ELSEVIER

The Drug Discovery Portal: a resource to enhance drug discovery from academia

Rachel L. Clark[1], Blair F. Johnston[1], Simon P. Mackay[1], Catherine J. Breslin[2], Murray N. Robertson[1] and Alan L. Harvey[1]

[1] Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 27 Taylor Street, Glasgow G4 0NR, UK
[2] Research & Knowledge Exchange Services, University of Strathclyde, 50 George Street, Glasgow G1 1QE, UK

Drug discovery in universities is usually associated with research on drug targets and mechanisms, but more recently there have been efforts to progress from target studies to proof of concept by applying commercially focussed medicinal chemistry. This creates more opportunities for novel interactions and partnering models between academic groups and pharmaceutical companies. We present a review of coordinated, multi-institutional drug discovery operations within academia that are engaging with industry nationally and internationally and describe how the Drug Discovery Portal at the University of Strathclyde enhances the possibilities for academic drug discovery.

Introduction

Research relevant to drug discovery carried out by scientists in academia is generally at the fundamental level, with a focus on drug mechanisms and the identification and validation of potential therapeutic targets. Academic research is also important in commercially unattractive therapeutic areas, such as rare disorders [1] and parasitic diseases [2], and in research into natural products [3]. More recently, increased funding for translational research and interest in finding small molecular weight inhibitors of pathways studied in chemical biology programmes have led to several academic groups investing in core infrastructure to undertake drug discovery [4,5]. There are now 65 screening centres listed in the Society for Biomolecular Sciences Academic Screening Facility Directory (http://www.sbsonline.org), which can be contacted to support chemical biology and drug discovery programmes. For example, the Molecular Library Screening Centres Network (MLSCN) incorporates ten US institutions with different facilities, therapeutic targets and expertise. The combination of fundamental and applied research being pursued in academia is creating more opportunities for novel interactions and partnering models with industry, as reviewed recently in Ref. [5]. This is a timely development for academic drug research because pharmaceutical companies can be interested in novel compounds with novel actions despite their early stage of development [6].

High-throughput screening (HTS) is the main approach to drug discovery, both in companies and in academic screening centres. Companies can screen their own compound libraries, which can contain millions of proprietary and commercially available compounds, to identify hits for new drug discovery programmes. The HTS facilities in academia generally screen smaller collections that have been sourced from various commercial suppliers. For example, the Rockefeller University (http://www.rockefeller.edu/highthroughput/highthroughput.php) and Michigan University (http://mhtsc.kvcc.edu/) screening centres have libraries in the region of 80,000–100,000 compounds sourced from commercial vendors such as ChemDiv (http://www.chemdiv.com/), ChemBridge (http://www.chembridge.com/), and Maybridge (http://www.maybridge.com/default.aspx), which were designed using their own algorithms for the analysis of diversity and coverage of chemical space.

Although company and academic screening follow the same general approach to hit finding, there is a marked difference thereafter. Companies will take promising hits and improve them with their in-house medicinal chemistry to find promising candidates for clinical development. This option is rarely available to academic screeners, however, because medicinal chemistry resources are not generally associated with the screening centres (with a few

Corresponding author:. Clark, R.L. (rachel.clark@strath.ac.uk)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2010.06.003

exceptions, such as the Drug Discovery Unit at the University of Dundee [4] and the Imperial College Drug Discovery Centre in London; http://www1.ic.ac.uk/medicine/about/institutes/drugdiscoverycentre/). More typically, the screening hits are used as experimental tools and to inform research on biochemical pathways. For example, the hits from all of the MLSCN screens are publicly available for cheminformatics purposes with the aim of enriching the data available for the biological targets and the molecules that interact with them (http://pubchem.ncbi.nlm.nih.gov/). The MLSCN network has generated more than 200 publications resulting directly from their screening services since 2006. There seems to be a bottleneck between the finding of interesting hits on novel potential therapeutic targets included in academic screening projects and the use of such hits for drug development purposes.

Thus, many of the hits identified in academic screening programmes are likely to lie fallow if there is not a similar network of medicinal chemists to exploit and develop these hits into optimized hits, leads and preclinical candidates [7]. There is a problem in finding medicinal chemists with the available time, resources and aptitude to provide hit optimization. Although the commercial vendors of the majority of compounds used to populate academic screening libraries usually offer custom medicinal chemistry support or a tailored follow-up library for any hit identified, this can be a costly service for academic groups which have used a free screening service and identified a chemical structure with associated activity data. A different approach might be to find a way to link the intrinsic interests of synthetic chemists in academia with the discovery of biological activity in screening programmes. This can be done if the chemists' own compounds are used in the screening libraries.

