JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

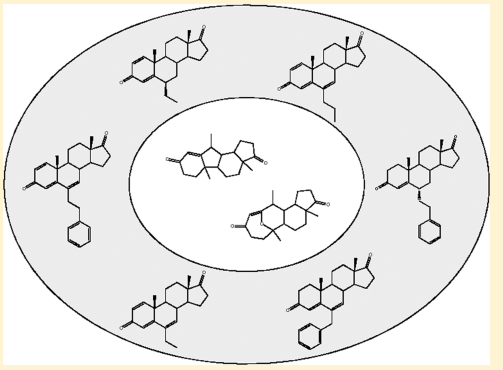
pubs.acs.org/jcim

REPROVIS-DB: A Benchmark System for Ligand-Based Virtual Screening Derived from Reproducible Prospective Applications

Peter Ripphausen, Anne Mai Wassermann, and Jürgen Bajorath*

Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Dahlmannstr. 2, D-53113 Bonn, Germany

ABSTRACT: Benchmark calculations are essential for the evaluation of virtual screening (VS) methods. Typically, classes of known active compounds taken from the medicinal chemistry literature are divided into reference molecules (search templates) and potential hits that are added to background databases assumed to consist of compounds not sharing this activity. Then VS calculations are carried out, and the recall of known active compounds is determined. However, conventional benchmarking is affected by a number of problems that reduce its value for method evaluation. In addition to often insufficient statistical validation and the lack of generally accepted evaluation standards, the artificial nature of typical benchmark settings is often criticized. Retrospective benchmark calculations generally overestimate the potential of VS methods and do not scale with their performance in prospective applications. In order to provide additional opportunities for benchmarking that more closely resemble practical VS conditions, we have designed a publicly available compound database (DB) of *repro*ducible *v*irtual *s*creens (REPROVIS-DB) that organizes information from successful ligand-based VS applications including reference compounds, screening databases, compound selection criteria, and experimentally confirmed hits. Using the currently available 25 hand-selected compound data sets, one can attempt to reproduce successful virtual screens with other than the originally applied methods and assess their potential for practical applications.



1. INTRODUCTION

For the evaluation of virtual screening (VS) methodologies, benchmark calculations are a general requirement.[1,2] Typical benchmarking in ligand-based VS (LBVS) involves the use of known compound activity classes from which reference molecules are selected. Then the remaining active molecules are added as potential hits to a background database assumed to exclusively consist of inactive compounds and searched for using the reference compounds as input. The retrieval of hits in database selection sets is monitored to assess LBVS performance. Conventional measures such as cumulative recall curves, recovery rates (i.e., the ratio of correctly retrieved active compounds and all available database hits), or hit rates (i.e., the ratio of correctly identified hits and the number of database compounds in a selection set) are often utilized to evaluate and compare the LBVS performance of different methods.[2] Despite their popularity, recovery rates are an artificial readout of VS performance, because they cannot be determined in practical (prospective) applications where the total number of available database hits is of course unknown.

If we wanted to state this case more provocatively, benchmarking could also be considered a "necessary evil" of VS, given its at least partly questionable nature. There are problems at different levels associated with conventional benchmarking, which make it difficult to accurately evaluate method performance. First, there are formal inconsistencies. Many published benchmark investigations are not reproducible because the descriptions of system setups and VS calculation protocols are insufficient or the studies involve proprietary compounds.[1] Another complication is that there are currently no generally accepted statistical validation standards,[3] which makes it often impossible to compare different methods on the basis of literature data. Second, other shortcomings exist that are intrinsic to benchmarking and perhaps even more difficult to address than formal inconsistencies. For example, activity classes assembled from the literature usually originate from advanced medicinal chemistry projects and consist of well-optimized molecules that are much more complex than average database compounds and do not resemble typical screening hits.[2] Thus, these optimized compounds are in general relatively easy to distinguish from the background database using similarity-based methods, reflecting a form of "complexity bias". In addition, such activity classes often contain large series of analogs that are also easily detected if reference compounds have been selected from these series, a problem often referred to as "analog bias" in LBVS.[1] Both effects lead to an artificial enrichment of active database compounds in LBVS calculations.[4,5] As a consequence, typical benchmark calculations, albeit necessary for method evaluation, generally overestimate the potential of LBVS methods, and benchmark performance

