Free online resources enabling crowd-sourced drug discovery

The availability of freely accessible online resources to enable and support drug discovery has blossomed in recent years. The PubChem platform is now accompanied by a myriad of other online databases including ChEBI, DrugBank, the Human Metabolome Database and ChemSpider. The access to the array of software tools and diverse data in public domain provides capabilities previously only available within the confines of organisations (eg, big Pharma) that could afford significant investments in cheminformatics. This paper provides an overview of the internet resources available to drug discovery scientists and discusses the advantages of such accessibility but also the potential risks that reside within the data. It also examines what the present resources continue to lack and sets a vision for future approaches to providing internet-based resources for drug discovery.

By Dr Antony J. Williams, Valery Tkachenko, Dr Chris Lipinski, Professor Alexander Tropsha and Dr Sean Ekins

The past five years have seen a mini revolution in the availability of resources to support drug discovery and, in particular, databases searchable by molecular structure (**Figure 1**). Chemistry information on the internet continues to become more widely accessible and at an increasing rate. There are many freely available chemical compound databases on the web[1,2]. These databases generally contain the chemical identifiers in the form of chemical names (systematic and trade) and registry numbers. Since the files in the databases are assembled in a heterogeneous manner, using variations in deposition processes and procedures to handle chemical structures, the resulting data are plagued with inconsistencies and quality issues. There are many databases available from which the drug discovery community can derive value. These databases generally have a specific focus based on the domain expertise of the hosting organisation; examples include databases of curated literature data, chemical vendor catalogues, patents, analytical data, biological data, etc. There are too many to include in this single article so only a small number will be discussed. For example, the authors recommend a recent article that assesses the expanding public and commercial databases containing bioactive compounds[3] and conclude that the commercial efforts are ahead of the public ones.

The availability of molecule databases such as PubChem (http://pubchem.ncbi.nlm.nih.gov/) has dramatically changed the landscape of publicly available cheminformatics resources, yet PubChem covers only a fraction of the chemical universe, mostly of interest to chemical genomics and pharmaceutical research. PubChem was

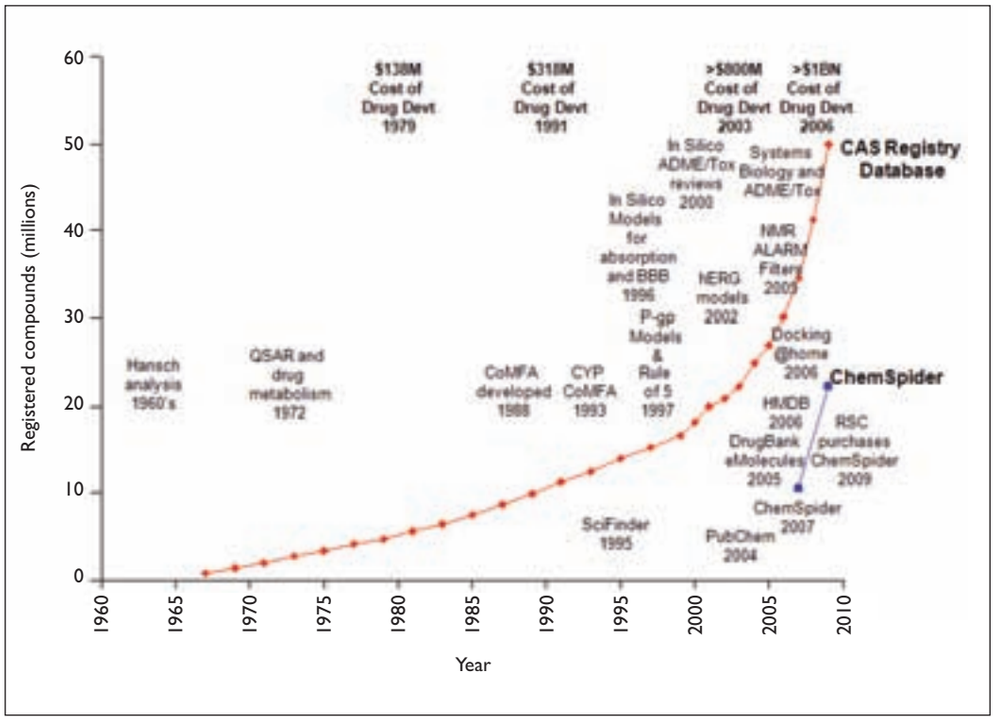


Figure 1
A graphical interpretation of the history of chem/bioinformatics software, model and database development, and increasing drug development costs versus registered compounds in the CAS Registry and the ChemSpider database