Only a few academic chemists advertise their compound collections for use in HTS, however; for example, the Gray Kinase Focused Library (187 compounds; http://iccb.med.harvard.edu/screening/compound_libraries/GrayKinaseInhib.htm), the Boston University Center for Chemical Methodology and Library Development (718 compounds; http://cmld.bu.edu/) [8], and the Marvel collection (University of Illinois; 7732 compounds) [9]. The French National Compound Library (FNCL) is the first example of a national, federated resource of available chemicals from academic laboratories. The collection was established in 2000 to catalyse scientific partnerships between chemists and biologists. It currently contains more than 37,000 molecules, which are plated for use on their HTS platform (PCBIS) (http://chimiotheque-nationale.enscm.fr/) [10]. Compounds are available to search online after signing a collaboration agreement allowing users to view all available structures. A single compound or plates of compounds suitable for HTS can be requested. A similar collection has been initiated in Australia between the Queensland State Government and Griffith University, which has recently established the Queensland Compound Library. Academic groups are requested to pay a membership fee for differing levels of access to the collection or to donate compounds. The aim is to synergize interactions between Australian chemists and biomedical researchers and their international colleagues, and the collection is available for HTS (http://www.griffith.edu.au/science/queensland-compound-library). In Japan, the Chemical Biology Research Initiative based at the University of Tokyo (http://www.cbri.u-tokyo.ac.jp/index_e.html) was launched to collate compounds of pharmaceutical relevance in Japanese universities and to harness the intellectual property for collaboration with industry. The group has both HTS and virtual screening (VS) capability to identify biologically relevant hits.

There are several VS service providers for biologists or chemists (some freely available, others commercial), where users provide information about their target of interest and compounds they would like to screen computationally (for a non-exhaustive but representative list, see Box 1). Recently, there have been attempts to fully automate VS processes, but results have thus far shown these to be inferior to those led by an expert [9,11].

Historically, chemists in university laboratories worldwide have synthesized many thousands of compounds while investigating or optimizing a synthetic process, producing compound libraries to challenge a particular biological target or developing a new route to a complex natural product. Collectively, this represents one of the most chemically diverse libraries in existence, and it would be a unique resource for drug discovery if it could be harnessed, although academic compound collections might raise user

BOX 1

Virtual screening service providers

Biognos AB, Göteborg, Sweden

(http://www.biognos.se/)
Services: Ligand- and structure-based virtual screening, conformational analysis, and solubility predictions.
Focus: Virtual screening service providers to pharmaceutical and biotech companies using an in-house commercial database and commercial software.

Dock Blaster, University of California, San Francisco
(http://blaster.docking.org/)
Services: Public access service for structure-based ligand discovery.
Focus: Users must provide protein (in .pdb or .mol2 format) with a bound ligand (mol2) or binding site specification (in .pdb). ZINC database used to match small molecules to user-defined sites.

BioFocus, A Galapagos Company
(http://www.biofocus.com/)
Services: *In silico* screening and molecular informatics.
Focus: Lead expansion and optimization to support chemical library synthesis. Uses StarDrop optimization platform.

Quantum Pharmaceuticals, Moscow
(http://q-pharm.com/)
Services: Drug hit identification, hit-to-lead optimization, drug toxicity and pharmacokinetic profiling.
Focus: Quantum and statistical physics methods used to calculate intermolecular interactions with the accuracy of *in vitro* experiments.

Drug Discovery Portal, Strathclyde, UK
(http://www.ddp.strath.ac.uk)
Services: Hit identification, hit-to-lead optimization (computational support and/or synthetic medicinal chemistry output), library analysis and design, and allosteric site identification.
Focus: Matching academic chemists and academic biologists through intelligent virtual screening performed by experienced researchers.