Received: July 7, 2011
Published: September 09, 2011

Table 1. Summary of LBVS Applications Utilized for Benchmark System Design[a]

data set	target	no. of ref compounds	no. of hits	applied LBVS method	screening database	compound selection criterion	ref
1	Myc-Max heterodimer	6	1	pharmacophore search (3D)	ZINC[43]	database ranking	Mustata et al.[18]
2	heat shock protein 90α	80	20	pharmacophore search (3D)	NCI[44]	database ranking	Al-Sha'er and Taha[19]
3	androgen receptor	5	2	pharmacophore search (3D)	Maybridge,[45] NCI	database ranking	Purushottamachar et al.[20]
4	cyclin-dependent kinase 1	23	8	pharmacophore search (3D)	NCI	database ranking	Al-Sha'er and Taha[21]
5	peroxisome proliferator-activated receptor γ	88	2	pharmacophore search (3D)	NCI	database ranking	Al-Najjar et al.[22]
6	sodium-dependent glucose cotransporter 2	8	3	pharmacophore search/ shape screen (3D)	Chembridge,[46] Maybridge, Specs[47]	database ranking	Wu et al.[23]
7	NAADP receptor	1	1	shape screen (3D)	ZINC	database ranking	Naylor et al.[24]
8	metabotropic glutamate receptor 1	6	1	fingerprint search (2D)	Asinex[48]	database ranking	Noeske et al.[25]
9	5α-reductase	16	1	pharmacophore search (3D)	NCI	expert knowledge	Chen et al.[26]
10	NR2B-selective NMDA receptor	16	1	pharmacophore search (3D)	InterBioScreen[49]	expert knowledge	Mony et al.[27]
11	T-type calcium channel	8	1	pharmacophore search (3D)	ChemDiv,[50] Maybridge	expert knowledge	Doddareddy et al.[28]
12	aromatase	6	3	pharmacophore search/ shape screen (3D)	NCI	expert knowledge	Neves et al.[29]
13	NK3 subtype of tachykinin receptor	33	1	shape screen (3D)	ZINC	expert knowledge	Geldenhuys et al.[30]
14	P90 ribosomal S6 protein kinase 2	2	12	shape screen (3D)	Specs, Maybridge	expert knowledge	Lu et al.[31]
15	L-type calcium channel	1	6	fingerprint search (3D)	Sigma-Aldrich,[51] BIONET,[52] Maybridge, Menai,[53] Peakdale[54]	expert knowledge	Carosati et al.[32]
16	L-type calcium channel	2	9	grid-based search (3D)	Sigma-Aldrich	expert knowledge	Carosati et al.[33]
17	5-lipoxygenase	43	2	fingerprint search (2D)	MEGx,[55] NATx,[56]	expert knowledge	Franke et al.[34]
18	endothelin-A receptor	21	2	pharmacophore search (3D)	Maybridge	unknown	Funk et al.[35]
19	autotaxin	8	17	pharmacophore search (3D)	NCI	unknown	North et al.[36]
20	dopamine transporter	5	2	pharmacophore search (3D)	NCI	unknown	Enyedy et al.[37]
21	α7 nicotinic receptor	6	2	pharmacophore search (3D)	Enamine,[57] Chembridge, Maybridge	unknown	Peng et al.[38]
22	microsomal prostaglandin E_2 synthase-1	6	9	pharmacophore search (3D)	NCI, Specs	unknown	Waltenberger et al.[39]
23	P-glycoprotein	26	8	pharmacophore search (3D)	DrugBank[58]	unknown	Palmeira et al.[40]
24	IκB kinase β	4	1	pharmacophore search/ shape screen (3D)	NCI	unknown	Noha et al.[41]
25	methionyl-tRNA synthetase	1	1	fingerprint search(2D)	ChemDiv[59]	unknown	Kim et al.[42]

[a] For the 25 benchmarks set available in REPROVIS-DB, target protein names ("target") and the numbers of reference compounds and hits are given. If additional hits with >10 μM potency were reported in these studies, they were not considered here. In addition, for each data set, the originally applied LBVS strategy ("applied LBVS method"), the screening database, and the primary compound selection criterion are reported. Furthermore, the number of each data set entry and the original literature reference from which all information was extracted are provided. Data sets are grouped on the basis of compound selection criteria.

