ELSEVIER



# Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery

**Moses Hohman[1], Kellan Gregory[1], Kelly Chibale[2], Peter J. Smith[3], Sean Ekins[4,5,6] and Barry Bunin[1]**

[1] Collaborative Drug Discovery, Inc. 1818 Gilbreth Road, Suite 220, Burlingame, CA 94403, USA
[2] Institute of Infectious Disease and Molecular Medicine and Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa
[3] Division of Pharmacology, Department of Medicine, University of Cape Town, Medical School, K45, OMB, Groote Schuur Hospital, Observatory, 7925, South Africa
[4] Collaborations in Chemistry, Jenkintown, PA 19046, USA
[5] University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, NJ 08854, USA
[6] Department of Pharmaceutical Sciences, University of Maryland, 20 Penn Street, Baltimore, MD 21201, USA

**A convergence of different commercial and publicly accessible chemical informatics, databases and social networking tools is positioned to change the way that research collaborations are initiated, maintained and expanded, particularly in the realm of neglected diseases. A community-based platform that combines traditional drug discovery informatics with Web2.0 features in secure groups is believed to be the key to facilitating richer, instantaneous collaborations involving sensitive drug discovery data and intellectual property. Heterogeneous chemical and biological data from low-throughput or high-throughput experiments are archived, mined and then selectively shared either just securely between specifically designated colleagues or openly on the Internet in standardized formats. We will illustrate several case studies for anti-malarial research enabled by this platform, which we suggest could be easily expanded more broadly for pharmaceutical research in general.**

## The networked revolution

Recent research suggests that open collaborative drug discovery will be the future paradigm of biomedical research [1–3]. Reviews in this journal have provided a perspective on the many publicly accessible, open access chemistry databases and Internet-based collaborative tools [4,5] that are likely to enhance scientific research in future. Some of these public databases are already being used for structure activity relationship (SAR) development [6] and rapid lead identification [7]. It takes a combination of biology and chemistry insight, however, to translate molecules into potential drugs and there has been little, if any, discussion of how collaborations between chemists and biologists are to be facilitated [8]. The challenges associated with bringing chemists and biologists together for virtual drug discovery projects for neglected diseases [8] provide an arena for testing new approaches that can perhaps be expanded more broadly to commercial drug discovery projects. The biological data available for sharing are frequently stored in single document or Excel™ files. Compilation of data is sporadic with no depth and little, if any, standardization of the data formats or crucial information such as experimental procedures and statistical analysis to quantify data quality to allow reproducibility and comparisons between groups. Before collaborations begin, data security and integrity should always be considered while intellectual property arrangements [Materials Transfer and intellectual property (IP) Rights Agreements] are often (at least in academia) seen as necessary, but generally as a hindrance to progress. As a collaboration progresses the needs of data users may change, so it is important to have flexibility in the use of systems for tracking or storage of data and between systems [8].

Any tool that can tap into a growing community of researchers becomes more valuable as a function of Metcalfe's law, which simply states the value of a network is equal to the square of the

*Corresponding author:* Bunin, B. (bbunin@collaborativedrug.com)

1359-6446/06/$ - see front matter  doi:10.1016/j.drudis.2008.11.015

number of connected nodes [expressed mathematically for the number of unique connections in a network of $n$ nodes as $n(n-1)/2$, which follows $n^2$ asymptotically, see: http://en.wikipedia.org/wiki/Metcalfe's_law]. So the impact and value of these tools would be expected to increase in an exponential, rather than linear, manner and as a function of the number of interconnected users. Examples of network-based technologies that we are all very familiar with and take for granted include telephones, fax machines, and the Internet. Networks are everywhere. Other examples such as the power grid, metabolic networks for molecule–target interactions [9] and our food supply chain, as well as collaborative networks of scientists [10], clearly demonstrate the scale-free nature of these networks [11]. Although we do not have evidence for whether software tools follow this pattern, as yet in the collaborative scientific domain recent studies have suggested productivity benefits of collaboration [12,13] and the formation of collaboration networks [14].

## The challenge and opportunity

Today, a diverse set of drug discovery informatics tools [15] and platforms (Box 1) are being joined by a new hybrid of chemistry-relevant and biology-relevant informatics technologies some of which have social networking capabilities (see Box 2). These platforms facilitate the development of a novel approach to neglected disease drug discovery.

Despite the numerous examples of network-based technologies, such as LinkedIn for business networking and Facebook, LabMeeting and many other examples for social and scientific networking (see Box 2), current networked technologies have only recently started to impact drug discovery directly. The general web collaboration tools have limited or no capability to archive and manage laboratory data and then mine them on the basis of chemical structure. Traditional commercially available chemistry and biology data management systems (see Box 1) do not have collaboration features nor enable data sharing (open source public data exchange). These commercially available tools do not foster community-based models for drug discovery and, in addition, are relatively costly to maintain and support. By contrast, open, public chemistry and biology data repositories (PubChem, ZINC, eMolecules, ChemSpider, etc. [4,5]) focus on publicly available data and are not designed for comprehensive data archiving by the user. Furthermore, these open repositories lack the ability to specify private data or limit sharing to specific groups. Traditionally, users have been forced to make a choice between sharing all or none of their data.