concerns of purity, availability and supporting documentation. In addition, academic groups remain the most likely means of sampling an important area of chemical space that is generally out of reach of HTS facilities – natural products. There are plated natural product extract libraries [12,13], and some commercial suppliers offer plated pure natural products in small numbers: for example, InterBioScreen NP collection (1464 compounds; http://www.ibscreen.com/products.shtml), SPECS (323 compounds; http://www.specs.net/page.php?pageid=2004111115353984&smenu=2008111411133023), ChemDiv NP (549 compounds; http://us.chemdiv.com/index.php?option=com_content&view=article&id=90&Itemid=189) and Greenpharma (part of the Prestwick collection; http://www.prestwickchemical.com/index.php?pa=27). A comprehensive virtual 3D library of commercially available natural products is available through the ZINC website, which comprises more than 80,000 natural products, derivatives of natural products and semi-synthetic analogues of natural products from seven commercial vendors (http://zinc.docking.org/vendor0/index_meta.shtml). Although pharma companies have generally abandoned their natural product research and compound collections, research into the isolation, identification, purification and synthesis of natural products continues in academic institutions. More often than not, only a few mgs or less of these compounds are isolated in their pure form, and – not surprisingly – academics are reluctant to, and often unable to, provide sufficient compound for indiscriminate random screening programmes. However, now that the techniques for such isolation, purification and identification have become more sophisticated, sensitive and to a certain extent automated, natural products are technically more accessible for use in HTS-driven drug discovery.

A recent example of the interplay between pharma and academia in the drug discovery arena is the launch by Eli Lilly of the Phenotypic Drug Discovery Initiative (PD^2). Its aim is to interact with universities, research institutes and biotech companies to identify novel and pharmaceutically interesting compounds to test in phenotypic assays currently relating to Alzheimer's disease, cancer (G2/M arrest and anti-angiogenesis), diabetes and bone formation. Researchers are invited to submit chemical structures of compounds they would like to screen, and the structures are then converted to 2D fingerprints and compared with the fingerprints of every compound in the company's library. A compound is rejected if it exceeds a similarity threshold of 85% against the Lilly collection, known drugs, compounds from PubChem or previous submissions and 90% for controlled substances. Compounds are then subjected to MedChem Rules designed to identify uninteresting or undesirable features, such as reactive groups and molecules that have no rings or are too large (+50 heavy atoms) or too small (<7 heavy atoms), which can also result in rejection. There is a further demerit system for compounds that have unfavourable structural motifs typically identified through QSAR models for ADME or toxicity, for example. A molecule exceeding the demerit score threshold will also be rejected (https://pd2.lilly.com/pd2Web/DefaultMenuItems/others/PD2_Structure_Evaluation_Details.pdf). By the use of the 2D fingerprint technique, Lilly assures users that no structural data are disclosed to the company, and if a novel compound is identified through this process, the owner is alerted and a sample is requested for primary screening. This is a novel interaction mechanism between a pharmaceutical company and academia, in both the approach to compound selection and the phenotypic screening cascade it is subjected to (https://pd2.lilly.com/). Furthermore, it demonstrates that academia, particularly its chemists, has an extremely valuable resource for pharma, namely its chemical diversity.

The Drug Discovery Portal concept

As described above, drug discovery resources in academia have been enhanced through acquisition of HTS facilities and large screening collections based on commercially available compounds. There has been little attempt to develop these university-based activities in ways that are compatible with the needs of the drug discovery process itself, or the drug discovery industry, and that make use of the inherent diversity of synthetic chemistry in academia to access areas of drug-like chemical space not readily available to pharma companies.

Medicinal chemists at the University of Strathclyde initiated the concept of the Drug Discovery Portal (DDP) to provide a resource that would enhance drug discovery collaborations between chemists and biologists in academia while addressing the issues of novel chemical space and protectable Intellectual property rights (IPR). The DDP has created a unique, secure database of compounds that is used in VS campaigns; samples of predicted hits can be rapidly provided for testing in real screens (Fig. 1). The DDP's chemical database contains structures of synthetic compounds and natural products from many different sources, meaning that it has high chemical diversity (as described below), and the database only contains compounds that can be made physically available for screening should VS predict them to be hits. To address issues of quality assurance and availability, a DDP chemist has a responsibility to ensure that the relevant analytical information is provided when the compound is requested for biological screening. To engender new drug discovery collaborations, biologists can propose therapeutic targets to the DDP and these can be used in VS campaigns with the DDP chemical database. Just as chemists are required to be able to provide samples of structures submitted to the DDP, biologists can only propose targets if the relevant assays are up and running.

