole ligand and well placed in the active site of the receptor. In the optimization process, the generated ligand should meet the given pharmacophoric features, which are represented by the important receptor-ligand interaction points. These interaction points are usually generated by the analysis of the 3D-structure of receptor.

Table 2 shows the representative *de novo* design methods applied in drug design. Besides TOPAS and LEA, the other approaches are developed before 2000, which have been reviewed in literature [36]. The methods can be roughly divided into two categories: atom-based and fragment-based. Atom-based techniques build up a molecule atom-by-atom, while fragment-based methods use sets of pre-defined molecular building blocks that are connected by a virtual synthesis scheme. Compared with atom-based techniques, fragment-based approaches seem more appealing, because a

Table 1. Examples of Bioactive Molecules Design Using Pharmacophore-Based Virtual Screening^a

Molecular structure	Target	Method	Activity (μM)	Ref.
	tRNA-guanine transglycosylase (TGT)	SuperStar, DrugScore, UNITY, FlexX	0.25 (K_i)	22
	Dopamine transporter (DAT)	Quanta, Chem-X	7.3 (K_i)	23
	Dopamine transporter (DAT)	Quanta, Chem-X	0.255 (K_i)	24
	Serine protease chymase	CATALYST	0.909 (K_i)	25
 <p data-bbox="358 1310 542 1335">R = leu-Asp-Val-OH</p>	antigen 4 1	CATALYST	0.0006 (IC_{50})	26
	EDG3	CATALYST	10 (IC_{50})	27
	Imidazole glycerol phosphate dehydratase	ISIS-3D	4.0 (IC_{50})	28
	Non-peptidyl endothelin converting enzyme	DISCO/ISIS-3D	10 (IC_{50})	29

(Table 1) contd....

Molecular structure	Target	Method	Activity (μm)	Ref.
	Mesangial cell proliferation	CATALYST	10 (IC_{50})	30
	Dopamine transporter (DAT)	Quanta, Chem-X	0.084 (IC_{50})	31
	Rat 5 -reductase	CATALYST	6.9 (IC_{50})	32
	Growth hormone secretagogues	DistComp/ISIS-3D	1 (EC_{50})	33
	E-selectin	ISIS-3D	1.2 nm (IC_{50})	34

virtual molecule can be easily constructed from combinatorial building blocks. In practical applications, atom-based techniques really do not show effective successes in drug discovery. In these *de novo* design methods shown in (Table 1), LUDI (Accelrys, <http://www.accelrys.com>) [14-15, 37],

Leapfrog (Tripos, <http://www.tripos.com>) [38] and SPROUT (SimBioSys, <http://www.simbiosys.ca>) [39-41] may be the most important ones, and all of these belong to fragment-based approaches. Fragment-based approaches can be further divided into two subclasses: (1) sequential growth, (2)

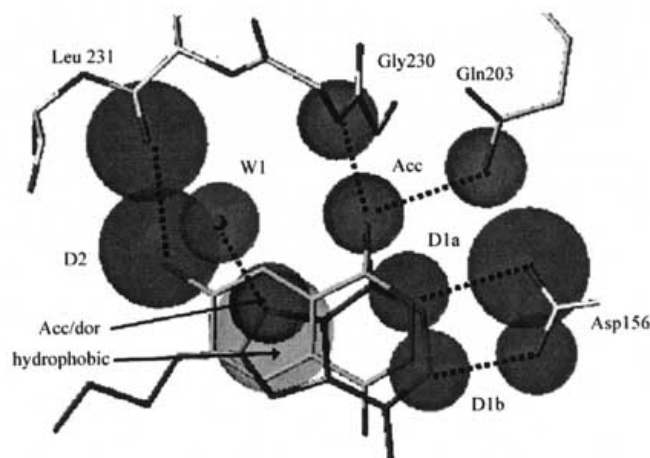


Fig. (3). Structure-based pharmacophore hypothesis.

Table 2. Examples of *De Novo* Design Methods

Method	Description	Ref.
CAVEAT	Fragment-based, fragment-based 3-D database search	42
CAVITY, FOUNDATION, DBMAKER, SPLICE	Fragment-based, generation of ligand by 3-D database search	43-46
CLIX	Fragment-based, fragment-based 3-D database search	47
CONCEPTS	Atom-based, stochastic search, MD optimization	48
CONCERTS	Fragment-based, stochastic search	49
DLD	Atom-based, structural sampling by Monte Carlo	50
GenStar	Atom-based, sequential growth	51
GroupBuild	Fragment-based, sequential growth	52
GROW	Peptide design, sequential growth	53
GrowMol	Fragment based, sequential growth	54
HOOK	Fragment-based linker search	55
LEA	Fragment-based, genetic algorithm	56
Leapfrog	Fragment-based, combinatorial search	37
LEGEND	Atom-based, sequential growth	57-58
LUDI	Fragment-based, sequential growth, combinatorial search	14-15
MCDNLG	Atom-based, stochastic search	59
MCSS	Fragment-based, multiple-copy simultaneous sampling	60
NEWLEAD	Fragment-based, connecting pharmacophoric pieces	61
PRO_LIGAND	Fragment-based, sequential growth, combinatorial search	62
SMOG	Fragment-based, sequential growth	63
SPROUT	Fragment-based, sequential growth, combinatorial growth	38-40
TOPAS	Fragment-based, evolutionary algorithm	64

fragment-location and linking. The latter approaches rely on the concept that a small number of well-placed fragments, each making very favorable interactions with receptor, can provide a significant overall binding affinity. Historically, these methods may be the most widely studied of all the *de novo* approaches. Certainly, we cannot give the final decision that fragment-location and linking techniques are superior to sequential growth ones because both of these two classes of methods bear inherent advantages and disadvantages. In many *de novo* design systems, these two approaches can be used as the user's demand, for example, LUDI and Leapfrog.

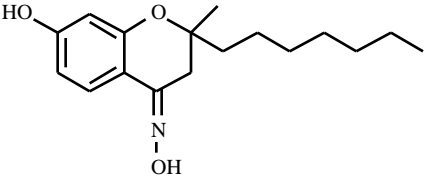
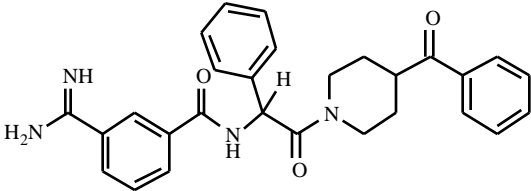
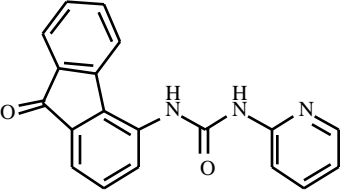
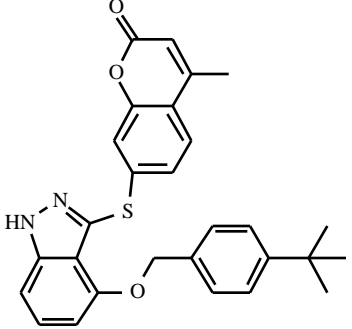
The big problem of most *de novo* design methods may be synthetic inaccessibility, which can be overcome by using the program CLIX, SEEDS, CAVEAT to search the

Available Chemicals Directory (MDL, Inc., San Leandro, CA) for scaffolds that were components of the designed ligands. The scaffold-based approach was used to discover novel inhibitors of Factor Xa and DNA gyrase (Table 3). It is interesting to find that the scaffold-based approach is quite similar to pharmacophore-based VS. These molecular fragments used as scaffolds can be considered as the pharmacophoric features in a pharmacophore model. In fact, sometimes, the borderline between two different approaches is not so obvious.

3.1. Recent Advances and Successful Applications of Combinatorial Drug Design

In the past several years, combinatorial drug design approaches based on genetic algorithm may be the most

Table 3. Examples of Bioactive Molecules Design Using *De Novo* Design

Molecular structure	Target	Method	Activity	Ref.
	Lanosterol 14 -demethylase (CYP51)	MCSS/LUDI	35.21 (IC ₅₀)	65
	Factor Xa	PRO_SELECT	16nm (K _i)	66
	Cdk4	LEGEND/SEEDS	7.6 (IC ₅₀)	67
	DNA gyrase	LUDI, CATALYST, MOLOC	0.03 (MNEC)	68

important advances in this area. In (Table 2), both of TOPAS and LEA are based on genetic algorithm. In TOPAS [64], a set of ~24,563 fragment structures serves as the building blocks, which were obtained by a straightforward fragmentation procedure applied to 36,000 known drugs. Eleven reaction schemes were implemented for both fragmentation and building block assembly. This combination of drug derived building blocks and a restricted set of reaction schemes proved to be a key for the automatic development of novel, synthetically tractable structures. TOPAS is based on a simple evolutionary algorithm (EA), a (1,) evolution strategy. In optimization process, molecules were generated from a parent structure by virtual synthesis, and the best structure of a generation was selected as the parent for the subsequent TOPAS cycle. The fitness function used in GA optimization is the pair-wise similarity between the template and a new variant structure. TOPAS was used to find the

potential thrombin inhibitors. The first application of TOPAS was tried to develop molecules mimicking the NAPAP structure. The Tanimoto index was used as the fitness function. After only 12 optimization cycles the process converged at a high fitness level (0.86), and the best hit is quite similar to NAPAP (Fig. 4). Moreover, TOPAS was used to the design of the peptide-analogues of the tripeptide D-Phe-Pro-Arg, an excellent natural thrombin substrate. The calculated results are encouraging.

The LEA approach is also based on evolutionary algorithm [56], but the basic operations of LEA are quite different from those of TOPAS. The present LEA is able to operate directly on the SMILES line notation. The algorithm takes an initial set of fragments and iteratively improves them by means of crossover and mutation operations. Figure 5 shows the crossover operations in LEA. In LEA, physico-

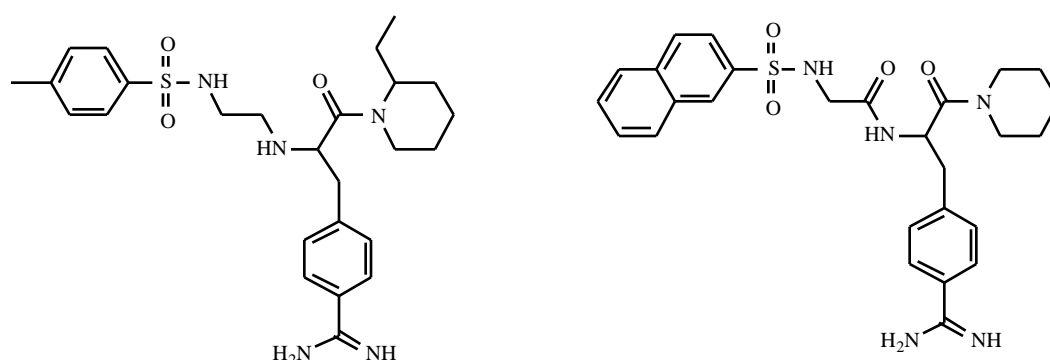


Fig. (4). The best structure afforded by TOPAS (a) and (b) NAPAP.

