Drugs of the Future 2009, 34(6): 501-508

CCC: 0377-8282/2009

HIGHLIGHTS OF THE SOCIETY FOR MEDICINES RESEARCH SYMPOSIUM: APPROACHES TO LEAD GENERATION

P. Jeffrey[1], R. Lock[2], J. Witherington[3]

[1]Inflammation and Immunity Centre of Excellence for Drug Discovery, GlaxoSmithKline, Gunnels Wood Road, Stevenage, SG1 2NY, U.K.; [2]Novartis Institutes for Biomedical Research, Wimblehurst Road, Horsham, RH12 5AB, U.K.; [3]Neurology and GI Centre of Excellence for Drug Discovery, GlaxoSmithKline, New Frontiers Science Park, Third Avenue, Harlow, CM19 5AW, U.K.

ABSTRACT

Drug discovery has become increasingly difficult in the last few decades and the cost of drug development has risen relentlessly. Given the high attrition rates the industry continues to face, there is an urgent and pressing need for quality developmental leads. Essential to tackling the observed critical productivity and attrition problems are learning from past mistakes and turning information into knowledge. This will ensure the efficient prosecution of tractable targets and optimization of compounds into quality drug candidates with the ability to make it to market. The Society for Medicines Research (SMR) meeting entitled "Approaches to Lead Generation", held on June 12, 2008, assembled a diverse program of expert speakers to discuss the exciting scientific and technological opportunities that will enable today's "drug hunters" to improve the efficiency and quality of future lead generation.

Dr. Paul Leeson (Head of Medicinal Chemistry, Respiratory and Inflammation Research, AstraZeneca R&D, Charnwood, U.K.) opened the meeting with the provocatively titled lecture "Drug-like concepts, 10 years on – what have we learned?" Highlighting the importance and impact of "Lipinski's rule of 5", a widely accepted rule of drug-like properties that has received over 1,700 citations since it was published in 1997, Dr. Leeson presented an exhaustive review on how drug-like concepts have potentially influenced the compounds that have been synthesized. Data were obtained from a variety of sources, including the FDA Orange Book, patented compounds for the top 12 pharmaceutical companies via Prous Science Integrity®, GVK Bio and Cerep Bioprint®, and then combined with calculated physical property data.

In this analysis of recently patented compounds originating from the major pharmaceutical companies, Dr. Leeson and his colleagues concluded that these more recent compounds have significantly different physicochemical properties compared to recently discovered oral drugs and those in early development (1). Molecular weight, O+N- and OH+NH counts have all increased sharply, whereas lipophilicity has remained relatively unchanged, and it has been observed that the more lipophilic compounds do not survive the development cycle and are discontinued (2). Current medicinal chemistry trends over the period 2003-2007 appear to be moving towards creating chemical libraries with both increased molecular weight and lipophilicity. Interestingly, Dr. Leeson noted that the cLogP range within a target class was greater than between target classes. Drug safety is also encoded, albeit empirically, by physicochemical properties and there is a broad trend showing that compounds with such profiles will have concomitantly increased risks with respect to their "developability profile", resulting in reduced developmental success.

The reasons for this surprising but apparent trend of physical property inflation are complex. Inevitably, there will be clear differences that exist in every company based on historical, cultural and scientific awareness and precedents, despite the wide acceptance of the rule of 5 for drug-like properties and the decline in new chemical entity (NCE) productivity. Dr. Leeson illustrated this point by reviewing the WO patent literature on chemokine receptor CCR5 antagonists designed around a common phenylpropylpiperidine template. Synthetic chemistry efforts by four major pharmaceutical companies resulted in very different outcomes: AstraZeneca registered 1,069 compounds with a mean cLogP of 3.13 and mean molecular weight of 579.9; GlaxoSmithKline registered 690 compounds with a mean cLogP of 5.28 and mean molecular weight of 598.1; Merck & Co. registered 2,457 compounds with a mean cLogP of 5.24 and mean molecular weight of 562.8; and Pfizer registered only 309 compounds with a mean cLogP of 2.76 and mean molecular weight of 488.4.

In concluding, Dr. Leeson emphasized that now more than ever medicinal chemistry must play its part in reducing pipeline risk and attrition. Physical properties are 100% controllable, there are numerous computational tools available which can assess these risks presynthesis, and reducing lipophilicity alone will improve productivity and thus increase the likelihood of success for future campaigns.