When hits are predicted from VS, they are sourced from the originating chemist and provided to the biologist for assaying. When hits are found in the real screening, there is a natural link with the originating chemist, who, presumably, has the background expertise and wherewithal to synthesize analogues or isolate related natural products as part of a hit identification or hit-to-lead process.

The DDP database

The DDP was initiated in 2007 and has collated more than 7000 compounds so far from chemists around the world, built initially from previous collaborators then extended through e-mail, seminar and conference exposure. We have links with the FNCL and the University of Illinois, which extends the searchable virtual screening database to 54,000 compounds that are associated with physical screening samples and links with the contributing chemists. The sister biological target database has 19 targets with multiple, optimized and validated 3D pharmacophores and binding sites. We continually seek to expand both databases, and we welcome contributions from any interested parties.

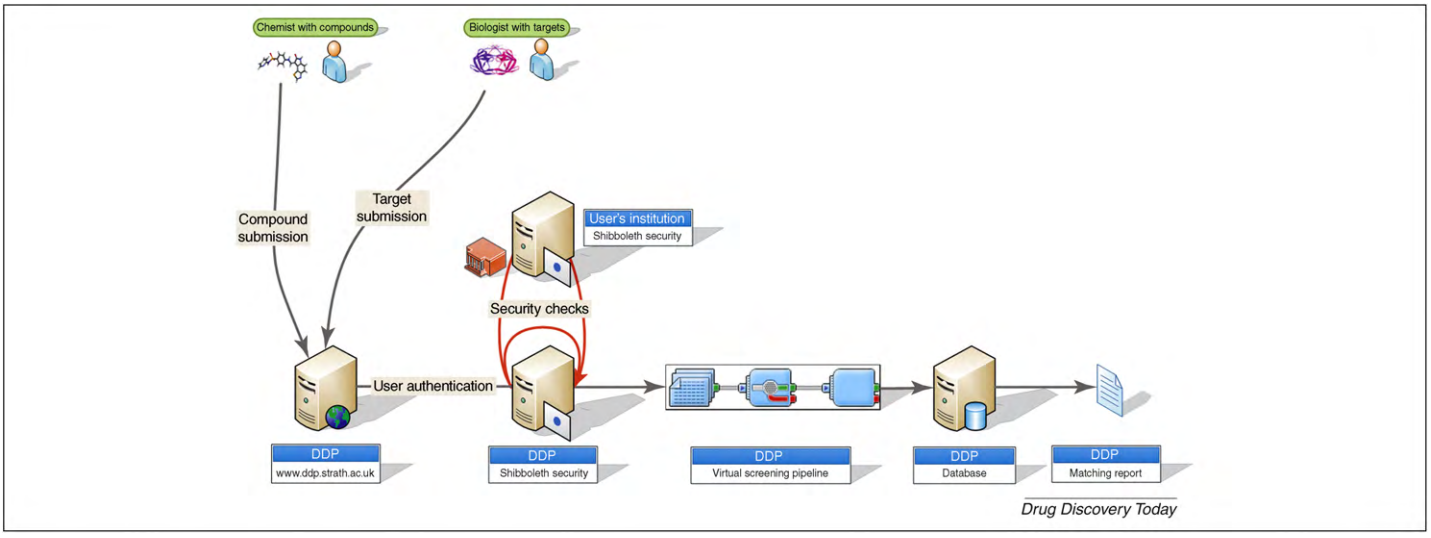

FIGURE 1

DDP security and virtual screening workflow. The VS pipeline is expert led, particularly in ligand preparation, target scoping and docked pose analysis.

