Send Orders of Reprints at reprints@benthamscience.org

Virtual Screening Methods as Tools for Drug Lead Discovery from Large Chemical Libraries

X.H. Ma[1,2], F. Zhu[1], X. Liu[2], Z. Shi[2], J.X. Zhang[2], S.Y. Yang[1], Y.Q. Wei[1] and Y.Z. Chen*[,1,2]

[1]*State Key Laboratory of Biotherapy, Sichuan University, Chengdu 610064, P.R.China;* [2]*Bioinformatics and Drug Design Group, Department of Pharmacy and Department of Computational Science, National University of Singapore, Blk S16, Level 8, 3 Science Drive 2, Singapore 117543*

Abstract: Virtual screening methods have been developed and explored as useful tools for searching drug lead compounds from chemical libraries, including large libraries that have become publically available. In this review, we discussed the new developments in exploring virtual screening methods for enhanced performance in searching large chemical libraries, their applications in screening libraries of ~ 1 million or more compounds in the last five years, the difficulties in their applications, and the strategies for further improving these methods.

Keywords: Machine learning, molecular docking, pharmacophore, quantitative structure activity relationship, similarity searching, support vector machines.

INTRODUCTION

High-throughput screening (HTS) [1] and virtual screening (VS) [2-6] methods have been developed for drug lead discovery from increasingly larger chemical space. In particular, VS complements HTS in improving hit rates and chemical space coverage [7]. The chemical libraries have been expanded to >1 million purchasable compounds [8], >30 million compounds [9] and >1 million bioactive compounds [10]. VS methods, particularly the combination of VS methods, have been increasingly explored for searching drug leads from large compound libraries at good yields and sufficiently low false hit rates so that the subsequent synthesis and testing tasks are at manageable levels. In this review, we discuss the new developments in exploring VS methods for screening large libraries at enhanced performances, their applications for screening large libraries of ~ 1 million or more compounds in the last five years, the difficulties in their applications, and the strategies for further improving these methods.

1. BRIEF OVERVIEW OF EXTENSIVE USED VIRTUAL SCREENING METHODS

The most extensively used VS methods are docking [11], pharmacophore [12], quantitative structure activity relationship (QSAR) [2, 3, 13], similarity searching [5], and machine learning [14]. These methods are often divided into structure-based virtual screening (SBVS) and ligand-based virtual screening (LBVS) depending on what is already known about a target and its ligands Fig. (**1**). Structure-based virtual screening involves docking of candidate ligands into a protein target followed by applying a scoring function to estimate the likelihood that the ligand will bind to the protein with high affinity. Ligand-based virtual screening applies computational descriptors of molecular structure, properties, or pharmacophore features to analyze relationships between active templates and compounds from chemical libraries. Some available annotated chemical libraries which are the sources of active compounds and molecular properties are listed in (Table **1**).

Docking method identifies active compounds by geometrically docking molecules to a pre-selected target site followed by binding configuration optimization and scoring [4, 6]. This method accurately describes molecular binding and requires no knowledge of known active compounds, allowing for an intuitive understanding of binding process and enabling the identification of novel active compounds with different binding modes against those of the known ligands [2, 15]. But it may have limited capability in modelling target structural flexibility [16] and salvation and entropic effects [17]. These limitations may be partly overcome by more comprehensive conformational sampling, improvement of scoring algorithms, and adequate modelling of solvation and entropic effects, but the increased computational cost may limit its use for virtual screening.

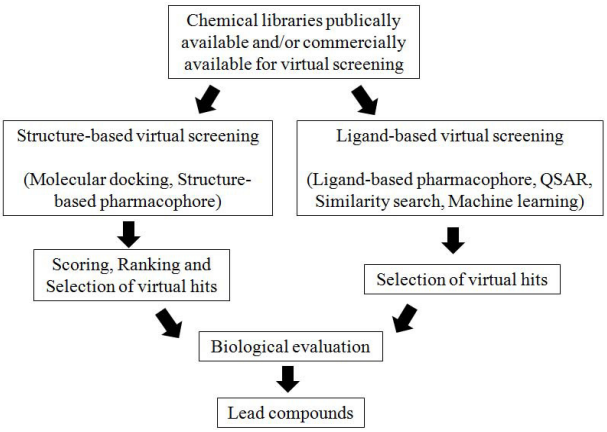

Fig. (1). The general workflow of structure-based virtual screening and ligand-based vrtual screening.

Pharmacophore method identifies active compounds by matching molecules to an assembly of steric and physicochemical features necessary for activity [17]. Pharmacophoric features can be derived by either ligand-based or structure-based methods [2]. Ligand-based methods generate pharmacophoric features by superposing multiple active molecules to extract essential features. Structure-based methods construct pharmacophoric features by probing possible interaction points between target and ligands. Pharmacophore method complements docking method by its ability to select active compounds of higher structural flexibility, but tends to be less effective in modelling detailed binding interactions than docking method. The performance of pharmacophore models is affected by sensitivity to training datasets [18], multiple choices of features [3], quality of conformational sampling and molecular overlay, anchoring points selection, and binding affinity estimation [3].

*Address correspondence to this author at the State Key Laboratory of Biotherapy, Sichuan University, Chengdu 610064, P.R.China; Tel: 65-6874-6877; Fax: 65-6774-6756; E-mail: phacyz@nus.edu.sg

1875-533X/12 $58.00+.00

QSAR method identifies active compounds by estimating their activities from a statistically significant correlation between molecular structures and activities [13, 19, 20]. Molecular structures are typically represented by specific sets of structural and physicochemical properties (molecular descriptors) relevant to their activities. QSAR reportedly outperforms docking in some VS tasks such as biogenic amine binding to G-protein coupled receptors [5]. The quality of QSAR models is affected by concept compatibility, representativeness of structure-activity data, influence of data outliers, fitness of developed quantitative relationships, starting geometry in 3D-QSAR, and multiple choices of solutions [14].

Similarity searching method identifies active compounds by measuring the level of their structural similarity to the known active compounds [16, 21, 22]. In this method, compounds are typically represented by molecular fingerprints [16, 21, 22] and in some studies by molecular descriptors [22], and molecular structural similarity are frequently determined by Tanimoto coefficient [4, 22, 23]. Similarity searching method is highly effective and fast, often producing superior VS performance compared to docking [22], but tends to be limited by the requirement of structural or sub-structural similarity to the known active compounds.

Machine learning methods such as binary kernel discrimination (BKD) [23, 24] and support vector machines (SVM) [25, 26] identify active compounds by statistical analysis of intrinsic correlations between activities and the structural and physicochemical profiles of known active and inactive compounds [27, 28]. Machine learning methods utilize nonlinear supervised learning algorithms to develop statistical models capable of predicting a more diverse spectrum of structural and physicochemical properties than conventional QSAR models [29] at high-CPU speed (100K data points per hour on 3GHz PC) [28], which is particularly useful for screening large compound libraries and for identifying novel scaffolds to complement docking, pharmacophore, QSAR and similarity searching methods [30, 31]. The performance of machine learning methods depends on such factors as training set diversity, their ability to deal with imbalanced datasets (inactive compounds typically outnumber active compounds), and parameter ranges in covering active and inactive chemical space [29].

2. VIRTUAL SCREENING METHODS USED IN THE SCREENING OF LARGE CHEMICAL LIBRARIES

In principle, the VS methods described in the previous section can be individually used for screening large chemical libraries. The significantly large library size places much higher demand on the screening speed and low false hit rates than those of screening small and medium sized libraries. Therefore, these VS methods have been typically combined with additional filters/constraints or among themselves for searching large chemical libraries to take their complementary advantages in screening speed and hit selection strategies. The literature-reported performance of these methods in screening large chemical libraries of >1 million compounds in the last 5 years is summarised in (Table **2**).