Dr. Mark Whittaker gave a lecture outlining Evotec's multiple approaches to lead generation. In agreement with Dr. Hann, the decision to adopt a particular screening strategy is strongly influ-

Table I. *Screening methods for lead generation.*

Screening method	Biochemical and cellular uHTS	Biochemical fragment screening	NMR fragment screening	Virtual screening
Chemical matter	> 250,000 compounds from commercial sources and in-house synthesis	20,000 fragments from commercial sources	20,000 fragments from commercial sources	About 6 million commercially available compounds
Technology	Proprietary EVOscreen™ systems	Proprietary EVOscreen™ systems	600 MHz NMR Licence to SAR-by-NMR™ technology	GOLD run on distributed computer grid of 300 PCs
Throughput	> 100,000 data points/day	> 100,000 data points/day	About 600 compounds/day	About 50,000 compounds/day

enced by whether there is structural information on the target and/or known ligands. Evotec's toolbox for hit finding ranges from an ability to conduct biochemical ultra-high-throughput screening (uHTS) through the use of NMR fragment screening and in silico virtual screens (Table I).

A good example of this approach was highlighted through the collaborative efforts with Oxagen in the identification of CRTH2 (GPR44) antagonists. Indomethacin was identified from the literature as a starting point as a weak CRTH2 agonist. Rapid optimization around the core scaffold led to the identification of a potent CRTH2 antagonist that displayed good bioavailability in rats. Subsequent structure–activity relationship (SAR) iterations then led to the identification of a development candidate (Fig. 1).

Hematopoietic prostaglandin D_2 synthase was highlighted as an example of a successful HTS-to-lead effort. PGD_2 synthase is involved in the conversion of PGH_2 to PGD_2, a process that has been implicated in allergic inflammation and a number of metabolic disorders (Fig. 2). The activity of PGD_2 is glutathione-dependent and this was utilized in the configuration of a fluorescence-based assay. Measuring the reaction of monochlorobimane (MCB) with glutathione (GSH) enabled the assay to be miniaturized to a 1,536-well assay format while retaining a Z′ of 0.88 and a reference hit threshold of ca. 15% (Fig. 3). In parallel to the HTS assay, x-ray crystallography was conducted to give the chemists access to a < 2 Å crystal structure of H-PGD_2S. Verified hits from the HTS campaign were then co-crystallized, providing the chemists with structural information that led to the rapid identification of a subnanomolar inhibitor from low- molecular-weight fragments (Fig. 4).

Dr. Mark Dowling (Novartis Institutes for BioMedical Research) introduced a pharmacological perspective to the meeting and delivered a lecture concerning how, why and when to determine an antagonist mechanism of action in the drug discovery setting. At what stage of the discovery process should you understand the location of drug binding to a target protein, since this governs the mechanism of inhibition and influences the maximal potential efficacy. Should it be introduced in the hit-to-lead phase? Typically, following validation of antagonist HTS campaign data, chemically attractive hits are routinely clustered into distinct structural classes. Chemically tractable classes are then promoted into the hit-to-lead phase of drug discovery, and at this stage the affinity of individual compounds is often defined by an IC_{50} or derived $K_{i/b}$. The mode of inhibition (competitive, noncompetitive, allosteric) is commonly only determined for late-stage compounds for information rather than as a means of candidate selection.

Dr. Dowling suggested two main reasons as to why you would want to promote mechanism of action studies: 1) pharmacological analysis and interpretation of data; and 2) potential for improved therapies. Regarding the second point, Dr. Dowling explained that allosteric modulators, for example, could maintain both spatial and temporal aspects of agonist signaling and provide the potential for enhanced efficacy. Whereas competitive antagonism is influenced by local agonist concentrations, an inhibitor acting via a noncompetitive mechanism will give effective blockade even in the presence of high agonist concentrations. Earlier identification of antagonists acting via a noncompetitive mechanism would be useful for certain respiratory targets – for example, targeting inhibition of neutrophil recruitment in the lung (local target) in respiratory disease states.

Dr. Dowling indicated that there can be a number of assay difficulties in establishing a definitive mechanism of action. For example, the appropriate assay format needs to be designed depending upon the type of inhibition being defined (i.e., competitive, noncompetitive, allosteric), and this information may not be available prior to initiation of such studies. In addition, potential assay artifacts can lead to the incorrect assignment of the inhibitory mechanism. The use of highly amplified/efficiently coupled systems can lead to misinterpretation of true noncompetitive inhibition. However, the use of a lower expression/less efficiently coupled system will "unveil" a noncompetitive mechanism of action. Compounds that are strongly allosteric may appear competitive at low concentrations. Testing a broad range of concentrations during mechanism of action studies will differentiate between allosteric and competitive inhibition. Finally, Dr. Dowling recommended that non-equilibrium-based assays (e.g., FLIPR) should not be used to determine mechanism of action as kinetic artifacts may lead to misclassification of the antagonist type.

The lecture closed with the question "When should mechanism of action be determined in the drug discovery process?" Ideally, in order to provide the most beneficial impact in the drug discovery process, Dr. Dowling felt it should be an integral part of lead finding, but prior to lead optimization. However, the time taken to determine mechanism of action must be balanced with the high-throughput nature required at this stage in the discovery process.