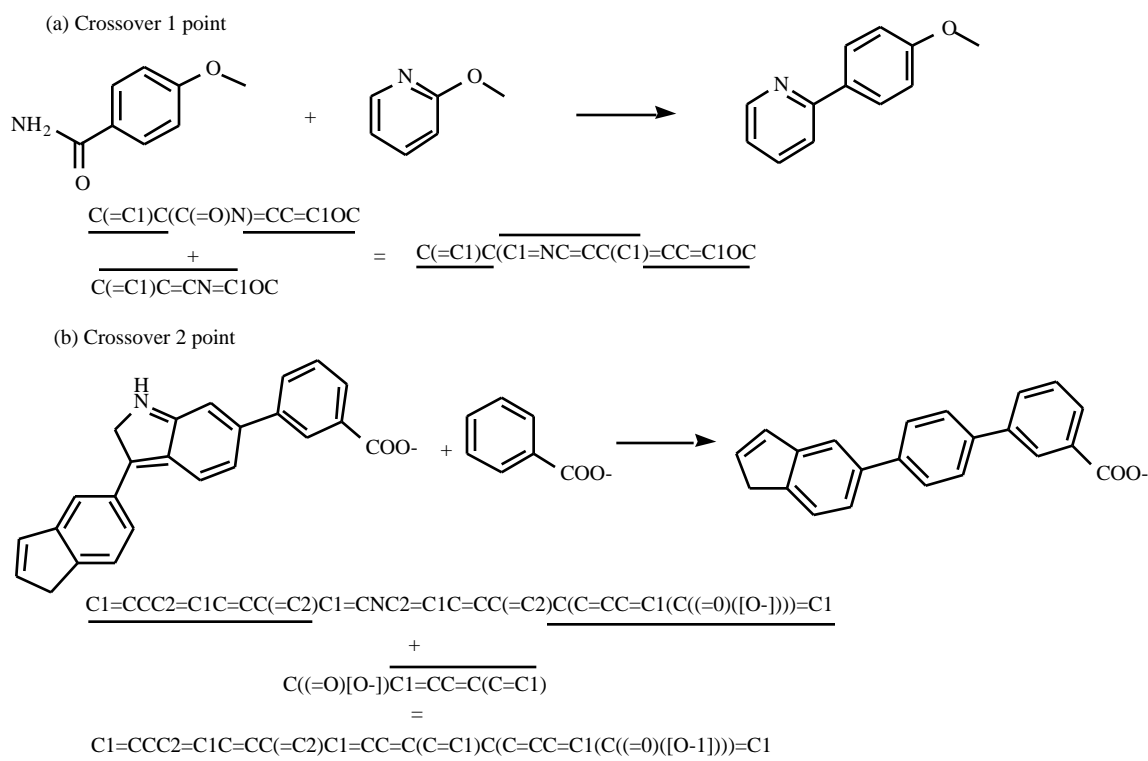


Fig. (5). Crossover operations in LEA.

chemical properties of molecules are employed as scoring function. LEA was applied to the *de novo* molecular design of new retinoids, and some promising retinoid analogs are obtained.

Compared with the other virtual screening approaches, the *de novo* design techniques do not show huge progress in practical applications. Certainly, *De novo* design approaches are beginning to show efficacy in the process of lead generation. Recently, Ji *et al.* proposed several non-azole lead molecules by *de novo* design [65]. All of the lead molecules exhibited strong inhibitory effects on CYP51 of *Candida albicans*. Table 3 shows the examples of bioactive molecules design using *de novo* design.

4. VIRTUAL SCREENING BASED ON MOLECULAR DOCKING

Molecular docking can fit molecules together in favorable configuration to form a complex system. It has been practically applied as a very efficient way in the studies of protein-ligand interactions. Structural information from the theoretically modeled complex may help us to clarify the mechanism of molecular recognition, and even can instruct us to discover novel ligand according to the predicted binding affinities between receptor and ligands [69]. The attraction of receptor-ligand docking is that it represents the most detailed and relevant computational model for identifying a receptor-focused subset of database. There are a large number of docking programs available for use in virtual screening, differing in the sampling algorithm, scoring functions, the treatment of flexibility of ligand and receptor, and the CPU time required to dock a molecule to a given target. The commonly used docking programs are listed in (Table 4).

4.1. Recent Advances of Molecular Docking and Successful Examples

Researches on methodology of molecular docking are very active in the past several years. Along with the further improvements of traditional methods, such as DOCK, FlexX (Tripos, <http://www.tripos.com>), AutoDock, Gold (Cambridge Crystallographic Data Centre, <http://www.ccdc.cam.ac.uk>), ICM (MolSof, <http://www.molsoft.com>), several new docking programs have been released recently, including EUDOC, LigandFit (accelrys, <http://www.accelrys.com>) and eHiTS (SimBioSys, <http://www.simbiosys.ca>).

EUDOC was designed by Pang *et al.* [72]. Compared with the other docking techniques, the EUDOC program has its own features. First, the molecular flexibility of both ligand and receptor is taken into account via either conformation selection theory or conformation substitution theory. Second, the program computes the intermolecular interaction energies of biologically important metal ions such as Zn^{2+} , Ca^{2+} , and Mg^{2+} that mediate the binding of ligand to receptor using a cationic dummy atom approach. Third, the program uses "spatial decomposition" to achieve 100% parallelism in computing. Using the EUDOC program, the authors performed a computational screening of the 1998 release of the ACD (Available Chemicals Directory) database to identify prototypic inhibitors of FTase containing a zinc divalent cation in the active site and identified 21 compounds as potential inhibitor leads of FTase. Among these 21 compounds, 18 compounds have inhibitory activities against FTase *in vitro* at the concentration of 500 μM ; and 4 compounds with *in vitro* IC_{50} values in the range from 25 to 100 μM . The most potent inhibitor also has inhibitory activity in human lung cancer cells. Furthermore, this lead ($IC_{50} = 25 \mu M$) identified with the aid of the EUDOC

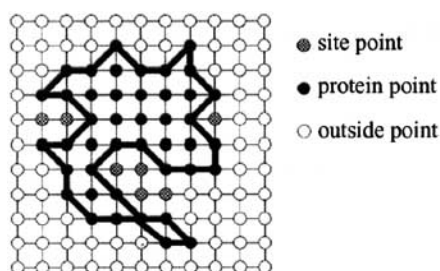
Table 4. Commonly Used Docking Programs for Virtual Screening

Method	Sampling method	Scoring function	Speed	Ref.
DOCK	Incremental build	Force field, contact score, chemical complementary score	Fast	70
eHiTS	Exhaustive search	Empirical score	Fast	71
EUDOC	Exhaustive search	Force field	Fast	72
FlexX	Incremental build	Empirical score	Fast	73
ICM-Dock	stochastic global optimization	Empirical score	Fast	74
LigandFit	Monte Carlo	Empirical score	Fast	75
FlexiDock	Genetic algorithm	Force field	Slow	76
Fred	Conformational ensembles	Shape complementarity, Gaussian score	Fast	77
Slide	Conformational ensembles	Empirical score	Fast	78
Affinity	Monte Carlo/MM/MD	Force field	Slow	79
AutoDock	Genetic algorithm	Empirical score	Slow	80
Glide	Exhaustive search	Empirical score	Slow	81
Gold	Genetic algorithm	Empirical score	Fast	82
QXP	Monte Carlo	Force field	Slow	83

program, is more active than the generally recognized most potent inhibitor lead, kurasoin A ($IC_{50} = 59 \mu M$).

LigandFit is developed by Venkatachalam *et al.* recently [75]. The method belongs to a new shape-based method, which employs a cavity detection algorithm for detecting invaginations in the protein as candidate active site regions (Fig. 6a). A shape comparison filter is combined with a Monte Carlo conformational search for generating ligand poses consistent with the active site shape. The initial docking of a ligand is obtained by alignment of the principal axes of the ligand to the principal axes of the site (Fig. 6b). Then the docking energy was used to explore the best docking conformation of ligand. The docking energy includes internal energy of the ligand and the interaction energy of ligand and receptor. In order to reduce the errors arising from grid interpolation, a new non-linear interpolation scheme was used. The authors applied LigandFit to 19 diverse protein-ligand complexes. The results appear quite encouraging, reproducing the X-ray structure ligand poses within an RMS of 2 Å in 14 out of 19 complexes. Moreover, high-throughput virtual screening using LigandFit was applied to the thymidine kinase receptor. These results show that combined with LigScore, LigandFit yields very good hit rates for a ligand pool seeded with known activities.

6a.



6b.

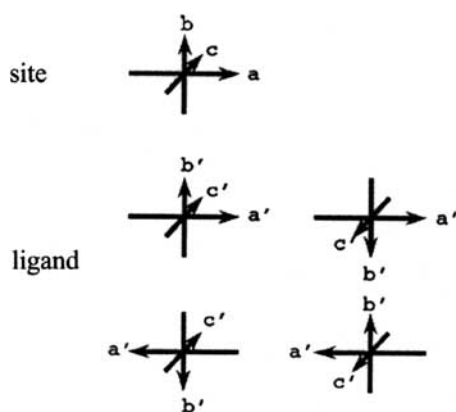


Fig. (6). (a) Schematic representation of the grid system enclosing the protein (black boundary), and (b) four orientations of the ligand consistent with the shape correspondence between the ligand and the site.

The eHiTs program was released by SimBioSys, Inc. recently [71]. In eHiTs, the search algorithm is based on exhaustive graph matching that rapidly enumerates all possible mappings of interacting atoms between receptor and ligand. Then dihedral angles of rotatable bonds are computed deterministically as required by the positioning of the interacting atoms. Consequently, the algorithm can find the optimal conformation even if unusual rotomers are required. The scoring function contains novel treatment of weak hydrogen bonds, aromatic π -stacking and penalties for conflicting interactions. Alternatively, the target cavity description can be based on a pharmacophore model or a CoMFA study of overlaid lead compounds. Another very important feature of the eHiTS system is its exhaustive nature. A systematic algorithm is used in eHiTs with no random, stochastic or evolutionary element. Therefore, eHiTS provides comprehensive search space coverage unlike other methods, which are limited to finding an arbitrary subset of possible solutions due to their use of sampling. eHiTs divides a ligand into rigid fragments or nodes that are docked separately in receptor and stores rigid fragment poses in DockTable, a SQL database that increases the speed of dockings by dynamically updating and retrieving molecular fragments during the docking process.

All docking procedures should meet two ongoing methodological challenges: adequate sampling of receptor-ligand conformations and accurate evaluation of receptor-ligand complementarity. In most docking programs, the flexibility of ligand is carefully treated with different computational techniques, for examples, incremental build and conformational search in DOCK and FlexX, genetic algorithm in AutoDock and GOLD and pseudo-Brownian sampling in ICM etc. But the treatment of receptor flexibility remains a major challenge. In many cases, the conformational changes induced by ligand binding are very significant. If one cannot consider the receptor flexibility properly, the docking results may be questionable. Recently, Hou *et al.* studied the binding mode of Quinazoline Type inhibitor complexed with EGF-R. DOCK calculations do not give the proper orientation of inhibitor in EGF-R. The reason lies in the fact that the present DOCK program does not consider the flexibility of the protein [84].

Most effects to incorporate receptor flexibility are the usage of the ensemble of receptor structures from NMR structural study or molecular dynamics simulations. Recently, the program AutoDock has been used to investigate several strategies for incorporating protein flexibility using an ensemble of receptor to generate Boltzmann-weighted grids with which docking function are generated [85]. In FlexE, an extension to FlexX, a predefined ensemble of receptor structures were used in molecular docking. The several available crystal structures were superimposed, and alternative conformations are recombined to create complete structure of receptor. Docking against this ensemble is two-fold faster than explicit docking against all conformations [86]. Naturally, the usage of the ensemble of receptor structures would be computational expensive, because the docking process should be performed to each individual receptor structure, with a view to identifying all ligands that bound to at least one conformational form of receptor.