Machine learning and similarity searching methods have been extensively used for screening large chemical libraries because of their speed, good yields and low false hit rates. SVM VS models for eight target classes muscarinic acetylcholine receptor subtype 1 (M1 receptor) agonists, N-methyl-D-aspartate receptor (NMDA receptor) antagonists, thrombin inhibitors, human immunodeficiency virus (HIV) protease inhibitors, cephalosporins, and renin inhibitors [32], Bcr-Abl tyrosine-kinase inhibitors [23] and histone deacetylase inhibitors [33] have been tested in in-silico screening of 9.997-13.56 million compounds. These SVM VS models have shown good to moderate yields (17.4%-92.4%) and low false hit rates (0.11%-2.9%). The performance of SVM may be further improved by adding a second layer SVM model that is developed based on the feedback of testing results of the initial SVM model [34]. A two-layer SVM model for searching AR ligands has identified 500 virtual hits from 19 million compounds, with 8 of the 12 purchasable virtual hits tested active [34]. In another case, a two-layer similarity searching model for phosphodiesterase type 5 inhibitors (the first and second layer uses a conventional and a refined similarity searching algorithm respectively) has identified 137 virtual hits from 25 million compounds with 25 hits tested active [35].

In the filter approach, rule-based filers such as drug-like filters (e.g. Lipinski's rule of five (RO5) [36], rotatable bond count, solubility), SciTegic HTS filter [37], and substructure filters have been used for pre-selecting compounds prior to the screening by a VS method. For instance, in identifying c-Met tyrosine kinase inhibitors, 13.5 million compounds have been screened by RO5 to pre-select 600,000 drug-like compounds, followed by docking to identify 175 virtual hits, with 3 of the 70 purchasable virtual hits tested active [38]. In identifying Glutamate Transporter 1 inhibitors from 26.4 million compounds, substructure and rotatable bond count filters have been used to pre-select 252,311 compounds, followed by docking to identify 5176 top-ranked compounds, with 3 of the 14 synthesized top-ranked compounds tested active [39].

The constraint approach targets selected compounds based on such structural constraints as ligand-centered binding-site and binding-modes selection in docking, applicability domain in QSAR, and excluded volumes in pharmacophore methods. In docking search of c-Met tyrosine kinase inhibitors from 13.5 million compounds, subsets of these compounds have been pre-selected based on the presence of specific hydrogen bond, hydrophobic/aromatic interactions, and aromatic π-stacking interactions at the binding site [38]. A combinatorial docking with pharmacophore constraints have been used for searching dihydrofolate reductase inhibitors from a randomly selected subset of 22,500 compounds from combinatorial libraries of 1,356,250 compounds, which identifies 4 actives in the top 50 hits [40]. In using the combinatorial QSAR modeling strategy for identifying geranylgeranyltransferase type I inhibitors [41] and histone deacetylase inhibitors [42] from 9.5 million compounds, subsets of compounds within the applicability domains (i.e. within certain distances to the training compounds) have been pre-selected before using these QSAR models, which led to the identification of 47 consensus virtual hits against geranylgeranyltransferase type I with 7 of the 7 selected hits tested active [41], and 45 novel consensus virtual hits against histone deacetylase with 3 of the 4 selected hits tested active [42]. Therefore, rationally defined structural constraints are highly effective in selecting relevant compounds from large chemical libraries.

In the combination approach, multiple VS methods that complement one another in screening speed and hit selection strategy (e.g. ligand-based and structure-based methods) are collectively used for enhanced performance, typically with a faster method used first followed by a slower method. The combination of pharmacophore, rule of five based drug-likeness filtering, and docking has been used for discovering potential DNA gyrase B inhibitors through screening ~5M chemical libraries and the two most active compounds with about 64 μM activity are selected for further investigation [43]. The two-stage pharmacophore and docking method has been used together with RO5, HTS and ADME-Tox filters for searching human tyrosyl-DNA phospodiesterase inhibitors from 27 million compounds to identify 178 virtual hits, with 1 of the 46 tested compounds found active [44]. The two-stage similarity searching and docking method has been used in three studies. In one study, it has been used together with additional RO5 and HTS filters for searching anthrax toxin lethal factor inhibitors from 35 million compounds to identify 602 virtual hits, with 12 of the 39 commercially available hits tested active [45]. In the second study, it has been used together with additional molecular weight and solubility filters for identifying MAPK-interacting Kinase 1 inhibitors from 18.8 million compounds, with 26 of the 1236 virtual hits tested active [46]. In searching phosphodiesterase type 5 inhibitors

Table 1. Listing of Available Annotated Libraries

Database and Ref	Web Pages	Number of Molecules	Free
ChEMBLdb [84]	https://www.ebi.ac.uk/chembl/	1,213,239 distinct bioactive compounds and 9003 targets	Yes
PubChem BioAssay [85]	http://pubchem.ncbi.nlm.nih.gov/	1,600,000 small molecules, 50,000 bioassay records, over 130,000,000 bioactivity results	Yes
BindingDB [86]	http://www.bindingdb.org/bind/index.jsp	910,836 binding data, 6,263 protein targets and 378,980 small molecules	Yes
WOMBAT [87]	http://www.sunsetmolecular.com/	331,872 molecules, 1966 unique targets	No
MDDR	http://symyx.com/	Over 180,000 biologically relevant compounds	No
DrugBank [88]	http://www.drugbank.ca/	6711 drugs and 4227 non-redundant protein	Yes
The Mother of All Databases (MOAD) [89]	http://bindingmoad.org/	16,948 protein-ligand structures, 5,630 structures with binding data, 8,140 different ligands	Yes
Therapeutic Target Database (TTD) [90, 91]	http://bidd.nus.edu.sg/group/ttd/ttd.asp	2025 targets, 17,816 drugs	Yes
PDB Bind [92, 93]	http://www.pdbbind.org/	3214 ligand-protein complexs	Yes
AffinDB [94]	http://www.agklebe.de/affinity	748 ligand-protein complexs	Yes
PDSP [95]	http://pdsp.med.unc.edu/pdsp.php	55,470 Ki values	Yes
IUPHARdb [96]	http://www.iuphar-db.org/index.jsp	627 human genes, 3177 distinct ligands	Yes
CARLSBAD	http://carlsbad.health.unm.edu/	755,329 compounds, 3613 protein targets, and 1,449,924 activities	Yes

Table 2. Performance of Different Methods Used in Screening Large Chemical Libraries of >1 Million Compounds in Last 5 Years

Type of VS Method	VS Method (Year of Study) [Reference]	VS Task and Number of Virtual Hits				VS Performance			
		Target Class	Database Screened	No of Compounds Screened	No of Virtual Hits	Testing Method	No of Virtual Hits Tested	Activity of the Most Active Compound	Hit Rates[a]
Individual Approach	Docking (2008)[97]	DNA Ligase	In-house database	1.5M	233		192	<50 μM	5.2%
	SVM (2008)[61]	Various target classes	PubChem	2.986M	299~9,502	*In silico* identification of known active compounds not in training datasets			4.7%~73.8%
	SVM (2008)[32]	Six MDDR classes (M1, NMDA, thrombin, HIV protease, cephalosporins, and renin)	PubChem	9.997 M	804~2,253	*In silico* identification of known active compounds not in training datasets			26.7%~90.9%
	Pharmacophore (2009)[98]	HCV RNA-dependent RNA polymerase	PharmoDB	3.5M			119	20 μM	2.5%
	Two layer SVM (2009)[34]	Androgen receptor	PubChem	19M	500	Bioassay of purchasable virtual hits	12	0.00374μM	67%
	SVM (2009)[23]	Bcr-Abl tyrosine-kinase	PubChem	13.56M	29,118	*In silico* identification of known active compounds not in training dataset			>0.16%
	Similarity searching (2009)[99]	NAADP	ZINC	2.7M	500		25	2 μM	84%

