FAF-Drugs: free ADME/tox filtering of compound collections

Maria A. Miteva, Stephanie Violas, Matthieu Montes, David Gomez[1], Pierre Tuffery[1] and Bruno O. Villoutreix*

Inserm U648, Paris 5 University, 45 rue des Sts Peres, 75006 Paris, France and [1]INSERM U726, EBGM, University Paris 7, France

ABSTRACT

In silico screening based on the structures of the ligands or of the receptors has become an essential tool to facilitate the drug discovery process but compound collections are needed to carry out such *in silico* experiments. It has been recognized that absorption, distribution, metabolism, excretion and toxicity (ADME/tox) are key properties that need to be considered early on, even during the database preparation stage. FAF-Drugs is an online service based on Frowns (a chemoinformatics toolkit) that allows users to process their own compound collections via simple ADME/Tox filtering rules such as molecular weight, polar surface area, logP or number of rotatable bonds. SMILES (Simplified Molecular Input Line Entry System), CANSMILES (canonical smiles) or SDF (structure data file) files are required as input and molecules that pass or do not pass the filters are sent back in CANSMILES format. This service should thus help scientists engaging in drug discovery campaigns. Other utilities and several compound collections suitable for *in silico* screening are available at our site. FAF-Drugs can be accessed at http://bioserv.rpbs.jussieu.fr/FAFDrugs.html.

INTRODUCTION

Drug discovery is a complex and expensive endeavor that usually requires seven major steps: disease selection, target hypothesis, lead compound identification (screening), lead optimization, pre-clinical trial, clinical trial and pharmacogenomic optimization. Among the various techniques used to facilitate the drug discovery process, virtual or *in silico* ligand screening (VLS) based on the structure of known ligands or

on the structure of the receptor is becoming a method of choice (1–11), as seen in several recent studies [reviewed in (12–16)]. All these investigations require suitable compound collections. It has been suggested that these libraries of purchasable small organic compounds should be filtered [ADME/tox (absorption, distribution, metabolism, excretion and toxicity) filtering] in an attempt to work with databases of molecules with acceptable physical properties and chemical functionalities, at least consistent with known drug profiles (17–27). Common filtering protocols can be variations of Lipinski's rule-of-five (or RO5, potential for oral bioavailability) (25): molecular weight (MW) (poor absorption is observed if MW is more than 500), computed log P (P = octanol/water partition coefficient) (should not be more than 5), H-bond donors (should not be more than 5) and H-bond acceptors (should not be more than 10). Filters can also include a limit on the number of rotatable bonds, on the polar surface area (a value correlated to the number of H-bond donors and acceptors) among others, or can remove compounds containing specific chemical substructures associated with poor chemical stability or toxicity and sometimes attempt to predict drug metabolism (e.g. cytochrome-mediated metabolism, Pgp efflux) (28–32). The selected molecules after applying Lipinski's RO5 or related filters based on physicochemical properties or investigation of chemical functionalities are erroneously called 'drug-like' while in fact, many organic compounds conform to the above listed rules but they are by no means drug-like (33). In fact, these rules define only some necessary conditions for a drug candidate (such as likely solubility, bio-availability) but not sufficient ones. Different levels of filtering could be applied in agreement with the aims of the project. For instance, soft filtering protocols are usually appropriate for cancer projects while, for some other studies, only small and rigid compounds/fragments (low MW, few rotatable bonds) are needed (e.g. fragment-based lead discovery projects or fraganomics) (34). Only few online ADME/tox tools are available, they can usually evaluate one compound at a

Table 1. Example of free online ADME/tox tools

http://www.molinspiration.com/	One molecule at a time
http://www.molsoft.com/	One molecule at a time
http://www.chemaxon.com/	One molecule at a time
http://zinc.docking.org (37)	All molecules—ADME/tox filtering with Filter (OpenEye)
http://www.syrres.com/	Computation of logP (one molecule at a time), Syracuse Research Corporation
http://www.logp.com/	Computation of logP (one molecule at a time)
http://146.107.217.178/lab/alogps/ (42)	Computation of logP (one molecule at a time)