In the last three years, the practical applications of VS based on molecular docking in drug design are very successful (Table 5), which means that VS based molecular docking is gradually stepping from a stage of technical development into a new stage of successful application in pharmaceutical industry.

4.2. Scoring Functions

As we discussed above, docking of small molecules into their biomolecular receptors is a useful tool in drug discovery. The reliability of molecular docking or even *de novo* design is eventually based on precise scoring functions. Scoring functions are used to estimate the binding affinity of a novel molecule or a molecular fragment in the active site of receptor. Scoring functions are a very active and rapidly advanced research area, and substantial progress has been made in 2000 and 2003, resulting in a broad spectrum of methods for estimation of binding affinities. Here, we only give a brief introduction to the new development in this field, including statistically-fit scoring functions, knowledge-based scoring functions, first-principle-based scoring functions and consensus scoring approach.

Statistically-fit scoring functions are widely used in molecular docking, which divide the total binding affinity of protein-ligand complexes to several empirical weighted-interaction terms, including hydrogen bonding, metal-ligand interaction, hydrophobic interaction, and rotational entropy etc [107]. The advantage of statistically-fit scoring functions are obviously: first, these kinds of scoring functions are computationally efficient; second, the individual terms give chemists more intuitionistic knowledge of which specific force largely contributing to binding free energies. Unfortunately, statistically-fit scoring functions are dependent on their training set. Each author has tried to make this as general as possible, but concerns remain as to the extensibility of these functions to new systems.

In recent years, knowledge-based potentials have made huge progresses. These kinds of approaches used the experimental determined protein-ligand complexes by statistical means to extract rules on preferred binding geometries. These rules are converted into 'pseudo-potentials', which, in turn, can be applied to score predicted ligand binding conformations. Compared with force-field potentials, these knowledge-based potentials implicitly incorporates physical effects not yet fully interpreted from only the theoretical point of view, for example, (de) solvation and polarization. As statistically-fit scoring functions, knowledge-based potentials may be greatly limited by experimental crystal structures for deriving these potentials. During the past several years, several knowledge-based approaches to ligand binding have been reported, which include SMOG [108, 109], PMF [110], BLEEP [111], DrugScore [112, 113] and others [114, 115]. All these approaches are based on the same formalism and rate observed frequency or probability distributions to pair-(pseudo-)potentials:

$$W_{ij}(r) = -\ln \frac{g_{ij}(r)}{g_{ref}} \quad (1)$$

where $g_{ij}(r)$ is the frequency or probability distribution of atom pairs of type i and j at a distance from each other, and g_{ref} corresponds to a reference distribution. The choices of

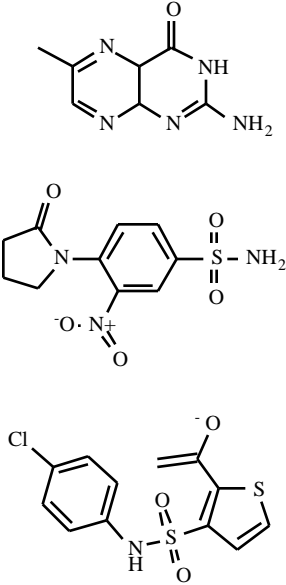
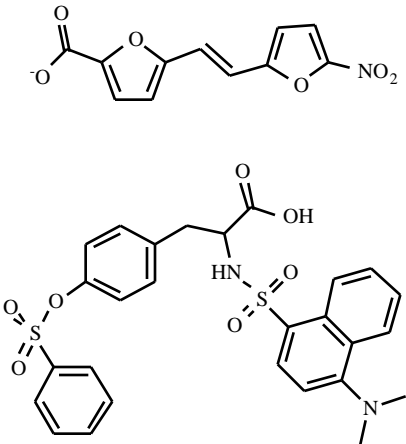
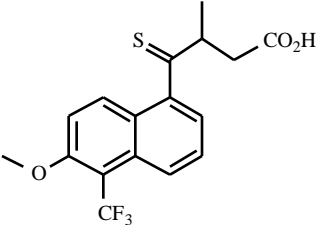
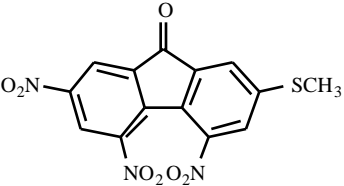
the reference distribution are treated differently in these approaches. Moreover, the treatments of solvation effects are different in these approaches.

In these three knowledge-based potentials, DrugScore proposed by Gohlke *et al.* may be the most popular one. The knowledge-based scoring function used in DrugScore was developed by converting structural information for 1376 protein-ligand complexes, extracted from Relibase [116], into distance-dependent pair-potentials and solvent-accessible surface-dependent singlet-potentials using 17 different atom types. The sum of both terms was used to score protein-ligand interactions. For two test sets of 91 and 68 complexes, DrugScore reevaluated the multiple solutions proposed by FlexX, and gave the best solution (rank 1) in 75% of all cases. Moreover, Gohlke *et al.* compared the binding geometries used DrugScore and 'chemical score' in DOCK 4.0, and found that DrugScore is superior. This proves the DrugScore is somewhat more reliable in predicting the binding affinities. For a test set with 56 crystal protein-ligand complexes, predicted power of $r^2=0.56$ was obtained for predicting binding free energies. Furthermore, DrugScore was used in objective function in docking optimizations. Compared to the AutoDock Scoring function, DrugScore yields slightly superior results in flexible docking [117]. Now, DrugScore is available in the SYBYL interface to FlexX.

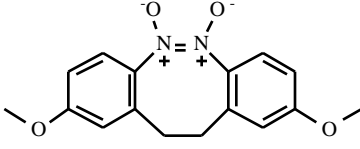
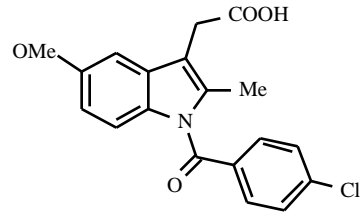
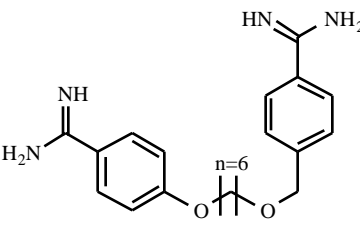
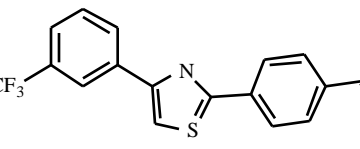
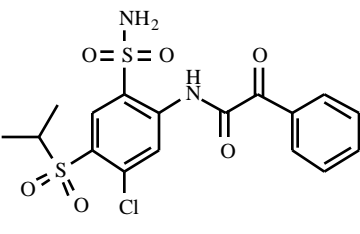
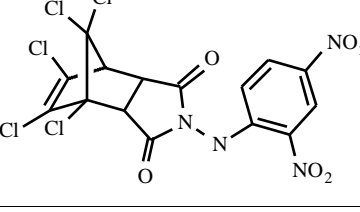
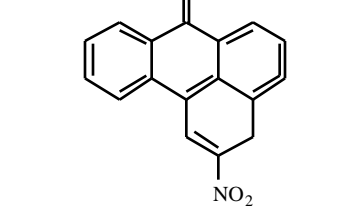
Recently, Ishchenko reported SmoG2001 [118], the improved knowledge-based scoring function of SMOG. In the latest version of the potentials, the authors defined the reference state that ensures proper normalization of contact probabilities (sum of all values over atom types is equal to 1) and introduced to distance intervals ("bins") over which the contact statistics are computed. SMOG2001 reproduces the experimental binding constants of the majority of 119 complexes of the training set with high accuracy. Moreover, SMOG2001 performs better than two other widely used scoring functions, PMF and SCORE1 in LUDI, and comparably to DrugScore. SMOG2001 poorly predicts the affinities of ligands interacted with metal ions and ligands that are large and flexible.

First-principle-based approaches may become more and more important in molecular docking in the near future. In contrast to statistically-fit scoring functions, first-principle-based approaches divided the total binding affinities to individual terms, and computed those terms from physico-chemical theory while not from fitting to experimental binding affinities. Compared with the two kinds of scoring functions above, first-principle-based approaches are the most theoretically rigorous, though they are relatively time-consuming and bear theoretical difficulties in the evaluations of certain contributions, especially entropy and de(solvation) effect. Recently, the researches in the Kuntz groups have ever introduced PBSA (Poisson-Boltzmann/surface area) and GBSA (Generalized Born/surface area) to evaluate de(solvation) contributions in molecular docking [119, 120]. In First-principle-based approaches, the 'MM-PBSA' approaches proposed by Kollman *et al.* may be the most important improvement in this field [120-124]. In MM-PBSA, the binding free energy of a noncovalent associating for a protein-ligand system can be computed as:

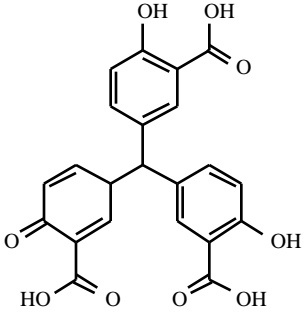
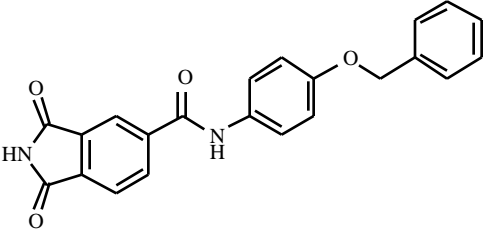
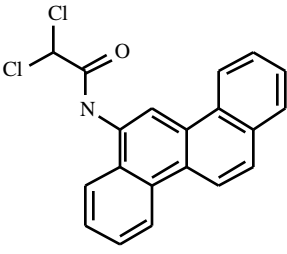
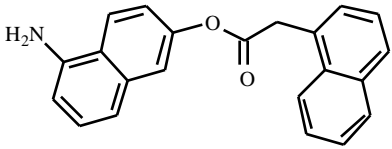
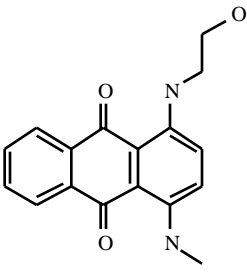
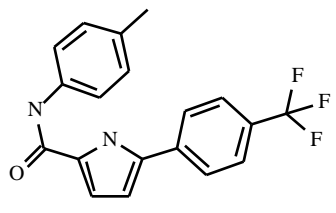
Table 5. Examples of Bioactive Molecules Design Using Docking-Based Virtual Screening

Molecular structure	Target	Method	Activity	Ref.
	<p>tRNA-guanine transglycosylase (TGT)</p> <p>Human carbonic anhydrase</p> <p>AmpC β-lactamase</p>	<p>SuperStar, DrugScore, UNITY, FlexX</p> <p>Unity, FlexS, FlexX</p> <p>DOCK (Northwest Univ. Version)</p>	<p>0.25 (K_i)</p> <p>0.6 nm (IC_{50})</p> <p>26 (K_i)</p>	<p>22</p> <p>87~88</p> <p>89</p>
	<p>β-adrenergic receptor kinase 1 (β-ARK1)</p> <p>Thymidylate synthase</p>	<p>ARCHER</p> <p>DOCK</p>	<p>126 (IC_{50})</p> <p>1.4 (K_i)</p>	<p>90</p> <p>91</p>
	<p>Aldose reductase (ALR2)</p>	<p>DOCK</p>	<p>0.10 (IC_{50})</p>	<p>92</p>
	<p>Adenovirus proteinase</p>	<p>EUDOC</p>	<p>3.09 (K_i)</p>	<p>93</p>

(Table 5) contd....