(Table 1) contd…

Type of VS Method	VS Method (Year of Study) [Reference]	VS Task and Number of Virtual Hits				VS Performance			
		Target Class	Database Screened	No of Compounds Screened	No of Virtual Hits	Testing Method	No of Virtual Hits Tested	Activity of the Most Active Compound	Hit Rates[a]
	Feature Trees (2009)[100]	TGFβ		500 billion	7,653~23,124				6.7%
	Feature Trees (2009)[101]	5-HT3 antagonists		1 trillion	2000				0.4%
	Bayesian Idea Generator (2009)[102]	H4		0.95M	96				1.0%
	Docking (2010)[103]	Dengue Virus Methyltransferase	ZINC, Schrodinger inhouse CACDB database	5M	750		35	4.47 μM	28.6%
	SVM (2010)[33]	Histone deacetylase	PubChem	13.56 M	74,761	*In silico* identification of known active compounds not in training dataset			>0.13%
	Combinatorial SVM (2011)[104]	Dual kinase inhibitors of 11 kinase pairs	PubChem	13.56 M	1,824~4,149	*In silico* identification of known active compounds not in training dataset			4.3%
	Combinatorial SVM (2011)[105]	Dual target serotonin reuptake inhibitors of 7 target pairs	PubChem	17M	1,154~7,601	*In silico* identification of known active compounds not in training dataset			12.6%
	Two layer similarity searching (2011)[35]	Phosphodiesterase type 5	ZINC	25M	137	Bioassay of virtual hits	137	0.1μM	18.2%
	Docking (2012)[106]	Cell division cycle 25 (Cdc25)	ZINC drug-like subset, NCI diversity set	2.1M	208		30	0.07 μM	13.3%
Filter Approach	RO5, substructure filter+docking (2009)[107]	Trypanosoma cruzi trans-sialidase	Evotec in-house supplier database	2.5M	1,819	Bioassay of purchasable virtual hits	23	0.12mM	0%
	RO5 filter + docking (2009)[38]	c-Met tyrosine kinase	ChemNavigator	13.5M	175	Bioassay of purchasable virtual hits	70	0.675μM	4.2%
	RO5, pharmacophore, ADME filters+docking (2009)[108]	Phosphoinositide-3-kinase (PI3K)	ZINC	2.5M	982		89	3 μM	7.9%
	Pharmacophore filter+docking (2010)[109]	Bcl-XL	ZINC 'big-n-greasy' subset	1.8M	277	Bioassay of purchasable virtual hits	45	2.48 μM	6.67%
	Substructure, rotatable bond count filters + docking (2010)[39]	Glutamate Transporter 1	GDB	26.4M	5,176	Bioassay of synthesizable top-ranked virtual hits	14	1.4 μM	21.4%

(Table 1) contd…

Type of VS Method	VS Method (Year of Study) [Reference]	VS Task and Number of Virtual Hits				VS Performance			
		Target Class	Database Screened	No of Compounds Screened	No of Virtual Hits	Testing Method	No of Virtual Hits Tested	Activity of the Most Active Compound	Hit Rates[a]
	Molecular weight filter+Similarity searching+functional group filter (2011)[110]	Insect Molting Hormone Receptor	Database from Namiki Shoji Co. Ltd	3M	237		24	13 μM	8.33%
	Pharmacophore+RO5 filter (2012)[111]	Na+-dependent Taurocholate Co-transporting Polypeptide (NTCP)	CoCoCo database, CAP database	10M	160		10	3.2 μM	60%
	Clean lead filter, RO5 filter+docking (2012)[112]	H. pylori urease enzyme	ZINC	5M	1000		1	~60 μM	
	Permeability and known toxic motifs filter+docking (2012)[113]	AmpC β-Lactamase	Chembridge, Enamine, ChemDiv, NCI, Interchim	6M			61	351.3 μM	11.5%
Constraint Approach	Consensus kNN QSAR (2009)[41]	Geranylgeranyltransferase type I	ZINC	9.5M	47	Bioassay of purchasable virtual hits	7	~10 μM	100%
	Consensus SVM and kNN QSAR (2009)[42]	Histone deacetylase	ZINC	9.5M	45	Bioassay of purchasable virtual hits	4	10 μM	75%
	FlexXc Docking +RO5+ pharmacophore constraints (2006)[40]	Bcr-Abl tyrosine-kinase		22.4M					
Combination Approach	Similarity searching + Pharmacophore searching+docking (2008)[114]	Fatty acid receptor 1 (FFAR1)	ZINC drug-like subset	2.6M	4529		52	3.6 μM	11.5%
	Pharmacophore + docking (2009)[115]	MurD/MurE ligases	Life Chemicals, Chemical Diversity, Vitas-M	2M		Bioassay of purchasable virtual hits	14	32 μM	14.3%
	Similarity searching + docking + RO5, HTS filters (2009)[45]	Anthrax toxin lethal factor	LeadQuest, MLSMR,GDB ,ZINC, NCI, University of Minnesota Institute for Therapeutics Discovery and Development (ITDD) inhouse vendor database	35M	602	Bioassay of purchasable virtual hits	39	49.5 μM	12.8%
	RO5, rotatable bond, functional group filter+pharmacophore+docking(2009)[116]	Carbonic anhydrase IX	ZINC	4.6M	6		6	0.29 μM	100%
	Pharmacophore + docking + RO5, HTS, ADME-Tox filters (2010)[44]	Human tyrosyl-DNA phospodiesterase	ChemNavigator iResearch Library	27M	178	Bioassay of purchasable virtual hits	46	7.94 μM	2.2%
	Pharmacophore + docking (2010)[117]	DNA gyrase B	ZINC			Bioassay of purchasable virtual hits	12	25 μM	41.7%

(Table 1) contd…

Type of VS Method	VS Method (Year of Study) [Reference]	VS Task and Number of Virtual Hits				VS Performance			
		Target Class	Database Screened	No of Compounds Screened	No of Virtual Hits	Testing Method	No of Virtual Hits Tested	Activity of the Most Active Compound	Hit Rates[a]
	Similarity searching + docking + molecular weight, solubility filters (2010)[46]	MAPK-interacting Kinase 1	CNIO library	18.8M	1,236	Bioassay	1236	0.83 μM	2.1%
	Pharmacophore + docking (2010)[118]	IKKβ	ChemDiv, Asinex-Gold, Asinex-Platinum	1.04M	100		60	6.7 μM	5%
	A softened version of the Lipinski rule filter+substructure fil-ter+pharmacophore+dockin g (2010)[119]	The malig-nant brain tumor (MBT) protein	iResearch Library	5.9M	67		51	5.7 μM	37.3%
	Similarity searching + docking (2010)[120]	Transthyreti n amyloid	ZINC	11M					
	Similarity searching + docking (2011)[121]	Human AP endonucle-ase	ZINC drug-like subset	2.6M	420		38	0.003 μM	
	Similarity searching + docking (2011)[35]	Phosphodi-esterase type 5	ZINC	25M	60	Bioassay of virtual hits	60	<1 μM	16%
	SVM + docking (2011)[48]	c-Met tyro-sine kinase inhibitors	PubChem, Specs, Enamine	18M	1,000	Bioassay of selected virtual hits	75	<10 μM	10.7%
	SVM + Pharmacophore (2011)[69]	mGluR1	PubChem, Specs, Enamine	20M	2,000				
	Hierarchical multistage SVM, pharmacophore and docking (2011)[51]	Pim-1 kinase	PubChem, Specs, Enamine	18M	935	Bioassay of selected virtual hits	47	0.263 μM	31.9%
	Pharmacophore + RO5+docking (2012)[43]	DNA gyrase B	Amri Global, Vitas M, Chembridge, Enamine, Chemical diversity	5M		Bioassay of selected virtual hits	26	64 μM	7.7%
	Pharmacophore + docking (2012)[122]	DNA gyrase B	Amri Global, Vitas M, Chembridge, Enamine, Chemical diversity	5M	400	Bioassay of selected virtual hits	12	5.5 μM	16.7%

a Hit rate after further selection of *in silico* hits by such considerations as top-rank scores, synthesizability and other intuitive considerations.

from 25 million compounds, it identified 60 virtual hits with 11 hits tested active (hit rate 16%) [35].