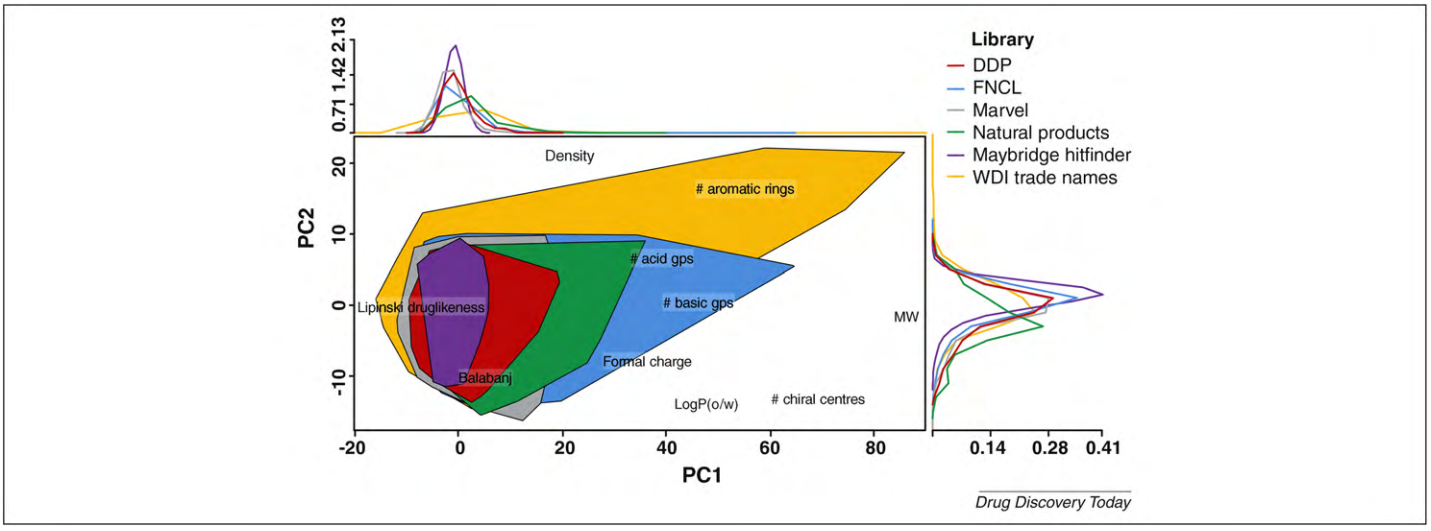

FIGURE 2

PCA score plot (with loadings superimposed). The regions of chemical space occupied by each library are shown, along with the distribution of compounds within each region (top and right). Left to right, molecular mass increases along with the number of chiral centres and number of rings within the structures. Log *P* dominates vertically; highly lipophilic compounds are found in the lower half of the scores plots, and hydrophilic molecules are found at the top. The DDP occupies Lipinski-like chemical space towards the left and natural product space, which extends to the right.

Principal components analysis (PCA) has been used to assess the diversity of the chemicals present in the DDP, FNCL and Marvel academic libraries with respect to commercially available natural products from ChemDiv, InterBioScreen and Specs (grouped as 'Natural Products' in Fig. 2), Maybridge's HitFinder library of lead-like compounds, and a subset of compounds from the World Drug Index (WDI) that have trade names (i.e. they are drugs that have made it to market). In total, 90 non-correlated (<90%) molecular descriptors were reduced to two principal components (PC1 and PC2) explaining 38% of the overall observed variance. The HitFinder and WDI libraries represent the extremes of chemical diversity: HitFinder's compounds occupy a Lipinski-like [14] area of small molecules to the left in Fig. 2, and the WDI extends to an area of heavier chemical space on the right. It should be noted that although the WDI occupies the largest area, the majority of compounds within this library can be found in areas occupied by the other libraries, as seen from the distribution curves above and to the right of the PCA score plot. The DDP and Marvel academic libraries occupy similar areas of chemical space, with slightly heavier molecules than those found in HitFinder, whereas the FNCL is relatively more diverse, containing even heavier compounds and pushing further away from the Lipinski region. The distribution curves show that on both PC1 and PC2, the natural product libraries are the most chemically distinct. In particular, this can be seen on the score distributions on PC2, which at low values are governed by several chiral centres and lipophilicity. Unsurprisingly, the natural product libraries occupy more of this region, not unlike the DDP library, which reflects the research interests of many of the contributing academic researchers involved with the DDP. Despite the apparent overlap of

libraries in the Lipinski-like region [14], the contents of each library are more than 99% unique and offer equal value for use in a screening campaign.