Table 2. Some online compound collections

http://chembank.med.harvard.edu (44)	Free collections
http://www.cermn.unicaen.fr/chimiotheque	Free collections
http://www.genome.ad.jp/dbget/ligand.html (45)	Free collections
http://Ligand.info (46)	Free collections
http://zinc.docking.org (37)	Free and ADME/tox filtered collections
http://bioserv.rpbs.jussieu.fr/FAFDrugs.html	Free and ADME/tox filtered collections
http://bioweb.ucr.edu/ChemMine (47)	Free collections
http://www.mdli.com	Available chemicals directory commercial collections—ADME/tox filtering possible
http://www.chemnavigator.com	Commercial collections
http://www.ebi.ac.uk/chebi/	Dictionary of small molecules
http://www.bindingdb.org/ (48)	Measured binding affinities
http://www.pdbbind.org/ (49)	Proteins with co-crystallized ligands and experimental binding affinities
http://kibank.iis.u-tokyo.ac.jp/ (50)	Proteins with co-crystallized ligands and experimental binding affinities

time (Table 1) while commercial packages are in general expensive [see review (35)]. Compound libraries can be found online, but they are usually not free nor filtered (Table 2) (36). Only recently, a free 3D database of compounds ready for VLS projects has been reported, ZINC: http://zinc.docking.org (37). It is also possible to perform ADME/tox computations via ZINC and in this case they are carried out by the program Filter (OpenEye Scientific Software, a program to remove undesirable molecules based on physicochemical properties and about 100 rules to eliminate unstable/reactive/dye chemical groups as well as to desalt the molecules). For the time being, the users of ZINC can only apply default thresholds for the various computed properties.

Because ADME/tox calculations are usually not available online, we have created FAF-Drugs, a tool to perform physicochemical filtering. Also, in order to make VLS experiments easier to perform to a broad community of users, we have interfaced several additional utilities (such as binding site prediction, OpenBabel...) and processed five major compound collections.

METHODS AND IMPLEMENTATION

ADME/tox filters

We use Frowns (developed by Brian Kelley), a chemoinformatics toolkit (http://frowns.sourceforge.net/) written in Python and C++ to parse/read SMILES (see explanations about the format at http://www.daylight.com/) or SDF files (see format at Molecular Design Limited).

We have implemented an algorithm in Python that make use of Frowns features to compute properties known to be important for filtering databases and that utilizes Xtool (38) to compute log *P*-values.

Because salts and counterions are often present in compound collections we recommend users to first apply the desalt utility that removes most salts and counterions prior to FAF-Drugs calculations.

Then, our program computes the following molecular properties:

(i) Molecular weight (part of Lipinski's RO5)
(ii) Hydrogen bond donors and acceptors (part of Lipinski's RO5)

Defined as the number of hydrogen bond acceptors (sum of N + O) and hydrogen bond donors (sum of OH + NH).

(iii) Number of rigid bonds
(iv) Number of rings
(v) Size of the rings
(vi) Number of rotatable bond

Defined as any single non-ring bond, bounded to non-terminal heavy atom (29). The amide C-N bonds are not considered because of their high rotational energy barrier.

(vii) Number of carbon atoms, number of heteroatoms and ratio.
(viii) Number of atom with a net charge
(ix) Sum of formal charges
(x) The Topological Polar Surface Area (TPSA)

The method described in (30) has been implemented. Briefly, the molecular polar surface area (PSA) (i.e. surface belonging to polar atoms) is a descriptor that was shown to correlate well with passive molecular transport through membranes. The calculation of PSA, however, is rather time-consuming because of the necessity to generate a reasonable 3D molecular geometry and the calculation of the surface itself. A new approach for the calculation of the PSA was developed by Erlt *et al.* (30) based on the summation of

tabulated surface contributions of polar fragments. This approach was called topological polar surface area, it provides results that are practically identical with the 3D PSA while the computation speed is 2–3 orders of magnitude faster.

(xi) Computation of XlogP (P = calculated octanol/water partition coefficient) (part of Lipinski's RO5)

We use the XScore package (sw16.im.med.umich.edu/software/xtool) to compute XlogP as described in (38). This method gives log P-values by summing the contributions of component atoms while making use of correction factors. About 90 atom types are used to classify carbon, nitrogen, oxygen, sulfur, phosphorus and halogen atoms, and 10 correction factors are used for some special substructures. The contributions of each atom type and correction factor were derived by multivariate regression analysis of about 1850 organic compounds with known experimental log P-values.