dx.doi.org/10.1021/ci200309j | J. Chem. Inf. Model. 2011, 51, 2467–2473

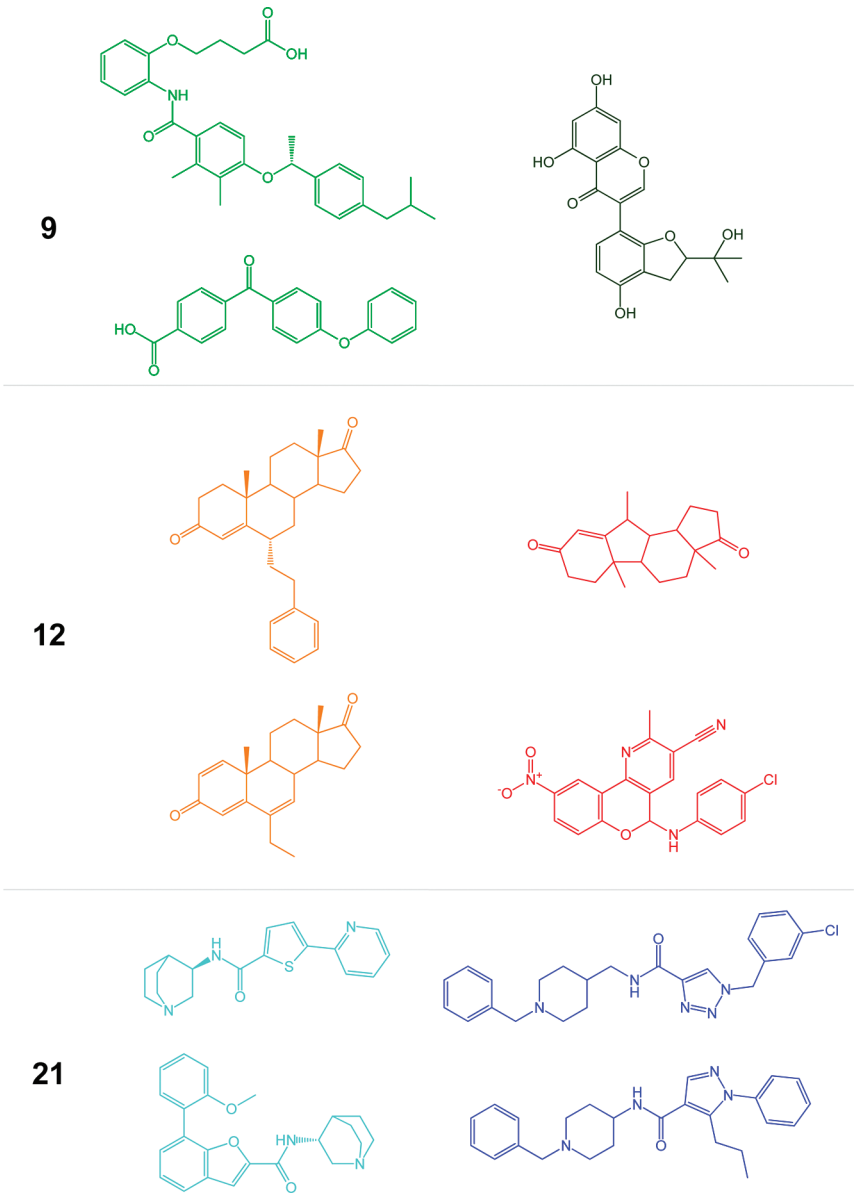


Figure 1. Exemplary compounds. Examples of reference compounds and hits active against different targets are shown that are available in REPROVIS-DB. Compounds are annotated with the numeric identifier of the data set from which they are taken. Reference compounds are shown in light colors on the left and hits in dark colors on the right. The following color code is applied: green, 5α-reductase inhibitors; orange/red, aromatase inhibitors; blue, α7 nicotinic receptor antagonists.

does not scale with the much lower success rates of LBVS observed in prospective applications.[2]

These types of problems associated with benchmarking have of course been recognized by investigators in this field and various attempts have been made to improve the basis for method evaluation and comparison. These efforts can essentially be divided into two categories, i.e., the design of meaningful performance metrics and specialized data sets. For example, statistical measures based on the receiver−operator characteristic (ROC)[6] and the area under the ROC curve (AUC)[6] have been adapted for the evaluation of VS calculations.[7−9] ROC AUC measures are well-established standards in other research fields, and their use is increasingly and positively impacting the VS field. Furthermore, a number of compound data sets have specifically been designed for benchmarking purposes consisting of, for example, compounds with suitable molecular properties, structural characteristics, and/or scaffold distributions.[10−12] Data sets with well-defined and balanced scaffold content provide the basis for a meaningful assessment of the scaffold-hopping[13] potential of LBVS. Importantly, advanced performance metrics and specifically designed data sets are utilized in attempts to further improve the reproducibility and relevance of benchmark investigations, but they do not provide conceptional alternatives to conventional benchmarking.