SVM is a relatively new VS method [23, 32-34, 47] increasingly explored in combination with other VS methods for the screening of large chemical libraries. The two-stage SVM and docking method has been used for searching novel c-Met tyrosine kinase inhibitors from 18 million compounds to identify 1000 top-ranked virtual hits, with 8 of the 75 selected hits tested active (hit rate 10.7% after additional selection) [48]. The two-stage SVM and pharmacophore method has been used for predicting novel metabotropic glutamate receptor 1 (mGluR1) antagonists from 20 million compounds, with 200 interesting virtual hits selected for further investigation [49]. SVM, pharmacophore and docking method have been hierarchically and multi-staged used for searching novel Pim-1 kinase inhibitors from 18 million compounds to identify 935 virtual hits, with 15 of the 47 tested hits showing nanomolar to submicromolar activities (hit rate 31.9% after additional selection) [50, 51].

Moreover, some VS methods have been used for developing consensus virtual screening models in which multiple virtual screening models are constructed from different training datasets, parameters, and/or scoring function, and these models are used for conducting multiple screening campaigns to collectively select virtual hits based on consensus scoring. This consensus modelling approach has been used for developing consensus kNN (K nearest neighbour) QSAR models of geranylgeranyltransferase type I in-

hibitors [41] and the consensus SVM and kNN QSAR models of histone deacetylase inhibitors [42] with hit rates of 100% and 75% (after additional hit selection) in searching geranylgeranyltransferase type I inhibitors [41] and histone deacetylase inhibitors [42] from 9.5 million compounds.

3. DIFFICULTIES IN THE APPLICATION OF VIRTUAL SCREENING METHODS AND STRATEGIES FOR FURTHER IMPROVEMENTS

The quality of training data is important for the development of ligand-based VS models, and the quality of the screening chemical libraries is also important for the successful applications of VS models. The publically available annotated chemical libraries listed in (Table **1**) and the public and commercial chemical libraries listed in (Table **2**) have been extensively used for developing and applying VS models. However, some public chemical libraries tend to contain higher level of noises [52, 53] because of errors in data curation, measurement variations and insufficient quality check. These data are subject to pre-processing quality and inconsistency check. Moreover, the same set of molecular preparation procedures, such as adding implicit hydrogen atoms and assigning the correct charges, should be applied for both the training data and the screening chemical libraries to ensure consistency [54].

For targets with available 3D structures, docking has been the most popularly used VS method for searching drug leads and their binding modes at target sites. However, a recent survey of over 15 years of docking studies has revealed that in these studies evidence has rarely been presented to establish the predicted binding modes are the correct ones [55]. Accurate estimation of binding affinity is another challenging task of docking [56]. The computational cost of docking is very high, making it less useful for screening large chemical libraries. It is often used in the late stage of virtual screening in the filter approach, the constraint approach and the combination approach.

The application range, accuracy and the computational cost of pharmacophore VS models depend on how the criteria of pharmacophore queries are set and how the activity-relevant structural features are recognized. Stricter setting typically lead to less coverage of scaffold diversity and more relaxed setting tend to significantly increase the false hit rates. A more balanced setting may be fine tuned with respect to the screening target and chemical libraries. Structural features relevant to the targeted activity may be selected by using feature selection methods and based on pharmacophore analysis of multiple active compounds [54]. Moreover, pharmacophore can be used as a follow-up screening tool [57].

Similarity searching and machine learning methods have been increasingly used individually or in combination with other methods in searching large chemical libraries partly due to their advantages of high screening speed and good yields. However, similarity searching methods rely on molecular features for comparing molecules [54] and their performance is thus highly sensitive to the molecular features and similarity measures used in a particular VS study [58, 59]. Retrospective analysis of the DUD datasets of multiple target classes has suggested that in some cases similarity searching based on 2D descriptors perform better than that based on 3D descriptors [60]. The quality of machine learning VS models depend heavily on the structural diversity of their training datasets which are constrained by the available known active and inactive compounds. Structurally diverse putative inactive compounds has been introduced in the training datasets of machine learning VS models to enhance their performance by reducing the false hit rates and improving the yields, and such a strategy has also shown some level of success in training datasets of highly sparse active compounds [32, 61].

The application of filters and structural constraints can efficiently narrow down the search scope of large chemical libraries. But it may in some cases reduce the yields. An analysis of 1000 approved drugs has shown that 8%~15% of these drugs have at least one violation of the filters of RO5, polar surface area and rotatable bond count [62]. Another analysis of 1070 approved drugs has shown that 26% of these drugs contain at least one "undesirable" molecular moiety (such as the appearance of an electrophilic reactive group) [63]. High quality structural constraints are not easily obtained. The applicability domain of QSAR is typically determined by the physicochemical properties of or structural similarity to the training molecules, and mechanistic understanding of their structure-activity relationships [64]. In cases of fewer known actives, it may be difficult to define an appropriate applicability domain to optimally cover the physicochemical properties of the active compounds against one or more targets, particularly the novel ones, without substantially increasing false hit rate.

The performance of VS methods in searching large chemical libraries can be further enhanced by improving VS algorithms. Examples of recent efforts for improving VS algorithms include adding pharmacophoric shims for generating highly predictive target-customized docking models [65, 66], most-frequent-feature guided development of pharmacophore models [19], and a two-step target binding and selectivity screening approach for searching target subtype selective ligands [67]. Moreover, consensus VS strategy has been increasingly used for collectively exploring the advantages of the VS methods and compensating their disadvantages. An analysis of consensus docking models has shown that specific consensus models consistently outperform individual models for predicting protein ligands [68]. Consensus models of docking [69], scoring functions [70], pharmacophore [71], QSAR [72], similarity searching [73], clustering [74], and machine learning models [47] have been reported to improve VS performance.

Another screening strategy is the integration of VS and HTS [75]. VS and HTS can be combined parallelly to screen the same chemical library and additively identify active hits for expanded the coverage of active chemical space. Parallel HTS and VS screening against such targets as tyrosine phosphatase-1B [76], cruzain [77], CDP-ME kinase [78] has identified substantially higher number of hits of distinct structures than each method alone, which suggests that combined HTS and VS complement each other in searching active compounds. HTS can also be combined sequentially with VS, with HTS as a post VS screening tool. In a sequential VS and HTS screening of 8.7 million compounds for ientifying hH4R binders, FlexX docking has been used for identifying 45,000 virtual hits which were then subject to a follow-up HTS that achieved a 7.2% hit rate [79].

4. PERSPECTIVES

VS methods have been adjusted for screening larger chemical libraries of ~ 1 million or more compounds, which have consistently shown promising performances. Given rapid expansion of chemical libraries [9, 10, 80] and the further development of library generation tools [81], there is a need for further improving VS methods to meet the challenge for searching larger libraries at higher performance levels, and for the identification of novel drug leads. In addition to the further improvement of individual VS methods and exploration of consensus modelling strategy and the integration of VS and HTS strategy, the exploration of VS methods can be facilitated by mining actives and inactives from literatures [10]. VS methods and the subsequent lead identification can also be further complemented by adding off-target identification and side effect prediction [82, 83]. The efficiency and predictive performance of the off-target search methods as well as the VS methods need to be further improved to meet this challenge.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

This work was supported by the Academic Research Funds Singapore [Grant R-148-000-081-112]; the National Natural Science Foundation of China [Grant 30772651]; and the Ministry of Science and Technology, 863 Hi-Tech Program China [Grant 2006AA020400].

REFERENCES

[1] Mayr, L.M.; Bojanic, D., Novel trends in high-throughput screening. Curr Opin Pharmacol, 2009, 9(5), 580-588

[2] Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov, 2004, 3(11), 935-949

[3] Yang, S.Y., Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug Discov Today, 2010, 15(11-12), 444-450

[4] Willett, P., Similarity-based virtual screening using 2D fingerprints. Drug Discov Today, 2006, 11(23-24), 1046-1053

[5] Clark, R.D., Prospective ligand- and target-based 3D QSAR: state of the art 2008. Curr Top Med Chem, 2009, 9(9), 791-810

[6] Li, H.; Yap, C.W.; Ung, C.Y.; Xue, Y.; Li, Z.R.; Han, L.Y.; Lin, H.H.; Chen, Y.Z., Machine learning approaches for predicting compounds that interact with therapeutic and ADMET related proteins. J Pharm Sci, 2007, 96(11), 2838-2860

[7] Bajorath, J., Integration of virtual and high-throughput screening. Nat Rev Drug Discov, 2002, 1(11), 882-894

[8] Monge, A.; Arrault, A.; Marot, C.; Morin-Allory, L., Managing, profiling and analyzing a library of 2.6 million compounds gathered from 32 chemical providers. Mol Divers, 2006, 10(3), 389-403

[9] Bolton, E.; Wang, Y.; Thiessen, P.A.; Bryant, S.H. PubChem: Integrated Platform of Small Molecules and Biological Activities; American Chemical Society: Washington, DC, Apr, 2008; pp 1365-1370.