"In silico lead generation: reality and aspiration" was discussed by Dr. Paul Finn (InhibOx, Oxford). The Human Genome Project and a

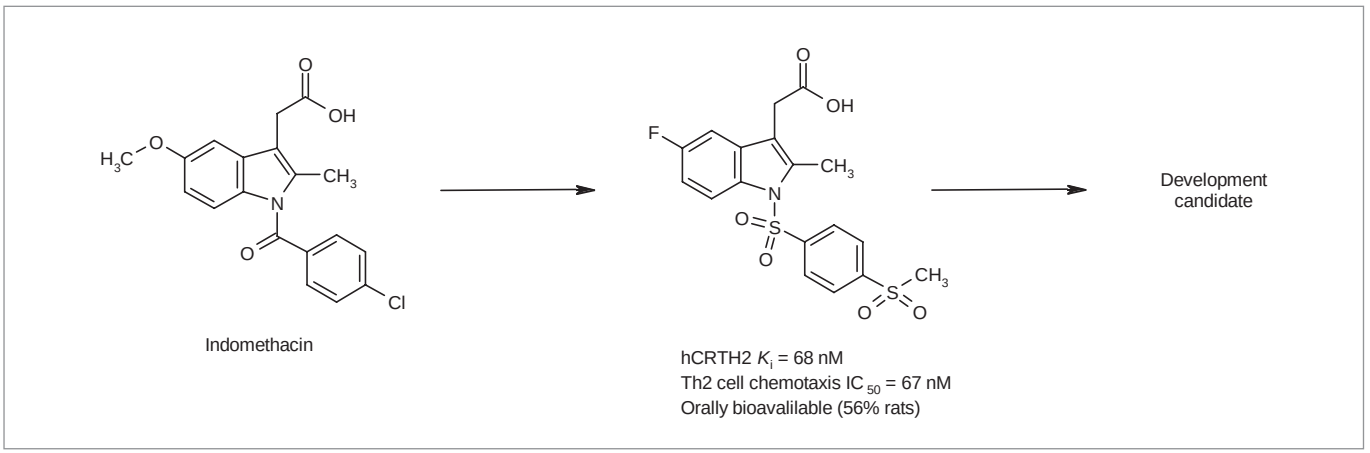


Figure 1. Optimization of indomethacin to a potent CRTH2 (GPR44) antagonist.

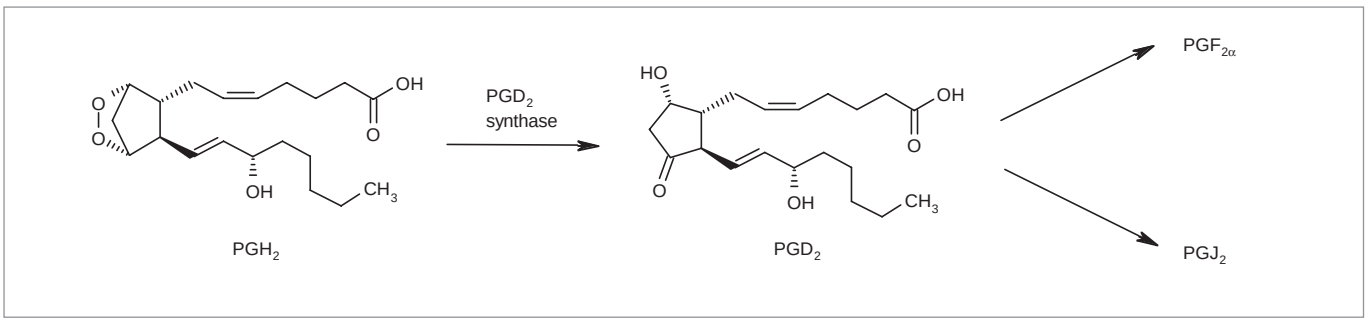


Figure 2. Conversion of PGH_2 to PGD_2 by hematopoietic prostaglandin D_2 (PGD_2) synthase.