**BOX 1**

### Examples of informatics technologies

***Chemistry Databases***: Beilstein, SPRESI, LUCIA, ACD, ACX, Scifinder, Derwent, Gene-family SAR Databases (GVK, Jubilant-Biosys, Eidogen-Sertanty)
***Registrations systems***: Accord, ChemOffice, IDBS, ISIS
***Modeling and prediction software***: CTC labs, Accelrys, Tripos, Partek, Hypercube, Virtual Computational Chemistry Laboratory (http://www.vcclab.org).
***Infrastructure and other underlying toolkits***: ChemAxon, Daylight, Spotfire, OpenEye, Pipeline Pilot, MOE, Oracle, MySQL

**BOX 2**

### Examples of technologies leveraging or modeling network effects

***Public Content Databases***: KEGG [36], NCI, PDSP [37,38], PubChem, ChemBank (see also http://depth-first.com/articles/2007/01/24/thirty-two-free-chemistry-databases), SureChem.
***Federated Databases***: ChemSpider [4,5] eMolecules, ZINC
***Systems Biology Tools***: Ariadne Pathway Studio, Cytoscape, Ingenuity Pathways Analysis, MetaCore, MetaDrug [39], WikiPathways [40], Systems Biology Research Tool [41] (also see http://www.biochemweb.org/systems.shtml)
***Drug Discovery Platforms***: Collaborative Drug Discovery (CDD) [15], SEURAT (Synaptic Science LLC), NextBio
***Service and Product Provider Databases***: Assay depot, R&D Chemicals
***Knowledge, Information and Social Networking***: Wikipedia, OpenWetWare, BioSpace, BioPortfolio, ACS Member Network, LabMeeting, Laboratree, SciLink, SciMeet, Nature Network, Ensembl (see also http://scitechnet.blogspot.com/, http://docs.google.com/View?docid=dhs5x5kr_572hccgvcct)

Tools that enable the selective sharing of diverse data would be a valuable asset, especially within the area of neglected disease drug development for which the need for collaborative efforts has been well documented [16,17]. The neglected disease marketplace also provides a venue for experimenting with new approaches [18] to discovery research, with ramifications for the efficiency of mainstream drug discovery efforts. For community-based drug discovery to work within the larger biopharmaceutical industry, a platform is required with strong privacy, security and collaborative software features that can work within these constraints. Researchers and funding organizations must have a way to protect their intellectual property. Furthermore, the transaction costs of collaboration (data transfer between laboratories, formal IP transfer agreements, Material Transfer Agreements, Confidential Disclosure Agreements, business negotiations, business contracts, etc.) often slow down the speed of progress. The networked drug discovery process would be faster if, instead, distributed scientists from academia and industry were to work together simultaneously in a similar way to how software programmers collaborate on open source projects or how researchers work together within a single company. This was observed to a degree when companies worked relatively closely together to develop the early HIV protease inhibitor drugs rapidly (http://www.sciencemag.org/cgi/content/summary/272/5270/1882). We envisage a shift from the limited private networks that are predominant today, towards a future vision of interconnected open scientific networks (Figure 1A) facilitated by scientific networking software.

What is probably required of new collaborative software for biologists and chemists is a combination of capabilities that ensure privacy but allow selective collaborations when intentionally desired. For mainstream applicability, the tool must handle free text and also complex, heterogeneous drug discovery data and molecular structures. Furthermore, this complex data must be presented so that humans can easily draw conclusions and prioritize experiments from the data, procedures and ancillary information.

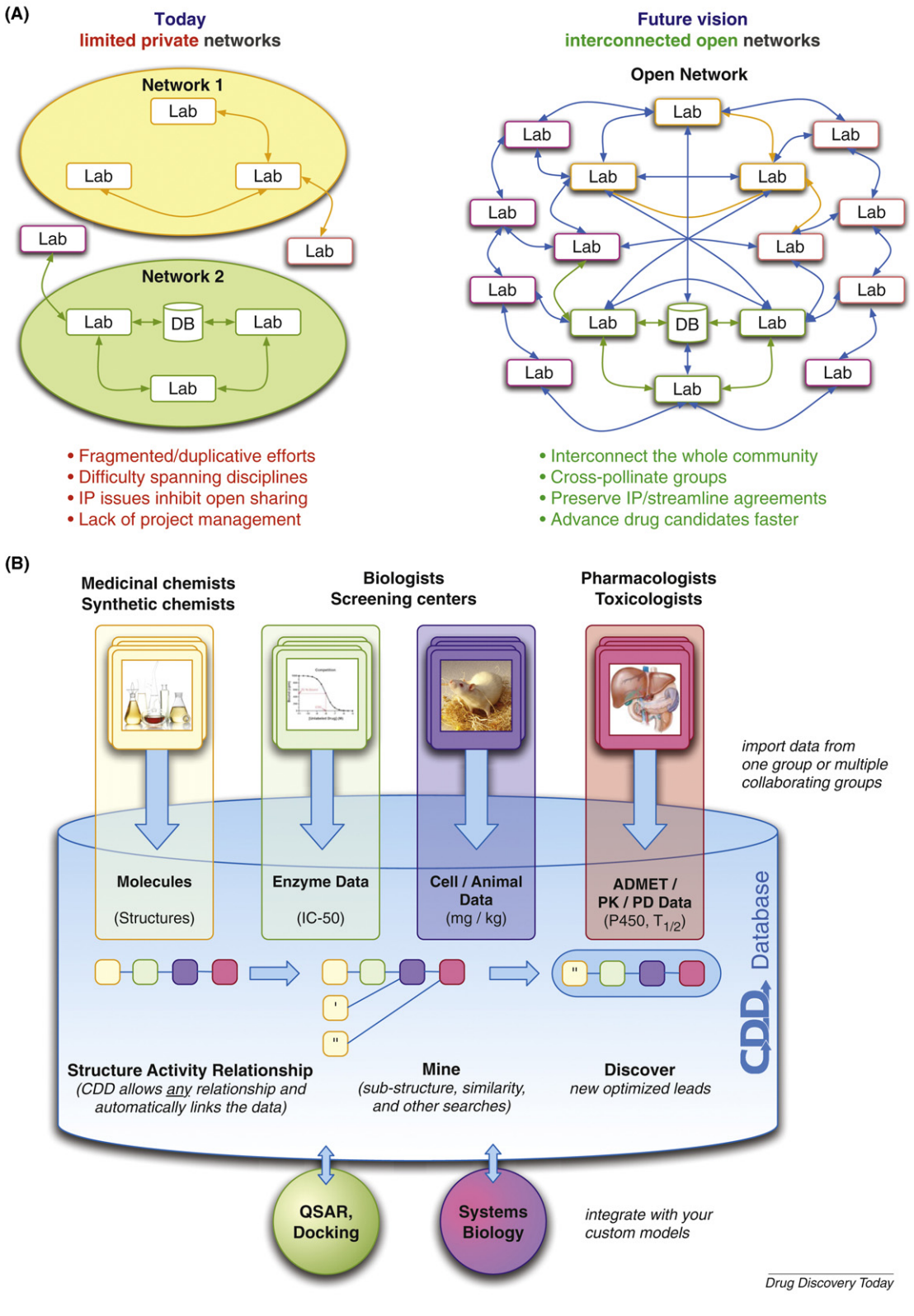



**FIGURE 1**

(A) Evolution of biomedical research from limited private networks to interconnected open networks. (B) Schematic overview of the CDD database.