[10] Bellis, L.J.; Akhtar, R.; Al-Lazikani, B.; Atkinson, F.; Bento, A.P.; Chambers, J.; Davies, M.; Gaulton, A.; Hersey, A.; Ikeda, K.; Kruger, F.A.; Light, Y.; McGlinchey, S.; Santos, R.; Stauch, B.; Overington, J.P., Collation and data-mining of literature bioactivity data for drug discovery. Biochem Soc Trans, 2011, 39(5), 1365-1370

[11] Weber, M.; Muthusubramaniam, L.; Murray, J.; Hudak, E.; Kornienko, O.; Johnson, E.N.; Strulovici, B.; Kunapuli, P., Ultra-high-throughput screening for antagonists of a Gi-coupled receptor in a 2.2-microl 3,456-well plate format cyclicAMP assay. Assay Drug Dev Technol, 2007, 5(1), 117-125

[12] Tanaka, K.; Koresawa, M.; Iida, M.; Fukasawa, K.; Stec, E.; Cassaday, J.; Chase, P.; Rickert, K.; Hodder, P.; Takagi, T.; Komatani, H., Multiplexed random peptide library and phospho-specific antibodies facilitate human polo-like kinase 1 inhibitor screen. Assay Drug Dev Technol, 8(1), 47-62

[13] Dudek, A.Z.; Arodz, T.; Galvez, J., Computational methods in developing quantitative structure-activity relationships (QSAR): a review. Comb Chem High Throughput Screen, 2006, 9(3), 213-228

[14] Verma, J.; Khedkar, V.M.; Coutinho, E.C., 3D-QSAR in drug design--a review. Curr Top Med Chem, 10(1), 95-115

[15] Gozalbes, R.; Simon, L.; Froloff, N.; Sartori, E.; Monteils, C.; Baudelle, R., Development and experimental validation of a docking strategy for the generation of kinase-targeted libraries. J Med Chem, 2008, 51(11), 3124-3132

[16] Evers, A.; Hessler, G.; Matter, H.; Klabunde, T., Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. J Med Chem, 2005, 48(17), 5448-5465

[17] Alonso, H.; Bliznyuk, A.A.; Gready, J.E., Combining docking and molecular dynamic simulations in drug design. Med Res Rev, 2006, 26(5), 531-568

[18] Camille-Georges, W.; Ganellin, C.R.; Lindberg, P.; M., L.A. Glossary of Terms Used in Medicinal Chemistry International Union of Pure and Applied Chemistry: 1998; pp 385-395.

[19] Zou, J.; Xie, H.Z.; Yang, S.Y.; Chen, J.J.; Ren, J.X.; Wei, Y.Q., Towards more accurate pharmacophore modeling: Multicomplex-based comp rehensive pharmacophore map and most-frequent-feature pharmacophore model of CDK2. Journal of molecular graphics & modelling, 2008, 27(4), 430-438

[20] Horvath, D., Pharmacophore-based virtual screening. Methods Mol Biol, 2011, 672, 261-298

[21] Scior, T.; Medina-Franco, J.L.; Do, Q.T.; Martinez-Mayorga, K.; Yunes Rojas, J.A.; Bernard, P., How to recognize and workaround pitfalls in QSAR studies: a critical review. Curr Med Chem, 2009, 16(32), 4297-4313

[22] Willett, P., Chemical Similarity Searching. J Chem Inf Comput Sci, 1998, 38, 983-996

[23] Liu, X.H.; Ma, X.H.; Tan, C.Y.; Jiang, Y.Y.; Go, M.L.; Low, B.C.; Chen, Y.Z., Virtual screening of Abl inhibitors from large compound libraries by support vector machines. J Chem Inf Model, 2009, 49(9), 2101-2110

[24] Zhang, Q.; Muegge, I., Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring. Journal of medicinal chemistry, 2006, 49(5), 1536-1548

[25] Harper, G.; Bradshaw, J.; Gittins, J.C.; Green, D.V.; Leach, A.R., Prediction of biological activity for high-throughput screening using binary kernel discrimination. J Chem Inf Comput Sci, 2001, 41(5), 1295-1300

[26] Chen, B.; Harrison, R.F.; Pasupa, K.; Willett, P.; Wilton, D.J.; Wood, D.J.; Lewell, X.Q., Virtual screening using binary kernel discrimination: effect of noisy training data and the optimization of performance. J Chem Inf Model, 2006, 46(2), 478-486

[27] Koike, A., Comparison of methods for chemical-compound affinity prediction. SAR QSAR Environ Res, 2006, 17(5), 497-514

[28] Yap, C.W.; Li, H.; Ji, Z.L.; Chen, Y.Z., Regression methods for developing QSAR and QSPR models to predict compounds of specific pharmacodynamic, pharmacokinetic and toxicological properties. Mini Rev Med Chem, 2007, 7(11), 1097-1107

[29] Ma, X.H.; Jia, J.; Zhu, F.; Xue, Y.; Li, Z.R.; Chen, Y.Z., Comparative analysis of machine learning methods in ligand-based virtual screening of large compound libraries. Comb Chem High Throughput Screen, 2009, 12(4), 344-357

[30] Li, H.; Yap, C.W.; Xue, Y.; Li, Z.R.; Ung, C.Y.; Han, L.Y.; Chen, Y.Z., Statistical learning approach for predicting specific pharmacodynamic, pharmacokinetic or toxicological properties of pharmaceutical agents. . Drug Development Research, 2006, 66(4), 245-259

[31] Lepp, Z.; Kinoshita, T.; Chuman, H., Screening for new antidepressant leads of multiple activities by support vector machines. J Chem Inf Model, 2006, 46(1), 158-167

[32] Ma, X.H.; Wang, R.; Yang, S.Y.; Li, Z.R.; Xue, Y.; Wei, Y.C.; Low, B.C.; Chen, Y.Z., Evaluation of virtual screening performance of support vector machines trained by sparsely distributed active compounds. J Chem Inf Model, 2008, 48(6), 1227-1237

[33] Liu, X.H.; Song, H.Y.; Zhang, J.X.; Han, B.C.; Wei, X.N.; Ma, X.H.; Cui, W.K.; Chen, Y.Z., Identifying Novel Type ZBGs and Nonhydroxamate HDAC Inhibitors Through a SVM Based Virtual Screening Approach. Mol. Inf., 2010, 29, 407-420

[34] Nagamine, N.; Shirakawa, T.; Minato, Y.; Torii, K.; Kobayashi, H.; Imoto, M.; Sakakibara, Y., Integrating statistical predictions and experimental verifications for enhancing protein-chemical interaction predictions in virtual screening. PLoS Comput Biol, 2009, 5(6), e1000397

[35] Tomori, T.; Hajdu, I.; Barna, L.; Lorincz, Z.; Cseh, S.; Dorman, G., Combining 2D and 3D in silico methods for rapid selection of potential PDE5 inhibitors from multimillion compounds' repositories: biological evaluation. Mol Divers, 2011, 16(1), 59-72

[36] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev, 2001, 46(1-3), 3-26

[37] This component filters out molecules that are likely to be poor candidates for high-throughput screening, including those containing non-organic atom types, reactive substructures, and those with a molecular weight >150.