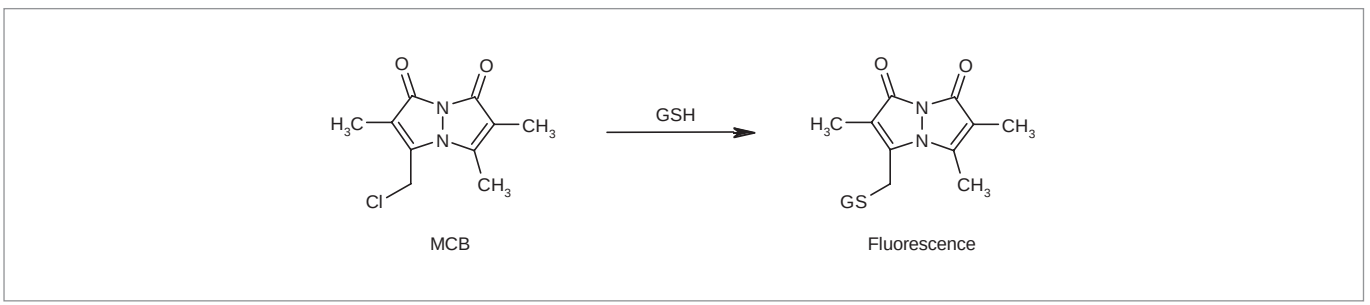


Figure 3. Utilization of monochlorobimane (MCB) to configure a 1,536-well assay for a PGD_2 synthase inhibitor high-throughput screening (HTS).

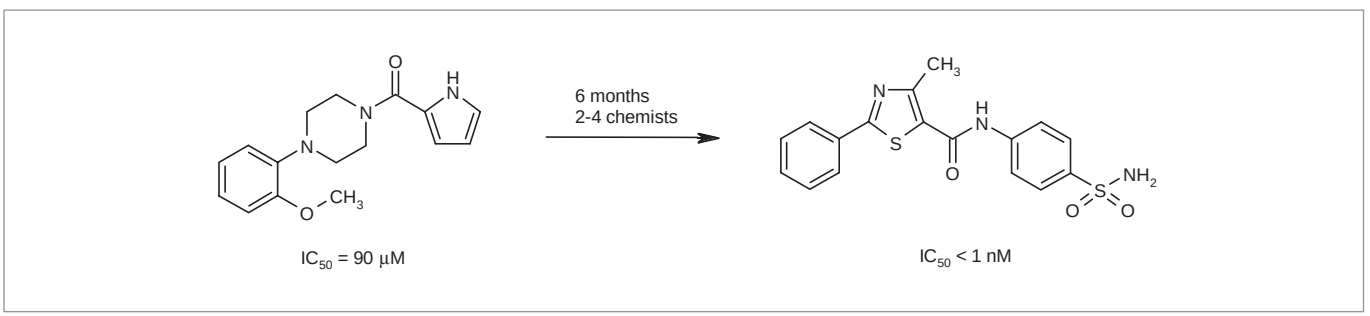


Figure 4. Utilization of x-ray crystallography to rapidly optimize a PGD_2 synthase inhibitor hit to subnanomolar lead.

greater understanding of disease processes have led to an increase in the number of potential therapeutic targets. Many of these represent new classes of targets for which chemical starting points are limited. Although there have been big investments in HTS technology and screening libraries, lead identification remains a bottleneck, especially for novel structural targets. The explosion in availability of structural data on target proteins, together with advances in computational performance, have led to the use of in silico lead generation technologies in order to try to alleviate this bottleneck. Dr. Finn described the two main categories of virtual screening: target-based (docking) and ligand-based. Within the literature there are many favorable examples of the use of virtual screening to generate leads (3-7). However, one of the issues of virtual screening is that the technique can lead to many false-positives, with a hit rate of > 90% being frequent. The complexity of the target is often one of the fundamental issues. In addition, there are few good comparative studies to identify which would be the best method to select (8). The best estimate is that average performance is similar across many methods, but with large variability (9). It is also usually difficult to predict the best method to use in advance. The high cost of commercial programs often means that only the largest companies can afford access to multiple methods. Dr. Finn pointed out that research into improved virtual screening methods is currently under way, with many groups worldwide being active in the development. However, in agreement with other experts, Dr. Finn felt that much of this effort is focused on incremental advances within current paradigms: "One can always hope that incremental improvements in current techniques will gradually lead to major advances. Such efforts are sensible, but they cannot be the only strategy; there is a call for more adventurous departures than are being published. For scoring in particular, the gap between what is required and the current methods is large." (10).

Dr. Finn detailed the steps that would be required to achieve the aspirations for virtual screening. These include: 1) better benchmarks to assist real-world choices; 2) more effort in the development of rigorous methods that better combine accuracy and speed; 3) faster methods to broaden the scope of chemistry space; and 4) wider access to methods, especially within academia. One approach to achieve improved accuracy of virtual screening is currently being examined by the DeZnIT project (Design of Zinc Metalloenzyme-Targeted Drugs Using an Integrated Technology Approach). This is a collaborative pan-European research project conducted by a consortium of seven partners involving leading experts in computer-aided design, synthetic medicinal chemistry, structural biology and the molecular biology of these enzymes. The project is coordinated by InhibOx (Paul Finn) and is aimed at improving technology for zinc metalloenzymes, therapeutically important but difficult targets for virtual screening.

