Essential Factors for Successful Virtual Screening

Markus H.J. Seifert* and Martin Lang

4SC AG, Am Klopferspitz 19A, D-82152 Planegg-Martinsried, Germany

Abstract: Virtual high-throughput screening (vHTS) is a powerful technique for identifying hit molecules as starting points for medicinal chemistry. Numerous successful applications of vHTS have been published using a large variety of methodologies. This review attempts to identify the essential factors for successful virtual screening in the hit identification phase.

Keywords: Virtual screening, high-throughput screening, docking, scoring, hit identification, similarity search.

1. INTRODUCTION

The initiation of a medicinal chemistry program in pharmaceutical research generally requires some rational starting point, usually represented by an initial hit structure. Most commonly today, such a hit compound derives from the screening of a physical or virtual compound library. High-throughput screening (HTS) directly furnishes applicable biological activity data, albeit at high cost in terms of money, time, and secondary assays to identify false-positives [1]. Replacing *in-vitro* experiments with suitable *in-silico* models, virtual high-throughput screening (vHTS) provides a time- and cost-efficient alternative to establish viable starting points for medicinal chemistry [2]. A multitude of success stories have been reported in literature, illustrating the validity of this approach. The methodology itself, however, still experiences rapid development, making it difficult for synthetic chemists to fully assess the current potential of vHTS in their projects. This review summarizes current case studies and aims to distil the essential factors for successful virtual screening, thus enabling the medicinal chemist to integrate the methodology into his work flow.

Virtual screening (VS) and virtual high-throughput screening (vHTS) in particular are defined as the search for the molecules within a database of compounds that match a given query structure which may be a pharmacophore [3], a structure of another active ligand [4], or a receptor structure [5, 6]. The choice of a particular method is often dictated by the level of information, e.g. with respect to structural data, and economic restraints on computing resources (see Fig. (**1**)).

While both pharmacophore-based and ligand-based screening identify the compounds most similar to the query structure, protein structure-based screening (e.g., protein-ligand docking) returns compounds which are complementary to the query structure, i.e. the receptor's ligand binding site. All types of vHTS require a fast scoring function to compute either similarity or complementarity to a given query structure. Obviously, one expects that the score calculated by a scoring function somehow translates into biological affinity in order to identify hits in a bioassay. This is, however, a rather intricate problem [7, 8, 9], leading to the paradox that vHTS is able to reveal suitable compounds from huge databases, but most often no significant correlation between calculated scores and experimentally determined binding affinities can be detected within a set of top scorers. Therefore, at least from a practical point of view, the identification of viable hits on a given target is the criterion for successful virtual screening. A correspondence between score and binding affinity within a set of predicted hit compounds is generally not to be expected.

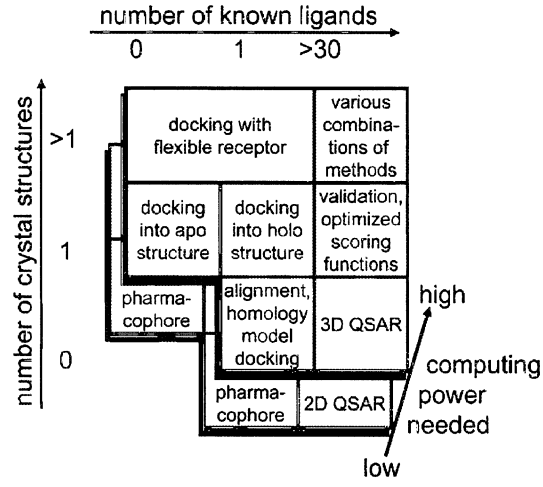


Fig. (1). The choice of virtual screening methods is guided by three parameters: number of known active small molecules, number of protein cystal structures available, and computing resources.

Virtual screening should enable the initiation of a medicinal chemistry program with a reasonable probability for identifying a lead compound. More precisely, VS has to identify at least one structure that (i) is biologically active, and (ii) allows for exploitation by medicinal chemistry. Using this definition, this review focuses on the hit identification phase of the drug design process. A timely update is given on recent developments of vHTS methodology and application. The results of this survey of current literature are discussed in terms of factors that increase the probability of success.

*Address correspondence to this author at 4SC AG, Am Klopferspitz 19A, D-82152 Planegg-Martinsried, Germany; Tel: +49-89-700763-0; Fax: +49-89-700763-29; E-mail: markus.seifert@4sc.com

1389-5575/07 $50.00+.00 © 2007 Bentham Science Publishers Ltd.

2. HIT IDENTIFICATION BY VIRTUAL SCREENING

2.1. Overview

The application of vHTS for hit and lead identification follows a typical sequence of processes (see Fig. (**2**)). The hit identification phase starts with vHTS of a database of choice. After an initial ranking list has been prepared, a selection of compounds is assembled for biological testing. Experiments then reveal the hit compounds that are biologically active. After the hit identification, the hit-to-lead process [10] usually starts by exploring the chemical space around the proven hits: hit-related compounds are identified by focused virtual screening or by manual database search, and subsequently assayed for biological activity. This back-screening process provides data to create an early QSAR which helps the medicinal chemists in finally identifying a lead compound.

The present review focuses on vHTS, selection of compounds, and identification of real positives. Table **1** summarizes the most important parameters of virtual screening as described in recently published case studies: target, virtual screening approach, type of screened database, and the software used. Additionally, filter criteria are listed. Methodology of the bioassay, hit criteria, and hit rates are given as well. A representative hit structure is shown in order to illustrate the chemistry involved. These data will be analyzed in the following sections, taking into account recent methodological advances and evaluation studies.

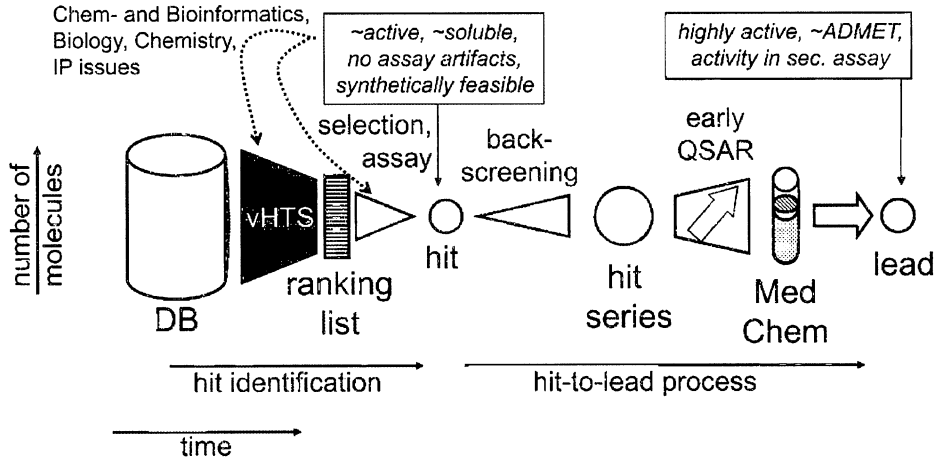


Fig. (2). General outline of typical hit identification and hit-to-lead processes.

