

Diversity screening versus focussed screening in drug discovery

Martin J. Valler and Darren Green

Which strategy is best for hit identification? Making the right choices in the capital-intensive world of modern drug discovery can make the difference between success and expensive failure. Keeping an open mind to all the options is essential. Two well-established strategies are diversity-based and focussed screening. This review will provide contrasting viewpoints highlighting the strengths and deficiencies of each approach, as well as some insights into why both strategies are likely to have a place in the research armoury of a successful drug company.

About five years ago, expectations of the hit identification process (Box 1) were very different from today's reality. Hundreds of new targets for drug discovery would be generated through genomics and modern molecular biology, with strong biological rationale and a wide variety of molecular mechanisms. Technology, in this context meaning the automation and miniaturization of traditional experimental processes such as synthetic chemistry and *in vitro* biochemical assay formats, would enable the production and screening of unprecedented numbers of molecules against these targets. So many potential lead molecules would be found that the bottleneck would be in deciding which lead series to optimize, and then optimizing several of these series in parallel to reduce the risk of failure in the development phase.

The 'rational' approach to drug discovery, taken to the extreme by *de novo* structure-based design but really an all-embracing term for quantitative medicinal chemistry, would be unable to compete with the speed and productivity of the new technology. In this environment, pharmaceutical companies invested heavily in HTS systems, automated compound inventories, and combinatorial chemistry technology. To be successful, it was large numbers of compounds that would be necessary.

Focussed screening

What constitutes diversity?

For chemists trained in the art of more rational approaches, the paradigm shift already described was a little disconcerting. There were, and still are, some uncomfortable facts that undermine the concept of mass random screening. For example, there is the vastness of chemical space. Estimates of the number of possible drug molecules vary widely, but average $\approx 10^{40}$. By comparison, the number of seconds since the Big Bang is only 10^{17} . Therefore, even if a company obtains access to 100 million molecules (far more than the 0.5–1.0 million most companies have achieved), the sampling of this massive space is still hopelessly inadequate. In short, aiming for an increase in screening capacity by one or two orders of magnitude does not even begin to tackle the problem. The latest high-throughput technologies, uHTS (ultra-HTS) and bead-based libraries are being developed. However, the mathematics suggests that these systems will be no more effective at solving the problem than their predecessors.

Computational methods

Focussed screening is now well established as a successful hit generation strategy. Although much attention has been

***Martin J. Valler**, Molecular Discovery Department and **Darren Green**, Computational Chemistry and Informatics Unit, GlaxoWellcome R&D, Medicines Research Centre, Gunnels Wood Road, Stevenage, UK SG1 2NY. tel: +44 1438 763819, fax: +44 1438 768206, e-mail: mjv1187@GlaxoWellcome.co.uk

Box 1. Interpretations of commonly used terms in drug discovery

The precise definition of the following terms varies widely between drug discovery companies. The meanings given here are aligned with the use of the terms within the lead discovery function at GlaxoWellcome.

Hit: A molecule with robust dose–response activity in a primary screen and known, confirmed structure. The output of most screening.

Progressible hit: A representative of a compound series with activity via an acceptable mechanism of action and some limited structure–activity relationship information.

Lead: A representative of a compound series with sufficient potential (as measured by potency, selectivity, pharmacokinetics, physicochemical properties, absence of toxicity and novelty) to progress to a full drug development programme.

Assay: Generically a bioassay where biological activity is derived; associated with a bioactive effector molecule. Within the screening discipline, an assay will probably be robust enough and have the capacity to enable testing of up to 10,000 samples, generally with limited chemical diversity.