We have investigated potential ways to generate alternative benchmark systems that might improve the practical relevance of LBVS test calculations. Therefore, we have searched for and identified fully reproducible LBVS applications that have led to the identification of novel hits and organized the compound and methodological information associated with these studies in a freely accessible database to enable benchmarking.

dx.doi.org/10.1021/ci200309j |J. Chem. Inf. Model. 2011, 51, 2467–2473

2. VIRTUAL SCREENING DATA

In 2010, we have carried out a literature survey of prospective structure-based VS and LBVS applications that successfully identified novel active compounds.[14] On the basis of a detailed analysis of original literature data, the success rates of practical VS applications have been determined and preferred methods identified.[14] For LBVS applications, we also analyzed how rigorously computational predictions were experimentally assessed.[15]

For our current investigation, the survey of prospective LBVS applications has been updated and further refined. All relevant publications were extracted from a total of 12 established journals (with a 2009 impact factor of at least 2) including journals where LBVS applications most frequently appear (such as the *Journal of Medicinal Chemistry, Journal of Chemical Information and Modeling*, or *Bioorganic Medicinal Chemistry*) and high-impact journals where such studies are occasionally published (for example, *Nature Chemical Biology* or *Angewandte Chemie, International Edition*). LBVS publications were initially identified on the basis of keyword searches and then individually examined, which resulted in a high level of confidence of the extracted information and compound data.

We selected publications reporting the identification of active compounds with <10 μM potency and core structures distinct from reference molecules. For core structure comparisons, graph frameworks[16] were extracted from active compounds to evaluate the presence of topologically distinct scaffolds.

As of June 01, 2011, a total of 132 prospective LBVS publications were collected that qualified for our subsequent analysis.

3. REPRODUCIBLE VIRTUAL SCREENS

This pool of original publications was then further analyzed to identify LBVS investigations that were fully reproducible. Specifically, we required that 2D structures of (1) reference compounds and (2) newly identified hits (with confirmed potency values) were reported, (3) the source database utilized for screening was specified and publicly available, and (4) the LBVS method(s) and calculation protocols were provided and understandable. Furthermore, we only considered studies further that (5) utilized fewer than 100 reference compounds (to avoid large amounts of redundant compound information). From publications meeting these criteria, compound selection information was also extracted. Here we were specifically interested in determining whether compounds were selected for testing only on the basis of calculated database ranks or by taking visual inspection or other subjective/knowledge-based criteria into account. We also confirmed that all reported hits were tested in target-relevant assays.

A total of 25 studies were identified that met all of our reproducibility criteria. These LBVS investigations representing successful scaffold-hopping exercises (vide supra) are reported in Table 1.

The majority of studies that did not qualify were not reproducible because proprietary or nonaccessible compound collections were screened. In other instances, reference molecules were not disclosed or it was not possible to rationalize the LBVS approaches or protocols that were utilized.

4. COMPOUND DATA SETS

Table 1 also describes the compound data sets we incorporated into our benchmark system. These 25 sets cover a wide variety of target receptors or enzymes and contain between 1 and 88 reference compounds and between 1 and 20 qualifying hits. In total, the sets contain 421 reference and 116 newly active compounds. Representative examples are shown in Figure 1. In five cases, only one or two reference compounds are available, which limits the range of applicable LBVS methods to approaches for similarity searching. For nine other sets, between 16 and 88 reference molecules are available, which also provides a sound basis for the application of machine learning methods. The number of newly identified active compounds is generally small, which is characteristic of many practical LBVS applications (but in marked contrast to hit rates typically achieved in standard benchmark calculations). In 17 cases, between one and three hits (with potency values <10 μM) are available. However, there are also seven sets containing eight to 20 hits. As also reported in Table 1, the LBVS applications involve 17 publicly available compound collections, with the NCI and Maybridge databases most frequently used.

From a methodological point of view, 3D approaches clearly dominate these reproducible virtual screens. In 22 of 25 cases, 3D methods are applied and in only three instances 2D fingerprints. Among the different 3D approaches, i.e., pharmacophore, shape, fingerprint, and grid searches, 3D pharmacophore modeling is by far the most frequently seen (often in combination with QSAR analysis). A total of 14 3D pharmacophore searches are available, three combinations of pharmacophore and shape searching, three shape screens, one 3D grid, and one 3D fingerprint search. Thus, given the dominance of 3D approaches among these LBVS investigations, in particular pharmacophore modeling and searching, many of the associated compound sets should also provide interesting test cases for different types of 2D methods.

In addition to methodological aspects, other distinguishing features of these test cases are differences in compound selection strategies. Here three categories are observed. The first eight data sets in Table 1 provide examples of hit identification on the basis of high database rank positions, which means that the most highly ranked candidate compounds were selected for testing. This is followed by nine examples where candidate compounds were selected on the basis of expert knowledge from rankings and eight cases where compound selection strategies were not reported in the original publications. Thus, although all 25 data sets present reproducible test cases on the basis of their compound and database information, the first eight data sets in Table 1 provide essentially unbiased test cases for method comparison because database rankings are primarily considered for compound selection. The addition of expert knowledge to this process introduces a hit identification bias (of undefined but likely varying magnitude). Hence, for LBVS methods yielding compound rankings, achieving high ranks for hits provided with our 25 benchmark sets should be considered a success, especially in those instances where knowledge-based selection was involved or no information provided.