Improving the speed of virtual screening could be achievable through the use of ultrafast shape recognition (USR). USR is more than 1,500 times faster than the fastest method reported in the literature. A search with USR that takes 21 h would take the next fastest method at least 3.6 years to complete (11).

In order to improve the access to virtual screening, InhibOx is collaborating with the National Foundation for Cancer Research (NFCR) to make some of its tools available via a web portal (www.nfcr.org). While Dr. Finn acknowledged that the aforementioned research projects are still in their early stages, such approaches offer hope that aspirations for in silico lead generation will be converted into reality.

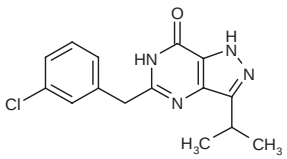
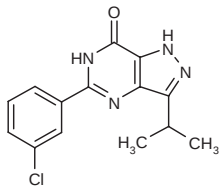
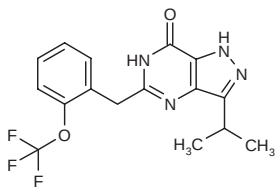
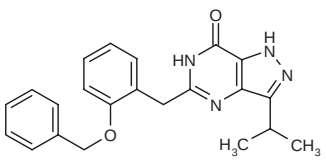
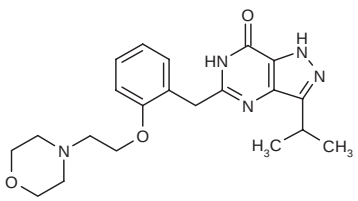
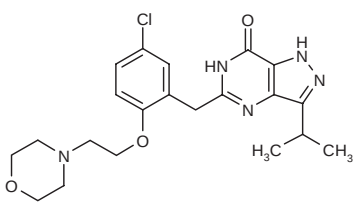
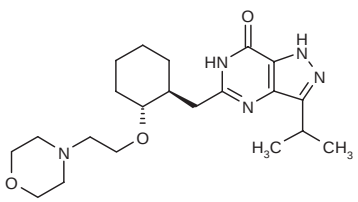
Dr. Andy Bell (Pfizer, Sandwich, U.K.) delivered an interesting lecture on enhancing HTS triage to deliver better-quality leads. Over the last 10 years the Pfizer file has expanded from 500,000 to 3.9 million compounds, largely as a result of a directed file enrichment initiative. From the outset, attrition was a key consideration for the file enrichment strategy. Knowledge that certain chemotypes are most likely to fail in development and that clinical candidates are similar to leads was incorporated into the library design in order to make and screen drug-like or lead-like compounds. This strategy was termed the "beautiful compound concept" (12).

Of the current Pfizer screening file, 68% is comprised of libraries, 27% of broadly available singletons and 5% of singletons with limited availability. While a file of this magnitude offers multiple options for many new HTS targets, there is a need to develop additional triage methods to cope with the volume of hits. By using historic screening data for compounds in its file, Pfizer has constructed activity models for all targets screened in-house. These have been used to subsequently identify false-positives and -negatives from primary screens and to prioritize compounds for IC_{50} testing. A reduction in the high proportion of false-positives/negatives and improvement in target specificity have been achieved by deriving multicategory Bayesian models. A similar approach utilized by Novartis was highlighted, in which they used in silico chemogenomics for defining false-positives in reporter gene assays and the methods for prioritizing cell-based HTS data (13). The Pfizer statistical models were derived using training sets from Cerep Bioprint, Inpharmatica StARLITe, Thomson IDDB and Pfizer in-house databases.

Dr. Bell gave an example of how the incorporation of statistical models has been used to avoid off-target pharmacology in a cell-based β-lactamase reporter gene assay (functional assay for an aminergic G-protein coupled receptor [GPCR] in antagonist format). A large number of confirmed hits (38,870) were run through their multicategory Bayesian model (predicting 515 human targets), generating 10,806 predictions. Predictions with a high Bayesian score (equating to a high confidence of any activity) were kept. It was found that the kinases dominated their high prediction hitters. These were subsequently discarded, to enable focus on compounds with predictions for no activity or aminergic activity. Pfizer frequent-hitter models were also generated in order to prioritize potential false-positives. The dominant targets that were predicted were the kinases (as observed previously), phosphodiesterases (PDEs; known mechanism) and peptide GPCRs (historic data and models contaminated with frequent hitters?).