In FAF-Drugs, the format for the input files has, for the time being, to be SDF, SMILES or CANSMILES while the compounds have to be in Mol2 format for XlogP computations. We use OpenBabel for file format conversion prior to XlogP calculations. Few compounds are found to have ambiguous atom types and in this case the log P is not computed. (Please see definitions about log P at: http://www.raell.demon.co.uk/chem/logp/logppka.html#Introduction)

(xii) Atom check

Molecules with some specific atoms can be filtered-out (for instance molecules containing H, C, N, O, F, S, P, Cl, Br, I atoms are kept when using default parameters).

RESULTS AND DISCUSSION

Online ADME/tox tools are usually not freely available, for this reason, we have developed FAF-Drugs. This latter stands for Free ADME/tox Filtering and 'Drug-like' compound collections. Our service can be used to filter collections available online as well as virtual libraries. Different levels of filtering have been reported in the literature, depending on the stage of the project, on the target and the disease types. For example, simple physicochemical property filtering could be used when searching for new hits on a new target while more complex ADME/tox models (39) [see for example a list of chemical groups incompatible with final drug development (36,40)] could be applied at a later stage. We chose to implement only simple physicochemical rules because they address the filtering process using widely understood molecular properties.

ADME/tox FILTERS

To start FAF-Drugs filtering, users can either write a molecule in SMILES or 2D/3D SDF format directly in the Web interface window or browse and upload a compound library. Salts and counterions are often present in compound collections and should be removed prior to ADME/tox calculations. If salts and counterions are present, we suggest users to run first our DeSalt utility. At present, the input formats for FAF-Drugs

calculations are CANSMILES, SMILES or SDF (please check our Web site for explanations about the required formats) but OpenBabel (http://openbabel.sourceforge.net/ or online at RPBS) can be used for file format conversion prior to the filtering step (Figure 1a). Then users can decide about the upper and lower limits of each investigated properties (adjustable thresholds) such as, to tailor the compound selection to a specific project. We also propose default parameters that are commonly used in the field (25,26,29,32,41).

Users obtain two files with molecules that pass and do not pass the filters in CANSMILES format together with the original (if available) compound ID provided by the chemical vendors. All computed properties (e.g. MW, TPSA, XlogP...) are also returned in a third file.

In order to test our program, we performed computations on 50 080 molecules extracted from the ChemBridge compound collection (Diversity set) with FAF-Drugs and Filter (version 1.0.2, OpenEye Scientific Software) with the same parameters with the same threshold values (MW, TPSA...). Both, Filter and FAF-Drugs compute TPSA using the approach of Erlt *et al*. (30) and log P using the method of Wang *et al*. (38). A total of 49 334 passed the filters with FAF-Drugs and 49 032 with Filter. Small differences could be due to the fact that some rules are implemented slightly differently, for instance TPSA or log P calculations or definition of flexible bond. Our tests on a Linux machine (Dell Precision 650, 3GHz, 2GB SDRAM) show that the standalone version of FAF-Drugs is able to process the above 50 080 molecules in about 20 min while equivalent computations on the same computer with Filter (OpenEye) took about 10 min. FAF-Drugs implementation is Python-based and is not presently optimized for speed. With regard to server implementation, similar computations took about 30 min, but it can be longer (about 3 h) depending on the server load.

We also compared FAF-Drugs with other online tools: Molinspiration (www.molinspiration.com) that allows evaluation of few physicochemical properties (one molecule at a time can be processed, they have implemented their own tools to calculate log P while they follow the Erlt *et al*. approach to compute polar surface), and the log P calculators provided by Syracuse Research Corporation (see Table 1) and by Tetko and Tanchuck, ALOGPS 2.1 (42). The method for log P prediction developed at Molinspiration (miLogP) is based on group contributions. These have been obtained by fitting calculated log P with experimental log P. ALOGPS uses a neural network approach to predict logP while Syracuse Research Corporation tool (LogKow) estimates log P using an atom/fragment contribution method. Over 100 diverse molecules were tested and in all cases we computed very similar values. To illustrate our calculations, results on four different molecules are reported in Table 3 and Figure 1b. Overall, we note a very good agreement among the different methods.