5. BENCHMARK SYSTEM DESIGN

The 25 sets described above were assembled in an online database. In each case, we regenerated 2D structures of reference compounds and hits from the original publication and provide them as connection tables and canonical SMILES[17] representations in an SD-file. In REPROVIS-DB, each data set entry also contains the target specification, original publication information, and potency values of hits. In addition, each entry provides

the source database and, wherever possible, a link to the version of the database that was screened in the original investigation. For some of the older studies, outdated source database versions are no longer publicly available online. Thus, if desired, these versions should be requested from the providers on the basis of original publication dates. Moreover, the database entries contain compound selection information according to Table 1 to help prioritize test cases for method comparison.

6. DISCUSSION

The ultimate goal in VS is to prospectively apply computational methods and experimentally evaluate putative hits. Recurrent success in practical hit identification without doubt assigns credence to a VS approach, regardless of benchmark performance. However, frequent experimental assessment of VS calculations is infeasible for many investigators who develop computational methods. Hence, despite their often questionable nature, benchmark calculations are an essential part of VS method evaluation. Common problems associated with benchmarking include the lack of generally accepted evaluation standards and, in many cases, proper statistical validation of benchmark calculations. These issues are independent of the test systems under investigation but specifically linked to benchmark protocols. However, many compound systems utilized for benchmarking are also problematic because they consist of highly optimized activity classes, often containing many analog series, which are added to background databases thought to be inactive. In these instances, analog bias and also complexity bias usually lead to an overestimation of LBVS performance. Complexity bias favors the preferential detection of complex compounds over average database molecules using similarity-based methods, regardless of bioactivity criteria.[2] Compound data sets with defined scaffold compositions, which have been designed for LBVS benchmark studies, typically limit analog bias, but not complexity bias, because these data sets usually also contain compounds taken from standard activity classes.

With the introduction of REPROVIS-DB, we have aimed to further extend the spectrum of currently available benchmark systems, with a special emphasis on more realistic LBVS method performance and practical application potential. Successful and reproducible LBVS scaffold-hopping applications provide a valuable resource for method evaluation because they mirror the state-of-the-art in the field and control complexity and analog bias better than artificial system setups. This is the case because newly identified hits originate from the source databases and have previously not been known to be active and usually have not been subjected to extensive chemical optimization efforts. The design idea of REPROVIS-DB also suggests considering an in part modified benchmark philosophy. Using this system, we can in principle not arrive at the conclusion that a new method would be superior to already existing ones. However, in successful cases, one can conclude that a given approach reaches current state-of-the-art performance levels and, importantly, has practical application potential. At the same time, failure to reproduce LBVS results using another method does not per se disqualify the potential of the approach, because it is well appreciated that there is often only little overlap between active compounds identified by alternative LBVS methods.[2] Hence, utilizing the REPROVIS-DB data sets provides a basis for what we would consider a form of "positive control" benchmarking, a previously unconsidered opportunity. The relatively small numbers of hits identified in most of the collected LBVS applications (even when many reference compounds were available) represent the state-of-the-art in the field, which should be taken into consideration in method evaluation. This has some consequences. For example, conventional statistical measures in benchmarking, such as the calculation of recall characteristics, are often not applicable in these instances. Rather, a detailed analysis of rank positions of active compounds will be more relevant. The data sets contained in REPROVIS-DB also provide realistic test cases for the comparison of 2D and 3D LBVS methods, an often discussed issue in the field. Another attractive feature of this benchmark system is its high level of data confidence. All information contained in REPROVIS-DB has been carefully extracted from original sources, and the database is manually curated. It is hoped that many investigators in the VS arena will find this system a useful asset to widen the scope of conventional benchmarking.

REPROVIS-DB is freely available via the following URL: http://www.lifescienceinformatics.uni-bonn.de (see "Downloads").

AUTHOR INFORMATION

Corresponding Author

*Tel: +49-228-2699-306. Fax: +49-228-2699-341. E-mail: bajorath@bit.uni-bonn.de.

ACKNOWLEDGMENT

The authors thank Dagmar Stumpfe for her help with representations and Mathias Wawer for helpful discussions.

REFERENCES

(1) Geppert, H.; Vogt, M.; Bajorath, J. Current Trends in Ligand-Based Virtual Screening: Molecular Representations, Data Mining Methods, New Application Areas, and Performance Evaluation. *J. Chem. Inf. Model.* **2010**, *50*, 205–216.

(2) Stumpfe, D.; Bajorath, J. Applied Virtual Screening: Strategies, Recommendations, and Caveats. In *Methods and Principles in Medicinal Chemistry. Virtual Screening. Principles, Challenges, and Practical Guidelines*; Sotriffer, C., Ed.; Wiley-VCH: Weinheim, 2011; pp 73−103.