Molecular structure	Target	Method	Activity	Ref.
	Bcl-2	DOCK	10 (IC ₅₀)	94
	Aldose reductase	ADAM&EVE, GREEN	4.3 (IC ₅₀)	95
	Matriptase	DOCK	0.92 (IC ₅₀)	96
	Retinoic acid receptor	ICM	2 (ED ₅₀)	97
	Dihydrodipicolinate reductase	FLOG	7.20 (K _i)	98
	Farnesyltransferase	EUDOC	25 (IC ₅₀)	99
	Protein tyrosine phosphatase (PTP1B)	DOCK	21 (K _i)	100

(Table 5) contd....

Molecular structure	Target	Method	Activity	Ref.
	Kinesin	DOCK	1.4 (IC ₅₀)	101
	phosphoribosyltransferase	DOCK	16 (IC ₅₀)	102
	Hypoxanthine Phosphoribosyltransferase (HPRT)	DOCK	0.5 (K _i)	103
	Serine protease	FRAGMENT++, DOCK	79 (pIC ₅₀)	104
	HIV-1 TAR	DOCK, ICM	1 (CD ₅₀)	105
	Nuclear hormone receptor (NR)	ICM	NR	106

$$\begin{aligned}
 G_{\text{bind}} &= G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} \\
 &= E_{\text{MM}} + G_{\text{PB}} + G_{\text{NP}} - T S
 \end{aligned}
 \quad (2)$$

where G_{PB} is the polar solvation energy in continuum solvent, usually computed using a finite-difference Poisson-Boltzmann (PB) model, and G_{NP} is the nonpolar solvation energy, which is often obtained from the solvent-accessible surface area (SA). E_{MM} denotes the sum of molecular mechanical (MM) energies of the molecules from internal, electrostatic and van der Waals energies. The last term in equation 2 is the solute entropy and can be estimated by a combination of classical statistical formulas and normal-mode analysis. The thermally averaged energy terms in equation 2 are obtained from the MD sampling. In recent paper of Wang's *et al.*, the author applied the MM/PBSA technique to determine the binding mode between HIV-1 RT and efavirenz. The authors calculated the binding free energies between HIV-1 RT and efavirenz for the five possible binding modes, and successfully recognized the correct one. The final structure of the MD simulations is very close to a 2.3 Å crystal structure solved by the Dupont Pharmaceuticals recently, and the rmsd of the ligand and its surroundings (about 50 residues) is 1.0 angstrom [125]. In 2003, Hou *et al.* applied MD simulations combined with MM-PBSA to determine the correct binding mode of the quinazoline inhibitor and EGF-R. The most favorable binding mode identified by MM/PBSA has a binding free energy about 10 kcal/mol more favorable than the second best one. When the authors began this project, the crystal structure of EGF-R is not available, so the homology model of EGF-R was applied. When the project was finished, the crystal structure of quinazoline/EGF-R was reported. It is interesting to find that the best binding model predicted by MM-PBSA is very similar to the structure [84]. Moreover, it is encouraging that the crystal water molecules crucial to ligand binding was produced by the predicted model. In the further research, the development of efficient approaches to represent the entropic effects precisely and quickly may be one of the most important tasks in MM/PBSA.

Since each scoring function has been derived from a different set of crystal structures, it is reasonable to use multiple functions when evaluating a protein-ligand pair. A consensus can be applied: structures with good fits in multiple scoring functions can be examined further, and just drop the ones without. In 1999, Charifson *et al.* totally used three scoring functions to rank docked conformations [126]. Compared with the performances of a single scoring function, the hit rates can be effectively improved. However, it has been questioned by the authors themselves whether such an approach is of general use to predict binding free energies of small sets compounds. Recently, Tripos proposed a module named CScore (Consensus Score) (Tripos, <http://www.tripos.com>) [127] to integrate a number of popular scoring functions for ranking the affinity of ligands bound to the active site of a receptor. The CScore module may be the first available commercial program that uses the consensus score. In CScore, five different scoring functions were used, including G_{score} [128], $\text{PMF}_{\text{score}}$ [129], D_{score} [130], ChemScore [131] and F_{score} [132]. In CScore, the range of scores for each scoring function are determined, above these the cutoff threshold are considered

"good", and the consensus score is the sum of the number of "good" results for each ligand in each scoring function. A publication by Tripos scientists indicates the reliability of molecular docking can be improved by combining results from functions in CScore [133]. In 2002, Paul and Rogan proposed a consensus docking approach (ConsDock) that takes advantage of three widely used docking tools (Dock, FlexX, and Gold) [134]. When applied to a test set of 100 protein-ligand complexes from the Protein Data Bank, ConsDock significantly outperforms single docking with respect to the docking accuracy of the top-ranked pose. In 60% of the investigated cases, ConsDock was able to rank as top solution a pose within 2 Å RMSD of the X-ray structure.

5. VIRTUAL SCREENING BASED ON ADME/TOX FILTERS

The significant failure rate of drug candidates in later developmental stages is driving the need for predictive tools that can eliminate inappropriate compounds before substantial time and money are invested in testing. It has been estimated that about 50% of such failures are caused by ADME/Tox deficiencies. Virtual screening should not be restricted to the scenarios with respect to optimizing binding affinity, and the pharmacokinetic properties should also be treated as important filtering protocol.

5.1. *In Silico* Prediction of ADME Properties

The simplest ADME-concerned filters may be "rule of 5" proposed by Lipinski *et al.* in 1997 [135]. Lipinski and coworkers analyzed a subset of 2245 drugs from the World Drug Index (WDI). They found that poor absorption and permeation are more likely to occur when (1) the molecular weight is over 500, (2) the octanol/water partition coefficient is over 5 (CLOGP) or 4.15 (MLOGP), (3) the number of hydrogen-bond donors (OH and NH groups) is more than 5, and (4) the number of hydrogen-bond acceptors (N and O atoms) is more than 10. The fast estimations of $\log P$ allow the "rule of 5" screening of library prior to enumeration. Moreover, based on screening results from Merck and Pfizer, Lipinski argues that it is much easier to optimize pharmacokinetic properties early on the process of drug discovery, and attempt to optimize the receptor binding affinity at a later stage [136].

Many computational approaches have been developed for ADME parameters [137-145], including bioavailability, aqueous solubility, intestinal permeability, blood-brain barrier penetration, metabolism, drug-drug interactions, drug transport and toxicity. The predictions of these properties are involved in two aspects of modeling methods: data modeling and molecular modeling. For molecular modeling, molecular mechanics, pharmacophore modeling, molecular docking, or even quantum mechanics are used to explore the potential interactions between the small molecules under consideration and proteins known to be involved in ADME processes, such as cytochrome P450s. For data modeling, quantitative structure-activity relationship (QSAR) approaches are typically applied. Based on appropriate descriptors, QSAR ranging from simple multiple linear regression to modern multivariate analysis techniques or machine-learning methods, such as partial least squares (PLS) [146], genetic algorithm

(GA) [147], neural networks (NN) [148], and support vector machines (SVM) [149], are now being applied to the analysis of ADME data.

The predictions of aqueous solubility, intestinal permeability, and blood-brain barrier penetration are usually based on the QSAR approaches. Most of the early predictive models for intestinal permeability or blood-brain barrier penetration are based on a multiple linear regression and many used physicochemical properties, such as polar surface area, $\log P$, volume, and hydrogen bonding capacity [138]. Additionally, other kinds of descriptors have been developed and applied to ADME predictions. Ghuloum *et al.* used a numerical molecular representation called molecular hashkey predict $\log P$ and intestinal absorption of a set of drugs [150]. In VolSurf, a variety of three-dimensional molecular field descriptors are transformed into a new set of descriptors, which was used as inputs for the construction of models for aqueous solubility, intestinal permeability and blood-brain barrier penetration [145].

The relationships between oral bioavailability and molecular structures have been investigated recently [151-154]. The prediction of oral bioavailability is relatively difficult, because it depends on a superposition of two processes: absorption and liver first-pass metabolism. Yoshida and Topliss constructed a QSAR model with a set of physicochemical parameters [151]. The oral bioavailability determined in human adults was assigned one of four ratings and analyzed in relation to physicochemical and structural factors by the ORMUCS (ordered multicategorical classification method using the simplex technique) method. The predictive power of the model was evaluated using a separate test set of 40 compounds, of which 60% (95% within one class) were correctly classified. Andrews *et al.* also proposed a QSAR model based on 591 compounds and 85 structural descriptors [152]. Moreover, compared with the Lipinski's rule of 5, the false negative predictions were reduced from 5% to 3% while the false positive predictions decreased from 78% to 53%. However, the predictability of this model of the false positive predictions are quite close to random, so it cannot be applied as a filter in virtual screening as the current form. Bans used genetic programming to predict oral bioavailability [153]. The results show a slight improvement than the Toshida's results, although a direct comparison is difficult owing to different selection of the bioavailability ranges of the four classes. Recently, Mandagere *et al.* proposed a graphical model integrating available *in vitro* ADME data, such as Caco-2 permeability and metabolic stability in liver S9 or microsomes, to estimate %F into groups of low, medium, or high regions [154]. For a large number of drug candidates, this graphical model provides a tool to estimate human oral bioavailability from *in vitro* ADME data. But the inputs for this model rely on other ADME-concerned properties, such as permeability and metabolic clearance, and thus cannot be used as high throughput fashion in virtual screening.

Theoretical studies of transporter proteins, especially P-Glycoprotein (P-gp), are active in recent years, because transport proteins are found in most organs involved in the uptake and elimination of endogenous compounds and xenobiotics, including drugs [155]. QSAR and pharmaco-

phore modeling were applied to study the common features of p-gp substrate, the interactions between substrate and P-gp [156-161]. On the basis of the obtained results, we can give some explanations to the broad structural variety of the P-gp substrates and inhibitors and give predicted models for discrimination between substrate and non-substrate, but at present the accuracy of those models is too limited to be applied in virtual screening.

The researches on metabolism are also very active in recent years, because the extent and rate of metabolism affect clearance, whereas the involvement of particular enzymes might lead to issues related to the polymorphic nature of some of these enzymes and to drug-drug interactions. Ekins *et al.* proposed 3D/4D QSAR and pharmacophore modeling to investigate the common structural features of cytochrome P450 (CYP) 2C9 inhibitors and interactions between competitive inhibitors and P4502C9 [162-163]. The obtained 3D- and 4D-QSAR models of CYP inhibition will aid in future prediction of drug-drug interactions. Cruciani *et al.* proposed a program named MetaSite to identify potential substrate of cytochrome P450s [164], which is based on pharmacophore model obtained from interaction fields for the protein structure and a pharmacophoric fingerprint for the potential substrate. Zuegge proposed a linear PLS-based prediction system for binary classification of drug-drug interaction liability caused by cytochrome P450 3A4 inhibition. It correctly predicts 95% of the training data and 90% of a semi-independent validation data set, and can be used as a valid filter in virtual screening [165]. Several approaches that use databases to predict metabolism are available or under development, including expert systems, such as MetabolExpert, META or Meteor and Metabolism [166, 167].