To further assess FAF-Drugs calculations, we compared over 100 computed log P-values [via our implementation of XlogP) with experimental log P (obtained via Syracuse Research Corporation and via the EDETOX database (http://edetox.ncl.ac.uk/)]. The computed values are in good agreement with the experimental data, indicating that our implementation of XlogP is appropriate and that this approach gives very good results (Figure 2).
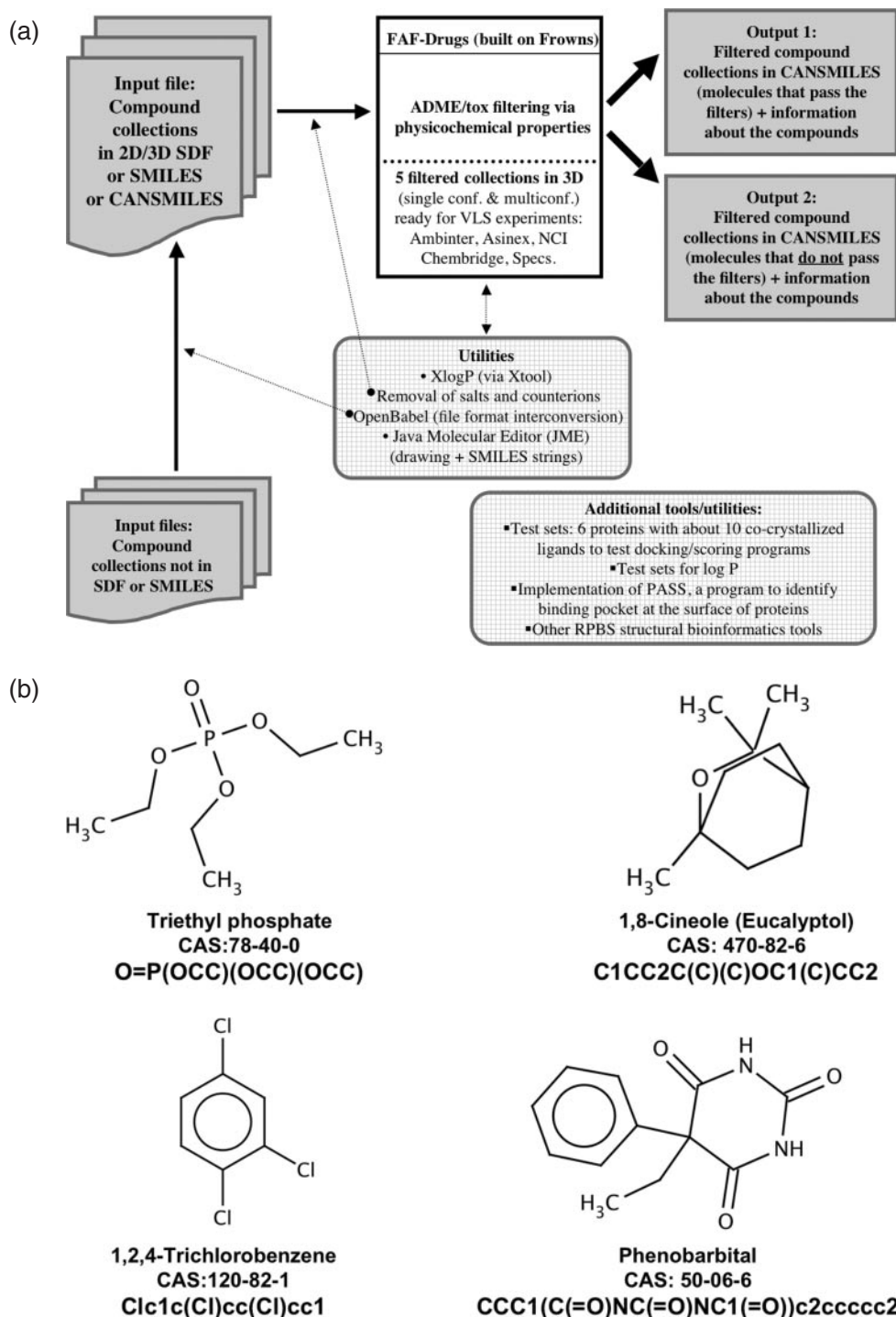
(a)