launched by the NIH in 2004 to support the 'New Pathways to Discovery' component of the Roadmap for Medical Research[4]. PubChem archives and organises information about the biological activities of chemical compounds into a comprehensive database and is the informatics backbone for the Molecular Libraries and Imaging Initiative, which is part of the NIH Roadmap. Pubchem is also intended to empower the scientific community to use small molecule chemical compounds in their research as molecular probes to investigate important biological processes or gene functions. The PubChem compound repository presently contains more than 25 million unique structures with biological property information provided for many of the compounds. For now, PubChem remains focused on its initial intent to support the Molecular Libraries Initiative and serves as an extremely valuable and authoritative resource for cheminformatics and chemical genomics. However, there are a number of constraints around the system, especially in its place as a repository of data and information without a special effort toward curating these data. Naturally, in the absence of data curation any errors in the data are transferred across many online databases that depend on PubChem and ultimately, the errors influence the quality of computational models based on this data.

The **Chemical Entities of Biological Interest**, or ChEBI database (http://www.ebi.ac.uk/chebi/) is a highly curated database of molecular entities focused on small chemical compounds. The entities are either natural products or synthetic products used to intervene in the processes of living organisms. ChEBI includes an ontological classification (**Figure 2**), whereby the relationships between molecular entities or classes of entities and their 'parents' and/or 'children' are specified. While the database presently offers access to close to 19,000 entities this is expected to expand to more than 440,000 by the end of October (http://www.ebi.ac.uk/chebi/newsForward.do#ChEMBL%20data%20integration). The database is available for download by anonymous FTP (ftp://ftp.ebi.ac.uk/pub/databases/chebi/).

The **Human Metabolome Database** (http://www.hmdb.ca)[5,6] (HMDB) is a comprehensive curated collection of human metabolite and human metabolism data. It contains records for more than 6,800 endogenous metabolites. In addition to its comprehensive literature-derived data, the HMDB also contains an extensive collection of experimental metabolite concentration data compiled from hundreds of mass spectra (MS) and Nuclear Magnetic resonance (NMR) metabolomic analyses performed on urine, blood

References

1 Williams, AJ (2008). A perspective of publicly accessible/open-access chemistry databases. Drug Discov Today 13 (11-12), 495-501.
2 Williams, AJ (2008). Internet-based tools for communication and collaboration in chemistry. Drug Discov Today 13 (11-12), 502-506.
3 Southan, C et al (2009). Quantitative assessment of the expanding complementarity between public and commercial databases of bioactive compounds. J Cheminformatics 1, 10.
4 Office of Portfolio Analysis and Strategic Initiatives, N.I.o.H (2008). The NIH Roadmap Initiative.
5 Wishart, DS et al (2007). HMDB: the Human Metabolome Database. Nucleic Acids Res 35 (Database issue), D521-526.
6 Wishart, DS et al (2009). HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res 37 (Database issue), D603-610.