Overall, the number of ADME properties that can be predicted computationally is very limited. Due to the predictability of these approaches, only several properties, including drug solubility, Caco-2 cell absorption, blood-brain barrier permeation, can be practically applied in virtual screening. At present, the widely-used programs for ADME predictions include VolSurf (tripos, <http://www.tripos.com>) [145, 168], C2.ADME (accelrys, <http://www.accelrys.com>) [169, 170] and QikProp (schrödinger, <http://www.schrödinger.com>) [171-173].

VolSurf reads or computes 3-D molecular interaction fields and uses image-processing methods to convert them into simple molecular descriptors that are easy to understand and interpret. These descriptors quantitatively characterize size, shape, polarity, and hydrophobicity of molecules, and the balance between them. Multivariate statistical methods within VolSurf enable the creation of models that relate its descriptors to biological properties. The ADME models included in VolSurf predict drug solubility, Caco-2 cell absorption, blood-brain barrier permeation, and drug distribution. These models have been developed from published experimental data from *in vitro* assays to *in vivo* behavior of drugs. Recent studies have shown that models of membrane partitioning generated using VolSurf's descriptors are significantly more predictive than those generated from other descriptors.

C2-ADME can give predictions for several ADME properties including passive intestinal absorption, blood-brain barrier (BBB) penetration, and aqueous solubility at 25°C. All three models are designed for analysis of combinatorial and virtual libraries, as well as individual molecules. The models were developed and validated using a large and diverse set of compounds with data obtained from the literature as well as data generated by Pharmacopeia Drug Discovery Services. C2-ADME does not require 3-D structure generation which can be the most time consuming part of the prediction process, usually slowing it down by making the analysis of large virtual libraries difficult and laborious.

Based on years of research, QikProp has been developed by Professor Bill Jorgensen at Yale University, specifically for drug discovery. QikProp results have been fitted to datasets of drug-like molecules, based on 2-D and 3-D descriptors reflecting Monte Carlo simulation studies as well as experiment. QikProp predictions are calculation-based, as opposed to fragment based. The calculated features of QikProp include Solubility, Blood/Brain barrier permeability, Caco-2 Cell Permeability Boehringer, Caco-2 Cell Permeability Affymax, MDCK Cell Permeability Affymax and Skin Permeability Coefficient.

5.2. *In Silico* Prediction of Toxicity

Toxicity is responsible for many compounds failing to reach the market and for the withdrawal of a significant number of compounds from the market once they have been approved. The current approaches for *in silico* prediction of toxicity can be divided into two basic categories: knowledge-based and statistically-based [174, 175]. Knowledge-based approaches use rules about generalized relationships between structure and biological activity that are derived from human expert opinion and interpretation of toxicological data to predict the potential toxicity of novel structure. On the other hand, statistically-based approaches use calculated parameters, structural connectivity and the application of various statistical methods to derive mathematical relationships for a training set of non-congeneric compounds in an unbiased manner. At present, several commercial tools of *in silico* prediction of toxicity are available, which include; DEREK (LHASA Limited, <http://www.chem.leeds.ac.uk>), Hazard Expert, COMPACT, CASE and MCASE (MultiCASE, <http://www.multicase.com>), OncoLogic (LogiChem, <http://www.logichem.com>) and TOPKAT (accelrys, <http://www.accelrys.com>). The features, main strength and limitations of these programs have recently been reviewed [174]. The primary emphasis of the current software packages is carcinogenicity and mutagenicity, although some packages do also include models and/or knowledge bases for other endpoints, such as teratogenicity, irritation, sensitization, immunotoxicology and neurotoxicity.

It should be noted that at present the predictability of the commercial tools for toxicity prediction is quite limited. The National Toxicology Program (NTP) has conducted several exercises to validate the predictability of the widely-used programs. The first exercise conducted by the NTP involved 44 chemicals [174]. The predictions made by Multi-CASE, TOPKAT, DEREK, and COMPACT were published in

advance of the bioassays being performed. The predictions were then compared to the results from the bioassays. The best prediction is given by DEREK, with a correction ratio of 59%. If we want to use effective filter of toxicity in virtual screening, the predicted precision of the current programs should be promoted greatly.

6. STRATEGIES FOR VIRTUAL SCREENING

When the crystal structure of receptor and a prepared 3-D database are available, virtual screening based on molecular docking can certainly be applied to obtain the best candidates. Usually, molecular docking is time-consuming, if the 3-D database, such as the ACDSC database, is very large, the cost of virtual screening will be very huge. Often, in a project, one can combine several approaches of virtual screening together, and reduce the screening scope gradually.

Figure 7 represents a typical strategy in virtual screening. First, one can use several ADME-concerned filters to reduce the size of the initial database. Common filtering protocols include “drug-like” property, Lipinski’s rule of five or ADME properties. Other physical filters could also be included, such as the number of rotatable bonds in molecules or polar surface areas. Additional filters are often applied at this stage to remove compounds containing specific chemical substructures associated with poor chemical stability or toxicity. All of these filters are computationally inexpensive and can be applied to very large databases.

Second, if the inhibitors of a protein are available, then some computational tools can be used, such as FlexS (tripos, <http://www.tripos.com>) or ROCS (OpenEye, <http://www.openeye.com>), to reduce the database size by shape-similarity with known inhibitors. FlexS can rapidly overlay rigid ligands to the template molecule and screen large lists of ligands that make use of rapid rigid scoring, followed by flexible alignment of the best scoring ligands.

Then, the chemical characteristics of the binding site can be analyzed to determine the functional group maps or hot spots for protein-ligand interactions, which can be used as receptor-based pharmacophore model. This pharmacophore model combined with some steric constraints from receptor can be used as query of UNITY for 3-D database screening. After screening, only these compounds with appropriate chemical features complementary to the receptor-based pharmacophore model and suitable steric features survived. The receptor-based pharmacophore model can be obtained from the analysis of the crystal complex structure or from computational programs such as GRID, MCSS, LUDI, SuperStar. Virtual screening based on receptor-based pharmacophore model is slower than virtual screening in the above two stages, but generally, it is also much faster than virtual screening based on molecular docking.

After the above three stages of virtual screening, the size of the initial database was greatly reduced, and this database can then be processed through atom-based molecular docking screening, using higher fidelity, but substantially slower, throughput methods.

In 2001, the group of Klebe *et al.* reported the *de novo* design of novel inhibitors against human carbonic anhydrase [87, 88]. In this project, the authors applied a protocol of

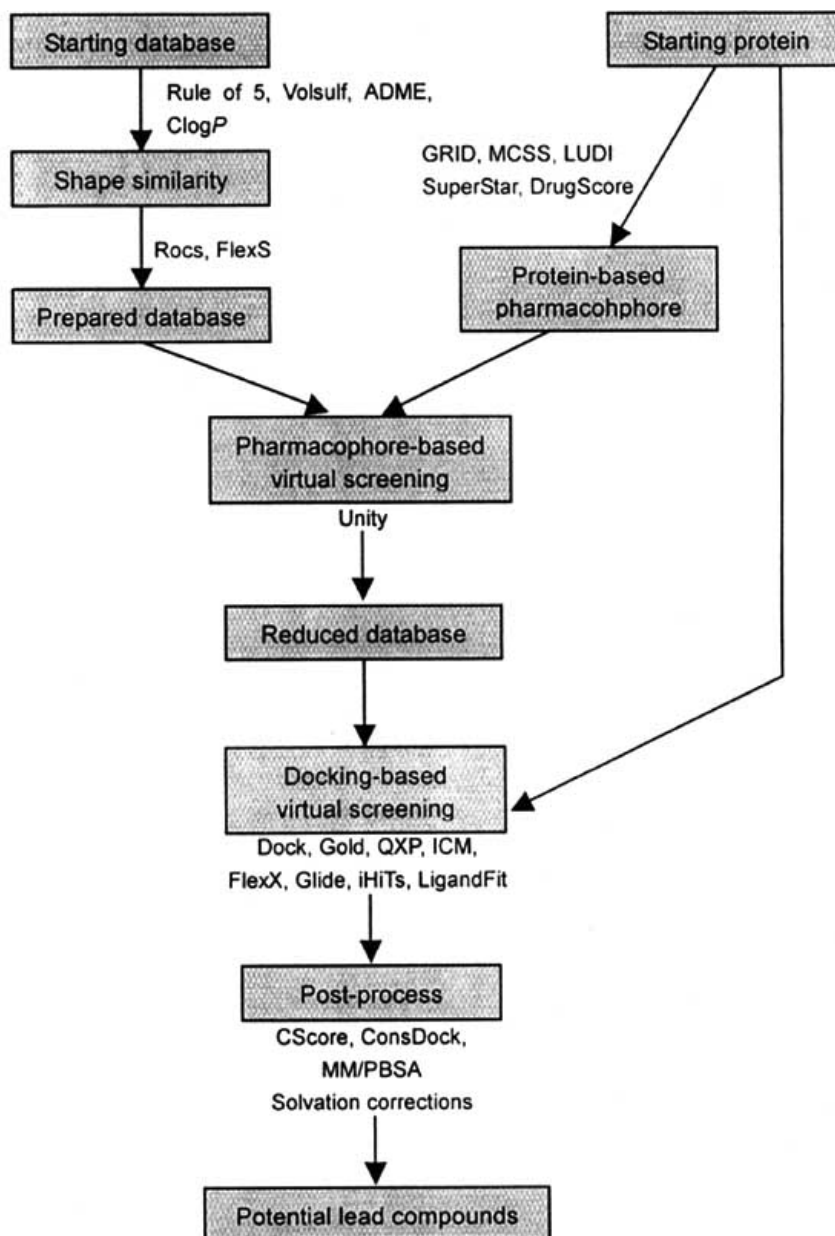


Fig. (7). Strategies for structure-based virtual screening.

several consecutive hierarchical screening protocols involving a preselection based on functional group requirements and fast pharmacophore matching (Fig. 8).

The search entries for virtual screening were taken from the Maybridge (61,186 entries) and the LeadQuest (37, 841 entries) databases. First, application 5-rules reduced this initial set to 98,850 entries. Then, UNITY search was performed to retrieve these compounds with possible zinc-binding groups, leading to a reduction to 5904 entries.

In the next step, the LUDI, GRID, SuperStar and DrugScore methods were applied to detect regions favorable for protein-ligand interactions. Flexible UNITY search were performed using the pharmacophore model based on the "hot spots" detected. Of the 5904 compounds selected in the first

step, only 3314 entries satisfied the pharmacophore query. Following, FlexS was used to superimpose all entries with two potent hCAII inhibitors and compute the similarity score with respect to these reference molecules. FlexS computed a superposition for 2237 of the previously selected 3314 compounds. In the final step, the 100 best-ranked hits from the FlexS filtering were docked into the binding pocket using FlexX. The final docked binding modes in hCAII were ranked for their expected binding affinity either by the statistically-fit scoring function implemented in FlexX or by DrugScore. Visual inspection of the suggested binding modes of FlexX, together with the scoring values of FlexX, FlexS, and DrugScore, was used to select a small set of compounds for ordering and subsequent biological testing. Finally, 13 compounds were selected for biological assays.

Of these 13 compounds, 3 compounds have the inhibitory activities at the level of subnanomolar, one is nanomolar, and while a further seven are micromolar inhibitors. The novelty of the discovered leads is best supported by the fact that a search in the patent literature showed the newly discovered subnanomolar compounds to comprise scaffolds not yet covered by existing patents.