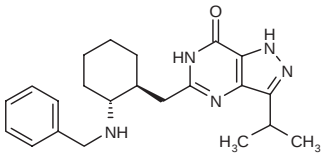
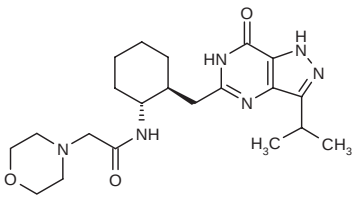
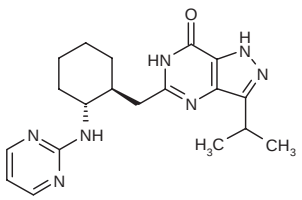
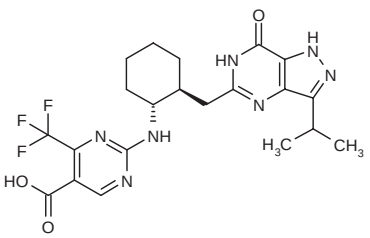
Although the majority of recent additions to the Pfizer corporate file are derived via parallel synthesis, a significant proportion of the HTS hits are still of singleton origin. Dr. Bell described how Pfizer has further used statistical methods to identify library protocols capable of delivering the closest analogues of hits, regardless of their origin. This strategy was used recently in PDE lead discovery and in the selection of a small-molecule PDE9 inhibitor for obesity and diabetes. HTS identified a hit not derived from a targeted library (see **1**, Table II).

Table II. *Selection of a potent and selective small-molecule PDE9 inhibitor.*

Compound	Structure	PDE9 IC_{50}	PDE1A IC_{50}	PDE1B IC_{50}	PDE1C IC_{50}
1		10 nM	45 nM	32 nM	5 nM
2		> 10,000 nM			
3		56 nM	2500 nM	2650 nM	520 nM
4		82 nM	1430 nM	2630 nM	1340 nM
5		40 nM	2200 nM	1650 nM	1530 nM
6		17 nM	2700 nM	953 nM	338 nM
7		87 nM	> 10,000 nM	> 10,000 nM	2050 nM

Continued

Table II (Cont.). *Selection of a potent and selective small-molecule PDE9 inhibitor.*

Compound	Structure	PDE9 IC_{50}	PDE1A IC_{50}	PDE1B IC_{50}	PDE1C IC_{50}
8		650 nM			
9		32 nM	> 10,000 nM	> 10,000 nM	1600 nM
10		36 nM	7700 nM	10,000 nM	4400 nM
11		7 nM	> 10,000 nM	> 10,000 nM	7200 nM

Two sets of libraries were prepared around this lead using the protocol described below, totaling 444 derivatives. The key finding was that *ortho*-substitution on the benzyl group at R resulted in analogues with higher levels of selectivity (Table II). Despite the improved selectivity of compounds **3** and **4**, they did not meet the required 100-fold window, and the solubility of these analogues was very poor. The addition of amine groups as in **5** was tolerated and improved the potency, selectivity and solubility. Further rapid optimization resulted in the identification of a potent and selective PDE9 inhibitor (compound **11**).

In "Marrying natural products and virtual screening to create new leads", Prof. Alan Harvey (Strathclyde Innovations in Drug Research, University of Strathclyde, Glasgow, U.K.) reminded the audience that natural products are the most consistently successful source of drug leads. While screening libraries may contain over a million compounds and are relatively simple to build, these collections are perhaps limited in their structural and chemical diversity. Natural products contain higher structural diversity and are a useful addition to libraries of synthetic compounds, since 40% of chemical skeletons in published natural product databases are not found in synthetic libraries. Natural products have the potential to operate outside of traditional synthetic space, enabling them to hit difficult targets, rescue stalled projects and create new chemical intellectual property.

Biodiversity is a key component to successful drug hunting. Past successes for natural products include atropine, ephedrine, morphine and quinine. More recent examples include ivermectin, cyclosporine and galanthamine. Since 1995, of the 244 different drug prototypes, nearly 60 originated from natural products compared with just 40 from synthetic chemistry. During the period 1981-2006, the FDA approved some 1,184 NCEs, of which over 50% were either derived from or inspired by natural products compared with only 30% from traditional synthetic campaigns. Despite the clear and recognizable structural diversity offered by natural products and their contributions to drug discovery research, the general trend has been for pharmaceutical companies to shy away from the use of natural products as part of an HTS strategy.

One of the reasons for this current trend, Prof. Harvey postulates, is the difficulty in accessing sufficient quantities of suitably diverse

chemicals, including those which are very rare. Working in association with other academic colleagues, the Drug Discovery Portal at Strathclyde University (www.ddp.strath.ac.uk) resolves these potential hurdles by linking structures and chemists with targets and biologists, thereby initiating and facilitating advanced in silico screening and hit identification. This approach provides easy access to a library containing a high range of chemical diversity, including natural products, derivatives and synthetic intermediates, that is constantly refined and updated. Highlighting the use of such an approach, Prof. Harvey presented a case history around the virtual screening of novel adenosine receptor A_{2A} antagonists. Using caffeine as a starting point, a pharmacophore model was created and 67 compounds were sourced from the database. Further profiling of these early hits against the A_{2A} receptor at 100 µM to determine the K_i for A_{2A} and A_1 receptors yielded a 10% hit rate of novel chemical skeletons with K_i values of < 1 µM at A_{2A} and > 100-fold selectivity versus A_1. Natural products and lead finding therefore have real potential for worldwide access to a unique collection with drug-like diversity and novel chemical intellectual property.