## Collaborative drug discovery

Collaborative Drug Discovery (CDD)© in conjunction with a growing community of hundreds of scientists has developed a novel collaborative web-based platform to advance drug candidates more effectively (Figure 1B). The CDD platform can archive and mine a broad range of diverse objects that can later be selectively and securely shared with other researchers (or permanently kept private, which is the default behavior). The CDD Database is a

**BOX 3**

### CDD Technical details

The CDD web application is hosted on a dual-Xeon, 4GB RAM server with a RAID-5 SCSI hard drive array with one online spare. In case of machine failure, there is an online failover machine with live database and application code replicates. These machines sit behind a hardware firewall allowing in only HTTP/S connections from the Internet. All HTTP requests are redirected to HTTPS, providing transport confidentiality from the user's browser to the server. CDD currently colocates servers at ColoServe in San Francisco (www.coloserve.com), which provides redundant power, HVAC and backbone connections, fire suppression and physical security. The CDD software is written in Ruby on Rails over a MySQL database. Ruby on Rails is a novel web application framework noted for enabling productive, 'quick and clean', well-factored object-oriented software development; strong web standards adherence; thorough automated software testing and horizontal scaling via a shared-nothing architecture. CDD does a full onsite and offsite backup of the production database on a nightly basis and backs up incremental changes every five minutes while application code is backed up nightly.

hosted collaborative system with an important advantage over traditional PC-based database systems because it can enable secure login into the database from any computer, using any common browser (e.g. Firefox, Internet Explorer or Safari). This unique capability for a database system provides flexibility for the users. The CDD web-based database architecture (Box 3; Figure 1B) handles a broad array of data types that can be archived and then selectively shared among colleagues or openly shared on the Internet in standardized formats. The CDD platform incorporates Marvin, calculated pluggins for physical chemical calculations and the JChem Cartridge for structure searching from ChemAxon (Budapest, Hungary) within the application as the chemistry engine. This allows one to do sophisticated SAR analysis, including chemical pattern recognition (e.g. similarity and substructure searching), physical chemical property calculations, Boolean search and save capabilities for potency, selectivity, toxicity and other experimentally derived properties. CDD technologies handle heterogeneous data files from instruments and individual experiments as well as standardized csv and sdf file convertible formats that represent the chemical and biological data (compatible with the NIH Pubchem initiative). CDD is tailored for common data formats used by biologists such as Microsoft Excel™ (.xls) and text (.txt) files. The technology can mine against a variety of values including concentration, time, percent, real, integer, textline, cpm, rlu, Z/Z' plate statistics and $IC_{50}$ (log $IC_{50}$, $R^2$ values, Hillslope, etc.).

The results of the mined queries can be saved, exported (both as excel and sdf files), emailed and securely shared with others selected in the database application via the web. A researcher can temporally select which data to keep 100% private, to share with groups of individual researchers or to share more generally with the public. A further unique capability to CDD is the ability to compare all or subsets of public access data with private data simultaneously together in a single container.

The power of this collaborative approach to discovery can be seen in different types of community-based research projects. These range from traditional completely private collaborations, to temporally private collaborations that may become more open following a privacy escrow period, to completely open collaborations where researchers literally blog about the experiments as they occur. The platform supports the full range of collaborations and the current community includes leading researchers working on neglected, developing world infectious diseases like malaria [19–21], tuberculosis [22], Chagas disease, leishmaniasis, African sleeping sickness (and others), as well as drug discovery projects on more confidential commercial targets.

In developing the CDD platform we focused on collaborative capabilities and quickly learned the importance of balancing privacy and security features. Other important functions were to allow researchers to extract their data from the application to use elsewhere perhaps in other software for QSAR or visualization (Spotfire, etc.). As the software is hosted on a remote server it also lowers the cost of software distribution and updates alone, provides easier software evolution and is better for instantaneous collaborations. It is important to note that all data are backed up automatically so the user does not have to invest any time in doing this. The scalability of the database enables researchers globally to use the software. While username/password protected groups ensure secure IP protection for private data (e.g. with online bank accounts routinely used by millions of people each day). The proven software-as-a-service (SAAS) subscription standard business model is used with the added bonus that as the platform grows the open access community data become more valuable to other subscribers, because these integrated datasets are not available elsewhere. Today, the majority of scientists use CDD in the private mode and access the open access data. Users without a subscription can upload an unlimited amount of data at no cost if they do not restrict access to that data. If access is restricted through user-controlled privacy settings, a subscription fee is charged. To maximize the impact for the whole community, researchers pay financially, or pay with data, to use the platform.

## Traditional private to private collaborations

The following examples illustrate different ways in which a collaborative platform can facilitate scientific results and in many cases the collaborations would not have been likely to occur without the database and corresponding community. The first example of the application and power of the CDD platform for drug discovery involves three geographically dispersed groups working together in a collaboration facilitated by the CDD software (Figure 2). The collaboration identified novel chemosensitizers for combination malarial therapy to overcome parasitic resistance to chloroquine, which is the most widely used and inexpensive treatment for malaria today [23]. Chloroquine is hypothesized to work by inhibiting the polymerization of heme to hemazoin, leading to heme concentrations toxic to *Plasmodium falciparum*. Historically, this drug has been the most inexpensive treatment for malaria, although resistant strains of malaria have emerged. Specific mutations (T76K, S163R) in the *P. falciparum* chloroquine resistance transporter protein have been associated with resistance in the literature.