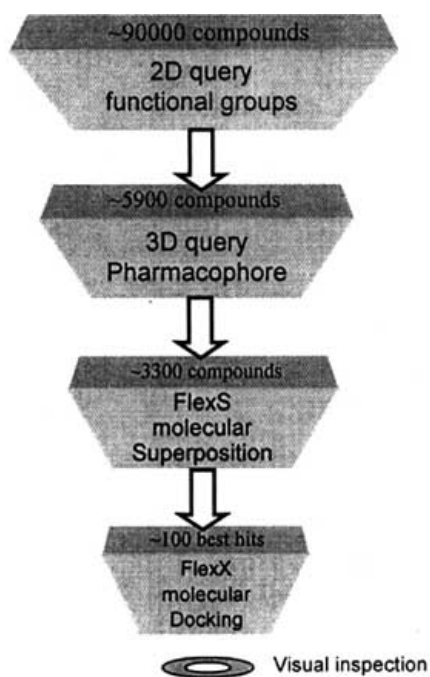


Fig. (8). The consecutive hierarchical screening protocols used in the design of inhibitors against human carbonic anhydrase.

Ultimately, virtual screening is knowledge-based and we should apply different strategy for different problem in hand. It is recommended to use different combinations of virtual screening techniques and consider the subset of the database that satisfied the defined criteria. Certainly, the usage of whether virtual screening techniques should be consistent with the available program and computational resources that can be used by researchers.

7. CONCLUSIONS AND OUTLOOK

This review gives a brief summary of recent development of techniques used in virtual screening, especially docking-based virtual screening, scoring functions and ADME-based virtual screening. Furthermore, successful examples of VS are provided, demonstrating that VS has become an essential part of tool to find new lead compounds for the pharmaceutical industry. The main issues still impeding the progress in this field include the quality of the scoring functions and ADME/Tox employed. In molecular docking or *de novo* design approach, scoring function is the central problem. At present, all docking programs and scoring functions have a tendency to generate a significant number of false positives. Although, this problem has been partially solved by using several scoring functions in parallel efforts to improve the quality of the scoring functions should be needed to progress this technology further. Current models

used to predict ADME/Tox are also very limited by low predictability. Further developments of virtual ADME/Tox are valuable for making this a reality and transforming more virtual molecules into real drugs.

ACKNOWLEDGEMENTS

This project is supported by National Natural Science Foundation of China (NSFC 29992590-2).

REFERENCES

References 176-178 are related articles recently published in Current Pharmaceutical Design.

- [1] Debouck C, Metdalf B. The impact of genomics on drug discovery. *Annu Rev Pharmacol Toxicol* 2000; 40: 193-207.
- [2] Hertzberg RP, Pope AJ. High-throughput screening: new technology for the 21st century. *Curr Opin Chem Biol* 2000; 4: 445-51.
- [3] Walters WP, Stahl MT, Murcko MA. Virtual screening-an overview. *Drug Discov Today* 1998; 3: 160-78.
- [4] Lyne PD Structure-based virtual screening: an overview. *Drug Discov Today* 2002; 7: 1047-1055.
- [5] Schneider G, Böhm HJ. Virtual screening and fast automated docking methods. *Drug Discov Today* 2002; 7: 63-70.
- [6] Gund P. Three-dimensional Pharmacophore Pattern Searching. *Prog Mol Subcell Biol* 1977; 5: 117-43.
- [7] Kurogi Y, Guner OF. Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr Med Chem* 2001; 8: 1035-55.
- [8] Langer T, Hoffmann RD. Virtual screening: an effective tool for lead structure discovery? *Curr Pharm Design* 2001; 7: 509-27.
- [9] CATALYST is available from Accelrys, 9685 Scranton Road, San Diego, CA 92121, US.
- [10] DISCO is available from Tripos Inc 1699 South Hanley Road, St Louis, MO 63144-2913, US.
- [11] GASP is available from Tripos Inc 1699 South Hanley Road, St Louis, MO 63144-2913, US.
- [12] Jones G, Willett P, Glen RC. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J Comput Aided Mol Des* 1995; 9: 532-49.
- [13] Structure Based Focusing is available from Accelrys, 9685 Scranton Road, San Diego, CA 92121, US.
- [14] Böhm HJ The computer program LUDI: A new method for the *de novo* design of enzyme inhibitors. *J Comp Aided Mol Des* 1992a; 6: 61-79.
- [15] Böhm HJ LUDI: rule based automatic design of new substituents for enzyme inhibitor leads. *J Comp Aided Mol Des* 1992b; 6: 593-606.
- [16] DISCO is available from Tripos Inc. 1699 South Hanley Road, St Louis, MO 63144-2913, USA.
- [17] ISIS Base, MDL Information Systems, Inc; San Leandro, CA, USA.
- [18] Chem-X version 96; Oxford Molecular Group, Inc; Hunt Valley, MD 21030, 2001
- [19] Fox T, Haaksma EEJ. Computer based screening of compound database: I preselection of benzamidine-based thrombin inhibitor. *J Comput Aided Mol Des* 2000; 14: 411-25.
- [20] Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 1985; 28: 849-57.
- [21] Hindle SA, Rarey M, Buning C, Lengauer T. Flexible docking under pharmacophore type constraints. *J Comput Aided Mol Des* 2002; 16: 129-49.
- [22] Brenk R, Naerum L, Grädler U, Gerber HD, Garcia GA, Reuter K, et al. Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on new unexpected binding model detected by crystal structure analysis. *J Med Chem* 2003; 46: 1133-43.
- [23] Sakamuri S, Enyedy IJ, Zaman WA, Tella SR, Kozikowski AP, Flippen-Anderson JL, et al. 2,3-disubstituted quinuclidines as a novel class of dopamine transporter inhibitors. *Bioorgan Med Chem* 2003; 11: 1123-35.
- [24] Enyedy IJ, Sakamuri S, Zaman WA, Johnson KM, Wang SM. Pharmacophore-based discovery of substituted pyridines as novel

- dopamine transporter inhibitors. *Bioorgan Med Chem Lett* 2003; 13: 513-17.
- [25] Koide Y, Tatsui A, Hasegawa T, Nurakami A, Satoh S, Yamada H, et al. Identification of a stable chymase inhibitor using a pharmacophore-based database search. *Bioorgan Med Chem Lett* 2003; 13: 25-9.
- [26] Singh J, van Vlijmen H, Liao YS, Lee WC, Cornebise M, Harris M, et al. Identification of potent and novel 4-1 antagonists using *in silico* screening. *J Med Chem* 2002; 45: 2988-93.
- [27] Koide Y, Hasegawa T, Takahashi A, Endo A, Mochizudi N, Nakagawa M, et al. Development of novel EDG3 antagonist using a 3-D database search and their structure-activity relationship. *J Med Chem* 2002; 45:4629-38.
- [28] Schweitzer BA, Loida PJ, CaJacob CA, Chott RC, Colantes EM, Hedge SG, et al. Discovery of imidazole glycerol phosphate dehydrate inhibitor through 3-D database searching. *Bioorgan Med Chem Lett* 2002; 12: 1743-46.
- [29] Yamazaki KT, Hasegawa H, Umekawa K, Ueki Y, Ohashi N, Kananoka M. Design, synthesis and biological activity and novel-non-peptidyl endothelin converting enzyme inhibitors, 1-phenyl-terazole-formaza analogues. *Bioorgan Med Chem Lett* 2002; 12: 1275-78.
- [30] Kurogi Y, Muiyata K, Okamura T, hashimoto K, Tsutsumi K, Nasu M, et al. Discovery of novel messangial cell proliferation inhibitors using a three-dimensional database searching method. *J Med Chem* 2001; 44: 2304-07.
- [31] Enyedy IJ, Zaman WA, Sakamuri S, Kozikowski AP, Johnson KM, Wang SM. Pharmacophore-based discovery of 3, 4-disubstituted pyrrolidines as a novel class of monoamine transporter inhibitors. *Bioorgan Med Chem Lett* 2001; 11: 1113-18.
- [32] Chen GS, Chang CS, Kan WM, Chang CL, Wang KC, Chern JW. Novel lead generation through hypothetical pharmacophore three-dimensional database searching: discovery of isoflavonoids as nonsteroidal inhibitors of Rat 5 -reductase. *J Med Chem* 2001; 44: 3759-63.
- [33] Huang P, Loew GH, Fonamizu H, Mimura M, Ishiyama N, Hayashida M, et al. Rational design, discovery, and synthesis of a novel series of potent growth hormone secretagogues. *J Med Chem* 2001; 44: 4082-91.
- [34] Hiramatsu Y, Tsukida T, Nakai Y, Inoue Y, Kondo H. Study on selection blocker 8 lead discovery of non-sugar antagonists using a 3D-pharmacophore model. *J Med Chem* 2000; 43: 1476-83.
- [35] Langer T, Hoffmann A. Virtual screening: an effective tool for lead structure discovery? *Curr Pharm Des* 2001; 7: 509-27.
- [36] Murcko M. Recent advances in ligand design methods; *Reviews in Computational Chemistry Vol 11*; Edited by Lipkowitz KB and Boyd DB Wiley-VCH inc, 1997; 1-66.
- [37] LUDI is available from Accelrys, 9685 Scranton Road, San Diego, CA 92121, USA.
- [38] Leapfrog is available from Tripos Inc 1699 South Hanley Road, St Louis, MO 63144-2913, USA.
- [39] Sprout is available from SimBioSym In 135 Queen's Plate Drive Unit 355 Toronto, Ontario M9W 6V1, Canada.
- [40] Gillet VJ, Myatt G, Zsoldos Z, Johnson AP. SPROUT, HIPPO and CAESA: Tools for *de novo* structure generation and estimation of synthetic accessibility. *Perspect Drug Discov Design* 1995; 3: 34-50.
- [41] Mata P, Gillet VJ, Johnson AP, Lampreia J, Myatt GJ, Sike S, et al. SPROUT-3D structure generation using templates. *J Chem Inf Comput Sci* 1995; 35: 479-93.
- [42] Lauri G, Barlett PA. CAVEAT: a program to facilitate the design of organic molecules. *J Comput Aided Mol Des* 1994; 8:51-66.
- [43] Ho CMW, Marshall GR. Cavity search: an algorithm for the isolation and display of cavity-like binding regions. *J Comput Aided Mol Des* 1990; 4: 337-54.
- [44] Ho CM, Marshall GR. FOUNDATION: a program to retrieve all possible structures containing a user-defined minimum number of matching query elements from three-dimensional databases. *J Comput Aided Mol Des* 1993; 7: 623-47.
- [45] Ho CM; Marshall GR. DBMAKER: a set of programs designed to generate three-dimensional databases based upon user-specific criteria. *J Comput Aided Mol Des* 1995; 9: 65-86.
- [46] Ho CM, Marshall GR. SPLICE: a program to assemble partial query solution from three-dimensional database searches into novel ligands. *J Comput Aided Mol Des* 1993; 7: 623.
- [47] Lawrence MC, Davis PC. CLIX: a search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure. *Proteins: Struct Funct Genet* 1992; 12: 31-41.
- [48] Pearlman DA, Murcko MA. CONCEPTS: new dynamics algorithm for *de novo* drug suggestions. *J Comput Aided Mol Des* 1993; 14: 1184.
- [49] Pearlman DA, Murcko MA. CONCERTS: dynamics connection of fragments as an approach to *De novo* ligand design. *J Comput Aided Mol Des* 1996; 39: 1651-63.
- [50] Miranker A, Karplus M. An automated-method for dynamic ligand design. *Proteins: Struct Funct Genet* 1995; 23: 472-90.
- [51] Rotstein SH, Murcko MA. GenStar: a method for *de novo* drug design. *J Comput Aided Mol Des* 1993; 7: 23-43.
- [52] Rotstein SH, Murcko MA. GroupBuild: a fragment-based method for *de novo* drug design. *J Med Chem* 1993; 36: 1700-10.
- [53] Moon J, Howe W. Computer design of bioactive molecules: a method for receptor-based *de novo* ligand design. *Proteins* 1991; 11: 314-28.
- [54] Bohacek RS, McMartin C. Multiple highly diverse structures complementary to enzyme binding sites: results of extensive application of a *de novo* design method incorporating combinatorial growth. *J Am Chem Soc* 1994; 116: 5560-71.
- [55] Eisen MB, Wiley DC, Karplus M, Hubbard RE. HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins* 1994; 19: 199-221.
- [56] Douguet D, Thoreau E, Grassy GA. Genetic algorithm for the automated generation of small organic molecules: drug design using an evolutionary algorithm. *J Comput Aided Mol Des* 2000; 14: 449-66.
- [57] Nishibata Y, Itai A. Automatic creation of drug candidate structures based on receptor structure-starting point for artificial lead generation. *Tetrahedron* 1991; 47: 8985-90.
- [58] Nishibata Y, Itai A. Confirmation of usefulness of a structure construction program based on three-dimensional receptor structure for rational lead generation. *J Med Chem* 1993; 36: 2921-28.
- [59] Gehlaar DK. *De novo* design of enzyme inhibitors by Monte Carlo ligand generation. *J Med Chem* 1995; 38: 466-72;
- [60] Miranker A, Karplus M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* 1991; 11: 29-34.
- [61] Tschinke V, Cohen NC. The NEWLEAD program: a new method for the design for the design of candidate structures from pharmacophoric hypotheses. *J Med Chem* 1993; 36: 3863-70.
- [62] Clark DE, Frenkel D, Levy SA, Li J, Murray CW, Robson B, et al. PRO_LIGAND: an approach to *de novo* drug design. *J Comput Aided Mol Des* 1995; 9: 13-32.
- [63] DeWitte R. SMOG, *de novo* design method based on simple, fast, and accurate free energy estimates 2 case studies in molecular design. *J Am Chem Soc* 1997; 119: 4608-17.
- [64] Schneider G, Lee ML, Stahl M, Schneider P. *De novo* design of molecular architectures by evolutionary assembly of drug-derived building blocks. *J Comput Aided Mol Des* 2000; 14: 487-94.
- [65] Ji TT, Zhang WN, Zhang M, Kudo M, Aoyama Y, Yoshida Y, et al. Structure-based *de novo* design, synthesis, and biological evaluation of non-azole inhibitors specific for lanosterol 14 -demethylase of fungi. *J Med Chem* 2003; 46: 474-85.
- [66] Liebeschuetz JW, Jones SD, Morgan PJ, Murray CW, Rimmer AD, Roscoe JME, et al. PRO_SELECT: combining structure-based drug design and array-based chemistry for rapid lead discovery 2 the development of a series of highly potent and selective factor Xa inhibitors. *J Med Chem* 2002; 45: 1221-32.
- [67] Honma T, Hayashi K, Aoyama T, Hashimoto N, Machida T, Fukasawa K, et al. Structure-based generation of a new class of potent Cdk4 inhibitors: new *de novo* design strategy and library design. *J Med Chem* 2001; 44: 4615-27.
- [68] Boehm HJ, Boehringer M, Bur D, Gmuender H, Huber W, Klaus W, et al. Novel inhibitors of DNA gyrase: 3-D structure based biased needle screening, hit validation by biophysical methods, and 3-D guided optimization A promising alternative to random screening. *J Med Chem* 2000; 43: 2664-74
- [69] Hou TJ, Wang JM, Chen LR, Xu XJ. Automatic docking of peptides and proteins using a hybrid method combined with genetic algorithm and tabu search. *Protein Eng* 1999; 12: 639-47.