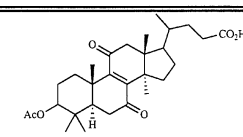
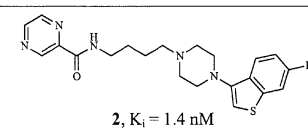
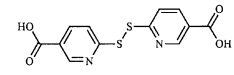
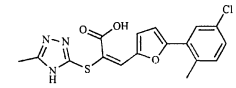
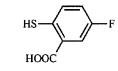
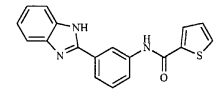
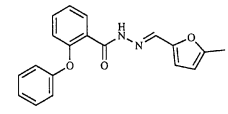
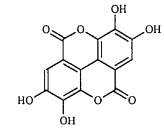
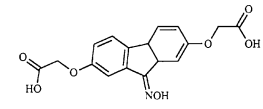
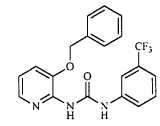
2.2. Software

It is apparent from Table **1** that a multitude of approaches can be applied for virtual screening. Most of the VS applications cited in Table **1** are based on protein-ligand docking. For docking, several established software packages exist, e.g. DOCK [11], AutoDock [12], FlexX [13], Gold [14], QXP [15], ICM [16], Slide [17], Glide [18], FRED [19], and Surflex [20], which are still being optimized: e.g. an extra-precision Glide protocol has been published along with some promising results [21] and the FlexX suite of programs was extended by FlexNovo [22]. In the last two years several new approaches have been reported: for example, eHITS [23] exhaustively searches the conformation space by first docking all rigid fragments of a ligand into all sub pockets of an active site, then identifying sets of docked fragments that can be matched by the complete ligand, and finally constructing ligand poses in the active site by joining the fragments correctly. This methodology is reported to improve the binding mode accuracy considerably compared to Gold and FlexX. MolDock [24] applies a differential evolution algorithm to identify docked binding modes. Again, it is reported to improve binding mode accuracy in comparison to Glide, Gold, Surflex, and FlexX. PSI-DOCK [25] uses a combination of taboo search and a genetic algorithm for docking resulting in a binding mode accuracy comparable to Gold and Glide. Recently, the protein-protein docking program ROSETTA-DOCK, which applies a Monte Carlo sampling of rigid body and side chain rotamer degrees of freedom, was extended for protein-ligand docking and achieved a binding mode reproduction rate of 71-80% within a database of 100 protein-ligand complexes [26]. In contrast to the above mentioned software, ProPose [27] uses an established incremental construction algorithm. The main focus of ProPose is to provide a highly flexible platform for virtual screening, allowing for an efficient combination of different approaches like docking, ligand alignment and pharmacophores [28].

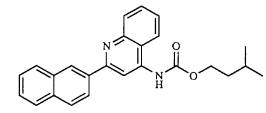
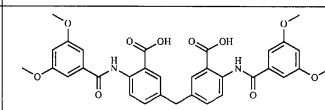
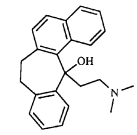
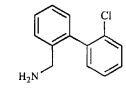
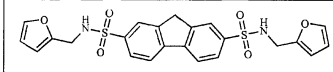
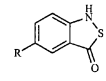
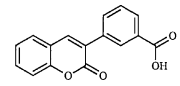
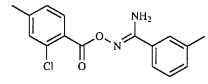
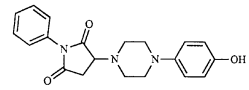
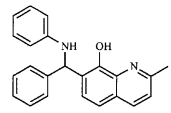
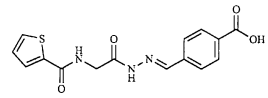
It should be noted at this point that – according to Warren *et al.* [9] – a good performance in reproduction of experimental binding modes does not necessarily impart success in virtual screening, although the contrary has been suggested as well [29]. Additionally, the study of Warren *et al.* clearly demonstrates that no single software tool performs well on all targets under consideration.

Ligand-based approaches for virtual screening are mainly used in case of no structural information about the target being available. During the last two years, some new developments have been reported in the field of molecular superpositioning, i.e. ligand-ligand alignment: for example, BRUTUS [30, 31] superimposes small molecules on molecular interaction fields (e.g., electrostatic fields) derived from a template. In principle, it generates a rigid molecule – rigid template alignment, and the conformational flexibility of the molecule has to be considered by the pre-generation of conformations. The algorithm is fast enough for virtual screen-

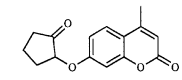
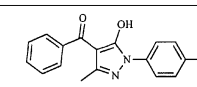
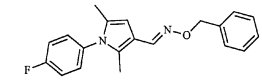
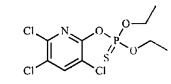
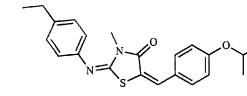
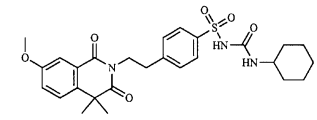
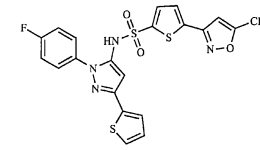
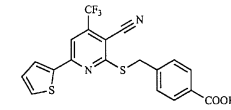
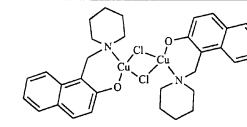
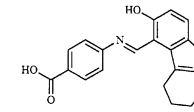
Table 1. Hit Identification by Virtual Screening: Recent Case Studies

Target (Class [a])	VS [b]	Database (Size)	Software Main Auxiliary	Filter [c]	Assay (Hit Criteria)	Active / Tested / Screened	Representative Hit cpd., Activity	Ref.
11β-hydroxy-steroid-dehydrogenase type 1, 11β-HSD 1 (E)	P	12 commercial libraries (1,776,579 in total)	Catalyst	h_don ≤ 5, h_acc ≤ 10, no match with hERG pharmacophore, clogP ≤ 5, logS ≥ -5, manual selection	radioactive assay (inhibition ≥70% at 10 μM)	7 / 30 / 1,776,579	 **1**, IC_{50} = 144 nM	[d]
α1A adrenergic receptor (GPCR)	R, H	proprietary	Gold, Catalyst [e]	b_rot ≤ 9, MW < 600, match of α1A pharmacophore	radioactive assay (inhibition > 50% at 10 μM)	38 / 80 / 22,950	 **2**, K_i = 1.4 nM	[f]
acetohydroxy acid synthase, AHAS (E)	R	ACD-3D [g] (~164,000)	Dock 4.0, AutoDock	no metals, XlogP -0.5 ... 3.0, visual inspection	colorimetric assay ($K_{i,app}$ < 100 μM)	3 / 14 / n.d.	 **3**, $K_{i,app}$ = 15 μM	[h]
angiotensin converting enzyme 2, ACE2 (E)	P	commercial compounds (3,787,666; unique ~2,500,000)	Catalyst, eHITS, Q-fit [i]	post-filtering: shape, exclusion volume, h_acc	fluorescence (inhibition > 70% at 200 μM)	7 / 17 / 3,787,666	 **4**, IC_{50} = 62 μM	[j]
antimalarial activity (U)	L, QSAR (LDA, MLR)	Merck Index [k] (~10,000)	BMDP (statistics software)	only drugs with therapeutic properties	*in-vitro* testing against *P. falciparum* (IC_{50} < 50 μM)	16 / 27 / 2000	monensin and nigericin with IC_{50}s of 0.327 and 0.425 nM, respectively	[l]
B. fragilis zinc β-lactamase, CcrA (E)	R	fragment-like subset of ZINC [m] (~10,000)	DOCK	MW < 250, logP –2 ... 3, h_don <3, h_acc <6, b_rot <3	spectrophotometric assay (K_i < 120 μM)	5 / 15 / 33,000	 **5**, K_i = 2 μM	[n]
B. subtilis RNA methyltransferase, *C. pneumonia* dimethyl-adenosine transferase (E)	R	SPECS, Maybridge [o] (~300,000)	FlexX	no reactive groups, no otherwise undesirable groups for drugs	cell-based *C. pneumoniae* assay (inhibition >50% at 50 μM)	8 / 33 / n.d.	 **6**, inhibition = 90% at 50 μM	[p]
β-catenin – Tcf4 antagonists (PP)	R	subset of Pharmacia & Upjohn collection (90,000)	FLO-QXP	Ro5, availability, low biological promiscuity (active in <4 non-antibiotic screens), solubility > 50 μM, visual inspection	NMR (mixtures at 50 μM), ITC (at 100 μM)	3 / 22 / 17,000	 **7**, K_D = 450 nM	[q]
casein kinase 2 CK2 (E)	R	proprietary database of naturally occurring compounds (~2000)	MOE-Dock [r], Glide, FRED, Gold	consensus scoring	radioactive assay (n.d.)	n.d. / n.d. / ~2000	 **8**, K_i = 20 nM	[s]
checkpoint kinase 1, Chk1 (E)	R	commercial compound libraries (1,600,000)	rDock (docking software)	pre-filter: MW 250 - 550, b_rot 0 - 6, no reactive groups that interfere with assay; post-filter: visual inspection, diversity	radioactive assay, scintillation readout (IC_{50} < 50 μM)	9 / 1179 / ~700,000	 **9**, IC_{50} = 13.4 μM	[t]
cyclophilin A, CypA (E)	R, P	ACD (296,387)	FlexX, ISIS Base [u]	pharmacophore, visual inspection, MW < 650	spectrophotometric assay (inhibition ≥80% at 10 μM)	5 / 31 / 3129	 **10**, IC_{50} = 303 nM	[v]