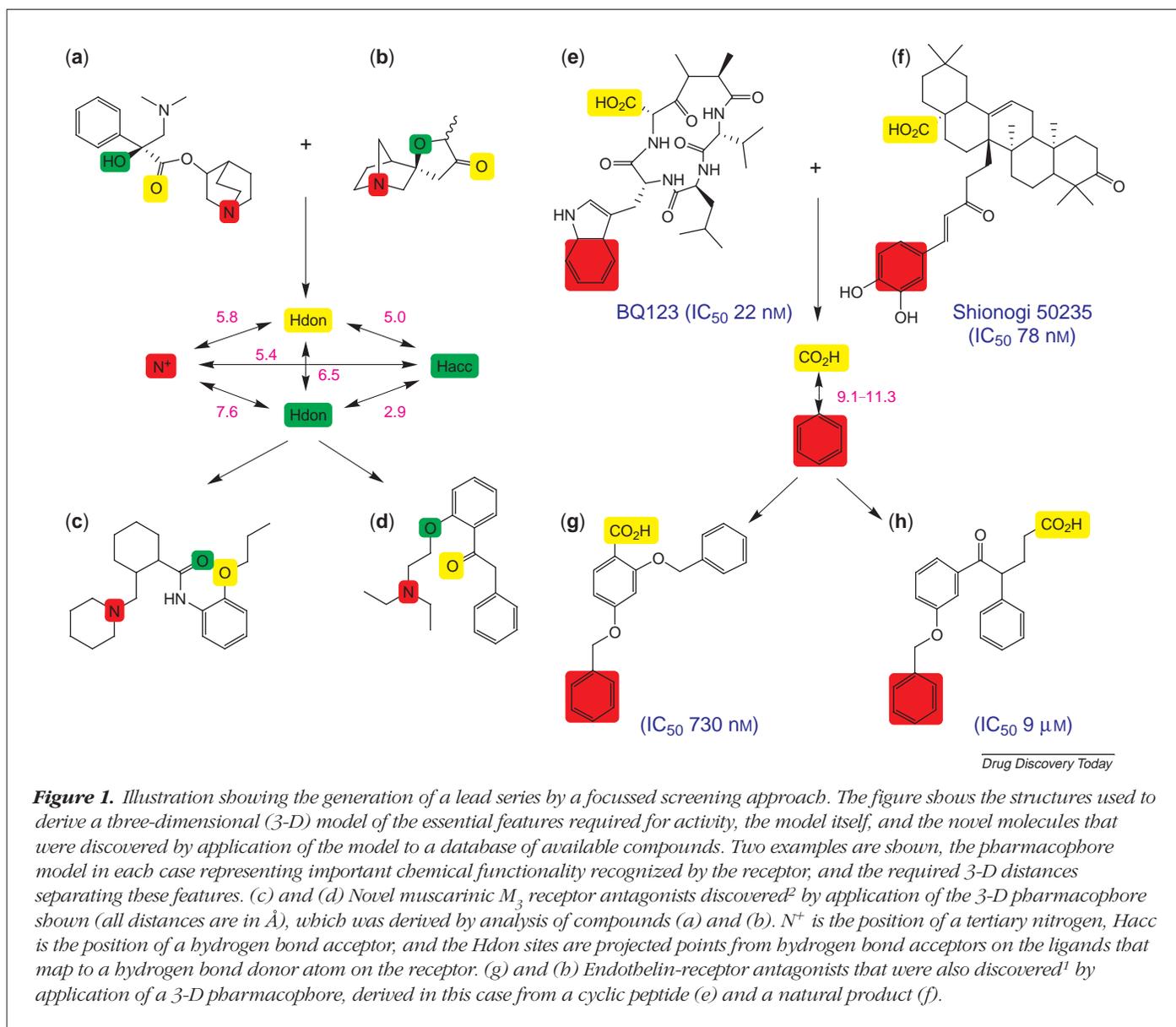
Screen: An optimized, streamlined assay format with characterized robustness to diverse chemical types and conditions such that testing of $\geq 10,000$ samples is both feasible and cost effective. The spectrum of low-throughput screening (10,000–50,000 assay points), medium-throughput screening (50,000–100,000 data points) and high-throughput screening (100,000–500,000+ data points) can be defined. The scale of implementation of a given screen is greatly influenced by format, application of technology (e.g. automation), time and resource constraints.

paid to the hardware of synthesis and screening, computational methodology has been steadily improving, resulting in several proven hit identification methodologies^{1,2} (Fig. 1). The most popular and successful of these methodologies involves the use of three-dimensional (3-D) information. For targets where a crystal structure is available, there are several validated docking algorithms that enable the selection of compounds from databases of available chemicals and virtual libraries³. If there are known ligands for the target, these can be used to construct a 3-D pharmacophore, which can be used to search a 3-D database system^{1,2}. These methods can provide hit rates of one

to more than ten percent², and can routinely show a 10–100-fold improvement over random screening (M.J. Valler, unpublished data). Indeed, even with access to a protein structure, it is sometimes useful to construct a 3-D pharmacophore from the active site and use it to search in preference to a docking method if the pharmacophore is simple or there are issues such as a flexible protein active site.