Traditionally, one would begin the search for novel chemosensitizers by either screening natural products or synthesized molecules. With web-based collaborative capabilities; however, researchers can now rapidly compare compounds across groups


**FIGURE 2**

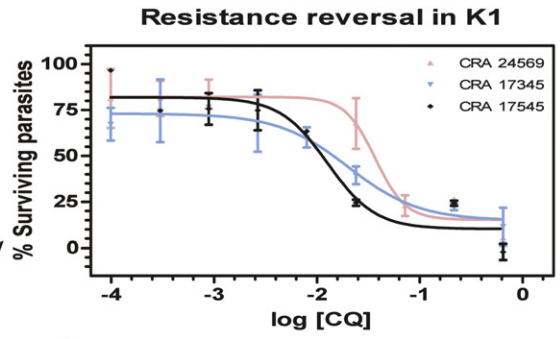
Collaborative research between three different groups sharing chemical structures of interest 'in house' with a biologist half-way around the world.

and sources from around the globe for novel or similar compounds. When one is not particularly interested in novel compositions of matter (as is frequently the case for neglected diseases), the efficiency of the research can be increased by tapping directly into data from the current generation and past generations of scientists.

The most promising compounds from this three way collaboration were shipped to the University of Cape Town, and then tested to identify novel compounds and several FDA-approved drugs that almost completely reversed the chloroquine resistance in resistant strains in human red blood cells (Figure 3).

In this case, there was a known chemotype (chemical substructure with an aromatic ring four atoms from a secondary nitrogen) that was conserved among chemosensitizers initially observed in verapamil [19,24] (Figure 4). Because groups were willing to work collaboratively, the compounds being screened at UCSF by Professor James McKerrow's group were shared in an 'invitation-only', username and password protected secure group to maintain IP protection. A substructure search for the known chemosensitizer substructure led to the identification of hundreds of compounds for laboratory evaluation by the laboratories of Dr. Peter Smith in Cape Town. Leading candidates were identified and sent for

| Compound | Dose (μg/ml) | CQ IC50 in K1 (nM) | RMI* |
|---|---|---|---|
| Chloroquine | - | 105.1 | - |
| CRA 10892 | 2 | 18.2 | 0.173 |
| CRA 17408 | 0.2 | 23.28 | 0.222 |
| CRA 24397 | 2 | 16.13 | 0.153 |
| CRA 24356 | 5 | 29.02 | 0.276 |
| CRA 24569 | 2 | 33.88 | 0.322 |
| CRA 17345 | 2 | 13.92 | 0.132 |
| CRA 17545 | 4 | 11.33 | 0.108 |
| 1500115 | 5 | 14.89 | 0.142 |
| 1500325 | 0.2 | 64.57 | 0.614 |
| 1500373 | 0.1 | 87.15 | 0.829 |
| 1501134 | 0.5 | 35.35 | 0.336 |




Drug Discovery Today

**FIGURE 3**

Resistance reversal experiments in *Plasmodium falciparum* K1 using molecules derived from a substructure search across multiple datasets through the CDD database. Inset shows example dose response curves. The highest concentration at which no antimalarial activity was observed was established for each compound. This concentration of each compound is included in a chloroquine dose response curve against the chloroquine resistant strain, K1. The ratio of the $IC_{50}$ in presence and absence of the compound (RMI) corresponds to the chloroquine reversal activity at the chosen concentrations. Several compounds almost completely reversed chloroquine resistance *in vitro* (7-fold), and these include the FDA-approved drugs pimozide, vinblastine, sertraline and dihydroergotamine mesylate.

evaluation of efficacy in assays using the resistant African malarial parasite strains in human red blood cells. Novel compounds that almost entirely reversed the resistance were identified (Figure 3). This process shaves months off a project timeline relative to synthesizing new compounds from scratch.

The same substructure query was used on the set of known FDA-approved and orphan-approved drug compounds (including structures) provided by Dr. Christopher Lipinski (www.collaborative-drug.com/register). Because the compounds are already approved for other indications, they could be developed rapidly if found to be efficacious. Eighteen compounds were identified with the conserved substructure and half a dozen were purchased, shipped to Africa and, when tested in the assay, these known drugs were shown to reverse (7-fold reversal) the resistance in human blood cells almost completely (Figure 4). Because the compounds in the Lipinski-CDD Database are drugs that are already known to be safe and efficacious in humans, the process could save years off the drug development timeline [25,26]. The repurposing of old drugs for malaria has also been indicated by others recently [27,28] as a generally useful strategy that can also be applied elsewhere.

## Temporarily restricted data sharing

A second example of how the CDD platform can be used involves a large set of anti-malarial animal SAR data that was intentionally kept private for 12 months before being released for use by the malaria researcher community by Professor R. Kiplan Guy (St. Jude Childrens Research Hospital). The data came from a two-volume collection of studies on malarial drugs published by the U.S. Army in 1946 [29]. This publication had contributions from a number of leading researchers of the time and was designed to help researchers develop effective anti-malarial drugs, and to serve as a model for how scientists could develop drugs for other infections. The corresponding SAR dataset consisted of over 12 000 hand-drawn molecules with bioactivity relative to known compounds tested in half a dozen animal species. The collection contains other pharmacological data, in addition to their level of toxicity (see Figure 5). Although the original studies were decades old, now, for the first time, the data are accessible in a format for computational model building and direct comparisons with recent experimental results. Professor Alex Tropsha's group at the University of North Carolina was able to build new predictive computational models using their combinatorial QSAR modeling techniques [30,31] with this 'new' data. Initially, 131 active and 228 inactive compounds (that were most chemically similar to actives) were selected from 3133 compounds screened for anti-*P falciparum* (3D7 strain) activity and used to develop preliminary combinatorial QSAR *k*-nearest neighbors (kNN) classification models with Dragon descriptors. Three hundred and eighty three internally validated models afforded a correct classification rate for an external dataset of 80.7%. Additionally, 674 compounds (with log activity −1.52–2.78) with *in vivo* data from Peking ducks inoculated with *Plasmodium lophurae* malaria were also used to generate 283 continuous kNN models ($R^2 = 0.80$ for an external test set of 80 molecules). These models enabled virtual screening of libraries of compounds to find further compounds for *in vitro* testing and repopulating the CDD database for selection of candidates for further *in vitro* testing.