[38] Peach, M.L.; Tan, N.; Choyke, S.J.; Giubellino, A.; Athauda, G.; Burke, T.R., Jr.; Nicklaus, M.C.; Bottaro, D.P., Directed discovery of agents targeting the Met tyrosine kinase domain by virtual screening. J Med Chem, 2009, 52(4), 943-951

[39] Luethi, E.; Nguyen, K.T.; Burzle, M.; Blum, L.C.; Suzuki, Y.; Hediger, M.; Reymond, J.L., Identification of selective norbornane-type aspartate analogue inhibitors of the glutamate transporter 1 (GLT-1) from the chemical universe generated database (GDB). J Med Chem, 2010, 53(19), 7236-7250

[40] Gastreich, M.; Lilienthal, M.; Briem, H.; Claussen, H., Ultrafast de novo docking combining pharmacophores and combinatorics. J Comput Aided Mol Des, 2006, 20(12), 717-734

[41] Peterson, Y.K.; Wang, X.S.; Casey, P.J.; Tropsha, A., Discovery of geranylgeranyltransferase-I inhibitors with novel scaffolds by the means of quantitative structure-activity relationship modeling, virtual screening, and experimental validation. J Med Chem, 2009, 52(14), 4210-4220

[42] Tang, H.; Wang, X.S.; Huang, X.P.; Roth, B.L.; Butler, K.V.; Kozikowski, A.P.; Jung, M.; Tropsha, A., Novel inhibitors of human histone deacetylase (HDAC) identified by QSAR modeling of known inhibitors, virtual screening, and experimental validation. J Chem Inf Model, 2009, 49(2), 461-476

[43] Brvar, M.; Perdih, A.; Hodnik, V.; Renko, M.; Anderluh, G.; Jerala, R.; Solmajer, T., In silico discovery and biophysical evaluation of novel 5-(2-hydroxybenzylidene) rhodanine inhibitors of DNA gyrase B. Bioorg Med Chem, 2012, 20(8), 2572-2580

[44] Weidlich, I.E.; Dexheimer, T.; Marchand, C.; Antony, S.; Pommier, Y.; Nicklaus, M.C., Inhibitors of human tyrosyl-DNA phospodiesterase (hTdp1) developed by virtual screening using ligand-based pharmacophores. Bioorg Med Chem, 2010, 18(1), 182-189

[45] Chiu, T.L.; Solberg, J.; Patil, S.; Geders, T.W.; Zhang, X.; Rangarajan, S.; Francis, R.; Finzel, B.C.; Walters, M.A.; Hook, D.J.; Amin, E.A., Identification of novel non-hydroxamate anthrax toxin lethal factor inhibitors

by topomeric searching, docking and scoring, and *in vitro* screening. J Chem Inf Model, 2009, 49(12), 2726-2734

[46] Oyarzabal, J.; Zarich, N.; Albarran, M.I.; Palacios, I.; Urbano-Cuadrado, M.; Mateos, G.; Reymundo, I.; Rabal, O.; Salgado, A.; Corrionero, A.; Fominaya, J.; Pastor, J.; Bischoff, J.R., Discovery of mitogen-activated protein kinase-interacting kinase 1 inhibitors by a comprehensive fragment-oriented virtual screening approach. J Med Chem, 2010, 53(18), 6618-6628

[47] Li, J.; Gramatica, P., Classification and virtual screening of androgen receptor antagonists. J Chem Inf Model, 2010, 50(5), 861-874

[48] Xie, Q.Q.; Zhong, L.; Pan, Y.L.; Wang, X.Y.; Zhou, J.P.; Di-Wu, L.; Huang, Q.; Wang, Y.L.; Yang, L.L.; Xie, H.Z.; Yang, S.Y., Combined SVM-based and docking-based virtual screening for retrieving novel inhibitors of c-Met. Eur J Med Chem, 2011, 46(9), 3675-3680

[49] Li, G.B.; Yang, L.L.; Feng, S.; Zhou, J.P.; Huang, Q.; Xie, H.Z.; Li, L.L.; Yang, S.Y., Discovery of novel mGluR1 antagonists: a multistep virtual screening approach based on an SVM model and a pharmacophore hypothesis significantly increases the hit rate and enrichment factor. Bioorg Med Chem Lett, 2011, 21(6), 1736-1740

[50] Blum, L.C.; Reymond, J.L., 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. J Am Chem Soc, 2009, 131(25), 8732-8733

[51] Ren, J.X.; Li, L.L.; Zheng, R.L.; Xie, H.Z.; Cao, Z.X.; Feng, S.; Pan, Y.L.; Chen, X.; Wei, Y.Q.; Yang, S.Y., Discovery of novel Pim-1 kinase inhibitors by a hierarchical multistage virtual screening approach based on SVM model, pharmacophore, and molecular docking. J Chem Inf Model, 2011, 51(6), 1364-1375

[52] Williams, A.J.; Ekins, S., A quality alert and call for improved curation of public chemistry databases. Drug Discov Today, 2011, 16(17-18), 747-750

[53] Kramer, C.; Kalliokoski, T.; Gedeck, P.; Vulpetti, A., The experimental uncertainty of heterogeneous public K(i) data. J Med Chem, 2012, 55(11), 5165-5173

[54] Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J.L.; Martinez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D.K., Recognizing Pitfalls in Virtual Screening: A Critical Review. J Chem Inf Model, 2012

[55] Kolb, P.; Irwin, J.J., Docking screens: right for the right reasons? Curr Top Med Chem, 2009, 9(9), 755-770

[56] Warren, G.L.; Andrews, C.W.; Capelli, A.M.; Clarke, B.; LaLonde, J.; Lambert, M.H.; Lindvall, M.; Nevins, N.; Semus, S.F.; Senger, S.; Tedesco, G.; Wall, I.D.; Woolven, J.M.; Peishoff, C.E.; Head, M.S., A critical assessment of docking programs and scoring functions. J Med Chem, 2006, 49(20), 5912-5931

[57] Peach, M.L.; Nicklaus, M.C., Combining docking with pharmacophore filtering for improved virtual screening. J Cheminform, 2009, 1(1), 6

[58] Bender, A., How similar are those molecules after all? Use two descriptors and you will have three different answers. Expert Opin Drug Discov, 2010, 5(12), 1141-1151

[59] Bender, A.; Jenkins, J.L.; Scheiber, J.; Sukuru, S.C.; Glick, M.; Davies, J.W., How similar are similarity searching methods? A principal component analysis of molecular descriptor space. J Chem Inf Model, 2009, 49(1), 108-119

[60] Venkatraman, V.; Perez-Nueno, V.I.; Mavridis, L.; Ritchie, D.W., Comprehensive comparison of ligand-based virtual screening tools against the DUD data set reveals limitations of current 3D methods. J Chem Inf Model, 2010, 50(12), 2079-2093

[61] Han, L.Y.; Ma, X.H.; Lin, H.H.; Jia, J.; Zhu, F.; Xue, Y.; Li, Z.R.; Cao, Z.W.; Ji, Z.L.; Chen, Y.Z., A support vector machines approach for virtual screening of active compounds of single and multiple mechanisms from large libraries at an improved hit-rate and enrichment factor. J Mol Graph Model, 2008, 26(8), 1276-1286

[62] Bade, R.; Chan, H.F.; Reynisson, J., Characteristics of known drug space. Natural products, their derivatives and synthetic drugs. Eur J Med Chem, 2010, 45(12), 5646-5652

[63] Axerio-Cilies, P.; Castaneda, I.P.; Mirza, A.; Reynisson, J., Investigation of the incidence of "undesirable" molecular moieties for high-throughput screening compound libraries in marketed drug compounds. Eur J Med Chem, 2009, 44(3), 1128-1134

[64] Dimitrov, S.; Dimitrova, G.; Pavlov, T.; Dimitrova, N.; Patlewicz, G.; Niemela, J.; Mekenyan, O., A stepwise approach for defining the applicability domain of SAR and QSAR models. J Chem Inf Model, 2005, 45(4), 839-849

[65] Martin, E.J.; Sullivan, D.C., Surrogate AutoShim: predocking into a universal ensemble kinase receptor for three dimensional activity prediction, very quickly, without a crystal structure. J Chem Inf Model, 2008, 48(4), 873-881