Several groups are experimenting with the characterization of general compound classes, in support of systems-based biology on kinases, seven-transmembrane domain (7TM) receptors and proteases. Examples of the utility of these techniques have been published, as applied to antimicrobial⁴ and CNS-penetrative⁵ compounds. These methods often use 1-D (MW or counts of aromatic rings, donors, acceptors or rotatable bonds) or 2-D (the presence or absence of certain chemical fragments) descriptors. Other methods that typically use 2-D descriptors include similarity measures⁶ (e.g. the widely used Tanimoto index) or the recursive partitioning approaches developed for use in sequential screening⁷.

Information concerning a target can come from the crystal structure of a protein with a ligand bound, from known inhibitors, homology models and sequence similarities. All of this information can be used to select compounds that have a high probability of interacting with the target. The key term here is probability: by contrast to *de novo* structure-based design, these techniques do not set out to find a ‘silver bullet’. Rather, the typical result is a selection of several hundreds or thousands of compounds that should be screened. The methods are not accurate enough to screen a handful of compounds with a high probability of success, but the deficiencies in the methods can be compensated by selecting and screening all the molecules within a set level of probability of being active. Thus, the screening technology must have a minimum throughput, and the compound handling systems must have a ‘cherry picking’ capability to select individual samples from plates. All this is ‘free’ with even a minimal HTS capability. The bottleneck for focussed, *in silico* methods is getting the correct level of information and, therefore, having a strategy to acquire it. This strategy could be as drastic as only working on targets for which the crystal structure is known [examples of this strategy are used by Agouron Pharmaceuticals (La Jolla, CA, USA) and Vertex Pharmaceuticals (Cambridge, MA, USA)]. More conventionally, a company often has an effective biophysics (crystallography and protein NMR) function and, more importantly, possesses strong computational chemistry and cheminformatics capabilities.



Drug Discovery Today

There are other ways of improving the hit generation success rate that are more closely associated with target selection. There is a much higher probability of finding hits for targets from those families that have successfully produced hits previously (7TM receptors, ion channels, kinases and proteases), often termed tractable targets. Therefore, if only tractable targets are screened, the chance of finding hits by random screening increases. These targets often produce successful outcomes because usually much is known about them because of previous work on related targets, not only on their structural characteristics, but more importantly on the type of molecules they usually recognize. Hence, screening only tractable targets is essentially performing a focussed screen, only rather inefficiently.

The advantages of focussed screening

With focussed screening, it should also be possible to use an assay that is more appropriate, rather than one that works well at a large scale. For example, it has been common for *in vitro* isolated receptor-binding assay formats to be developed for high-throughput screens, even when greater value could be derived from a whole-cell functional approach. However, one concern is often how quickly a screen can be brought online.

The logistics of developing a high-throughput screen, delivering hundreds of thousands of compounds to the screen, screening the plates and producing possibly thousands of follow-up samples, requires an extensive infrastructure and involved planning, as well as time. By

contrast, a focussed screen can progress a smaller number of compounds, pre-filtered by the medicinal chemists, so that the screen only selects compounds that would make a good starting point, through an assay that might not be robust enough to be a screen (Box 1). On the basis of these results, or competitor publications, further compounds can be screened, made or purchased. This can be significantly quicker overall than a comparable high-throughput screen. In the current environment of aggressive patenting strategies, particularly around families of tractable targets like kinases, the speed factor is potentially a major advantage of focussed screening.

In theory, focussed screening should also have a significant advantage in cost terms because of sample, reagent and handling costs. However, there are still high costs associated with the focussed strategy. Firstly, a flexible compound storage and handling system is required – it is more expensive to continually select bespoke sets of samples from a store than to simply resupply the same sets of samples from the same sets of plates. Secondly, the cheminformatics software required is expensive to buy. Thirdly, it is often necessary to write your own software to complement commercially available software and to gain a competitive advantage – skilled people to do this are in short supply. Finally, there is a risk: focussed screening often involves making decisions concerning what data to use and how to use it. No selection method can be perfect, but at worst the focussed strategy is still as good as the random approach, sample for sample. The only way to minimize this risk is to invest in new or evolved algorithms, to use human intuition where it is most useful and in combination with software (not artificial intelligence, but computer-aided decision making).