In this case, the group only has access to data with the permission of the data owners to generate and refine a master combinatorial model. Moreover, the exchange of data is governed by

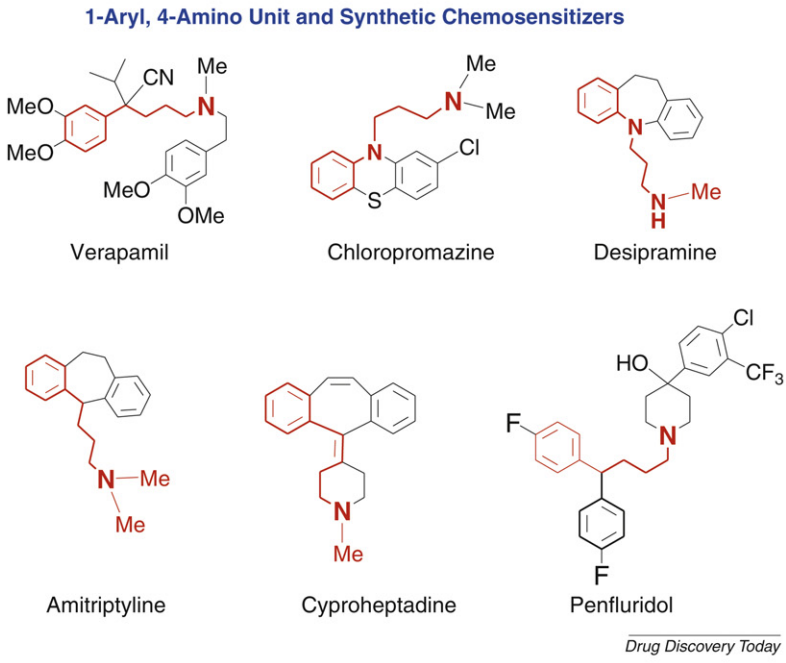



**FIGURE 4**

Conserved chemotype with chemosensitizer activity consisting of an aromatic ring four atoms from a secondary nitrogen.

appropriate intellectual property rights agreements or, even simpler, can be sidestepped entirely by having a mutually trusted intermediary working under confidentiality until a discovery is found worthy of the cost and hassle of additional agreements.

## Putting 'open source drug discovery' philosophy into practice

The third case study highlights not only the collaborative data and technology but also the value of a growing community of networked scientists. Professor Jean-Claude Bradley (Drexel University) is particularly committed to open science and neglected disease by providing all his experimental results openly online (http://usefulchem.blogspot.com/).

Professor Bradley approached CDD to find a collaborative partner to screen the products of his UGI-4CC libraries for anti-malarial activity. Via the CDD Community (Figure 6) he started collaborations with Dr. Rosenthal's Group at the UCSF General Hospital to access their enzyme and cell-based anti-malarial screens. A third group (Dr. Guha at Indiana University) also provided the computational expertize to help select which compounds to screen. All synthetic products and screening results were provided openly via the software platform for other community researchers to see and use (Figure 7). In addition Professor Bradley blogged about these experiments at his own site at: http://usefulchem.blogspot.com/2007/08/usefulchem-on-cdd.html.

## Pharmaceutical organizations: virtual Pharma—FIPCO and FIPNET

A web-based database technology that incorporates chemistry, biology and social networking components should appeal to those in the scientific community who are focused on 'virtual drug discovery' in contrast to the traditional pharmaceutical brick-and-mortar discovery organizations. Anecdotal evidence points to a growing segment of scientists forming small, focused discovery groups and who outsource as much of the drug discovery and development process as possible. In addition, many orphan or other disease foundations conduct their research in a decentralized manner. Virtual discovery organizations are highly cost sensitive and often delocalized geographically. To track the data and ensure that all members of an organization are fully informed requires a flexible database system. Ideally, the vendors or contractors should be able to upload their data easily and others in the organization should be able to mine it readily for new relationships. Larger pharmaceutical organizations are also in the process of transition from fully integrated pharmaceutical companies (FIPCO) to fully integrated pharmaceutical networks (FIPNET), such that discovery and development will be 'just in time' analogous to the supply chain approach for manufacturing industries. In such cases the social networking component of a database will be valuable for connecting members of the network as well as project tracking and planning.

## Discussion

It is becoming widely appreciated that it is more efficient when collaborating researchers share information and work together [8,32] and in addition there are other benefits of collaboration [12,13] and collaboration networks [14]. In neglected disease and orphan drug research [33], as embodied by the NIH anti-convulsant screening program [34], it has been suggested that best practices include creative application of technologies, collaboration and flexibility. Translational research that can move preclinical research into the clinic [35] also requires collaborative researchers and a supportive infrastructure. A software platform that allows researchers to easily toggle between private and shared