(b)

Triethyl phosphate
CAS:78-40-0
O=P(OCC)(OCC)(OCC)

1,8-Cineole (Eucalyptol)
CAS: 470-82-6
C1CC2C(C)(C)OC1(C)CC2

1,2,4-Trichlorobenzene
CAS:120-82-1
Clc1c(Cl)cc(Cl)cc1

Phenobarbital
CAS: 50-06-6
CCC1(C(=O)NC(=O)NC1(=O))c2ccccc2

Figure 1. (**a**) Schema of the FAF-Drugs service. Compound collections in SMILES, CANSMILES or SDF format are needed as input. Users can select a threshold for each investigated physicochemical properties. XlogP calculations are performed with Xtool (see text). Users obtain two output files, one with molecules that pass the filters and the other with compounds that do not pass the filters. A third file with all the computed properties can also be downloaded. Several other utilities are available at FAF-Drugs, these involve online XlogP calculations (38) computed with Xtool, online OpenBabel for file format conversion and implementation of the Java Molecular Editor from Dr P. Ertl (Novartis Pharma AG, Basel, Switzerland) to draw molecules and obtain the corresponding SMILES string. In addition, at FAF-Drugs, users can find five ADME/tox filtered compound collections ready for VLS computations. Three levels of filtering were applied (see our web site for further details) in order to better suit the needs of potential users. The OpenEye's Omega program was used to generate 3D models, either single conformation or up to 50 conformations, for each molecule that passed the ADME/tox filters. The compound collections can be downloaded in Mol2 format or in SMILES format. Other utilities consist of a Test Set that contains six protein targets (PDB format) and about 10 corresponding ligands (Mol2 format, see information about the format at http://www.tripos.com) to facilitate evaluation of docking/scoring methods and an interface to PASS (43), a program that predicts binding pocket at the surface of a receptor. Many additional tools pertaining to the field of structural bioinformatics are also available at RPBS such as protein electrostatic computations, loop search, solvent accessibility prediction...(see RPBS services). (**b**) FAF-Drugs results. Four molecules with different physicochemical properties were selected in order to compare FAF-Drugs calculations with other online tools.

Table 3. Comparison of FAF-Drugs with several online tools

Some computed data	FAF-Drugs	Molinspiration	Syracuse_logP	ALOGPS	logP experimental	Compound
MW	154.1	154.25	154.25	154.25		1,8-Cineole CAS: 470-82-6
HD (OH+NH)	0	0	—	—		
HA (O+N)	1	1	—	—		
Rot_bond	0	0	—	—		
TPSA	9.23	9.23	—	—		
log P	2.59	2.71	3.13	3.37	2.50	
MW	182	182.15	182.16	182.16		Triethyl Phosphate CAS: 78-40-0
HD	0	0	—	—		
HA	4	4	—	—		
Rot_bond	6	6	—	—		
TPSA	54.57	44.77	—	—		
log P	0.58	0.69	0.87	0.71	0.80	
MW	232.1	232.23	232.24	232.24		Phenobarbital CAS: 50-06-6
HD	2	2	—	—		
HA	5	5	—	—		
Rot_bond	2	2	—	—		
TPSA	75.27	75.26	—	—		
log P	1.32	0.79	1.33	1.41	1.47	
MW	181.4	181.45	181.45	181.45		1,2,4-Trichlorobenzene CAS: 120-82-1
HD	0	0	—	—		
HA	0	0	—	—		
Rot_bond	0	0	—	—		
TPSA	0	0	—	—		
log P	3.89	3.89	3.93	4.08	4.02	