(3) Jain, A.; Nicholls, A. Recommendations for Evaluation of Computational Methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133–139.

(4) Good, A. C.; Hermsmeier, M. A.; Hindle, S. A. Measuring CAMD Technique Performance: A Virtual Screening Case Study in the Design of Validation Experiments. *J. Comput.-Aided. Mol. Des.* **2004**, *18*, 529–536.

(5) Good, A.; Oprea, T. Optimization of CAMD Techniques 3. Virtual Screening Enrichment Studies: A Help or Hindrance in Tool Selection? *J Comput. Aided Mol. Des.* **2008**, *22*, 169–178.

(6) Witten, I. H.; Frank, E. *Data Mining − Practical Machine Learning Tools and Techniques*, ed. 2. Morgan Kaufmann: San Francisco, 2005, pp 161−176.

(7) Nicholls, A. What Do We Know and When Do We Know It? *J Comput. Aided Mol. Des.* **2008**, *22*, 239–255.

(8) Truchon, J.; Bayly, C. I. Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem. *J. Chem. Inf. Model.* **2007**, *47*, 488–508.

(9) Clark, R.; Webster-Clark, D. Managing Bias in ROC Curves. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 141–146.

(10) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.

(11) Vogt, M.; Stumpfe, D.; Geppert, H.; Bajorath, J. Scaffold Hopping Using Two-Dimensional Fingerprints: True Potential, Black Magic, or a Hopeless Endeavor? Guidelines for Virtual Screening. *J. Med. Chem.* **2010**, *53*, 5707–5715.

(12) Rohrer, S. G.; Baumann, K. Impact of Benchmark Data Set Topology on the Validation of Virtual Screening Methods: Exploration and Quantification by Spatial Statistics. *J. Chem. Inf. Model.* **2008**, *48*, 704–718.

(13) Schneider, G.; Schneider, P.; Renner, S. Scaffold-Hopping: How Far Can You Jump. *QSAR Comb. Sci.* **2006**, *25*, 1162–1171.

(14) Ripphausen, P.; Nisius, B.; Peltason, L.; Bajorath, J. Quo Vadis, Virtual Screening? A Comprehensive Survey of Prospective Applications. *J. Med. Chem.* **2010**, *53*, 8461–8467.

(15) Ripphausen, P.; Nisius, B.; Bajorath, J. State-of-the-Art in Ligand-Based Virtual Screening. *Drug Discovery Today* **2011**, *16*, 372–376.

(16) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.

(17) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.

(18) Mustata, G.; Follis, A. V.; Hammoudeh, D. I.; Metallo, S. J.; Wang, H.; Prochownik, E. V.; Lazo, J. S.; Bahar, I. Discovery of Novel Myc-Max Heterodimer Disruptors with a Three-Dimensional Pharmacophore Model. *J. Med. Chem.* **2009**, *52*, 1247–1250.

(19) Al-Sha'er, M. A.; Taha, M. O. Elaborate Ligand-Based Modeling Reveals New Nanomolar Heat Shock Protein 90α Inhibitors. *J. Chem. Inf. Model.* **2010**, *50*, 1706–1723.

(20) Purushottamachar, P.; Khandelwal, A.; Chopra, P.; Maheshwari, N.; Gediya, L. K.; Vasaitis, T. S.; Bruno, R. D.; Clement, O. O.; Njar, V. C. O. First Pharmacophore-Based Identification of Androgen Receptor Down-Regulating Agents: Discovery of Potent Anti-Prostate Cancer Agents. *Bioorg. Med. Chem.* **2007**, *15*, 3413–3421.

(21) Al-Sha'er., M. A.; Taha, M. O. Discovery of Novel CDK1 Inhibitors by Combining Pharmacophore Modelling, QSAR Analysis and in Silico Screening Followed by in Vitro Bioassay. *Eur. J. Med. Chem.* **2010**, *45*, 4316–4330.

(22) Al-Najjar, B. O.; Wahab, H. A.; Muhammad, T. S. T.; Shu-Chien, A. C.; Noruddin, N. A. A.; Taha, M. O. Discovery of New Nanomolar Peroxisome Proliferator-Activated Receptor γ Activators via Elaborate Ligand-Based Modeling. *Eur. J. Med. Chem.* **2011**, *46*, 2513–2529.

(23) Wu, J.-S.; Peng, Y.-H.; Wu, J.-M.; Hsieh, C.-J.; Wu, S.-H.; Coumar, M. S.; Song, J.-S.; Lee, J.-C.; Tsai, C.-H.; Chen, C.-T.; Liu, Y.-W.; Chao, Y.-S.; Wu, S.-Y. Discovery of Non-Glycoside Sodium-Dependent Glucose Co-Transporter 2 (SGLT2) Inhibitors by Ligand-Based Virtual Screening. *J. Med. Chem.* **2010**, *53*, 8770–8774.