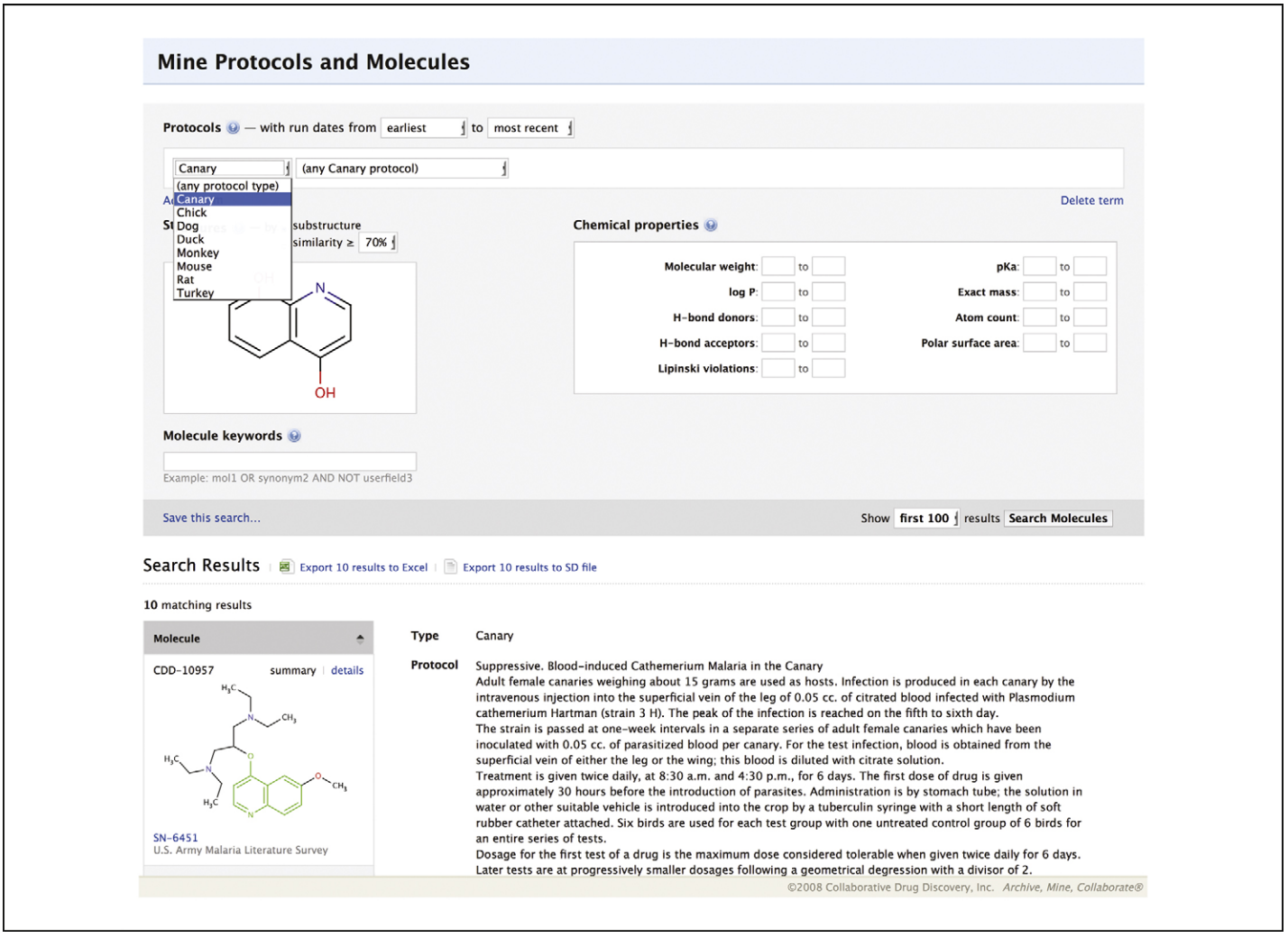



**FIGURE 5**

Community-based anti-malarial animal data. These data were released for general use in a public group following a 12 months escrow period when the data were exclusively only in a private group.

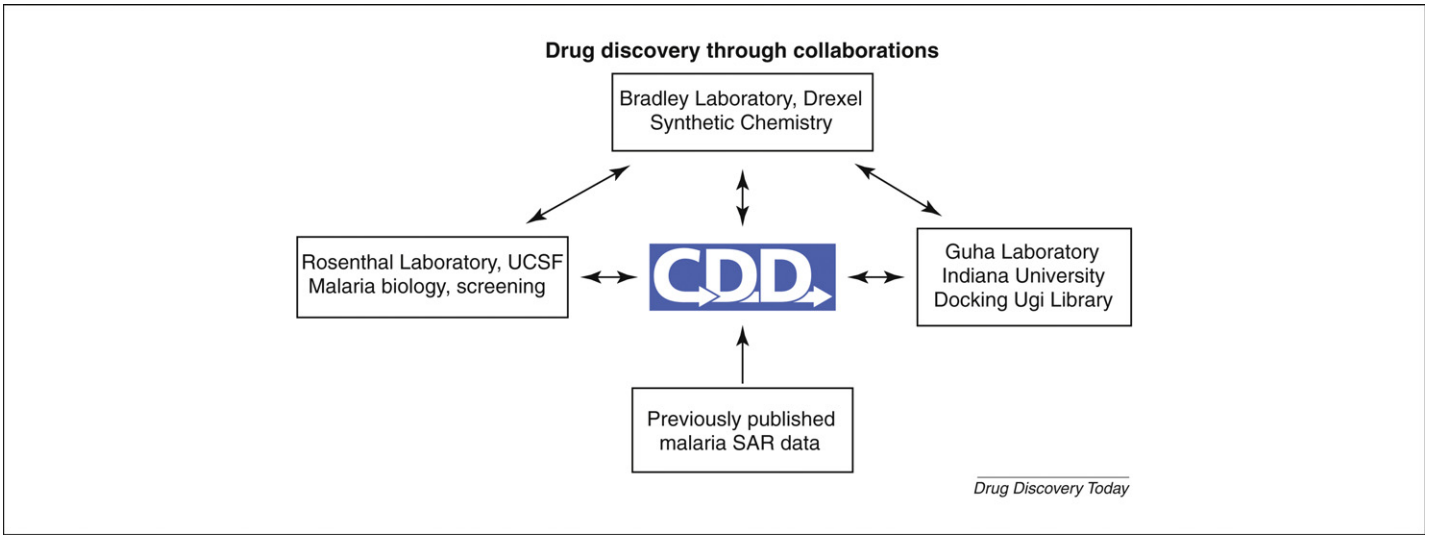



**FIGURE 6**

Chemists, biologists and computational scientists can privately or openly share structures, SAR and predictions via CDD Database as part of a growing community of scientists.