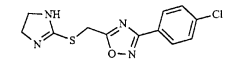
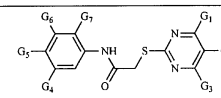
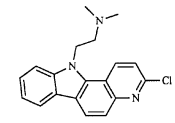
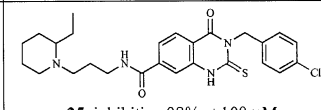
(Table 1. Contd....)

Target (Class [a])	VS [b]	Database (Size)	Software Main Auxiliary	Filter [c]	Assay (Hit Criteria)	Active / Tested / Screened	Representative Hit cpd., Activity	Ref.
cyclophilin A, CypA (E)	R	focused library (255)	LD1.0 (docking software)	Ro5, MW < 600	surface plasmon resonance (K_d < 50 μM)	16 / 16 / 255	 **11**, K_d = 76 nM	[w]
cyclophilin A, CypA (E)	R	SPECS (~280,000)	Dock 4.0, FlexX	drug-likeness	surface plasmon resonance (K_d < ~50 μM)	15 / 82 / 85,000	 **12**, K_d = 0.58 μM	[x]
cytochrome 2D6, CYP2D6 (E)	R, H	proprietary (5,760; incl. tautomers and stereo-isomers: 19,619)	Gold	diversity	fluorescence based assay (IC_{50} < 10 μM)	5 / 8 / 19,619	 **13**, IC_{50} = 0.06 μM	[y]
dipeptidyl peptidase IV, DPP-IV (E)	R	ACD, in-house database (n.d.)	FlexX-Pharm	primary, aliphatic amines, other filters n.d.	n.d. (IC_{50} < 50 μM)	6. / n.d. / ~10,000	 **14**, IC_{50} = 2.3 μM	[z]
falcipain 2 and 3, FP2, FP3 (E)	R, H	ACD (355,000)	Gold	pre-filter: no metals, no toxic functional groups, ADME filters, Ro5; post-filter: visual inspection	fluorescence based assay (FP2 IC_{50} < 55 μM)	18 / 100 / 80,000	 **15**, IC_{50} = 1.4 μM (FP2) and 11.4 μM (FP3)	[aa]
glycogen synthase kinase 3, GSK-3 (E)	R	proprietary (16,299)	FlexX, Flex-Pharm, FlexE	MW 200-800, logP ≤ 7, b_rot < 16, diversity filter	luminescence (inhibition >70% at 5μM)	21 / 162 / 16299	 **16**, inhibition = 90% at 5 μM	[bb]
Inducers of foetal haemoglobin (U)	P, PR	SPECS [cc] (subset, 13,000)	FLO-QXP	one acid group and HAC ≤ 25	reporter gene assay (relative gene activity > 100%)	23 / 26 / 630	 **17**, relative gene activity = 522%	[dd]
Kv1.5 (C)	P, H	proprietary	UNITY [ee]	b_rot ≤ 10, Ro5, visual inspection	voltage-clamp (IC_{50} < 10 μM)	5 / 244 / n.d.	structure not disclosed IC_{50} = 0.9 μM	[ff]
lipoxygenase 12-hLO, 15-hLO (E)	R, H	ChemBridge [gg] (50,000)	Glide	visual inspection	spectrophoto-metric assay (IC_{50} < 20 μM)	3 / 20 / 50,000	 **18**, 15-hLO IC_{50} = 6.8 μM	[hh]
malic enzyme, ME (E)	R, H	combinator. library (112,000)	FlexX	ADMET properties, diversity	continuous spectrophoto-metric assay (IC_{50} < 10 μM)	6 / n.d. / 10,000	 **19**, IC_{50} = 0.15 μM	[ii]
MDM2-p53 antagonists (PP)	R, P	NCI (~250,000)	Gold, in-house pharmaco-phore query tool	pharmacophore, cpd. availability, drug-likeness, visual inspection, MW 200-600, b_rot 1-10, h_acc < 10, #N/#O > 1	fluorescence-polarization based assay (K_i < 10 μM)	10 / 67 / 2599 (R), 110,000 (P)	 **20**, K_i = 120 nM	[jj]
methionyl-tRNA synthase, MetRS (E)	L, P	ChemDiv [kk] (508,143)	Chemo-soft [ll], Sybyl	n.d.	scintillation proximity assay (IC_{50} < 50 μM)	4 / 91 / 508,143	 **21**, IC_{50} = 0.237 μM	[mm]

(Table 1. Contd....)