Dr. Mike Hann (Director of Structural Biology, GlaxoSmithKline, U.K.) gave a highly interesting and thought-provoking lecture on the challenges facing the pharmaceutical industry on the potential numbers of "druggable molecules" that can be made and how this impacts "hit" identification. In reality, chemists struggle with balancing numerous properties, even within the druggable space. The need to focus on diversity is inversely proportional to the knowledge that is available on each target. The extremes of this are the use of structure-based drug design, where a protein structure is available, and its converse, the need to prepare diverse libraries where zero knowledge of the target is available. Ideally, drug discovery should attempt to make the journey as short and efficient as possible by choosing the optimal starting point carefully, maximizing the rate of analogue synthesis (use of automation and tractable chemistry) and ensuring that the size of the step in desired properties is as large as possible by utilizing in silico predictive models which have been correlated to measured biophysical data.

Dr. Hann presented the advantages of exploring multiple approaches to lead identification, in particular the benefits of fragment-based drug discovery, which usually involves the detection (usually at high concentration) and elucidation of the binding mode (usually by x-ray crystallography or NMR) of small, low-molecular-weight compounds (i.e., "fragments", "scaffolds", "templates", "privileged cores", "monomers", "building blocks", etc). In practice, this can involve screening of customized libraries of hundreds to thousands of "fragments" that might be expected to ultimately be a key part of a more fully developed and optimized drug-like molecule.

Dr. Hann then introduced the audience to his five fortes of fragments. Firstly, the combinatorial explosion of chemistry space means that fragments can sample more of the available chemistry space at that level of complexity than is possible with more complex molecules. Secondly, with lower complexity there is a higher probability of compounds matching the receptor even though they may be harder to detect. More complex molecules are more likely to have more "clashes" and thus do not fit. Thirdly, medicinal chemists like to build molecules and fragments are thus a great starting point for structure-based design. Analysis of Sneaders' book on "Drug Prototypes and Their Exploitation" indicates that from the 470 drug case histories, there are the following changes in property values during the lead to drug process:

Av # arom	Δ arom	%	Av cLog P	Δ cLog P	%	Av CMR	Δ CMR	%
1.3	0.2**	15	1.9	0.5**	26	7.6	1.0**	14.5

Av # HBA	Δ HBA	%	Av HBD	Δ HBD	%	Av heavy	Δ Heavy	%
2.2	0.3**	14	0.85	−0.05+	(4)	19.0	3.0**	16

Av # MW	Δ MW	%	Av MV	Δ MV	%	Av Rot B	Δ Rot B	%
272	42.0**	15	289	38.0**	13	3.5	0.9**	23

Interestingly, analysis of the molecular weight of the libraries synthesized by GlaxoSmithKline indicated that these were far greater than Sneaders' leads, probably as a result of designing large libraries with many points of diversity, which ultimately leads to an increase in complexity and potentially a decrease in the probability of any given molecule being a hit. Dr. Hann suggested that a more efficient approach is to stop making large libraries and instead focus on identifying lower molecular weight, less complex molecules that, while they may also be less potent, can be optimized into the druggable area by adding both affinity and the required pharmacokinetic properties. A fourth guide for drug discovery is to focus on ligand efficiency (14). While there are many ways to define this approach, it generally involves identifying and optimizing the potency of leads normalized for a number of calculated properties (molecular weight, polar surface area and lipophilicity).

Finally, Dr. Hann discussed the idea that by focusing on fragments due to their required properties of high solubility and ligand efficiency, the final compound tends to retain many of these properties. Furthermore, although these fragments might have low specificity against a given target, this gain can be built in as required during the optimization process. In contrast, many hits from HTS collections that have already been optimized for other programs may have significant molecular weight and complexity added during subsequent optimization for the new target. A further point to this approach is that if polypharmacology is required it is again more desirable to start with a small ligand-efficient fragment as a starting point.

Chemogenomics comprises a systems-based approach that establishes the relationship between targets and ligands that are used as target modulators in living systems, from individual cells to whole organisms. This approach requires the integration of the basic disciplines of chemistry, biology, genetics, informatics, structural biology and chemical screening into a collective study of complex chemical and cellular responses. Dr. John Overington (Biofocus DPI, London, U.K.) espoused an in-depth understanding of such interactions between small molecules and specific proteins, as this can enhance the development of new biological tools and the identification of new drug targets. Such knowledge can be obtained using chemogenomic screening, coupling biological and chemistry spaces at the

genomic level, thus leading to a gene family-led approach in generating novel and important therapies.