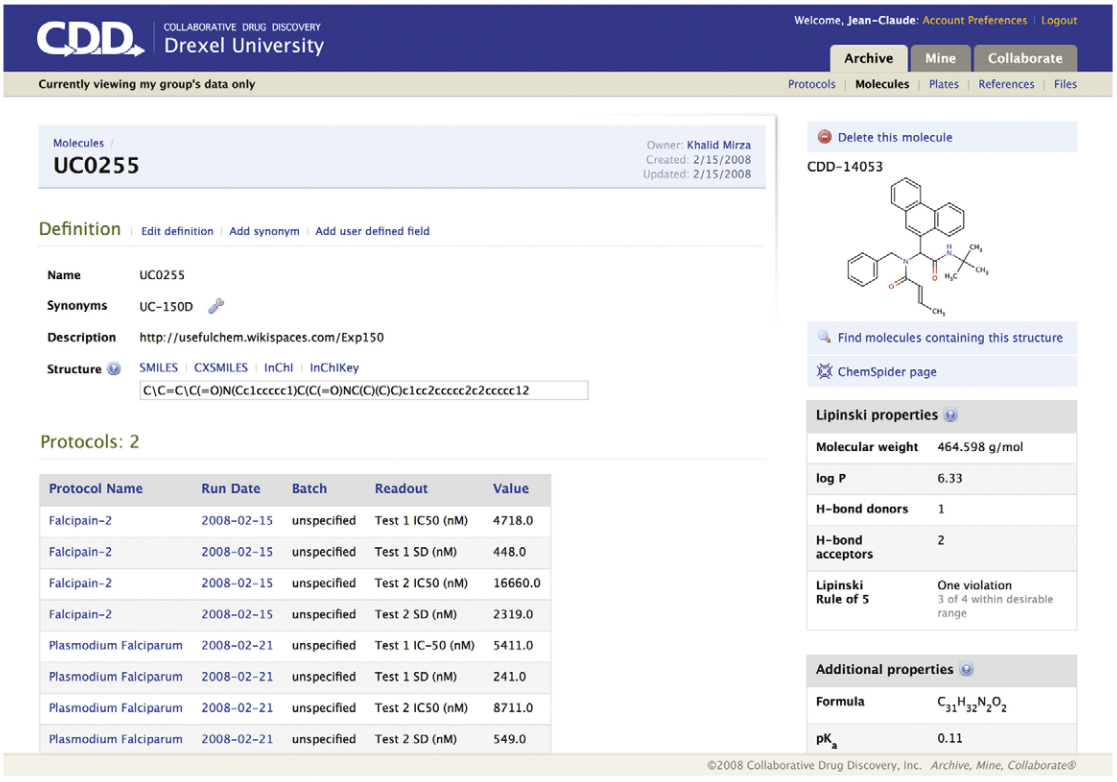



**FIGURE 7**

Representative drug-centric view with structural information, bioactivity data and calculated properties. Target centric views are also supported for target validation.

datasets would provide a technical solution to enable them to make decisions if, when and with whom they want in order to share some, all or none of their data. The same model could also be applied to enable greater efficiency for pharmaceutical and biotechnology companies as well as academic-driven or foundation-driven drug discovery in secure, private collaborative groups. In all cases, collaborative researchers can go far beyond what they would normally be able to do with just their own limited laboratory networks, ideas and resources.

## Conclusion

What do these trends mean for the future of drug discovery research and development informatics technologies and where is this field headed? To date, there has not been an extensive assessment of previously developed web semantic tools and their utilization. Yet, even in the absence of this, the trend across all industries towards SAAS web-hosted applications will become more prevalent for the drug discovery industry too. The incorporation of more private to private collaborative features and web2.0 social networking features would provide an integrated platform (CDD or similar sets of technologies) as a personal e-lab notebook for capturing organizing and collaborating with other scientists in the growing community, while maintaining the required security and privacy features. These new capabilities provide a useful, secure environment for any research and development organization to tap immediately into collective expertize whether it is connecting academic postdoctoral researchers, employees at a research organization half-way around the world with their colleagues in other countries or employees within a large biopharmaceutical company. Participating laboratories contribute in aggregate to the generation of datasets and predictive models, yet no specific data or approaches need to be exposed to the other research groups. Each group is then able to exploit the model to help guide its own screening activities or explore other scaffolds *in silico* without revealing any aspects of its intellectual approach.

New informatics tools that incorporate biology and chemistry with social networking technologies should enable a better, faster, cheaper mechanism to discover and advance drug candidates in a collaborative manner, regardless of whether they are for neglected, orphan or potential 'blockbuster' diseases.

## Conflicts of interest

Moses Hohman, Kellan Gregory and Barry Bunin are employed by Collaborative Drug Discovery Inc. Sean Ekins is a consultant for Collaborative Drug Discovery Inc. Kelly Chibale and Peter J. Smith have no conflicts of interest to declare.

## Acknowledgements

The authors gratefully acknowledge Jim Wikel and Deborah Bunin for comments and the support of the following researchers without whom none of this would have been possible: C. Lipinski (Melior Discovery), J. McKerrow (UCSF), E. Hansell (UCSF), K. Guy (St Jude CRH), A. Shelat (St. Jude CRH), A.Tropsha (Univ. of North

Carolina), L. Zhang (Univ. of North Carolina), H. Zhu (Univ. of North Carolina), J. Claude Bradley (Drexel), E. Messner (Drexel), K. Mirza (Drexel), R. Guha (Indiana Univ.), P. Rosenthal (UCSF General Hospital), J. Gut (UCSF General Hospital), and the rest of the CDD community. We also kindly acknowledge the reviewers suggestions.