Target (Class [a])	VS [b]	Database (Size)	Software Main Auxiliary	Filter [c]	Assay (Hit Criteria)	Active / Tested / Screened	Representative Hit cpd., Activity	Ref.
monoamine oxidase A, MAO-A (E)	L, QSAR (LDA)	small library of coumarins (n.d.)	March-Inside [nn]	different levels of structural complexity	spectrophotometric assay (IC_{50} < 25 μM)	9 / 15 / n.d.	 **22**, IC_{50} = 0.04 μM	[oo]
monoglyceridelipase (E)	R, H, P	Maybridge, LeadQuest [pp] (n.d.)	UNITY, Gold, BRUTUS	MW ≤ 450	substrate consumption (HPLC)	0 / 51 / n.d.	no inhibitor of MGL-like enzyme found, but hits were shown to be FAAH inhibitors	[qq]
Myco-bacterium tuberculosis, MTB (U)	L, P, QSAR (RP)	Asinex [rr] (~200,000)	Cerius [ss] Catalyst	pharmacophore, Ro5	full inhibition of MTB growth (MIC)	1/ 9 / 43,845	 **23**, MIC = 25 μg/ml	[tt]
Myco-bacterium tuberculosis, MTB (U)	L, P, QSAR (RP)	combinatorial library (5,177,717)	Cerius, Catalyst	diversity, Ro5, OTFO library design [uu]	full inhibition of MTB growth (MIC)	1/ 4 / 5000	 **24**, MIC = 25 μg/ml	[v v]
P. diminuta phospho-triesterase, PTE (E)	R	ACD (167,000)	DOCK	n.d.	spectrophotometric assay (K_m < 2 mM)	8 / 8 / 167,000	 **25**, K_m = 30 μM	[ww]
peroxisome proliferator-activated receptor, PPAR α, γ (R)	L, QSAR (NN)	SPECS (229,658)	CDK [xx]	low lipophilicity (SlogP), visual inspection	reporter gene assay (relative activation of PPARγ at 100 μM > 0)	4 / 9 / 229,658	 **26**, relative act. 0.54 at 100 μM	[yy]
peroxisome proliferator-activated receptor, PPAR γ (R)	R	TheraSTrat in-house database, Chembank (~14,000)	AutoDock	post-filtering: focus on sulfadimidine and sulfonylureas	FP-based PPAR assay, reporter gene assay (activation of PPARγ at 100 μM > 1)	6 / 6 / ~14,000	 **27**, IC_{50} = 8 nM, act. ~11	[zz]
peroxisome proliferator-activated receptor, PPARγ agonists, (R)	L, R	Maybridge (62,000)	Catalyst (shape-based search), Gold	n.d.	scintillation proximity assay (n.d.)	1 / 163 / 62,000	 **28**, IC_{50} = 175 nM	[aaa]
phospho-diesterase PDE-1, PDE-5 (E)	L, P, QSAR CART [bbb]	SPECS (subset 50,520)	SPSS [ccc] Answer-Tree Catalyst, Cerius	Ro5, MW < 500, H-acceptors < 10, H-donors < 5, AlogP < 6, b_rot < 12, visual inspection	n.d. (IC_{50} < 10 μM)	7 / 19 / 43,365	 **29**, PDE-1: IC_{50} = 1.9 μM, PDE-5: IC_{50} = 0.7 μM	[ddd]
protein phosphatase 2C, PP2C (E)	R	NCI Diversity Set [eee] (1,990)	AutoDock	logP ≤ 6.0 (established after testing the first 40 cpds)	radioactive assay (inhibition >50% at 100 μM)	3 / 100 / 1990	 **30**, inhibition 82% at 100 μM	[fff]
SARS-CoV protease, $3CL^{pro}$ (E)	R	n.d. (361,413)	EUDOC [ggg]	commercial availability, number of chiral centres, poor solubility or cell permeability	SARS-CoV cell-based inhibition assay	1 / 12 / 361,413	 **31**, EC_{50} = 23 μM	[hhh]

(Table 1. Contd....)

Target (Class [a])	VS [b]	Database (Size)	Software Main Auxiliary	Filter [c]	Assay (Hit Criteria)	Active / Tested / Screened	Representative Hit cpd., Activity	Ref.
SARS-CoV protease $3CL^{pro}$ (E)	R	Maybridge (58,855)	Gold	ranking: external H-bond energy term	FRET (inhibition >50% at 10 μM)	2 / 50 / 58,855	 **32**, inhibition 61 % at 10 μM	[iii]
SARS-CoV protease $3CL^{pro}$ (E)	R	Maybridge (59,363)	DOCK, Idea 2.0 [jjj]	Ro5, post-filter: H-bonding pattern	fluorescence based inhibition assay (IC_{50} < 30 μM)	21 / 93 / 59,363	 **33**, only core structure given	[kkk]
Tat – TAR RNA interaction (RNA)	P	SPECS (229,658)	CATS3D / SQUID (pharma-cophore query method)	selected 20,000 most drug-like	FRET (IC_{50} < 500 μM)	2 19 / 20,000	 **34**, IC_{50} = 46 μM	[lll]
T-type calcium channel α_{1H} (C)	L, P	Maybridge, comm. avail. ion channel database (63,495)	Catalyst	visual inspection	voltage clamp assay (inhibition >50% at 100 μM)	4 / 25 / 63,495	 **35**, inhibition 98% at 100 μM	[mmm]