The scope of sample selection for a focussed screen can be cast very widely, because the same methods that can search compounds available in a local, physical store can search databases of compounds available for purchase from the wider global supply network. This is a powerful way of targeting expenditure for sample acquisition but one which ties the hit identification process into the more extended timelines of the supplier network. The ability to search and select from virtual libraries of molecules that could potentially be made is an effective route to targeted synthesis. The generation of new progressible hit series by searching databases of virtual compounds is common practice. In fact, many focussed screening techniques were designed to do just that. The benefits of the focussed approach are therefore in steering the hit identification process towards a knowledge-driven environment where ever-expanding throughput is exchanged for target insight

and selective intuition built on enhanced computational methods.

The case for diversity screening

The drivers behind the current ethos of large-scale diversity HTS are rooted in the desire to build an improved hit identification process, and are based on the simple model of testing everything. The key activity over the past five or so years has been scaling: taking the existing model and increasing capacity by application of technology. As a strategy, the aim is to be able to synthesize, access and physically test (as opposed to *in silico* modeling) all the molecules that could constitute drug candidates. In many ways, this has been highly successful: the impact of technology has driven average sample numbers per screen up from a few tens of thousands in the early 1990s to the several hundred thousand per screen that is currently routine, and with relatively stable costs over this period.

In most pharmaceutical companies, diversity HTS is seen as an operational definition governed by the maximum extent of the accessible sample collection and the screen resources that can be assembled. Because of this, the actual number of compounds that constitutes HTS can range from tens of thousands up to hundreds of thousands (Box 1). The vision of uHTS currently extends this range up to one million samples and more, but in no case is this seen as an absolute limit, either technologically or procedurally. Hence the question arises – how many samples is enough?

Chemical space: the final frontier

The heartland of this debate centres on the definition, and hence extent, of chemical space. More precisely, it focusses on the extent of chemical space that is accessible by chemical synthesis and which could be described as drug-like. Although, as already described, the number of possible drug molecules is immense, what is much more difficult to define is the proportion of this huge expanse of molecules that must be sampled to ensure success in a lead generation programme. It is important to remember that the aim of screening is not to discover the lead molecule itself but to find progressible hits (see Box 1 for definitions) such that an iterative targetted synthetic optimization process (hit progression, in the terminology of GlaxoWellcome) could usefully be initiated.

As appropriate, the hit progression stage encompasses many more properties of a molecule than simply potency against the primary target. Aspects such as selectivity against related targets, physical properties and metabolic/toxicological testing are all highly relevant at this stage

and this breadth of study, coupled with the inherently interactive nature of bioassay guided synthesis, means that 'screening' as defined here is not appropriate.

Therefore, if the goal of a high-throughput screen is to derive a sufficient number of varied structural types to initiate and sustain a hit progression programme, how many (diverse) samples must be tested? The evidence within GlaxoWellcome is that a sample set of up to 500,000 compounds can be successful in generating hit molecules in many cases, despite the poor representation of chemical space that it might appear to have from theoretical considerations. Active selection is more often applied, in practice, as part of the hit progression criteria, often because the capacity of the downstream progression systems becomes the bottleneck. As discussed below, the ongoing improvement of this situation is being driven by importing more of the useful techniques and automation practices learnt from large-scale HTS into these processes, for example into toxicity and selectivity testing.

Success versus which targets?

Although computational methodologies are improving rapidly, any sensible estimation of the potential success criteria for sample selection has to take into account the diversity of drug targets themselves. The range of target types that lend themselves to structural and/or computational analysis has several important limitations. Macromolecular interactions, intracellular signal transduction pathways or cascade targets are largely off limits to predictive computational methods. It is also true, however, that for these types of targets, it is notoriously difficult to find low-MW effectors using any strategy. It is easy to be blinded to the realistic prospects of success for these so-called low-tractability target classes by the desire to find a molecule.

With the current and predicted growth in diverse target sequences derived from the Human Genome Project, the ability to define an analytical, predictive relationship between any novel drug target and its potential effectors is likely to remain a challenging objective. This is the logic that has driven pharmaceutical companies towards studying families of related targets within system-based research programmes, to generate and use comparative data to improve predictive capabilities. The jury is still out as to whether this type of strategy will predominate in the field of drug discovery. However, the most likely result is that system-based and single target-based aspects will continue to coexist for several years.