Continued on page 36

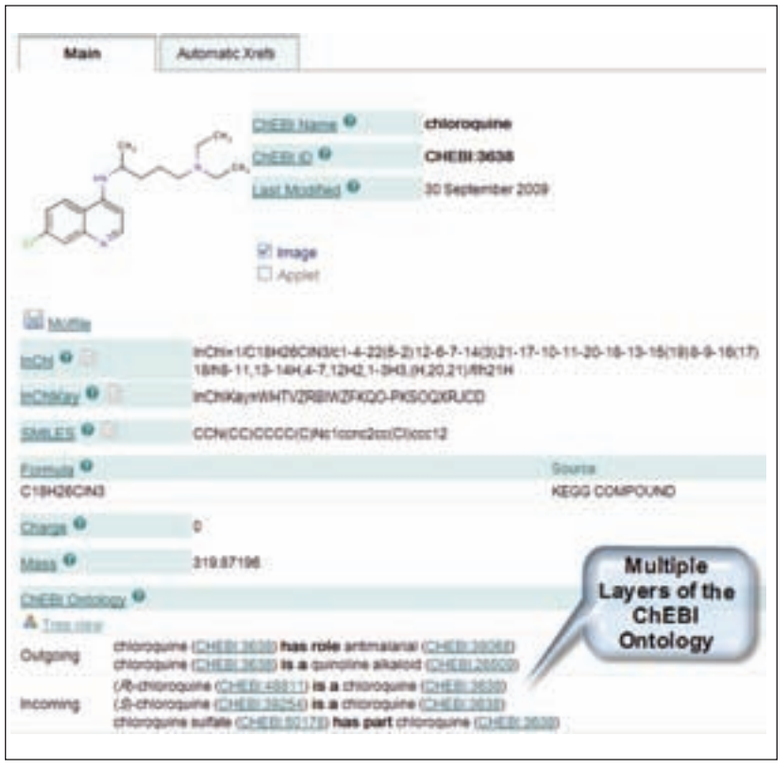
and cerebrospinal fluid samples. This is further supplemented with thousands of NMR and MS spectra collected on purified, reference metabolites. Each metabolite entry in the HMDB contains data fields including a comprehensive compound description, names and synonyms, structural information, physicochemical data, reference NMR and MS spectra, biofluid concentrations, disease associations, pathway information, enzyme data, gene sequence data, SNP and mutation data as well as extensive links to images, references and other public databases. Recent improvements have included spectra and substructure searching.

DrugBank (http://www.drugbank.ca/) is a manually curated resource[7] assembled from a series of other public domain databases (KEGG, PubChem, ChEBI, PDB, Swiss-Prot and GenBank) and enhanced with additional data generated within the laboratories of the hosts. The database aggregates both bioinformatics and cheminformatics data and combines detailed drug data with comprehensive drug target (ie protein) information. The database contains FDA approved small molecule and biotech drugs as well as experimental drugs, representing nearly 5,000 molecules[8]. The database supports extensive text, sequence, chemical structure and relational query searches of the nearly 100 data fields. The data from DrugBank has been used to show that the drug to drug-target relationship is scale-free and several classes of proteins are selectively enriched as drug targets for FDA approved drugs[9].

ZINC (http://zinc.docking.org/index.shtml) is a free, searchable database of commercially available compounds for virtual screening[10,11]. The library contains more than 20 million molecules, each with a 3D structure and gathered from the catalogues of compounds from vendors. All molecules in the databases are assigned biologically-relevant protonation states and annotated with molecular properties.

ChemSpider (http://www.chemspider.com/)[1,2] is a community resource for chemists provided by the Royal Society of Chemistry (**Figure 3**). It offers a number of facilities that distinguishes the service from many of the other databases listed in this article. At the time of writing it contains more than 23 million unique chemical entities aggregated from more than 200 diverse data sources, including government databases, chemical vendors, commercial database vendors, publishers, all of the databases listed above and from individual chemists. ChemSpider has also integrated the SureChem patent database collection (http://www.surechem.org/) of structures to facilitate structure-based linking to patents between the two data collections. ChemSpider can be queried using structure/substructure searching and alphanumeric text searching of both intrinsic as well as predicted molecular properties. Unique capabilities relative to other public chemistry databases include real time curation of the data, association of analytical data with chemical structures, real-time deposition of single or batch chemical structures (including with activity data) and transaction-based predictions of physicochemical data. A series of web services are provided to allow integration to the system for the purpose of searching and linking with other online databases from other groups (academia or industry). The integration can be with free or commercial resources. For example, Collaborative Drug Discovery, Inc (http://www.collaborativedrug.com) recently provided links to ChemSpider for molecules in its CDD database[12] thereby providing an integration path between a commercial resource and a public domain database. CDD is a highly secure, commercial collaborative drug discovery informatics platform and a new type of collaborative system that handles a broad array of