## References


1 Carpy, A.J. and Marchand-Geneste, N. (2006) Structural e-bioinformatics and drug design. *SAR QSAR Environ. Res.* 17, 1–10
2 Ertl, P. and Jelfs, S. (2007) Designing drugs on the internet? Free web tools and services supporting medicinal chemistry. *Curr. Top Med. Chem.* 7, 1491–1501
3 Munos, B. (2006) Can open-source R&D reinvigorate drug research? *Nat. Rev. Drug Discov.* 5, 723–729
4 Williams, A.J. (2008) Internet-based tools for communication and collaboration in chemistry. *Drug Discov. Today* 13, 502–506
5 Williams, A.J. (2008) A perspective of publicly accessible/open-access chemistry databases. *Drug Discov. Today* 13, 495–501
6 Ekins, S. *et al.* (2008) Molecular characterization of CYP2B6 substrates. *Curr. Drug Metab.* 9, 363–373
7 Ekins, S. *et al.* (2008) Computational discovery of novel low micromolar human pregnane X receptor antagonists. *Mol. Pharmacol.* 74, 662–672
8 Nwaka, S. and Ridley, R.G. (2003) Virtual drug discovery and development for neglected diseases through public-private partnerships. *Nat. Rev. Drug Discov.* 2, 919–928
9 Ekins, S. *et al.* (2007) In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br. J. Pharmacol.* 152, 9–20
10 Newman, M.E.J. (2003) Random graphs as models of networks. In *Handbook of Graphs and Networks* (Bornholdt, S. and Schuster, H.G., eds), pp. 35–68, Wiley-VCH
11 Barabasi, A.-L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113
12 Lee, S. and Bozeman, B. (2005) The impact of research collaboration on scientific productivity. *Soc. Stud. Sci.* 35, 673–702
13 Katsouyanni, K. (2008) Collaborative research: accomplishments & potential. *Environ. Health* 7, 3
14 Guimera, R. *et al.* (2005) Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308, 697–702
15 Bunin, B.A. *et al.* eds (2007) *Cheminformatics Theory, Practice, and Products*, Springer–Verlag
16 Radish, J. (2007) More medicines for neglected and emerging infectious diseases. *Bull. World Health Organ.* 85, 569–648
17 Grace, C. (2006) Developing new technologies to address neglected diseases: the role of product development partnerships and advanced market commitments, Department for International Development (DFID) Health Research Center Report
18 Wechsler, J. (2007) Manufacturers Tackle Neglected Diseases in http://biopharminternational.findpharma.com/biopharm/Article/Manufacturers-Tackle-Neglected-Diseases/ArticleStandard/Article/detail/439419
19 Chouteau, F. *et al.* (2005) Exploiting a basic chemosensitizing pharmacophore hypothesis. Part 1: synthesis and biological evaluation of novel arylbromide and bicyclic chemosensitizers against drug-resistant malaria parasites. *Bioorg. Med. Chem. Lett.* 15, 3024–3028
20 Chiyanzu, I. *et al.* (2005) Design, synthesis and anti-plasmodial evaluation *in vitro* of new 4-aminoquinoline isatin derivatives. *Bioorg. Med. Chem.* 13, 3249–3261
21 Weisman, J.L. *et al.* (2006) Searching for new antimalarial therapeutics amongst known drugs. *Chem. Biol. Drug Des.* 67, 409–416
22 Gold, B. *et al.* (2008) Identification of a copper-binding metallothionein in pathogenic mycobacteria. *Nat. Chem. Biol.* 4, 609–616
23 Tran, C.V. and Saier, M.H., Jr (2004) The principal chloroquine resistance protein of Plasmodium falciparum is a member of the drug/metabolite transporter superfamily. *Microbiology* 150 (Pt 1), 1–3
24 Warhurst, D.C. (2003) Polymorphism in the Plasmodium falciparum chloroquine-resistance transporter protein links verapamil enhancement of chloroquine sensitivity with the clinical efficacy of amodiaquine. *Malar. J.* 2, 31
25 O'Connor, K.A. and Roth, B.L. (2005) Finding new tricks for old drugs: an efficient route for public-sector drug discovery. *Nat. Rev. Drug Discov.* 4, 1005–1014
26 Chong, C.R. and Sullivan, D.J., Jr (2007) New uses for old drugs. *Nature* 448, 645–646
27 Chong, C.R. *et al.* (2006) A clinical drug library screen identifies astemizole as an antimalarial agent. *Nat. Chem. Biol.* 2, 415–416
28 Sannella, A.R. *et al.* (2008) New uses for old drugs. Auranofin, a clinically established antiarthritic metallodrug, exhibits potent antimalarial effects *in vitro*: mechanistic and pharmacological implications. *FEBS Lett.* 582, 844–847
29 Wiselogle, F.Y., ed. (1946) *A Survey of Antimalarial Drugs 1941–1945*, J.W. Edwards
30 de Cerqueira Lima, P. *et al.* (2006) Combinatorial QSAR modeling of P-glycoprotein substrates. *J. Chem. Inf. Model* 46, 1245–1254
31 Kovatcheva, A. *et al.* (2005) QSAR modeling of datasets with enantioselective compounds using chirality sensitive molecular descriptors. *SAR QSAR Environ. Res.* 16, 93–102
32 Nwaka, S. and Hudson, A. (2006) Innovative lead discovery strategies for tropical diseases. *Nat. Rev. Drug Discov.* 5, 941–955
33 Clark, J.E. *et al.* (2002) Novel trends in orphan market drug discovery: amyotrophic lateral sclerosis as a case study. *Front. Biosci.* 7, c83–c96
34 White, H.S. *et al.* (1998) The National Institutes of Health Anticonvulsant Drug Development program: screening for efficacy. *Adv. Neurol.* 76, 29–39
35 Escolar, D.M. *et al.* (2002) Collaborative translational research leading to multicenter clinical trials in Duchenne muscular dystrophy: the Cooperative International Neuromuscular Research Group (CINRG). *Neuromuscul. Disord.* 12 (Suppl. 1), S147–154
36 Kanehisa, M.A. *et al.* (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res.* 30, 42–46
37 Strachan, R.T. *et al.* (2006) Screening the receptorome: an efficient approach for drug discovery and target validation. *Drug Discov. Today* 11, 708–716
38 Roth, B.L. *et al.* (2004) Screening the receptorome to discover the molecular targets for plant-derived psychoactive compounds: a novel approach for CNS drug discovery. *Pharmacol. Ther.* 102, 99–110
39 Ekins, S. *et al.* (2007) Pathway mapping tools for analysis of high content data. *Methods Mol. Biol.* 356, 319–350
40 Pico, A.R. *et al.* (2008) WikiPathways: pathway editing for the people. *PLoS Biol.* 6, e184
41 Wright, J. and Wagner, A. (2008) The systems biology research tool: evolvable open-source software. *BMC Syst. Biol.* 2, 55