- [70] Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 40: search strategies for automated molecular docking of flexible molecule database. *J Comput Aided Mol Des* 2001; 15: 411-28.
- [71] Zsoldos Z, Johnson AP, Simon A, Szabo I, Szabo Z. eHITS: Novel algorithm for fast, exhaustive flexible ligand docking and scoring. Presentation at the: 224th ACS National Meeting, Boston, MA, USA, 2002.
- [72] Pang YP, Perola E, Xu K, Predergast FG. EDUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J Comput Chem* 2001; 22: 1750-71.
- [73] Rarey M, Kramer B, Lengauer T, Klebe GA. Fast Flexible Docking Method using an Incremental Construction Algorithm. *J Mol Biol* 1996; 261: 470-89.
- [74] ICM is available from MolSoft LLC, 3366 North Torrey Pines Court Suite 300 La Jolla, CA 92037, USA.
- [75] Venkatachalam CM, Jiang X, Oldfield T, Waldman M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J Mol Graphics Mol Mod* 2003; 21: 289-307.
- [76] Fred is available from OpenEye Inc, 3600 Cerrillos Rd, Suite 1107 Santa Fe, NM 87507, USA.
- [77] FlexiDock is available from Tripos Inc 1699 South Hanley Road, St Louis, MO 63144-2913, USA.
- [78] Schnecke V, Kuln LA. Virtual screening with solvation and ligand-induced complementarity. *Perspect Drug Des Discov* 2000; 20: 171-90.
- [79] Affinity is available from Accelrys, 9685 Scranton Road, San Diego, CA 92121, USA.
- [80] Morris GM, Goodsell DS, Huey R, Olson AJ. Distributed automated docking of flexible ligands to proteins: parallel applications of Autodock 2.4. *J Comput Aided Mol Des* 1996; 10: 293-304.
- [81] Glide is available from Schrödinger, 1500 SW First Avenue, Suite 1180, Portland.
- [82] Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J Mol Biol* 1997; 267: 727-48.
- [83] McMartin C, Bohacek RS. QXP: powerful, rapid computer algorithms for structure-based drug design. *J Comput Aided Mol Des* 1997; 11: 333-44.
- [84] Hou TJ, Zhu LL, Chen LR, Xu XJ. Mapping the binding site of a large set of quinazoline type EGF-R inhibitors using molecular field analyses and molecular docking studies. *J Chem Inf Comput Sci* 2003; 43: 273-87.
- [85] 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* 2002; 46: 34-40.
- [86] Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol* 2001; 308: 377-95.
- [87] Grünber S, Stubbs M, Klebe G. Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J Med Chem* 2002; 45: 3588-02.
- [88] Grüneberg S, Wendt B, Klebe G. Subnanomolar Inhibitors from Computer Screening: A Model Study Using Human Carbonic Anhydrase II. *Angew Chem Int Ed Engl* 2001; 40: 389-93.
- [89] Power RA, Morandi F, Shoichet BK. Structure-based discovery of a novel nonvalent inhibitor of AmpC β -lactamase. *Structure* 2002; 10: 1013-23.
- [90] Iino M, Furugori T, Mori T, Moriyama S, Fukuzawa A, Shibano T. Rational design of new lead compounds: structures for selective ARK1 inhibitors. *J Med Chem* 2002; 45: 2150-59.
- [91] Costi MP, Tondi D, Rinaldi M, Barlocco D, Pecorari P, Soragni F, et al. Structure-based studies on species-specific inhibition of thymidylate synthase. *BBA-Mol Basis Dis* 2002; 1587: 206-14.
- [92] Rastelli G, Ferrari AM, Costantino L, Gamberini MC. Discovery of new inhibitors of aldose reductase from molecular modeling and database searching. *Bioorgan Med Chem* 2002; 10: 1437-50.
- [93] Pang YP, Xu K, Kollmeyer TM, Perola E, McGrath WJ, Green DT, et al. Discovery of a new inhibitor lead of adenovirus proteinase: steps towards selective, irreversible inhibitors of cysteine proteinases. *FEBS Lett* 2001; 502: 93-97.
- [94] Enyedy I, Ling Y, Nacro K, Tomita Y, Wu XH, Cao YY, et al. Discovery of small molecule inhibitors of Bcl-2 through structure-based computer screening. *J Med Chem* 2001; 44: 4313-24.
- [95] Iwata Y, Arisawa M, Hamada R, Kita Y, Mizutani MY, Tomioda N, et al. Discovery of novel aldose reductase inhibitors using a protein structure-based approach: 3D-database search followed by design and synthesis. *J Med Chem* 2001; 44: 1718-28.
- [96] Enyedy IJ, Lee SL, Kuo AH, Dickson RB, Lin CY, Wang S. Structure-based approach for the discovery of bis-benzamides as novel inhibitors of matriptase. *J Med Chem* 2001; 44: 1349-55.
- [97] Schapira M, Raaka BM, Samuels HH, Abagyan R. Rational discovery of novel nuclear hormone receptor antagonists. *Proc Natl Acad Sci USA* 2000; 97: 1008-13.
- [98] Paiva AM, Vandervall DE, Blanchard JS, Kozarich JW, Williamson JM, Kelly TM. Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of mycobacterium tuberculosis. *BBA* 2001; 1545: 67-77.
- [99] Perola E, Xu K, Kollmeyer TM, Kaufmann SH, Predergast FG, Pang YP. Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J Med Chem* 2000; 43: 401-08.
- [100] Sarmiento M, Wu L, Keng YF, Song L, Luo ZW, Huang ZW, et al. Structure-based discovery of small inhibitors targeted to protein tyrosine phosphatase 1B. *J Med Chem* 2000; 3: 146-55.
- [101] Hopkins SC, Vale RD, Kuntz ID. Inhibitors of kinesin activity from structure-based computer screening. *Biochemistry* 2000; 39: 2805-18.
- [102] Aronov AM, Munagala NR, Ortizde Montellano PR, Kuntz ID, Wang CC. Rational design of selective submicromolar inhibitors of tritrichomonas foetus hypoxanthine-guanine-xanthine phosphoribosyltransferase. *Biochemistry* 2000; 39: 4684-91.
- [103] Freymann DM, Wenck MA, Engel JC, Feng J, Focia PJ, Eakin AE, et al. Efficient identification of inhibitors targeting the close active site conformation of the HPRT from trypanosoma cruzi. *Chem Biol* 2000; 7: 957-68.
- [104] Makino S, Kayahara T, Tashiro K, Takahashi M, Tsuji T, Shoji M. Discovery of a novel serine protease inhibitor utilizing a structure-based and experimental selection of fragments technique. *J Comput Aided Mol Des* 2000; 15: 553-59.
- [105] Filikov AV, Mohan V, Vickers TA, Griffey RH, Cook PD, Abagyan RA, et al. Identification of ligands for RNA targets via structure-based virtual screening: HIV TAR. *J Comput-Aided Mol Des* 2000; 14: 593-610.
- [106] Schapira M, Raaka BM, Samuels HH, Abagyan R. Rational discovery of novel nuclear hormone receptor antagonists. *Proc Natl Acad Sci USA* 2000; 93: 1008-13.
- [107] Gohlke H, Klebe G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed* 2002; 41: 2644-76.
- [108] Grzybowski BA, Ishchenko AV, Shimada J, Whitesides GM, Shakhnovich EI. From knowledge-based potentials to combinatorial lead design *in silico*. *Acc Chem Res* 2002; 5: 261-69.
- [109] Ishchenko AV, Shakhnovich EI. Small molecule growth (SMoG2001): an improved knowledge-based scoring function for protein-ligand interactions. *J Med Chem* 2002; 45: 2770-80.
- [110] Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem* 1999; 42: 791-804.
- [111] Mitchell JBO, Laskowski RA, Alex A, Thornton JM. BLEEP-potential of mean force describing protein-ligand interactions: I Generating potential. *J Comput Chem* 1999; 20: 1165-76.
- [112] Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* 2000; 20: 337-56.
- [113] Gohlke H, Hendlich M, Klebe G. Predicting binding modes, binding affinities and 'hot spots' for protein-ligand complexes using a knowledge-based scoring function. *Perspect Drug Discovery Des* 2000; 20: 115-44.
- [114] Verkhiver G, Appelt K, Freer ST, Villafranca JE. Empirical free energy calculations of protein-ligand crystallographic complexes I knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Eng* 1995; 8: 677-91.
- [115] Charifson PS, Corkery JJ, Murcko MA, Walters WP. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* 1999; 42: 5100-09.
- [116] Hendlich M. Databases for protein-ligand complexes. *Acta Crystallogr* 1998; D54: 1178-82.