[a] GPCR G-protein coupled receptor, R receptor, E enzyme, PP protein-protein interaction, C channel, U unknown, n.d. not disclosed.
[b] R receptor-based, PR pseudo-receptor-based, L ligand-based, P pharmacophore-based, H homology model-based.
[c] ADMET absorption-distribution-metabolism-excretion-toxicity, b_rot number of rotatable bonds, h_acc hydrogen bond acceptors, h_don hydrogen bond donors, HAC heavy atom count, logP log of water/octanol partition coefficient, MW molecular weight, Ro5 Lipinski's rule of five.
[d] Schuster, D.; Maurer, E.M.; Laggner, C.; Nashev, L.G.; Wilckens, T.; Langer, T.; Odermatt, A. *J. Med. Chem.,* **2006**, *49*, 3454.
[e] http://www.accelrys.com/products/catalyst/
[f] Evers, A.; Klabunde, T. *J. Med. Chem.*, **2005**, *48*, 1088.
[g] http://www.mdl.com/products/experiment/available_chem_dir/key_features.jsp
[h] Wang, J.G.; Xiao, Y.J.; Li, Y.H.; Ma, Y.; Li, Z.M. *Bioorg. Med. Chem.*, **2007**, *15*, 374.
[i] Jackson, R.M. *J. Comput. Aided. Mol. Des.,* **2002**, *16*, 43.
[j] Rella, M.; Rushworth, C.A.; Guy, J.L.; Turner, A.J.; Langer, T.; Jackson, R.M. *J. Chem. Inf. Model*, **2006**, *46*, 708.
[k] http://www.cambridgesoft.com/databases/details/?db=1
[l] Mahmoudi, N.; de Julian-Ortiz, J.V.; Ciceron, L.; Galvez, J.; Mazier, D.; Danis, M.; Derouin, F.; Garcia-Domenech, R. *J. Antimicrob. Chemother.*, **2006**, *57*, 489.
[m] http://blaster.docking.org/zinc/
[n] Irwin, J.J.; Raushel, F.M.; Shoichet, B.K. *Biochemistry*, **2005**, *44*, 12316.
[o] http://www.maybridge.com/default.aspx
[p] Alvesalo *et al.* [43].
[q] Trosset, J.Y.; Dalvit, C.; Knapp, S.; Fasolini, M.; Veronesi, M.; Mantegani, S.; Gianellini, L.M.; Catana, C.; Sundstrom, M.; Stouten, P.F.; Moll, J.K. *Proteins*, **2006**, *64*, 60.
[r] http://www.chemcomp.com/software-sbd.htm
[s] Cozza, G.; Bonvini, P.; Zorzi, E.; Poletto, G.; Pagano, M.A.; Sarno, S.; Donella-Deana, A.; Zagotto, G.; Rosolen, A.; Pinna, L.A.; Meggio, F.; Moro, S. *J. Med. Chem.*, **2006**, *49*, 2363.
[t] Foloppe, N.; Fisher, L.M.; Howes, R.; Potter, A.; Robertson, A.G.; Surgenor, A.E. *Bioorg. Med. Chem.*, **2006**, *14*, 4792.
[u] http://www.mdli.com/
[v] Guichou, J.F.; Viaud, J.; Mettling, C.; Subra, G.; Lin, Y.L.; Chavanieu, A. *J. Med. Chem.*, **2006**, *49*, 900.
[w] Li, J.; Zhang, J.; Chen, J.; Luo, X.; Zhu, W.; Shen, J.; Liu, H.; Shen, X.; Jiang, H. *J. Comb. Chem.*, **2006**, *8*, 326.
[x] Li, J.; Chen, J.; Gui, C.; Zhang, L.; Qin, Y.; Xu, Q.; Zhang, J.; Liu, H.; Shen, X.; Jiang, H. *Bioorg. Med. Chem.*, **2006**, *14*, 2209.
[y] de Graaf, C.; Oostenbrink, C.; Keizers, P.H.J.; van der Wijst, T.; Jongejan, A.; Vermeulen, N.P.E. *J. Med. Chem.*, **2006**, *49*, 2417.
[z] Rummey, C.; Nordhoff, S.; Thiemann, M.; Metz, G. *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 1405.
[aa] Desai, P.V.; Patny, A.; Gut, J.; Rosenthal, P.J.; Tekwani, B.; Srivastava, A.; Avery, M.. *J. Med. Chem.*, **2006**, *49*, 1576.
[bb] Polgar, T.; Baki, A.; Szendrei, G.I.; Keseru, G.M. *J. Med. Chem.*, **2005**, *48*, 7946.
[cc] http://www.specs.net/
[dd] Bohacek, R.; Boosalis, M.S.; McMartin, C.; Faller, D.V.; Perrine, S.P. *Chem. Biol. Drug Des.*, **2006**, *67*, 318.
[ee] http://www.tripos.com/index.php?family=modules,SimplePage,sybyl_unity
[ff] Pirard *et al.* [41].
[gg] http://chembridge.com/chembridge/
[hh] Kenyon, V.; Chorny, I.; Carvajal, W.J.; Holman, T.R.; Jacobson, M.P. *J. Med. Chem.*, **2006**, *49*, 1356.
[ii] Zhang *et al.* [52].
[jj] Lu, Y.; Nikolewska-Coleska, Z.; Fang, X.; Gao, W.; Shangary, S.; Qiu, S.; Qin, D.; Wang, S.; *J. Med. Chem.*, **2006**, *49*, 3759.
[kk] http://www.chemdiv.com/
[ll] http://www.chemosoft.com/
[mm] Kim, S.Y.; Lee, Y.S.; Kang, T.; Kim, S.; Lee, J. *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 4898.
[nn] González, D.-H.; Molina-Ruiz, R.; Hernández, I. *MARCH-INSIDE version 2.0 (Markovian Chemicals "In Silico" Design)*, Chemicals Bio-actives Center, Central University of "Las Villas", Cuba, **2003**.
[oo] Santana, L.; Uriarte, E.; Gonzalez-Diaz, H.; Zagotto, G.; Soto-Otero, R.; Mendez-Alvarez, E. *J. Med. Chem.*, **2006**, *49*, 1149.

[pp] http://www.leadquest.com/
[qq] Saario *et al.* [44].
[rr] http://www.asinex.com/
[ss] http://www.accelrys.com/products/cerius2/
[tt] Manetti, F.; Magnani, M.; Castagnolo, D.; Passalacqua, L.; Botta, M.; Corelli, F.; Saddi, M.; Deidda, D.; De Logu, A. *ChemMedChem*, **2006**, *1*, 973.
[uu] Jamois, E.A.; Lin, C.T.; Waldman, M.. *J. Mol. Graph. Model.*, **2003**, *22*, 141.
[vv] Manetti, F.; Magnani, M.; Castagnolo, D.; Passalacqua, L.; Botta, M.; Corelli, F.; Saddi, M.; Deidda, D.; De Logu, A. *ChemMedChem*, **2006**, *1*, 973.
[ww] Irwin, J.J.; Raushel, F.M.; Shoichet, B.K. *Biochemistry*, **2005**, *44*, 12316.
[xx] http://almost.cubic.uni-koeln.de/cdk/cdk_top
[yy] Derksen, S.; Rau, O.; Schneider, P.; Schubert-Zsilavecz, M.; Schneider, G. *ChemMedChem*, **2006**, *1*, 1346.
[zz] Scarsi, M.; Podvinec, M.; Roth, A.; Hug, H.; Kersten, S.; Albrecht, H.; Schwede, T.; Meyer, U.A.; Rucker, C. *Mol. Pharmacol.*, **2006**, doi:10.1124/mol.106.024596.
[aaa] Lu, I.L.; Huang, C.F.; Peng, Y.H.; Lin, Y.T.; Hsieh, H.P.; Chen, C.T.; Lien, T.W.; Lee, H.J.; Mahindroo, N.; Prakash, E.; Yueh, A.; Chen, H.Y.; Goparaju, C.M.; Chen, X.; Liao, C.C.; Chao, Y.S.; Hsu, J.T.; Wu, S.Y. *J. Med. Chem.*, **2006**, *49*, 2703.
[bbb] http://www.salford-systems.com/cart.php
[ccc] http://www.spss.com/
[ddd] Yamazaki, K.; Kusunose, N.; Fujita, K.; Sato, H.; Asano, S.; Dan, A.; Kanaoka, M. *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 1371.
[eee] http://dtp.nci.nih.gov/branches/dscb/diversity_explanation.html
[fff] Rogers, J.P.; Beuscher, A.E.; Flajolet, M.; McAvoy, T.; Nairn, A.C.; Olson, A.J.; Greengard, P. *J. Med. Chem.*, **2006**, *49*, 1658.
[ggg] Pang, Y.P.; Perola, E.; Xu, K.; Prendergast, F.G. *J. Comput. Chem.*, **2001**, *22*, 1750.
[hhh] Dooley, A.J.; Shindo, N.; Taggart, B.; Park, J.G.; Pang, Y.P. *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 830.
[iii] Lu, I.; Mahindroo, N.; Liang, P.H.; Peng, Y.H.; Kuo, C.J.; Tsai, K.C.; Hsieh, H.P.; Chao, Y.S.; Wu, S.Y. *J. Med. Chem.*, **2006**, *49*, 5154.
[jjj] http://www.breadth.com.tw
[kkk] Tsai, K.C.; Chen, S.Y.; Liang, P.H.; Lu, I.L.; Mahindroo, N.; Hsieh, H.P.; Chao, Y.S.; Liu, L.; Liu, D.; Lien, W.; Lin, T.H.; Wu, S.Y. *J. Med. Chem.*, **2006**, *49*, 3485.
[lll] Renner, S.; Ludwig, V.; Boden, O.; Scheffer, U.; Gobel, M.; Schneider, G. *Chembiochem*, **2005**, *6*, 1119.
[mmm] Doddareddy, M.R.; Choo, H.; Cho, Y.S.; Rhim, H.; Koh, H.Y.; Lee, J.H.; Jeong, S.W.; Pae, A.N. *Bioorg. Med. Chem.*, **2007**, *15*, 1091.

ing and has been applied to HIV-1 protease and Cox2 screening. Compared to UNITY 3D pharmacophore fingerprints [32], a reduced dependency on the query structure and a larger diversity of the hits was observed. PharmID [33], a program designed to generate pharmacophore models, superimposes sets of small molecules by determining their best matching conformations. This can be regarded as a flexible molecule – flexible template algorithm, although the conformations of the molecules have to be pre-calculated as well. Such a tool may improve the generation of search templates significantly.