Figure 2
The ChEBI database offers a detailed ontology including subdivision into (1) Molecular Structure, in which molecular entities or parts thereof are classified according to composition and structure (2) Role, which classifies entities either on the basis of their role within a biological context, eg antibiotic, antiviral agent, coenzyme, hormone, or on the basis of their intended use by humans, eg pesticide, antirheumatic drug, fuel. The structure shown is for chloroquine, identified as an antimalarial quinoline alkaloid in the ChEBI ontology



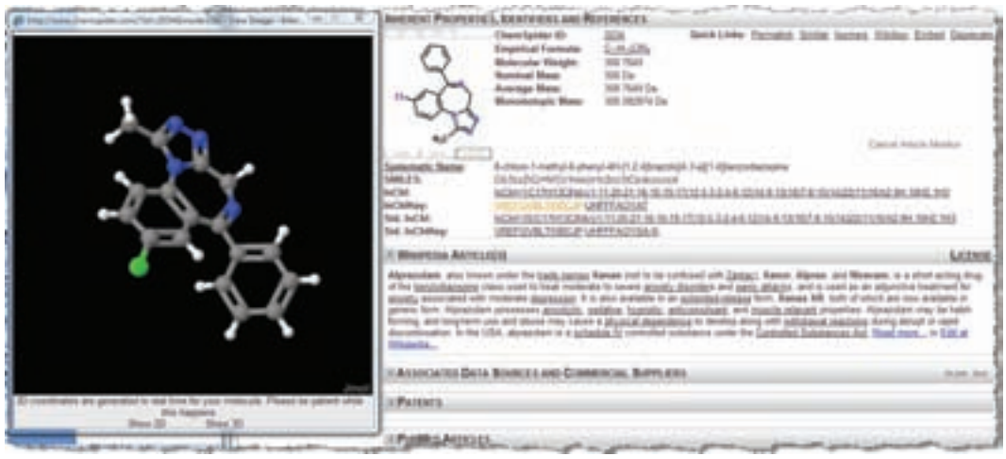


Figure 3
ChemSpider provides links to Wikipedia articles, links out to the original data sources and commercial suppliers, links out to patents and articles on PubMed. Flexible search capabilities are available, together with visualisation tools such as a real time 3D optimisation engine and display module

Continued from page 34

7 Wishart, DS et al (2006). DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. Nucleic Acids Res 34 (Database issue), D668-672.
8 Wishart, DS et al (2008). DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 36 (Database issue), D901-906.
9 Ma'ayan, A et al (2007). Network analysis of FDA approved drugs and their targets. Mt Sinai J Med 74 (1), 27-32.
10 Irwin, JJ and Shoichet, BK (2005). ZINC – a free database of commercially available compounds for virtual screening. J Chem Inf Model 45 (1), 177-182.
11 Irwin, JJ et al (2005). Virtual screening against metalloenzymes for inhibitors and substrates. Biochemistry 44 (37), 12316-12328.
12 Hohman, M et al (2009). Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery. Drug Disc Today 14, 261-270.
13 Tetko, IV et al (2008). Critical assessment of QSAR models of environmental toxicity against Tetrahymena pyriformis: focusing on applicability domain and overfitting by variable selection. J Chem Inf Model 48 (9), 1733-1746.

Continued on page 37

data types that can be archived and then selectively shared among colleagues or openly shared in standardised formats, at each research group's discretion. A focus of CDD is facilitating the growth of global collaborative research networks for neglected diseases such as malaria, African sleeping sickness, Chagas disease and tuberculosis. Subsequently there are currently 50 datasets available to the public upon registration which can be readily substructure or similarity searched.