- [117] Sotriffer CA, Gohlke H, Klebe G. Docking into knowledge-based potential fields: a comparative evaluation of DrugScore. *J Med Chem* 2002; 45: 1967-70.
- [118] Ischchenko AV, Shakhnovich EI. Small molecular growth 2001 (SMoG2001): an improved knowledge-based scoring function for protein-ligand interactions. *J Med Chem* 2002; 45: 2770-80.
- [119] Shoichet BK, Leach AR, Kuntz ID. Ligand solvation in molecular docking. *Proteins* 1999; 34: 4-16.
- [120] Zou X, Sun Y, Kuntz ID. Inclusion of solvation in ligand binding free energy calculations using the generalized-Born model. *J Am Chem Soc* 1999; 121: 8033-43.
- [121] Donini OAT, Kollman PA. Calculation and prediction of binding free energies for the matrix metalloproteinases. *J Med Chem* 2000; 43: 4180-88.
- [122] Kuhn B, Kollman PA. Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J Med Chem* 2000; 43: 3786-91.
- [123] Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J Am Chem Soc* 1998; 120: 9401-509.
- [124] Kollman PA, Massova I, Reyes C, Kuhn B, Huo SH, Chong L, et al. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc Chem Res* 2000; 33: 889-97.
- [125] Wang JM, Morin P, Wang W, Kollman PA. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J Am Chem Soc* 2001; 123: 5221-30.
- [126] Charifson PS, Corkery JJ, Murcko MA, Walters WP. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* 1999; 42: 5100-09.
- [127] CScore is available from Tripos Inc 1699 South Hanley Road, St Louis, MO 63144-2913, USA.
- [128] Jones G, Willett P, Glen R, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 1997; 267: 727-48.
- [129] Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *J Med Chem* 1999; 42: 791-804.
- [130] Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 1982; 161: 269-88.
- [131] Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions 1 The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 1997; 11: 425-45.
- [132] Rarey M, Kramer B, Lengauer T, Klebe GA. fast flexible docking method using an incremental construction algorithm. *J Mol Biol* 1996; 261: 470-89.
- [133] Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB. Consensus scoring for ligand/protein interactions. *J Mol Graph Mod* 2002; 20: 281-95.
- [134] Paul N, Rognan D, ConsDock: A new program for the consensus analysis of protein-ligand interactions. *Proteins* 2002; 47: 521-33.
- [135] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997; 23: 3-25.
- [136] Lipinski CA. Druglike properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 2000; 44: 235-49.
- [137] Stenberg P, Luthman K, Artursson P. Virtual screening of intestinal drug permeability. *J Control Release* 2000; 65: 231-43.
- [138] Stenberg P, Norinder U, Luthman K, Artursson P. Experimental and computational screening models for the prediction of intestinal drug absorption. *J Med Chem* 2001; 44: 1927-37.
- [139] Chaturvedi PR, Decker CJ, Odinecs A. Prediction of pharmacokinetic properties using experimental approaches during early drug discovery. *Curr Opin Chem Bio* 2001; 5: 452-63.
- [140] Ekins S, Ring BJ, Grace J, McRobie-Belle DJ, Wrighton SA. Present and further *in vitro* approaches for drug metabolism. *J Pharmacol Toxicol Methods* 2000; 44: 313-24.
- [141] Ekins S, Waller CL, Swaan PW, Cruciani G, Wrighton SA, Wikel JH. Progress in predicting human ADME parameters *in silico*. *J Pharmacol Toxicol Methods* 2001; 44: 251-272.
- [142] Ekins S, Wrighton SA. Application of *in silico* approaches to predicting drug-drug interactions. *J Pharmacol Toxicol Methods* 2001; 45: 65-9.
- [143] Beresford AP, Selick HE, Tarbit MH. The emerging importance of predictive ADME simulation in drug discovery. *Drug Disc Today* 2002; 7: 109-16.
- [144] Boobis A, Gundert-Remy U, Kremers P, Macheras P, Pelkonen O. *In silico* prediction of ADME and pharmacokinetics - Report of an expert meeting organised by COST B15. *Eur J Pharm Sci* 2002; 17: 183-93.
- [145] van de Waterbeemd H, Gifford E. ADMET *in silico* modeling: towards prediction paradise? *Nat Rev Drug Discov* 2003; 2: 192-204.
- [146] Cruciani G, Crivori P, Carrupt PA, Testa B. Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. *Theochem* 2000; 53: 17-30.
- [147] Hou TJ, Xu XJ. ADME Evaluation in Drug Discovery: applications of genetic algorithms on the prediction of blood-brain partitioning of a large set drugs. *J Mol Model* 2002; 8: 337-49.
- [148] Wessel MD, Jurs PC, Tolan JW, Muskal SM. Prediction of human intestinal absorption of drug compounds from molecular structure. *J Chem Info Comp Sci* 1998; 38: 726-35.
- [149] Doniger S, Hofmann T, Yeh J. Predicting CNS permeability of drug molecules: comparison of neural network and support vector machine algorithms. *J Comput Biol* 2002; 9: 849-64.
- [150] Ghuloum AM, Sage CR, Jain AN. Molecular hashkeys: a novel method for molecular characterization and its application for predicting important pharmaceutical properties of molecules. *J Med Chem* 1999; 42: 1739-48.
- [151] Yoshica F, Topliss JG. QSAR model for drug human oral bioavailability. *J Med Chem* 2000; 43: 2575-85.
- [152] Andrews CW, Bennett L, Yu LX. Predicting human oral bioavailability of a compound: development of a novel quantitative structure-bioavailability relationship *Pharm Res* 2000; 17: 639-44.
- [153] Bains W, Gilbert R, Sviridenko L, Gascon JM, Scoffin R, Birchall K, et al. Evolutionary computational methods to predict oral bioavailability QSPRs. *Curr Opin Drug Disc Dev* 2002; 5: 44-51.
- [154] Mandagere AK, Thompson TN, Hwang KK. A graphical model for estimating oral bioavailability of drugs in humans and other species from their Caco-2 permeability and *in vitro* liver enzyme metabolic stability rates. *J Med Chem* 2002; 45: 304-11.
- [155] Ayrton A, Morgan P. The role of transport proteins in drug absorption, distribution and excretion. *Xenobiotica* 2001; 31: 469-97.
- [156] Wiese M, Pajeva IK. Structure-activity relationships of multidrug resistance reversers. *Curr Med Chem* 2001; 8: 685-713.
- [157] Seelig A, Landwojtowicz E. Structure-activity relationship of P-glycoprotein substrates and modifiers. *Eur J Pharm Sci* 2000; 12: 31-40.
- [158] Österberg T, Norinder U. Theoretical calculation and prediction of P-glycoprotein-interacting drugs using MolSurf parametrization and PLS statistics. *Eur J Pharm Sci* 2000; 10: 295-303.
- [159] Ekins S, Kim RB, Leake BF, Dantzig AH, Schuetz EG, Lan LB, et al. Three-dimensional quantitative structure-activity relationships of inhibitors of P-glycoprotein. *Mol Pharmacol* 2002; 61: 964-73.
- [160] Penzotti JE, Lamb ML, Evensen E, Grootenhuys PDJ. A computational ensemble pharmacophore model for identifying substrates of P-glycoprotein. *J Med Chem* 2002; 45: 1737-40.
- [161] Pajeva IK, Wiese M. Pharmacophore model of drugs involved in P-glycoprotein multidrug resistance: explanation of structural variety (hypothesis). *J Med Chem* 2002; 45: 5671-86.
- [162] Ekins S, Bravi G, Binkley S, Gillespie JS, Ring BJ, Wikel JH, et al. Three- and four-dimensional quantitative structure-activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors. *Drug Metab Dispos* 2000; 28: 994-1002.
- [163] Ekins S, De Groot M, Jones JP. Pharmacophore and three-dimensional quantitative structure-activity relationship methods for modelling cytochrome P450 active sites. *Drug Metab Dispos* 2001; 29: 936-44.
- [164] Cruciani G. Abstract presented at Euro QSAR, 2002.
- [165] Zuegge J, Fechner U, Roche O, Parratt NJ, Engkvist O, Schneider G. A fast virtual screening filter for cytochrom450 3A4 inhibition liability of compound libraries. *QSAR* 2002; 21: 249-56.

- [166] Langowski J, Long A. Computer systems for the prediction of xenobiotic metabolism. *Adv Drug Delivery Rev* 2002; 54: 407-15.
- [167] Ehrhardt PW. *Drug Metabolism Databases and High-Throughput Testing During Drug Design and Development*, IUPAC, Blackwell Science, Malden, Massachusetts, 1999.
- [168] VolSurf is available from Tripos Inc 1699 South Hanley Road, St Louis, MO 63144-2913, USA.
- [169] Egan WJ, Merz KM Jr, Baldwin JJ. Prediction of Drug Absorption Using Multivariate Statistics. *J Med Chem* 2000; 43: 3867-77.
- [170] C2ADME is available from Accelrys, 9685 Scranton Road, San Diego, CA 92121, USA.
- [171] Duffy EM, Jorgensen WL. Prediction of properties from simulations: Free energies of solvation in hexadecane, octanol, and water. *J Am Chem Soc* 2000; 122: 2878-88.
- [172] Jorgensen WL, Duffy EM. Prediction of drug solubility from Monte Carlo simulations. *Bioorg Med Chem Lett* 2000; 10: 1155-58.
- [173] QikProp is available from Schrödinger, 1500 SW First Avenue, Suite 1180 Portland, OR 97201, USA.
- [174] Greene N. Computer systems for the prediction of toxicity: an update. *Adv Drug Deliv Rev* 2002; 54: 417-31.
- [175] Durham SK, Pearl GM. Computational methods to predict drug safety liabilities. *Drug Disc Dev* 2001; 4: 110-15.
- [176] Jiang S, Zhao Q, Debnath AK. Peptide and non-peptide HIV fusion inhibitors. *Curr Pharm Design* 2002; 8(8): 563-80.
- [177] Langer T, Hoffmann RD. Virtual screening: an effective tool for lead structure discovery? *Curr Pharm Design* 2001; 7(7): 509-27.
- [178] Mason JS, Good AC, Martin EJ. 3-D pharmacophores in drug discovery. *Curr Pharm Design* 2001; 7(7): 567-97.