The use of chemogenomic screening is on the increase but generates large amounts of data requiring novel and sophisticated analytical techniques. In utilizing this approach, Dr. Overington and his colleagues at Biofocus DPI have developed a suite of databases that aim to select the best biological targets for novel drugs based on the most appropriate drug-like chemistry starting points. This is in line with the strategy originally advocated by Prof. James Black in his famous maxim "the most fruitful basis for the discovery of a new drug is to start with an old drug". Drugstore™ is a database of known drugs, StARLITe™ contains data on known compounds and their effects, Strudle™ contains binding site and "druggability" information, and Kinase SARfari™ and GPCR SARfari™ are informatics systems for the most widely used target classes in drug discovery. (On July 23, 2008, Biofocus DPI announced the transfer of these databases for predictive drug discovery to the European Molecular Biology Laboratory's European Bioinformatics Institute [EMBL-EBI], which will provide access to the data under a grant from the Wellcome Trust.)

One of the problems that Dr. Overington encountered in the development of these tools was that the literature is still relatively polluted, containing numerous errors in data and even compound structures. By correctly abstracting the pure and correct data from the literature, maintaining and constantly updating these high-quality databases, the systems for chemogenomic data mining can be optimized and aligned with other databases and informatics systems. This "machine learning approach" represents an exciting new paradigm for drug discovery and new lead generation.

DISCLOSURE

Phil Jeffrey (GlaxoSmithKline, U.K.), Ruth Lock (Novartis, U.K.) and Jason Witherington (GlaxoSmithKline, U.K.) are members of the Society for Medicines Research Committee (SMR; http://www.smr.org.uk/), which organizes conferences on behalf of the SMR. The SMR would like to thank Evotec (http://www.evotec.com/en/) and Peakdale Molecular (http://www.peakdale.co.uk/) for their generous sponsorship of this meeting.

REFERENCES

1. Leeson, P.D., Springthorpe, B. *The influence of drug-like concepts on decision-making in medicinal chemistry*. Nat Rev Drug Discov 2007, 6(11): 881-90.
2. Wenlock, M.C., Austin, R.P., Barton, P., Davis, A.M., Leeson, P.D. *A comparison of physiochemical property profiles of development and marketed oral drugs.* J Med Chem 2003, 46(7): 1250-6.
3. Park, H., Jeon, Y.H. *Toward the virtual screening of Cdc25A phosphatase inhibitors with the homology modeled protein structure*. J Mol Model 2008, 14(9): 833-41.
4. Gundla, R., Kazemi, R., Sanam, R., Muttineni, R., Sarma, J.A., Dayam, R., Neamati, N. *Discovery of novel small-molecule inhibitors of human epidermal growth factor receptor-2: Combined ligand and target-based approach*. J Med Chem 2008, 51(12): 3367-77.
5. Gozalbes, R., Simon, L., Froloff, N., Sartori, E., Monteils, C., Baudelle, R. *Development and experimental validation of a docking strategy for the generation of kinase-targeted libraries*. J Med Chem 2008, 51(11): 3124-32.
6. Fells, J.I., Tsukahara, R., Fujiwara, Y. et al. *Identification of non-lipid LPA3 antagonists by virtual screening.* Bioorg Med Chem 2008, 16(11): 6207-17.
7. Kiss, R., Kiss, B., Könczöl, A. et al. *Discovery of novel human histamine H4 receptor ligands by large-scale structure-based virtual screening*. J Med Chem 2008, 51(11): 3145-53.
8. Nicholls, A. *What do we know and when do we know it?* J Comput Aided Mol Des 2008, 22(3-4): 239-55.
9. Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., Corbeil, C.R. *Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go.* Br J Pharmacol 2008, 153(Suppl. 1): S7-26.
10. Leach, A.R., Shoichet, B.K., Peishoff, C.E. *Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps.* J Med Chem 2006, 49(20): 5851-5.
11. Ballester, P.J., Finn, P.W., Hermsmeier, M.A., Richards, W.G. *Ultrafast shape recognition: Evaluating a new ligand based virtual screening technology*. J Chem Inf Model, Submitted for publication.
12. Lipinski, C.A. *Chris Lipinski discusses life and chemistry after the Rule of Five.* Drug Discov Today 2003, 8(1): 12-6.
13. Crisman, T.J., Parker, C.N., Jenkins, J.L. et al. *Understanding false positives in reporter gene assays: In silico chemogenomics approaches to prioritize cell-based HTS data*. J Chem Inf Model 2007, 47(4): 1319-27.
14. Abad-Zapatero, C., Metz, J.T. *Ligand efficiency indices as guideposts for drug discovery*. Drug Discov Today 2005, 10(7): 464-8.