(24) 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*, 220–226.

(25) Noeske, T.; Jirgensons, A.; Starchenkovs, I.; Renner, S.; Jaunzeme, I.; Trifanova, D.; Hechenberger, M.; Bauer, T.; Kauss, V.; Parsons, C. G.; Schneider, G.; Weil, T. Virtual Screening for Selective Allosteric mGluR1 Antagonists and Structure−Activity Relationship Investigations for Coumarine Derivatives. *ChemMedChem* **2007**, *2*, 1763–1773.

(26) Chen, G. S.; Chang, C.-S.; Kan, W. M.; Chang, C.-L.; Wang, K. C.; Chern, J.-W. Novel Lead Generation through Hypothetical Pharmacophore Three-Dimensional Database Searching: Discovery of Isoflavonoids as Nonsteroidal Inhibitors of Rat 5α-Reductase. *J. Med. Chem.* **2001**, *44*, 3759–3763.

(27) Mony, L.; Triballeau, N.; Paoletti, P.; Acher, F. C.; Bertrand, H.-O. Identification of a Novel NR2B-Selective NMDA Receptor Antagonist Using a Virtual Screening Approach. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 5552–5558.

(28) Doddareddy, M. R.; Choo, H.; Cho, Y. S.; Rhim, H.; Koh, H. Y.; Lee, J.-H.; Jeong, S.-W.; Pae, A. N. 3D Pharmacophore Based Virtual Screening of T-type Calcium Channel Blockers. *Bioorg. Med. Chem.* **2007**, *15*, 1091–1105.

(29) Neves, M. A. C.; Dinis, T. C. P.; Colombo, G.; Sáe Melo, M. L. An Efficient Steroid Pharmacophore-Based Strategy To Identify New Aromatase Inhibitors. *Eur. J. Med. Chem.* **2009**, *44*, 4121–4127.

(30) Geldenhuys, W. J.; Kuzenko, S. R.; Simmons, M. A. Virtual Screening To Identify Novel Antagonists for the G Protein-Coupled NK_3 Receptor. *J. Med. Chem.* **2010**, *53*, 8080–8088.

(31) Lu, W.; Liu, X.; Cao, X; Xue, M.; Liu, K.; Zhao, Z.; Shen, X.; Jiang, H.; Xu, Y.; Huang, J.; Li, H. SHAFTS: A Hybrid Approach for 3D Molecular Similarity Calculation. 2. Prospective Case Study in the Discovery of Diverse p90 Ribosomal S6 Protein Kinase 2 Inhibitors to Suppress Cell Migration. *J. Med. Chem.* **2011**, *54*, 3564–3574.

(32) Carosati, E.; Cruciani, G.; Chiarini, A.; Budriesi, R.; Ioan, P.; Spisani, R.; Spinelli, D.; Cosimelli, B.; Fusi, F.; Frosini, M.; Matucci, R.; Gasparrini, F.; Ciogli, A.; Stephens, P. J.; Devlin, F. J. Calcium Channel Antagonists Discovered by a Multidisciplinary Appraoch. *J. Med. Chem.* **2006**, *49*, 5206–5216.

(33) Carosati, E.; Budriesi, R.; Ioan, P.; Ugenti, M. P.; Frosini, M.; Fusi, F.; Corda, G.; Cosimelli, B.; Spinelli, D.; Chiarini, A.; Cruciani, G. Discovery of Novel and Cardioselective Diltiazem-like Calcium Channel Blockers via Virtual Screening. *J. Med. Chem.* **2008**, *51*, 5552–5565.

(34) Franke, L.; Schwarz, O.; Müller-Kuhrt, L.; Hoernig, C.; Fischer, L.; George, S.; Tanrikulu, Y.; Schneider, P.; Werz, O.; Steinhilber, D.; Schneider, G. Identification of Natural-Product-Derived Inhibitors of 5-Lipoxygenase Activity by Ligand-Based Virtual Screening. *J. Med. Chem.* **2007**, *50*, 2640–2646.

(35) Funk, O. F.; Kettmann, V.; Drimal, J.; Langer, T. Chemical Function Based Pharmacophore of Endothelin-A Selective Receptor Antagonists. *J. Med. Chem.* **2004**, *47*, 2750–2760.

(36) North, E. J.; Howard, A. L.; Wanjala, I. W.; Pham, T. C. T.; Baker, D. L.; Parrill, A. L. Pharmacophore Development and Application Toward the Identification of Novel, Small-Molecule Autotaxin Inhibitors. *J. Med. Chem.* **2010**, *53*, 3095–3105.