FAF-Drugs computes several descriptors, such as molecular weight (MW), hydrogen bond donors (HD), hydrogen bond acceptors (HA), number of rotatable bonds (Rot_bond), TPSA and log P (see text). Similar/identical results were obtained via the Molinspiration website and by FAF-Drugs. Experimental log *P*-values and CAS registry numbers were found in the EDETOX database (http://edetox.ncl.ac.uk/) and at the http://www.syrres.com/ (Syracuse Research Corporation) server. The corresponding molecules are shown in Figure 1B.
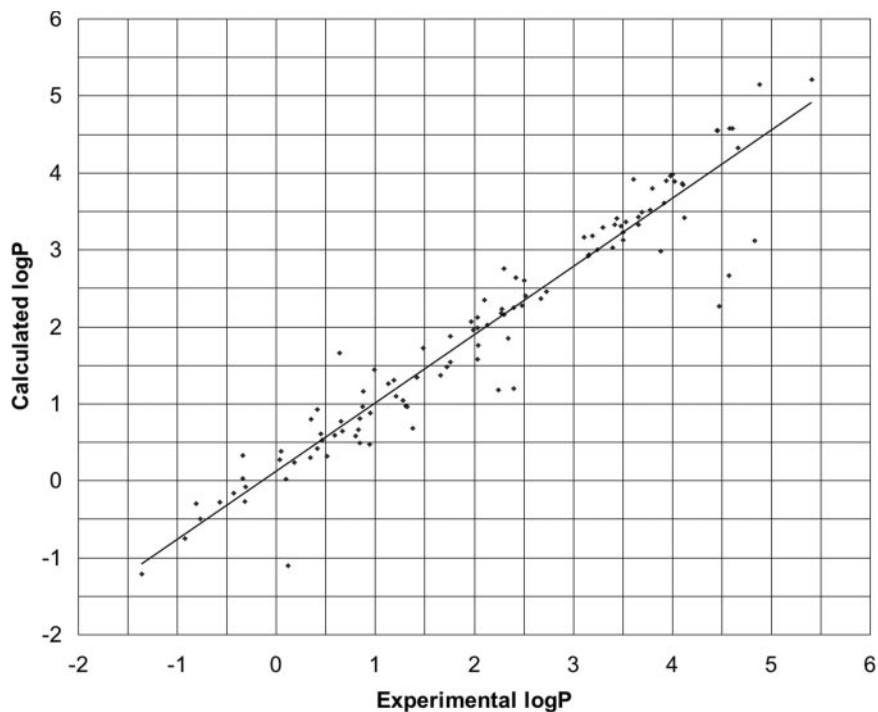
Figure 2. Experimental versus computed logP. Correlation between experimental and calculated log *P*-values for over 100 compounds.

Taken together, the above data suggest that our ADME/tox program is robust. Once users obtain the CANSMILES output, they can decide about adjusting the filters and run additional computations or use 1D/2D to 3D conversion programs such as Corina (http://www2.chemie.uni-erlangen.de/software/corina/free_struct.html), Omega (OpenEye Scientific Software), Converter (Accelrys) and start a VLS project.

For the time being, to protect our server from intensive use, we suggest scientists to upload files with less than 30 000 molecules. In the present version of the service, computations for several tens of thousands of compounds remain time consuming (e.g. several hours depending on the number of jobs in the queue) but work is in progress to improve this point. For this reason and in order to save CPU time and disk space, we also provide five filtered compound collections (Figure 1a).

CONCLUSION AND FUTURE DIRECTIONS

A rational approach to increase the efficiency of finding new drugs and reduce the R&D cost is to reduce the attrition rate in the costly downstream stages (e.g. clinical trials). Several important methods toward this goal have been developed, involving early computations of ADME/tox properties. We have developed FAF-Drugs to help modelers and biologists to embark into drug discovery projects. Users can filter their own compound libraries and adapt the thresholds to a specific project. Other tools pertaining to the field of drug design/ compound collections are also available at our Web site. We are presently working on improving the speed of the calculations on our server.