Most frequently, a subclass of ligand-based screening methods is applied that utilizes molecular descriptors and fingerprints encoding for molecular structure, pharmacophores, and physico-chemical properties (see Table **1**). A promising development in this field has been reported by Bonachera *et al.*, who introduced 2D pharmacophore fingerprints which take into account the concentration of differently protonated microspecies of a compound [34]. These fingerprints have the potential to avoid "activity cliffs" – i.e., compounds that are very similar with respect to their fingerprints, but exhibit largely different biological activity – often seen in classical descriptor spaces. Another obstacle usually associated with descriptor-based similarity searches is their limited ability for "scaffold hopping". Zhang and Muegge investigated this issue and came to the conclusion that atom pair descriptors and 3D pharmacophore fingerprints combined with consensus scoring perform well in finding novel bioactive scaffolds [35]. Notably, some hits depicted in Table **1** have been identified by using generic statistics software in combination with molecular descriptors, e.g. compounds (**22**) and (**29**). For a recent study of some statistical methods for virtual screening please read [36].

2.3. Drug-Likeness and Target-Specific Filtering of Databases

Despite all methodological differences, nearly all VS approaches include some kind of filtering before or after the actual VS run, usually implemented as cut-offs for physicochemical properties or pharmacophores. The filters are applied in order to (i) enhance the probability to find hits comprising reasonable, drug-like structures, (ii) adjust the screening results to target-specific requirements, and (iii) reduce the computational demands by pre-filtering large databases.

Most of the virtual screening applications cited in Table **1** have been performed on databases of available compounds. Since these virtual compound collections usually contain a lot of non-druglike structures, a filter ensuring at least some degree of drug-likeness is inevitable. For example, if drug-likeness is a criterion, structures like (**30**) would have to be examined carefully.

Extrapolating the study of Fink *et al.* [37] to larger molecular weights, it is expected that databases of known compounds contain a larger proportion of acyclic and aromatic structures, a much lower proportion of fused heterocycles, but a larger proportion of heteroaromatic structures compared to the unbiased, theoretical chemical space. Screening such databases, either *in-vitro* or *in-silico*, will produce sets of compounds with a similar bias. For example, nearly all hit structures shown in Table **1** contain two or more (hetero-) aryl groups. The property distribution suitable for a specific target may be rather different, requiring a filter focusing the compounds to be screened towards other skeletons.

This issue was investigated in-depth by Emanuele Perola for the target class of kinases [38]. He showed that true and false positives from kinase virtual screening, and otherwise known active molecules differ significantly: (i) true and false positives tend to have a larger molecular weight than known active compounds, (ii) false positives tend be more polar than true positives and false negatives, (iii) false positives showed an increased number of pseudo-*cis* conformations of acyclic amides, and most important (iv) false positives fail to form essential hydrogen bonds with the protein backbone. Using the last two criteria as post-filters for virtual screening, the ranking lists were dramatically down-sized, while at

the same time true positives were highly enriched. As 96.5% of 444 experimental ligand-kinase complex structures in the PDB are assigned to three types of ligand motifs which are hydrogen bonded to the kinase hinge region, Perola recommended a two step protocol for kinase virtual screening based on protein-ligand docking: first, using constrained docking into the kinase active site, enforcing two of three essential hydrogens bonds with the protein backbone. Second, prioritizing the resulting hit list according to the three preferred protein-ligand binding motifs. The author concluded that false positives are mostly generated because docking programs and scoring functions are unable to identify critical interactions and to treat them accordingly.

In general, this example highlights the importance of integrating target-specific knowledge into the virtual screening runs, either by tweaking the parameters of virtual screening, pre-filtering of databases to be screened, or enforcing specific interaction motifs during protein-ligand docking. Finally, a convenient side-effect of pre-filtering is the reduction of computing power needs by limiting the number of compounds to be screened.

2.4. Homology Models and Induced Fit

Usually the choice of a particular method for virtual screening is dictated by the data available. Uncertainties, however, remain regarding the value of homology models and the modelling of induced fit of protein-ligand complexes. With both these methods, the results of virtual screening may be compromised by potentially incorrect protein conformation.

Recently some evaluation studies concerning that issue have been published: for example, Bissantz *et al.* compared receptor-based and ligand-based methods for designing a small focused 5-HT_{2C} library starting from ~10,000 prefiltered compounds (MW $\leq$ 400 D, one basic amine, no negative charge, at least one aromatic ring, 207 active compounds) [39]. Different ligand-based similarity searches showed a large variation in enrichment of the active compounds, ranging from less-than-random to about 3.2-fold among the 2000 top-scoring molecules. Docking into a homology model of 5-HT_{2C} using different programs and scoring functions showed more consistent enrichment, ranging from 1.4-fold to 2.2-fold, and retrieved more diverse compounds. The authors concluded that one should not base a compound selection for experimental testing on a single similarity search. Evers *et al.* carried out a similar investigation on the α1A, $5HT_{2A}$, D2, and M1 receptors [40]. Again, the conclusion was that receptor-based methods achieve a somewhat lower enrichment than ligand-based methods.

Similar findings were reported for potassium channels. Pirard *et al.* compared similarity-based *vs.* ligand- and protein-based pharmacophores for virtual screening for Kv1.5 potassium channel blockers (see Table **1**) [41]. A query using a pharmacophore derived from a homology model revealed 3102 virtual hits, but did not return any known Kv1.5 blockers. Filtering (removal of reactive groups etc.) and visual inspection reduced the number of virtual hits to 244. These compounds were tested *in-vitro*, 19 of them were active, and 5 showed an IC_{50} below 10 μM. Similarity searches within the same database, using the verified hits as a query yielded a significant number of analogs for the five hits, which had not been found by the protein-based screening. The authors note that the structure-based approach outperformed ligand-based approaches in terms of hit rates and chemotypes identified. However both methods are complementary and should be applied in parallel.

Kairys *et al.* showed that there is no clear correlation between target-template similarity in homology modelling and enrichment in virtual screening [42]. The authors conclude that this is likely to be caused by details within the binding site, which are not detected by broad measures of molecular similarity. Notably, they found that docking in the template protein itself is often as successful as docking into the corresponding homology model. This is inline with the result of Alvesalo *et al.* (see Table **1** and [43]) who performed receptor-based virtual screening using FlexX on *Bacillus subtilis* RNA methyltransferase, a protein homologous to the actual target, *Chlamydia pneumoniae* dimethyladenosine transferase. A small library of 2000 top-scoring compounds was inspected visually, 33 compounds were selected for testing, and 8 turned out to be active on the actual target. This demonstrates that off-target hits – usually regarded as an annoying artefact – can be used to perform database queries for targets with unknown 3D structure. However, one has to be aware that homology model screening might lead to unexpected results as well (see e.g. Table **1** and [44]).