(37) Enyedy, I. J.; Sakamuri, S.; Zaman, W. A.; Johnson, K. M.; Wang, S. Pharmacophore-Based Discovery of Substituted Pyridines as Novel Dopamine Transporter Inhibitors. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 513–517.

(38) Peng, Y.; Zhang, Q.; Snyder, G. L.; Zhu, H.; Yao, W.; Tomesch, J.; Papke, R. L.; O'Callaghan, J. P.; Welsh, W. J.; Wennogle, L. P. Discovery of Novel Alpha7 Nicotinic Receptor Antagonists. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 4825–4830.

(39) Waltenberger, B.; Wiechmann, K.; Bauer, J.; Markt, P.; Noha, S. M.; Wolber, G.; Rollinger, J. M.; Werz, O.; Schuster, D.; Stuppner, H. Pharmacophore Modeling and Virtual Screening for Novel Acidic Inhibitors of Microsomal Prostaglandin E_2 Synthase-1 (mPGES-1). *J. Med. Chem.* **2011**, *54*, 3163–3174.

(40) Palmeira, A.; Rodrigues, F.; Sousa, E.; Pinto, M.; Vasconcelos, M. H.; Fernandes, M. X. New Uses for Old Drugs: Pharmacophore-Based Screening for the Discovery of P-Glycoprotein Inhibitors. *Chem. Biol. Drug Des.* **2011**, *78*, 57–72.

(41) Noha, S. M.; Atanasov, A. G.; Schuster, D.; Markt, P.; Fakhrudin, N.; Heiss, E. H.; Schrammel, O.; Rollinger, J. M.; Stuppner, H.; Dirsch, V. M.; Wolber, G. Discovery of a Novel IKK-β Inhibitor by Ligand-Based Virtual Screening Techniques. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 577–583.

(42) Kim, S. Y.; Lee, Y.-S.; Kang, T.; Kim, S.; Lee, J. Pharmacophore-Based Virtual Screening: The Discovery of Novel Methionyl-tRNA Synthetase Inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4898–4907.

(43) Irwin, J. J.; Shoichet, B. K. ZINC. A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.

(44) Milne, G. W. A.; Nicklaus, M. C.; Driscoll, J. S.; Wang, S.; Zaharevitz, D. W. National Cancer Institute Information System 3D Database. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1219–1224.

(45) *MayBridge Screening Collection*; Maybridge Chemicals Company Ltd.: Trevillet, Tintangel, Cornwall, U.K; http://www.maybridge.com/ (accessed July 5, 2011).

(46) *ChemBridge*; ChemBridge Corp.: San Diego, CA, http://www.chembridge.com/ (accessed July 5, 2011).

(47) *Specs Screening Compounds*; Specs: Delft, The Netherlands; http://www.specs.net/ (accessed July 5, 2011).

(48) *Asinex Gold Collection*; Asinex Ltd.: Moscow, Russia; http://www.asinex.com/library-gold.html (accessed July 5, 2011).
(49) *InterBioScreen Libraries*; InterBioScreen Ltd.: Moscow, Russia; http://www.ibscreen.com/ (accessed July 5, 2011).
(50) *ChemDiv diversity set*; ChemDiv Inc.: San Diego, CA; http://www.us.chemdiv.com/ (accessed July 5, 2011).
(51) *Sigma-Aldrich*; Sigma-Aldrich Inc.: St. Louis, MO; http://www.sigmaaldrich.com/ (accessed July 5, 2011).
(52) *BIONET*; Key Organics Ltd: Camelford, Cornwall, U.K.; http://www.keyorganics.co.uk/ (accessed July 5, 2011).
(53) *Menai Organics*; Menai Organics Ltd.: Bangor, U.K.; http://www.menaiorganics.co.uk/ (accessed July 5, 2011).
(54) *Peakdale Molecular*; Peakdale Molecular Ltd.: Chapel-en-le-Frith, PeakDistrict, U.K.; http://peakdale.co.uk/ (accessed July 5, 2011).
(55) *MEGx collection*; AnalytiCon Discovery GmbH: Potsdam, Germany; http://ac-discovery.com/ (accessed July 5, 2011).
(56) *NATx compound library*; AnalytiCon Discovery GmbH: Potsdam, Germany; http://ac-discovery.com/ (accessed July 5, 2011).
(57) *Enamine*; Enamine Ltd.: Kiev, Ukraine; http://www.enamine.net/ (accessed July 5, 2011).
(58) 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*, D901–906.
(59) *ChemDiv ion channel focused library*; ChemDiv Inc.: San Diego, CA; http://www.us.chemdiv.com/ (accessed July 5, 2011).