The importance of chemical data curation in QSAR modelling

Molecular modellers and cheminformaticians alike typically analyse data generated by other researchers providing, in general, experimental data. Consequently, when it comes to the quality of these data modellers are always at the mercy of the providers. Practically any modelling cheminformatics study entails the calculation of chemical descriptors that are expected to accurately reflect the intricate details of the underlying chemical structures. Obviously, any error in the structure translates into either an inability to calculate the descriptors for erroneous chemical records or into erroneous descriptors. Naturally, the models built with this data are either restricted to only a fraction of the formally available data or, worse, they are merely inaccurate. As both data and models of the data, as well as the body of scholarly publications in cheminformatics, continue to grow, it becomes increasingly important to address the issue of data quality that inherently effects the quality of models.

How significant is the problem of accurate structure representation as it concerns the adequacy and accuracy of cheminformatics models? A few recent reports indicate that this problem should be given serious attention. For instance, benchmarking studies by a large group of collaborators from six laboratories[13,14] have clearly demonstrated that the type of chemical descriptors has much greater influence on the prediction performances of QSAR models than the nature of the model optimisation techniques. Furthermore, in another recent seminal publication[15], the authors clearly pointed out the importance of chemical data curation in the context of QSAR modelling (eg incorrect structures generated from either correct or incorrect SMILEs). Their main conclusions were that small structural errors within a dataset could lead to significant losses in the predictive abilities of QSAR models. At the same time they further demonstrated that manual curation of the structural data leads to a substantial increase in the model predictivity[15].

In their report highlighting the importance of gathering accurate information to build the WOMBAT and WOMBAT-database, Oprea et al[16] discussed the error rate in medicinal chemistry publications. They found an average of approximately two errors per publication in the almost 6,800 papers indexed in the WOMBAT database. With a median of 25 compounds per series in a publication this implied an overall error rate of 8% with errors including[17]: incorrectly drawn or written structures, unspecified position of attachment of substituents, structures with the incorrect backbone, incorrect generic names or chemical names or duplicates.

The basic steps to curate a dataset of compounds have been either considered trivial or ignored by the experts in the field. For instance, several years ago a group of experts in QSAR modelling developed what is now known as OECD QSAR modeling and validation principles[18,19] that the

researchers should follow to achieve the regulatory acceptance of QSAR models. The need to curate the primary data from which the models are derived was not mentioned. The Journal of Chemical Information and Modeling published a special editorial highlighting the requirements for QSAR papers that should be followed by authors considering publishing their results in the journal[20] and recent publications addressing common mistakes and criticising faulty practices in the QSAR modelling field[21-23] have appeared, yet none of these sources have explicitly described and discussed the importance of chemical record curation for developing robust QSAR models.

There is an obvious trend within the community of QSAR modellers to develop and follow the standardised guidelines for developing statistically robust and externally predictive QSAR models[24]. The importance of developing best practices for data preparation prior to initiating the modelling process is obvious. There is therefore a pressing need to amend the five OECD principles by adding a sixth rule that would request careful data preparation prior to model development. There is a need to develop and systematically employ standard chemical record curation protocols that should be helpful in the pre-processing of any chemical dataset and these could be automated using existing software packages (many of which are free for academic investigators). The essential procedures include the removal of inorganic compounds, counterions and mixtures (because for the most part the current chemical descriptors do not account for such molecular records), ring aromatisation, normalisation of specific chemotypes, curation of tautomeric forms and the deletion of duplicates.

Data analytical studies are impossible without trusting the original data sources. It is important, whenever possible, to verify the accuracy of the primary data before developing any model. We believe that this approach could be summarised by a famous proverb 'Trust, but verify' that was frequently used by the late president Ronald Reagan during the cold war era and that traces back to the founder of the Russian KGB Felix Dzerzhinsky who invented it almost 100 years ago (http://en.wikipedia.org/wiki/Trust_but_Verify). Our hope is that other experts will also contribute their expertise and best practices to this effort.