Induced fit of protein-ligand complexes is a phenomenon currently not considered in most virtual screening applications due to limitations in computing power and uncertainty about the effects on enrichment. In principle, the same question arises in homology modelling and induced fit modelling: is the technique really accurate enough to improve the results of virtual screening, or does it just add some noise? The general importance of induced fit is apparent from an investigation of the PDB by Boström *et al.* who concluded that – although similar ligands bind in a similar fashion in most cases – it is likely to find altered receptor conformations [45]. Sherman *et al.* reported that an induced fit virtual screening protocol improved the retrieval of known ligands of, for example, p38 MAP kinase from 3 to 14 within the top-scoring 1% of a database containing 25,000 decoy molecules [46, 47]. Moitessier *et al.* reported an induced fit docking and scoring method for β-Secretase (BACE1) which showed a promising predictive power [48]. In contrast, Polgár and Keserü concluded from an investigation using FlexE [49] with β-Secretase and JNK-3 as targets, that the incorporation of protein side-chain flexibility, and even the consideration of slight loop movements does not lead to a significant improvement in the success rate of virtual screening [50]. In general, induced fit modelling is one of the major directions of current research in virtual screening.

3. CONCLUSIONS

As demonstrated by a large number of publications, vHTS is able to generate useful hit molecules for a large variety of targets, ranging from unknown targets to targets with several protein structures being available. A variety of methods have demonstrated their utility for successful virtual screening. There is no virtual screening-specific preference for any biological assay format. So, what conclusions can be

drawn from the surveyed literature with respect to success of virtual screening?

Essential: Filtering for Drug-Likeness

Medicinal chemists usually expect the vHTS approach to yield hits which allow for a further development with respect to their chemical structures. Most of the databases available contain a significant amount of undesirable structures. For example, only compounds matching Lipinski's Rule-of-Five criteria [51] may be extracted from the chosen database prior to the vHTS run, metal complexes may be removed, or compounds containing undesired functional groups may be excluded. As a post-filter of the VS run, a visual inspection is recommendable to remove compounds which (i) are synthetically not feasible, (ii) may cause assay artefacts (fluorescent compounds, fluorescence quenchers, scintilators, reactive compounds, etc.) or (iii) are very unlikely to be true actives due to unreasonable binding modes or similar. The application of filters is not limited to databases consisting of known compounds, but extends to combinatorial libraries as well [52]. Additionally, filters can be applied in order to exclude compounds which are likely to interact with anti-targets, e.g., with the hERG channel [53].

Essential: Target-Specific Optimization

Using target-specific information to optimize the screening setup is indispensable for using virtual screening in a reliable fashion. For example, as a consequence of linear models implemented in most vHTS methods, their application leads to an artificial preference for some specific chemotypes – e.g. they may prefer a high number of H-bonds or large molecular weights [8, 38, 54] – which have to be countered by appropriate constraints, e.g. on MW, or number of donors and acceptors, and/or other suitable pre- and post-filters. This is done in basically all of the case studies cited in Table **1**. The detection of false positives from protein-ligand docking may be performed automatically as well, e.g. by using statistical methods as post-filters [55] or by applying target-specific pharmacophore constraints during screening [28]. This is very important for some target classes, e.g. kinases [38]. In general, well characterized targets allow for the design of more successful virtual screening approaches (see e.g., [56]).

Beneficial: Using Multiple Methodologies

In general, it is hard to predict which virtual screening methodology or software tool performs best for a particular target. In particular, despite all methodological advances there is still no gold standard tool for protein-ligand docking [9, 57, 58]. Applying multiple methodologies clearly limits the effect of choosing a potentially less successful approach, although merging the results of too many different screening runs may dilute the true positives within the final ranking list. For some targets, e.g., homology models, it is highly advisable to use several different types of virtual screening approaches, most commonly by combining pharmacophores and protein-ligand docking. As a result, a flexible software basis is desirable for combining different approaches and optimizing the virtual screening setup. Regarding induced fit modelling, no recommendation can be given since this methodology is still in the evaluation phase. Future methods, however, that model this effect appropriately will be highly beneficial.

Useful: Screening Databases of Commercially Available Compounds

Databases containing compounds which are either commercially available or have been described in literature differ significantly from the theoretical chemical space in terms of composition and size [37]. In the majority of the case studies published in the last year, databases consisting of commercially available compounds were screened. It is reasonable to assume that these databases were chosen for the reason of the straightforward accessibility of the compounds. In contrast to molecules from de-novo design [59], hits from databases containing available compounds allow for a "rapid prototyping", i.e., a fast experimental testing. Virtual combinatorial libraries might be regarded as an alternative to databases of known compounds. Combinatorial libraries can be constructed in such a way that they e.g., systematically cover the scope of particular structural classes. If enumeration of a combinatorial library, i.e., the generation of molecular structures for all members of the library, is necessary, its size has to be limited and some focusing is inevitable. The availability of hit compounds on the respective target – potentially obtained by screening databases of known compounds – facilitates that task.

Less Important: Target Class

The target class determines the choice of particular methods for *in-silico* screening and experimental testing, but does not determine the outcome of virtual screening. Table **1** shows that virtual screening is able to identify hits for a large variety of target classes, including receptors, channels, metallo-enzymes, protein-protein interaction, and even unknown targets. If the methodology is carefully adjusted to the intrinsic properties of a specific target, virtual screening offers a high probability for success. For example, homology models can be suitable targets, if several active compounds are known that help to validate the homology model. More surprisingly, even cross-screening approaches, e.g. reported by Alvesalo *et al.* [43], are able to deliver reasonable results.

In summary, the probability for being successful with vHTS increases with knowledge about the target, which allows to optimize virtual screening setup and parameters. Therefore virtual screening is not a technology that works out-of-the-box. However, with expertise from chemistry and biology being available, virtual screening is one of the fastest and most cost-effective methods to generate hit compounds which enable the initiation of medicinal chemistry programs, irrespective of the target class. Hence a close interaction of all disciplines in a team is a key prerequisite for exploiting the full potential of this method.

ACKNOWLEDGEMENTS

We would like to apologize to all authors whose papers have not been cited due to space limitations. We thank Daniel Vitt and Andrea Aschenbrenner for support and proofreading of the manuscript.

CONFLICT OF INTEREST

Both authors are employed by 4SC AG, a commercial venture offering vHTS.

ABBREVIATIONS

CART = Classification and regression tree

CoMFA = Comparative molecular field analysis

LDA = Linear discriminant analysis

MLR = Multiple linear regression

NN = Neural network

QSAR = Quantitative structure-activity relationship

Ro5 = Lipinski's Rule of Five

RP = Recursive portioning

vHTS = Virtual high-throughput screening

VS = Virtual screening

REFERENCES

[1] Malo, N.; Hanley, J.A.; Cerquozzi, S.; Pelletier, J.; Nadon, R. *Nat. Biotechnol.*, **2006**, *24*, 167.

[2] Klebe, G. *Drug Discov. Today*, **2006**, *11*, 580.

[3] Langer, T.; Hoffmann, R. *Expert Opin. Drug. Discov.*, **2006**, *1*, 261.

[4] Stahura, F.L.; Bajorath, J. *Curr. Pharm. Des.*, **2005**, *11*, 1189.

[5] Mohan, V.; Gibbs, A.C.; Cummings, M.D.; Jaeger, E.P.; DesJarlais, R.L. *Curr. Pharm. Des.*, **2005**, *11*, 323.

[6] Sousa, S.F.; Fernandes, P.A.; Ramos, M.J. *Proteins*, **2006**, *65*, 15.