ACKNOWLEDGEMENTS

REFERENCES

1. Shoichet,B.K. (2004) Virtual screening of chemical libraries. *Nature*, **432**, 862–865.
2. Lyne,P.D. (2002) Structure-based virtual screening: an overview. *Drug Discov. Today*, **7**, 1047–1055.
3. Bleicher,K.H., Bohm,H.J., Muller,K. and Alanine,A.I. (2003) Hit and lead generation: beyond high-throughput screening. *Nature Rev. Drug Discov.*, **2**, 369–378.
4. Jennings,A. and Tennant,M. (2005) Discovery strategies in a BioPharmaceutical startup: maximising your chances of success using computational filters. *Curr. Pharm. Des.*, **11**, 335–344.
5. McConkey,B.J., Sobolev,V. and Edelman,M. (2002) The performance of current methods in ligand-protein docking. *Current Science*, **83**, 845–856.
6. Lengauer,T., Lemmen,C., Rarey,M. and Zimmermann,M. (2004) Novel technologies for virtual screening. *Drug Discov. Today*, **9**, 27–34.
7. Kitchen,D.B., Decornez,H., Furr,J.R. and Bajorath,J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Rev. Drug Discov.*, **3**, 935–949.
8. Oprea,T.I. and Matter,H. (2004) Integrating virtual screening in lead discovery. *Curr. Opin. Chem. Biol.*, **8**, 349–358.
9. Jain,A.N. (2004) Virtual screening in lead discovery and optimization. *Curr. Opin. Drug Discov. Devel.*, **7**, 396–403.
10. Fradera,X. and Mestres,J. (2004) Guided docking approaches to structure-based design and screening. *Curr. Top. Med. Chem.*, **4**, 687–700.
11. Stahura,F.L. and Bajorath,J. (2005) New methodologies for ligand-based virtual screening. *Curr. Pharm. Des.*, **11**, 1189–1202.
12. Congreve,M., Murray,C.W. and Blundell,T.L. (2005) Structural biology and drug discovery. *Drug. Discov. Today*, **10**, 895–907.
13. Hardy,L.W. and Malikayil,A. (2003) The impact of structure-guided drug design on clinical agents. *Curr. Drug. Discov.*, **15**, 15–20.
14. Abagyan,R. and Totrov,M. (2001) High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.*, **5**, 375–382.
15. Alvarez,J.C. (2004) High-throughput docking as a source of novel drug leads. *Curr. Opin. Chem. Biol.*, **8**, 365–370.
16. Schneider,G. and Bohm,H.J. (2002) Virtual screening and fast automated docking methods. *Drug. Discov. Today*, **7**, 64–70.
17. Kassel,D.B. (2004) Applications of high-throughput ADME in drug discovery. *Curr. Opin. Chem. Biol.*, **8**, 339–345.
18. Muegge,I. (2003) Selection criteria for drug-like compounds. *Med. Res. Rev.*, **23**, 302–321.
19. Hou,T. and Xu,X. (2004) Recent development and application of virtual screening in drug discovery: an overview. *Curr. Pharm. Des.*, **10**, 1011–1033.
20. Beresford,A.P., Segall,M. and Tarbit,M.H. (2004) *In silico* prediction of ADME properties: are we making progress? *Curr. Opin. Drug Discov. Devel.*, **7**, 36–42.
21. Lombardo,F., Gifford,E. and Shalaeva,M.Y. (2003) *In silico* ADME prediction: data, models, facts and myths. *Mini. Rev. Med. Chem.*, **3**, 861–875.
22. Gasteiger,J. (2003) Physicochemical effects in the representation of molecular structures for drug designing. *Mini. Rev. Med. Chem.*, **3**, 789–796.
23. Martin,Y.C. (2005) A bioavailability score. *J. Med. Chem.*, **48**, 3164–3170.
24. Rishton,G.M. (2003) Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discov. Today*, **8**, 86–96.
25. Lipinski,C.A., Lombardo,F., Dominy,B.W. and Feeney,P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, **46**, 3–26.
26. Li,A.P. (2001) Screening for human ADME/Tox drug properties in drug discovery. *Drug Discov. Today*, **6**, 357–366.
27. Roche,O. and Guba,W. (2005) Computational chemistry as an integral component of lead generation. *Mini. Rev. Med. Chem.*, **5**, 677–683.
28. Vermeulen,N.P. (2003) Prediction of drug metabolism: the case of cytochrome P450 2D6. *Curr. Top. Med. Chem.*, **3**, 1227–1239.
29. Veber,D.F., Johnson,S.R., Cheng,H.Y., Smith,B.R., Ward,K.W. and Kopple,K.D. (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.*, **45**, 2615–2623.
30. Ertl,P., Rohde,B. and Selzer,P. (2000) Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.*, **43**, 3714–3717.
31. Nassar,A.E., Kamel,A.M. and Clarimont,C. (2004) Improving the decision-making process in structural modification of drug candidates: reducing toxicity. *Drug Discov. Today*, **9**, 1055–1064.
32. Oprea,T.D. (2002) Virtual screening in lead discovery. *Molecules*, **7**, 51–62.
33. Kubinyi,H. (2003) Drug research: myths, hype and reality. *Nature Rev. Drug Discov.*, **2**, 665–668.
34. Yu,H. and Adedoyin,A. (2003) ADME-Tox in drug discovery: integration of experimental and computational technologies. *Drug Discov. Today*, **8**, 852–861.
35. van de Waterbeemd,H. and Gifford,E. (2003) ADMET *in silico* modelling: towards prediction paradise? *Nature Rev. Drug Discov.*, **2**, 192–204.
36. Baurin,N., Baker,R., Richardson,C., Chen,I., Foloppe,N., Potter,A., Jordan,A., Roughley,S., Parratt,M., Greaney,P. *et al.* (2004) Drug-like annotation and duplicate analysis of a 23—supplier chemical database totalling 2.7 million compounds. *J. Chem. Inf. Comput. Sci.*, **44**, 643–651.
37. Irwin,J.J. and Shoichet,B.K. (2005) ZINC–a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.*, **45**, 177–182.