[66] Mukherjee, P.; Martin, E., Development of a minimal kinase ensemble receptor (MKER) for surrogate AutoShim. J Chem Inf Model, 2011, 51(10), 2697-2705

[67] Zhang, J.; Han, B.; Wei, X.; Tan, C.; Chen, Y.; Jiang, Y., A two-step target binding and selectivity support vector machines approach for virtual screening of dopamine receptor subtype-selective ligands. PLoS One, 2012, 7(6), e39076

[68] Kukol, A., Consensus virtual screening approaches to predict protein ligands. Eur J Med Chem, 2011, 46(9), 4661-4664

[69] Kim, J.H.; Lim, J.W.; Lee, S.W.; Kim, K.; No, K.T., Ligand supported homology modeling and docking evaluation of CCR2: docked pose selection by consensus scoring. J Mol Model, 2011, 17(10), 2707-2716

[70] Yang, J.M.; Chen, Y.F.; Shen, T.W.; Kristal, B.S.; Hsu, D.F., Consensus scoring criteria for improving enrichment in virtual screening. J Chem Inf Model, 2005, 45(4), 1134-1146

[71] Lagisetti, C.; Pourpak, A.; Jiang, Q.; Cui, X.; Goronga, T.; Morris, S.W.; Webb, T.R., Antitumor compounds based on a natural product consensus pharmacophore. J Med Chem, 2008, 51(19), 6220-6224

[72] Shao, L.; Wu, L.; Fan, X.; Cheng, Y., Consensus ranking approach to understanding the underlying mechanism with QSAR. J Chem Inf Model, 2010, 50(11), 1941-1948

[73] Medina-Franco, J.L.; Martinez-Mayorga, K.; Bender, A.; Marin, R.M.; Giulianotti, M.A.; Pinilla, C.; Houghten, R.A., Characterization of activity landscapes using 2D and 3D similarity methods: consensus activity cliffs. J Chem Inf Model, 2009, 49(2), 477-491

[74] Perez-Nueno, V.I.; Ritchie, D.W.; Borrell, J.I.; Teixido, J., Clustering and classifying diverse HIV entry inhibitors using a novel consensus shape-based virtual screening approach: further evidence for multiple binding sites within the CCR5 extracellular pocket. J Chem Inf Model, 2008, 48(11), 2146-2165

[75] Polgar, T.; Keseru, G.M., Integration of virtual and high throughput screening in lead discovery settings. Comb Chem High Throughput Screen, 2011, 14(10), 889-897

[76] Doman, T.N.; McGovern, S.L.; Witherbee, B.J.; Kasten, T.P.; Kurumbail, R.; Stallings, W.C.; Connolly, D.T.; Shoichet, B.K., Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. J Med Chem, 2002, 45(11), 2213-2221

[77] Ferreira, R.S.; Simeonov, A.; Jadhav, A.; Eidam, O.; Mott, B.T.; Keiser, M.J.; McKerrow, J.H.; Maloney, D.J.; Irwin, J.J.; Shoichet, B.K., Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. Journal of medicinal chemistry, 2010, 53(13), 4891-4905

[78] Tidten-Luksch, N.; Grimaldi, R.; Torrie, L.S.; Frearson, J.A.; Hunter, W.N.; Brenk, R., IspE inhibitors identified by a combination of in silico and *in vitro* high-throughput screening. PLoS One, 2012, 7(4), e35792

[79] Kiss, R.; Kiss, B.; Konczol, A.; Szalai, F.; Jelinek, I.; Laszlo, V.; Noszal, B.; Falus, A.; Keseru, G.M., Discovery of novel human histamine H4 receptor ligands by large-scale structure-based virtual screening. J Med Chem, 2008, 51(11), 3145-3153

[80] Irwin, J.J.; Shoichet, B.K., ZINC--a free database of commercially available compounds for virtual screening. J Chem Inf Model, 2005, 45(1), 177-182

[81] Funatsu, K.; Miyao, T.; Arakawa, M., Systematic generation of chemical structures for rational drug design based on QSAR models. Curr Comput Aided Drug Des, 7(1), 1-9

[82] Schultz, N.; Marenstein, D.R.; De Angelis, D.A.; Wang, W.Q.; Nelander, S.; Jacobsen, A.; Marks, D.S.; Massague, J.; Sander, C., Off-target effects dominate a large-scale RNAi screen for modulators of the TGF-beta pathway and reveal microRNA regulation of TGFBR2. Silence, 2011, 2, 3

[83] Kulemina, L.V.; Ostrov, D.A., Prediction of off-target effects on angiotensin-converting enzyme 2. J Biomol Screen, 2011, 16(8), 878-885

[84] Overington, J., ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). Interview by Wendy A. Warr. J Comput Aided Mol Des, 2009, 23(4), 195-198

[85] Wang, Y.; Xiao, J.; Suzek, T.O.; Zhang, J.; Wang, J.; Zhou, Z.; Han, L.; Karapetyan, K.; Dracheva, S.; Shoemaker, B.A.; Bolton, E.; Gindulyte, A.; Bryant, S.H., PubChem's BioAssay Database. Nucleic Acids Res, 2012, 40(Database issue), D400-412

[86] Chen, X.; Lin, Y.; Liu, M.; Gilson, M.K., The Binding Database: data management and interface design. Bioinformatics, 2002, 18(1), 130-139

[87] Olah, M.; Mracec, M.; Ostopovici, L.; Rad, R.; Bora, A.; Hadaruga, N.; Olah, I.; Banda, M.; Simon, Z.; Mracec, M.; Oprea, T.I., Eds. WOMBAT: World of Molecular Bioactivity. Wiley-VCH: New York, 2004.

[88] Wishart, D.S.; Knox, C.; Guo, A.C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M., DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res, 2008, 36(Database issue), D901-906

[89] Hu, L.; Benson, M.L.; Smith, R.D.; Lerner, M.G.; Carlson, H.A., Binding MOAD (Mother Of All Databases). Proteins, 2005, 60(3), 333-340

[90] Chen, X.; Ji, Z.L.; Chen, Y.Z., TTD: Therapeutic Target Database. Nucleic Acids Res, 2002, 30(1), 412-415

[91] Zhu, F.; Shi, Z.; Qin, C.; Tao, L.; Liu, X.; Xu, F.; Zhang, L.; Song, Y.; Zhang, J.; Han, B.; Zhang, P.; Chen, Y., Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. Nucleic Acids Res, 2012, 40(Database issue), D1128-1136

[92] Wang, R.; Fang, X.; Lu, Y.; Wang, S., The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. J Med Chem, 2004, 47(12), 2977-2980

[93] Wang, R.; Fang, X.; Lu, Y.; Yang, C.Y.; Wang, S., The PDBbind database: methodologies and updates. J Med Chem, 2005, 48(12), 4111-4119

[94] Block, P.; Sotriffer, C.A.; Dramburg, I.; Klebe, G., AffinDB: a freely accessible database of affinities for protein-ligand complexes from the PDB. Nucleic Acids Res, 2006, 34(Database issue), D522-526

[95] Roth, B.L.; Kroeze, W.K.; Patel, S.; Lopez, E., The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrassment of riches? Neuroscientist, 2000, 6, 252-262

[96] Sharman, J.L.; Mpamhanga, C.P.; Spedding, M.; Germain, P.; Staels, B.; Dacquet, C.; Laudet, V.; Harmar, A.J., IUPHAR-DB: new receptors and tools for easy searching and visualization of pharmacological data. Nucleic Acids Res, 2011, 39(Database issue), D534-538

[97] Zhong, S.; Chen, X.; Zhu, X.; Dziegielewska, B.; Bachman, K.E.; Ellenberger, T.; Ballin, J.D.; Wilson, G.M.; Tomkinson, A.E.; MacKerell, A.D., Jr., Identification and validation of human DNA ligase inhibitors using computer-aided drug design. J Med Chem, 2008, 51(15), 4553-4562