[7] Wallqvist, A.; Huang, R.; Thanki, N.; Covell, D.G.; *J. Chem. Inf. Model.*, **2006**, *46*, 430.

[8] Ferrara, P.; Gohlke, H.; Price, D.J.; Klebe, G.; Brooks, C.L. *J. Med. Chem.*, **2004**, *47*, 3032.

[9] Warren, G.W.; 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. *J. Med. Chem.*, **2006**, *49*, 5912.

[10] Herz, T.; Wolf, K.; Kraus, J.; Kramer, B. *Exp. Opin. Drug Metabolism Toxicol.*, **2006**, *2*, 471.

[11] Shoichet, B.K.; Kuntz, I.D. *Protein Eng.*, **1993**, *6*, 723.

[12] Goodsell, D.S.; Morris, G.M.; Olson, A.J. *J. Mol. Recognit.*, **1996**, *9*, 1.

[13] Rarey, M.; Kramer, B.; Lengauer, T. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **1995**, *3*, 300.

[14] Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. *J. Mol. Biol.*, **1997**, *267*, 727.

[15] McMartin, C.; Bohacek, R.S. *J. Comput. Aided Mol. Des.*, **1997**, *11*, 333.

[16] Fernandez-Recio, J.; Totrov, M.; Abagyan, R. *Proteins*, **2003**, *52*, 113.

[17] Zavodszky, M.I.; Sanschagrin, P.C.; Korde, R.S.; Kuhn, L.A. *J. Comput. Aided Mol. Des.*, **2002**, *16*, 883.

[18] Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; Shaw, D.E.; Francis, P.; Shenkin, P.S. *J. Med. Chem.*, **2004**, *47*, 1739.

[19] McGann, M.; Almond, H.; Nicholls, A.; Grant, J.A.; Brown, F. *Biopolymers*, **2003**, *68*, 76.

[20] Jain, A.N. *J. Med. Chem.*, **2003**, *46*, 499.

[21] Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. *J. Med. Chem.*, **2006**, *49*, 6177.

[22] Degen, J.; Rarey, M. *ChemMedChem*, **2006**, *1*, 854.

[23] Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S.B.; Johnson, A.P. *J. Mol. Graph. Model.*, **2006**, *7*, 421.

[24] Thomsen, R.; Christensen, M.H. *J. Med. Chem.*, **2006**, *49*, 3315.

[25] Pei, J.; Wang, Q.; Liu, Z.; Li, Q.; Yang, K.; Lai, L. *Proteins*, **2006**, *62*, 934.

[26] Meiler, J.; Baker, D. *Proteins*, **2006**, *65*, 538

[27] Seifert, M.H.; Schmitt, F.; Herz, T.; Kramer, B. *J. Mol. Model.*, **2004**, *10*, 342

[28] Seifert, M.H. *J. Chem. Inf. Model.*, **2005**, *45*, 449.

[29] Kontoyianni, M.; Sokol, G.S.; McClellan, L.M. *J. Comput. Chem.*, **2005**, *26*, 11.

[30] Tervo, A.J.; Ronkko, T.; Nyronen, T.H.; Poso, A. *J. Med. Chem.*, **2005**, *48*, 4076.

[31] Ronkko, T.; Tervo, A.J.; Parkkinen, J.; Poso, A. *J. Comput. Aided Mol. Des.*, **2006**, *20*, 227.

[32] Tripos, Inc. 1699 South Hanley Road, St. Louis, MO 63144-2319, USA

[33] Feng, J.; Sanil, A.; Young, S.S. *J. Chem. Inf. Model.*, **2006**, *46*, 1352.

[34] Bonachéra, F.; Parent, B.; Barbosa, F.; Froloff, N.; Horvath, D. *J. Chem. Inf. Model.*, **2006**, *46*, 2457.

[35] Zhang, Q.; Muegge, I. *J. Med. Chem.*, **2006**, *49*, 1536.

[36] Plewczynski, D.; Spieser, S.A.; Koch, U. *J. Chem. Inf. Model.*, **2006**, *46*, 1098.

[37] Fink, T.; Bruggesser, H.; Reymond, JL. *Angew. Chem. Int. Ed. Engl.*, **2005**, *44*, 1504.

[38] Perola, E. *Proteins*, **2006**, *64*, 422.

[39] Bissantz, C.; Schalon, C.; Guba, W.; Stahl, M. *Proteins*, **2005**, *61*, 938.

[40] Evers, A.; Hessler, G.; Matter, H.; Klabunde, T. *J. Med. Chem.*, **2005**, *48*, 5448.

[41] Pirard, B.; Brendel, J.; Peukert, S. *J. Chem. Inf. Model.*, **2005**, *45*, 477.

[42] Kairys, V.; Fernandes, M.X.; Gilson, M.K. *J. Chem. Inf. Model.*, **2006**, *46*, 365.

[43] Alvesalo, J.K.; Siiskonen, A.; Vainio, M.J.; Tammela, P.S.; Vuorela, P.M. *J. Med. Chem.*, **2006**, *49*, 2353.

[44] Saario, S.M.; Poso, A.; Juvonen, R.O.; Järvinen, T.; Salo-Ahen, O.M.H. *J. Med. Chem.*, **2006**, *49*, 4650.

[45] Boström, J.; Hogner, A.; Schmitt, S. *J. Med. Chem.*, **2006**, *49*, 6716.

[46] Sherman, W.; Day, T.; Jacobson, M.P.; Friesner, R.A.; Farid, R. *J. Med. Chem.*, **2006**, *49*, 534.

[47] Sherman, W.; Beard, H.S.; Farid, R. *Chem. Biol. Drug Des.*, **2006**, *67*, 83.

[48] Moitessier, N.; Therrien, E.; Hanessian S. *J. Med. Chem.*, **2006**, *49*, 5885.

[49] Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. *J. Mol. Biol.*, **2001**, *308*, 377.

[50] Polgar, T.; Keseru, G.M. *J. Chem. Inf. Model.*, **2006**, *46*, 1795.

[51] Lipinski, C.L.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. *Adv. Drug Dev. Rev.*, **2001**, *46*, 3.

[52] Zhang, Y.J.; Wang, Z.; Sprous, D.; Nabioullin, R. *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 525.

[53] Aronov, A.M. *Drug Discov. Today*, **2005**, *10*, 149.

[54] Verdonk, M.L.; Berdini, V.; Hartshorn, M.J.; Mooij, W.T.; Murray, C.W.; Taylor, R.D.; Watson, P. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 793.

[55] Springer, C.; Adalsteinsson, H.; Young, M.M.; Kegelmeyer, P.W.; Roe, D.C. *J. Med. Chem.*, **2005**, *48*, 6821.

[56] Miteva, M.A.; Lee, W.H.; Montes, M.O.; Villoutreix, B.O. *J. Med. Chem.*, **2005**, *48*, 6012.

[57] Cummings, M.D.; DesJarlais, R.L.; Gibbs, A.C.; Mohan, V.; Jaeger, E.P. *J. Med. Chem.*, **2005**, *48*, 962.

[58] Chen, H.; Lyne, P.D.; Giordanetto, F.; Lovell, T.; Li, J. *J. Chem. Inf. Model.*, **2006**, *46*, 401.

[59] Schneider, G.; Fechner, U. *Nat. Rev. Drug Discov.*, **2005**, *4*, 649.

Received: 14 December, 2006 **Revised: 04 June, 2007** **Accepted: 04 June, 2007**