Improving the quality of putative hits and leads

Hits or leads in rare, orphan and neglected diseases (or for that matter many pharmaceutically relevant targets) can arise from phenotypic or mechanistic screening against commercially available screening libraries. Often the screening efforts arise in an academic setting. Because of the disconnect between academic biology and expert medicinal chemistry it is essential to carry out a medicinal chemistry annotation of putative hits or leads before expenditure of significant drug discovery effort. The early stages of the annotation process can be done using known filters and guidelines for acceptable chemistry functionality. A more detailed analysis asking questions about the chemistry of the hit or lead, and what is known biologically and chemically about substructures and similar compounds to the hit or lead currently requires a medicinal chemistry expert and takes on average about 20 minutes per compound. The in-depth data available through CAS SciFinder was used in the annotation of 64 putative tools and probes from the NIH Roadmap MLSCN effort[25]. Progress towards public sector tools for chemistry annotation might allow for a more affordable and accessible process in the future. For example, many companies have instituted filters (usually SMARTS queries) to remove undesirable molecules, false positives and frequent hitters from their HTS screening libraries or to filter vendor compounds. Early examples include REOS from Vertex[26], basic, hard and soft filters from GSK[27] and functional group compound filters from BMS[28]. These are in addition to the many proprietary filters at companies. A particular issue is chemical reactivity towards protein thiol groups. A group from Abbott reported a sensitive assay to detect reactive molecules by NMR (ALARM NMR)[29,30]. A follow up study used 8,800 compounds with data from this assay to create a Bayesian classifier model with extended connectivity fingerprints (ECFP_6) with good classification accuracy to predict reactivity[31]. This also identified 175 substructures that were likely of interest as potentially causing reactivity. Currently there is no freely accessible automated method for filtering compounds or alerting users to reactivity issues. If we were to take this further, how could we encode the knowledge of many medicinal chemists with drug discovery expertise into a piece of software or database that would identify chemical 'trash' or undesirable molecules for biologists? There is certainly some scope here to influence the quality of hits and leads that are published and annotate such molecules in public databases.

Discussion

Freely available databases and tools supporting drug discovery and chemistry in particular are

Continued from page 36

14 Zhu, H et al (2008). Combinatorial QSAR modeling of chemical toxicants tested against Tetrahymena pyriformis. J Chem Inf Model 48 (4), 766-784.
15 Young, D et al (2008). Are the chemical structures in your QSAR correct? QSAR Comb Sci 27, 1337-1345.
16 Oprea, TI et al (2007). WOMBAT and WOMBAT-PK: Bioactivity Databases for Lead and Drug Discovery, Chemical Biology: From Small Molecules to Systems Biology and Drug Design. Schreiber, SL, Kapoor TM and Wess, G (Eds), Wiley-VCH, New York, 2007, pp. 760-786.
17 Oprea, TI et al (2003). On the propagation of errors in the QSAR literature in EuroQSAR 2002 – Designing drugs and crop protectants: Processes, problems and solutions. Eds Ford, M, Livingstone, D, Dearden, J and Van de Waterbeemd H (Eds), New York, Blackwell Publishing, 2003, 314-315.
18 Dearden, JC et al (2009). How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). SAR QSAR Environ Res 20 (3-4), 241-266.
19 Group, QE (2004). The report from the expert group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the principles for the validation of (Q)SARs. OECD Series on Testing and Assessment No. 49. ENV/JM/MONO(2004)24. Organization for Economic Cooperation and Development, Paris, France. 206 pp.
20 Jorgensen, WL (2006). QSAR/QSPR and proprietary data. J Chem Inf Model 46, 937.
21 Maggiora, GM (2006). On outliers and activity cliffs – why QSAR often disappoints. J Chem Inf Model 46 (4), 1535.