[98] Ryu, K.; Kim, N.D.; Choi, S.I.; Han, C.K.; Yoon, J.H.; No, K.T.; Kim, K.H.; Seong, B.L., Identification of novel inhibitors of HCV RNA-dependent RNA polymerase by pharmacophore-based virtual screening and *in vitro* evaluation. Bioorg Med Chem, 2009, 17(8), 2975-2982

[99] Naylor, E.; Arredouani, A.; Vasudevan, S.R.; Lewis, A.M.; Parkesh, R.; Mizote, A.; Rosen, D.; Thomas, J.M.; Izumi, M.; Ganesan, A.; Galione, A.; Churchill, G.C., Identification of a chemical probe for NAADP by virtual screening. Nat Chem Biol, 2009, 5(4), 220-226

[100] Lessel, U.; Wellenzohn, B.; Lilienthal, M.; Claussen, H., Searching Fragment Spaces with feature trees. J Chem Inf Model, 2009, 49(2), 270-279

[101] Boehm, M.; Wu, T.Y.; Claussen, H.; Lemmen, C., Similarity searching and scaffold hopping in synthetically accessible combinatorial chemistry spaces. J Med Chem, 2008, 51(8), 2468-2480

[102] van Hoorn, W.P.; Bell, A.S., Searching chemical space with the Bayesian Idea Generator. J Chem Inf Model, 2009, 49(10), 2211-2220

[103] Podvinec, M.; Lim, S.P.; Schmidt, T.; Scarsi, M.; Wen, D.; Sonntag, L.S.; Sanschagrin, P.; Shenkin, P.S.; Schwede, T., Novel inhibitors of dengue virus methyltransferase: discovery by *in vitro*-driven virtual screening on a desktop computer grid. J Med Chem, 2010, 53(4), 1483-1495

[104] Ma, X.H.; Wang, R.; Tan, C.Y.; Jiang, Y.Y.; Lu, T.; Rao, H.B.; Li, X.Y.; Go, M.L.; Low, B.C.; Chen, Y.Z., Virtual Screening of Selective Multitarget Kinase Inhibitors by Combinatorial Support Vector Machines. Mol Pharm, 2010

[105] Shi, Z.; Ma, X.H.; Qin, C.; Jia, J.; Jiang, Y.Y.; Tan, C.Y.; Chen, Y.Z., Combinatorial support vector machines approach for virtual screening of selective multi-target serotonin reuptake inhibitors from large compound libraries. J Mol Graph Model, 2012, 32, 49-66

[106] Lavecchia, A.; Di Giovanni, C.; Pesapane, A.; Montuori, N.; Ragno, P.; Martucci, N.M.; Masullo, M.; De Vendittis, E.; Novellino, E., Discovery of new inhibitors of Cdc25B dual specificity phosphatases by structure-based virtual screening. J Med Chem, 2012, 55(9), 4142-4158

[107] Neres, J.; Brewer, M.L.; Ratier, L.; Botti, H.; Buschiazzo, A.; Edwards, P.N.; Mortenson, P.N.; Charlton, M.H.; Alzari, P.M.; Frasch, A.C.; Bryce, R.A.; Douglas, K.T., Discovery of novel inhibitors of Trypanosoma cruzi trans-sialidase from in silico screening. Bioorg Med Chem Lett, 2009, 19(3), 589-596

[108] Frederick, R.; Mawson, C.; Kendall, J.D.; Chaussade, C.; Rewcastle, G.W.; Shepherd, P.R.; Denny, W.A., Phosphoinositide-3-kinase (PI3K) inhibitors: identification of new scaffolds using virtual screening. Bioorg Med Chem Lett, 2009, 19(20), 5842-5847

[109] Mukherjee, P.; Desai, P.; Zhou, Y.D.; Avery, M., Targeting the BH3 domain mediated protein-protein interaction of Bcl-xL through virtual screening. J Chem Inf Model, 2010, 50(5), 906-923

[110] Harada, T.; Nakagawa, Y.; Ogura, T.; Yamada, Y.; Ohe, T.; Miyagawa, H., Virtual screening for ligands of the insect molting hormone receptor. J Chem Inf Model, 2011, 51(2), 296-305

[111] Greupink, R.; Nabuurs, S.B.; Zarzycka, B.; Verweij, V.; Monshouwer, M.; Huisman, M.T.; Russel, F.G., In Silico Identification of Potential Cholestasis-Inducing Agents via Modeling of Na+-Dependent Taurocholate Cotransporting Polypeptide Substrate Specificity. Toxicol Sci, 2012, 129(1), 35-48

[112] Azizian, H.; Nabati, F.; Sharifi, A.; Siavoshi, F.; Mahdavi, M.; Amanlou, M., Large-scale virtual screening for the identification of new Helicobacter pylori urease inhibitor scaffolds. J Mol Model, 2012, 18(7), 2917-2927

[113] Chan, F.Y.; Neves, M.A.; Sun, N.; Tsang, M.W.; Leung, Y.C.; Chan, T.H.; Abagyan, R.; Wong, K.Y., Validation of the AmpC beta-lactamase binding site and identification of inhibitors with novel scaffolds. J Chem Inf Model, 2012, 52(5), 1367-1375

[114] Tikhonova, I.G.; Sum, C.S.; Neumann, S.; Engel, S.; Raaka, B.M.; Costanzi, S.; Gershengorn, M.C., Discovery of novel agonists and antagonists of the free fatty acid receptor 1 (FFAR1) using virtual screening. J Med Chem, 2008, 51(3), 625-633

[115] Perdih, A.; Kovac, A.; Wolber, G.; Blanot, D.; Gobec, S.; Solmajer, T., Discovery of novel benzene 1,3-dicarboxylic acid inhibitors of bacterial MurD and MurE ligases by structure-based virtual screening approach. Bioorg Med Chem Lett, 2009, 19(10), 2668-2673

[116] Thiry, A.; Ledecq, M.; Cecchi, A.; Frederick, R.; Dogne, J.M.; Supuran, C.T.; Wouters, J.; Masereel, B., Ligand-based and structure-based virtual screening to identify carbonic anhydrase IX inhibitors. Bioorg Med Chem, 2009, 17(2), 553-557

[117] Brvar, M.; Perdih, A.; Oblak, M.; Masic, L.P.; Solmajer, T., In silico discovery of 2-amino-4-(2,4-dihydroxyphenyl)thiazoles as novel inhibitors of DNA gyrase B. Bioorg Med Chem Lett, 2010, 20(3), 958-962

[118] Nagarajan, S.; Choo, H.; Cho, Y.S.; Oh, K.S.; Lee, B.H.; Shin, K.J.; Pae, A.N., IKKbeta inhibitors identification part II: ligand and structure-based virtual screening. Bioorg Med Chem, 2010, 18(11), 3951-3960

[119] Kireev, D.; Wigle, T.J.; Norris-Drouin, J.; Herold, J.M.; Janzen, W.P.; Frye, S.V., Identification of non-peptide malignant brain tumor (MBT) repeat antagonists by virtual screening of commercially available compounds. J Med Chem, 2010, 53(21), 7625-7631

[120] Simoes, C.J.; Mukherjee, T.; Brito, R.M.; Jackson, R.M., Toward the discovery of functional transthyretin amyloid inhibitors: application of virtual screening methods. J Chem Inf Model, 2010, 50(10), 1806-1820

[121] Mohammed, M.Z.; Vyjayanti, V.N.; Laughton, C.A.; Dekker, L.V.; Fischer, P.M.; Wilson, D.M., 3rd; Abbotts, R.; Shah, S.; Patel, P.M.; Hickson, I.D.; Madhusudan, S., Development and evaluation of human AP endonuclease inhibitors in melanoma and glioma cell lines. Br J Cancer, 2011, 104(4), 653-663

[122] Brvar, M.; Perdih, A.; Renko, M.; Anderluh, G.; Turk, D.; Solmajer, T., Structure-based discovery of substituted 4,5'-bithiazoles as novel DNA gyrase inhibitors. J Med Chem, 2012, 55(14), 6413-6426

Received: May 06, 2012 **Revised: September 13, 2012** **Accepted: September 14, 2012**