38. Wang,R., Gao,Y. and Lai,L. (2000) Calculating partition coefficient by atom-additive method. *Perspectives in Drug Discovery and Design*, **19**, 47–66.

39. Frimurer,T.M., Bywater,R., Naerum,L., Lauristen,L.N. and Brunak,S. (2000) Improving the odds in discriminating drug-like from non-drug-like compounds. *J. Chem. Inf. Comput. Sci.*, **40**, 1315–1324.

40. Walters,W.P., Stahl,M.T. and Murcko,M.A. (1998) Virtual screening. *Drug Discov. Today*, **3**, 160–178.

41. Baurin,N., Aboul-Ela,F., Barril,X., Davis,B., Drysdale,M., Dymock,B., Finch,H., Fromont,C., Richardson,C., Simmonite,H. *et al.* (2004) Design and characterization of libraries of molecular fragments for use in NMR screening against protein targets. *J. Chem. Inf. Comput. Sci.*, **44**, 2157–2166.

42. Tetko,I.V. and Tanchuk,V.Y. (2002) Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. *J. Chem. Inf. Comput. Sci.*, **42**, 1136–1145.

43. Brady,G.P. Jr. and Stouten,P.F. (2000) Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Des.* **14**, 383–401.

44. Strausberg,R.L. and Schreiger,S.L. (2003) From knowing to controlling: a path from genomics to drugs using small molecule probes. *Science*, 294–295.

45. Kanehisa,M. (1997) Linking databases and organisms. *Trends Biochem. Sci.*, **22**, 442–444.

46. Grotthuss,v.M., Pas,J. and Rychlewski,L. (2003) Ligand-Info, searching for similar small compounds using index profiles. *Bioinformatics*, **19**, 1041–1042.

47. Girke,T., Cheng,L.C. and Raikhel,N. (2005) ChemMine. A compound mining database for chemical genomics. *Plant Physiol.*, **138**, 573–577.

48. Chen,X., Lin,Y. and Gilson,M.K. (2002) The binding database. *Biopolymers Nucleic Acid. Sci.*, **61**, 127–142.

49. Wang,R., Fang,X., Lu,Y., Yang,C.Y. and Wang,S. (2005) The PDBbind database: methodologies and updates. *J. Med. Chem.*, **48**, 4111–4119.

50. Zhang,J., Aizawa,M., Amari,S., Iwasawa,Y., Nakano,T. and Nakata,K. (2004) Development of KiBank, a database supporting structure-based drug design. *Comput. Biol. Chem.*, **28**, 401–407.