Continued on page 38

Continued from page 37

22 Zvinavashe, E et al (2008). Promises and pitfalls of quantitative structure-activity relationship approaches for predicting metabolism and toxicity. Chem Res Toxicol 21 (12), 2229-2236.
23 Johnson, SR (2008). The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). J Chem Inf Model 48 (1), 25-26.
24 Tropsha, A and Golbraikh, A (2007). Predictive QSAR modeling workflow, model applicability domains, and virtual screening. Curr Pharm Des 13 (34), 3494-3504.
25 Oprea, TI et al (2009). A crowdsourcing evaluation of the NIH chemical probes. Nat Chem Biol 5 (7), 441-447.
26 Walters, WP and Murcko, MA (2002). Prediction of 'drug-likeness'. Adv Drug Del Rev 54, 255-271.
27 Hann, M et al (1999). Strategic pooling of compounds for high-throughput screening. J Chem Inf Comput Sci 39 (5), 897-902.
28 Pearce, BC et al (2006). An empirical process for the design of high-throughput screening deck filters. J Chem Inf Model 46 (3), 1060-1068.
29 Huth, JR et al (2005). ALARM NMR: a rapid and robust experimental method to detect reactive false positives in biochemical screens. J Am Chem Soc 127 (1), 217-224.
30 Huth, JR et al (2007). Toxicological evaluation of thiol-reactive compounds identified using a la assay to detect reactive molecules by nuclear magnetic resonance. Chem Res Toxicol 20 (12), 1752-1759.
31 Metz, JT et al (2007). Enhancement of chemical rules for predicting compound reactivity towards protein thiol groups. J Comput Aided Mol Des 21 (1-3), 139-144.
32 Louise-May, S et al (2009). Towards integrated web-based tools in drug discovery. Touch Briefings – Drug Discovery in Press.

Continued on page 39

becoming increasingly available. In parallel we are seeing more discussion about the need for more pre-competitive[32-35], competitive[36] and collaborative approaches[12,32] in drug discovery and the pharmaceutical industry in general, covering areas such as informatics, ADME/tox and clinical. This raises the question: "What could we achieve by just making more software and data resources available on the web?" There is currently little in the way of freely available resources for computational ADME/Tox (apart from efforts like the ToxCast project[37,38] at the EPA where several hundred compounds have been screened in more than 600 biological assays and the results have been made public, representing a resource for future models) so when will this change? Perhaps, as more data is placed in the public domain by companies that are holding on to it closely. If more computational tools and biological data were freely available it would facilitate crowd-sourced drug discovery and basically level the playing field for small (or one-person) virtual companies versus other pharma and biotech without requiring expensive tools and databases (eg CAS SciFinder). In this case, anyone with access to a computer anywhere in the world can contribute to drug discovery regardless of whether they belong to a company, research institute or not. Young gamers are already contributing to the optimised folding of proteins as evidenced by the success in the Community-Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction, or CASP. (http://www.wired.com/medtech/genetics/magazine/17-05/ff_protein). Such efforts represent truly distributed discovery and could contribute to fully integrated pharmaceutical networks. When this occurs there will be more of a need to work with highly dispersed individual researchers, store their data and possibly take molecules to the next step, eg enabling preclinical testing, animal studies etc. This will then require companies such as AssayDepot (http://www.assaydepot.com/) and CDD to help generate and store data needed for progressing molecules to clinical studies and finding larger companies or organisations to take these further. We are seeing a shift from requiring powerful computers within insular organisations to do drug discovery to using resources on the web, and so this opens up being able to use cheap portable and mobile devices to search databases and generate predictions from computational models. Of course the quality of the output will be highly dependent on the initial data quality.

Surprisingly, the investigations into how primary data quality influences the quality of published cheminformatics models have been almost absent in the published literature. It appears that cheminformaticians and molecular modellers tend to take published chemical and biological data at their face value and launch calculations without carefully examining the accuracy of data records. However, there should be much less disagreement concerning the exact chemical structure of compounds in the databases except for arguably difficult issues such as tautomers. Thus, the accuracy of the chemical structure representation could be addressed directly in most cases.

Both common sense and the recent QSAR investigations described above indicate that chemical record curation should be viewed as a separate and perhaps critical component of cheminformatics research. By comparison, the community of protein x-ray crystallographers has long recognised the importance of structural data curation; indeed the Protein Data Bank (PDB) team includes a large group of structure curators whose only job is to process and validate primary data submitted to the PDB by experimental crystallographers[39]. Furthermore, the NIH recently awarded a significant Center grant to a group of scientists from the University of Michigan (http://www.genomeweb.com/informatics/nigms-allots-5m-new-database-house-protein-ligand-data-pharma-contribute) to curate primary data on protein-ligand complexes deposited to the PDB. Conversely, the largest publicly funded cheminformatics project, ie, PubChem, is considered a data repository and no special effort is dedicated to the curation of structural information deposited to PubChem by the various contributors. Chemical data curation has been addressed whenever possible by the privately funded, but publicly available, ChemSpider project as well as by several other projects reviewed above. It is critical that scientists who exploit and build models of datasets derived from current databases or extracted from publications dedicate their own effort to the task of data curation.