To match academic chemists' compounds to biological targets to discover new bioactivity, the unique and highly diverse DDP database is used for VS or similarity searching. VS and other informatics approaches are widely used by the pharmaceutical industry; they have aided in the development of numerous leads [15] and are commonly used to focus HTS and enrich screening libraries [11]. They play a particularly important part in the repurposing of chemists' compounds and in flagging scarce natural products for assays, when the natural products are unlikely to be available in sufficient quantity to supply for random HTS campaigns on many targets. To date, the DDP has screened 19 targets: 5 are in the optimization stage and 14 form the basis of ongoing collaborative research projects. Analogues of hit compounds have been identified, sourced and tested to improve the accuracy of the virtual screening process, and hit optimization is underway. Data from these studies will be used either to apply for funding for further drug development with DDP contributors now identified as co-investigators on grant applications or as the basis for discussions with development companies. The University of Strathclyde technology transfer office is in the process of establishing a spin-out company to drive projects through to preclinical proof of concept. Actar (http://www.actar.se/) is a drug discovery company based in the Karolinska Institute campus that has successfully adopted this approach. A new joint company is formed for every drug discovery project in which Actar, the Institute and researchers are shareholders. Because the DDP operates through formal collaboration agreements that preserve intellectual property rights, companies can have access to secure and proprietary information for their future development programmes. Despite increasing talk of 'open source pharma' [16], pharmaceutical companies need protected intellectual property as the basis for commercial developments.

Concluding remarks

Historically, chemists in university laboratories worldwide have synthesized many thousands of compounds, and collectively this might be the most chemically diverse library in existence. The DDP at the University of Strathclyde is a facility whose ambition is to harness and develop this resource. HTS libraries and facilities accessible to academics are generally populated with compounds from commercial vendors and are more suited to the discovery of probes for chemical biology programmes than to drug discovery. The missing element is the availability of medicinal chemistry to support hit-to-lead development. This is addressed by the DDP, which uses VS to create links between the biologist behind the therapeutic targets and the chemists who originally made the hit compounds. Use of VS enables compounds to be screened *in silico* at no cost to the chemist and with no wastage of potentially rare chemicals. The DDP database also has the advantage that it is constructed from many academic laboratories: its inherent structural diversity is greater than conventional screening collections.

Time will tell whether the DDP succeeds in its ambition to revolutionize academic drug discovery. Because intellectual property is protected throughout the DDP process, the outputs of novel bioactive compounds on potentially novel therapeutic targets could be of immediate interest to pharmaceutical development companies.

Disclosure statement

All authors are employees of the University of Strathclyde, spending a proportion of time working within the DDP, which is not an independent entity from the University.

Acknowledgements

The authors thank Richard Sinnott and Christopher Bayliss at the National e-Science Centre for their input into the construction of the DDP.

References

1 Glicksman, M. (2008) The drug discovery process in academics. Opportunity knocks. *BIOForum Europe* 6, 34–35
2 Shah, S. (2009) Drug discovery dilemma and Cura Quartet collaboration. *Drug Discov. Today* 14, 1006–1010
3 Glaser, K.B. and Mayer, A.M.S. (2009) A renaissance in marine pharmacology: from preclinical curiosity to clinical reality. *Biochem. Pharmacol.* 78, 440–448
4 Frearson, J.A. and Collie, I.A. (2009) HTS and hit finding in academia – from chemical genomics to drug discovery. *Drug Discov. Today* 14, 1150–1158
5 Tralau-Stewart, C.J. *et al.* (2009) Drug discovery: new models for industry–academic partnerships. *Drug Discov. Today* 14, 95–101
6 Hughes, B. (2008) Pharma pursues novel models for academic collaboration. *Nat. Rev. Drug Discov.* 7, 631–632
7 Kozikowski, A.P. *et al.* (2006) Why academic drug discovery makes sense. *Science* 313, 1235–1236
8 Simeonov, A. *et al.* (2008) Fluorescence spectroscopic profiling of compound libraries. *J. Med. Chem.* 51, 2363–2371
9 Hergenrother, P.J. (2006) Obtaining and screening compound collections: a user's guide and a call to chemists. *Curr. Opin. Chem. Biol.* 10, 213–218
10 Hibert, M.F. (2009) French/European academic compound library initiative. *Drug Discov. Today* 14, 723–725
11 Phatak, S.S. *et al.* (2009) High-throughput and *in silico* screenings in drug discovery. *Expert Opin. Drug Discov.* 4, 947–959
12 Harvey, A.L. (2007) Natural products as a screening resource. *Curr. Opin. Chem. Biol.* 11, 480–484
13 Harvey, A.L. (2008) Natural products in drug discovery. *Drug Discov. Today* 13, 894–901
14 Lipinski, C.A. *et al.* (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26
15 Ghosh, S. *et al.* (2006) Structure-based virtual screening of chemical libraries for drug discovery. *Curr. Opin. Chem. Biol.* 10, 194–202
16 Talaga, P. (2009) Open innovation: share or die. . .. *Drug Discov. Today* 14, 1003–1005