Of course the hope of using cheminformatics and databases in drug discovery is to increase the efficiency and quality of molecules that progress to later stages. Just identifying reactive molecules and false positives could be of great utility to the many groups that are not aware of this problem and avoid dead ends. If we really are to empower the user and do crowd-sourced drug discovery, we will create issues with IP and the ownership of the collaborative discovery. This consideration could be one of the reasons why this approach has not been followed before. Additionally, if we are to identify gaps in the free tools to crowd-sourced

drug discovery vision then it is perhaps related to having the molecules in a database but not physically having free access to them for testing. So, the next big step will be how to make the physical molecules more widely available to all, by making them on demand or a centralised storage facility funded by the NIH etc, a topic which is outside the scope of this article but worth considering.

Conflicts of interest statement

Sean Ekins consults for Collaborative Drug Discovery Inc and is on the advisory board for AssayDepot. Antony J. Williams and Valery Tkachenko are employed by the Royal Society of Chemistry which owns ChemSpider and associated technologies. Alexander Tropsha and Chris Lipinski have no conflicts of interest. DDW

Dr Antony Williams *is Vice-President, Strategic development, for ChemSpider at the Royal Society of Chemistry. He has authored more than 100 peer reviewed papers and book chapters on NMR, predictive ADME methods, internet-based tools, crowd-sourcing and database curation. He is an active blogger and participant in the internet chemistry network.*

Valery Tkachenko *is Chief Technical Officer for ChemSpider at the Royal Society of Chemistry. He was intimately involved with the development of the PubChem platform during his time with NIH and has been involved with the development of enterprise level web-based software applications for the Life Sciences for well over a decade.*

Dr Christopher Lipinski *is a Scientific Advisor to Melior Discovery. An ACS, AAPS and SBS member, he is author of the 'rule of five', a member of the ACS 'Medicinal Chemistry Hall of Fame' and winner of multiple awards. An adjunct professor at UMass Amherst, he has 235 publications and invited presentations and 17 issued US patents.*

Professor Alexander Tropsha *is K.H. Lee Distinguished Professor and Chair of the Division of Medicinal Chemistry and Natural Products in the Eshelman School of Pharmacy, UNC-Chapel Hill. His research interests are in the areas of Computer-Assisted Drug Design, Computational Toxicology, Cheminformatics, and Structural Bioinformatics.*

Dr Sean Ekins *is a Computational Chemist and has authored more than 130 peer reviewed papers and book chapters as well as edited three books on computational applications in pharmaceutical R&D and computational toxicology. His areas of interest are* in vitro *and computational ADME/Tox, systems biology, cheminformatics and computer-aided drug discovery.*

Continued from page 38

33 Ekins, S and Williams, AJ (2009). Precompetitive Preclinical ADME/Tox Data: Set It Free on The Web to Facilitate Computational Model Building to Assist Drug Development. Lab on a Chip in Press.
34 Hunter, AJ (2008). The Innovative Medicines Initiative: a pre-competitive initiative to enhance the biomedical science base of Europe to expedite the development of new medicines for patients. Drug Discov Today 13 (9-10), 371-373.
35 Barnes, MR et al (2009). Lowering industry firewalls: pre-competitive informatics initiatives in drug discovery. Nat Rev Drug Discov 8 (9), 701-708.
36 Bingham, A and Ekins, S (2009). Competitive Collaboration in the Pharmaceutical and Biotechnology Industry. Drug Disc Today Submitted.
37 Judson, R et al (2009). The toxicity data landscape for environmental chemicals. Environ Health Perspect 117 (5), 685-695.
38 Dix, DJ et al (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95 (1), 5-12.
39 Dutta, S et al (2008). Data deposition and annotation at the worldwide protein data bank. Methods Mol Biol 426, 81-101.