As an illustration, the contrast between the knowledge-rich environment of 7TM receptors and the disease area-

driven requirements of novel antibacterial research demonstrates the conflicting demands faced by the hit identification discipline. In the former case, much information has been produced over many years and can fuel targeted sample selection and screening highly effectively. In the latter case, the targets themselves are ill-defined and success using a knowledge-based strategy is difficult. Even so, the desire to find effective drugs in vital therapeutic areas such as these does not dissipate. If anything, the difficulty in achieving the goal makes success all the more potentially valuable in the marketplace. This is not to say that drugs derived from system-based research will be inherently of lower value, although competition is inevitably likely to be more intense, but novel, unclassified targets associated with important disease areas will always exist as a challenge to the pharmaceutical industry.

The technology of diversity HTS

The diversity approach to screening has, at its heart, a strong commitment to technology, especially automation, to drive the ongoing increases in capacity and throughput. The infrastructure of sample handling, bioassay and data manipulation defines the way that large numbers of data points can be handled.

Samples strategy

One of the realities that faces any screening effort, both focussed or diversity-based, is the access to samples for testing. Two key components of this are generation and presentation: essentially where would the molecules come from and how would they be applied to a biological target?

Any type of screening will use samples from either an internal or an external source. The in-house collections represent the fruit of many person-years of expensive chemistry time. The diversity represented by these collections is ultimately dependent on the projects that have been resourced within the company over its history and can show strong bias, particularly where there have been specific therapeutic franchises. The internal sources have the key advantage for speed purposes that they are immediately accessible.

Sample collections will increase by organic growth but this is a relatively slow process. A more rapid and more direct option is the acquisition of samples in bulk numbers from specialist companies. This route provides the potential for tens or hundreds of thousands of compounds but with the issues of cost, speed of access and novelty (as all companies have access to these suppliers). Sample acquisition is, however, an effective way of filling gaps in the chemical diversity of an in-house collection.

An alternative strategy focussed on HTS is that pioneered by Affymax Research (Palo Alto, CA, USA) and GlaxoWellcome, whereby the scale of chemical synthesis is massively upgraded by the power of combinatorial chemistry. Split-pool solid-phase bead libraries enable tens of thousands of structures to be created simply and cheaply. Because the synthesis, formatting and screening approaches are so intimately joined in this approach, it can be seen as spanning both strategies, capable of both diversity and focussed implementation.

Sample presentation

The mode of presentation of samples fundamentally defines the optimal mode of screening. The focus on plate-based or tube-based storage is the key factor for the way sample handling facilities are constructed. In both cases, a large sample storage capacity is required, but for focussed work, this must enable access to a wide range of individual samples as required while for diversity, it must enable delivery of the bulk of the store contents easily to each target. A sample handling system set up on a plate basis (plate store concept) will always be fundamentally inefficient in supplying individual samples compared with one designed specifically to do this (tube store concept) and vice versa.

Building a better screen

Independent of screen strategy, the emphasis on time-to-market and the progression of a target through the drug development process becomes ever more crucial. For diversity screening, this has led to the rise of automation solutions and the trend to miniaturized formats. Importantly, the solutions to the pressing concerns of speed and cost can already be found in systems, formats and technologies that are already known. Plate formats such as the NanoWell™ (Aurora Biosciences Corporation, San Diego, CA, USA), miniaturizable homogeneous readouts for radioactive (LEADseeker; Amersham Pharmacia Biotech, Amersham, UK) and non-radioactive signals and cell-based reporter systems⁸ based on luciferase (e.g. Lucscreen™; PE Biosystems, Foster City, CA, USA), β -galactosidase (e.g. Gal-Screen™; PE Biosystems) and miniaturized fluorescence resonance energy transfer⁹ (FRET; Aurora Biosciences Corporation) are now at the stage where a screening environment in which an economically viable, one million sample screen is normal, can be confidently predicted.

As well as the acknowledged advantages of automated screening as a means to enable scales of screening that would not be viable with conventional technology and

resources, there is a series of valuable extra benefits that come as part of the package, including improvements in reproducibility, precision and quality of data. The development of screens that work well on automated systems is a more extensive process than that for lower scale manual assay development, but the extra time required is more than repaid by speed and efficiency savings throughout a large screening campaign. The 'account' typically moves out of the 'red' and into the 'black' when a scale of approximately 50,000 samples is reached.

In the design phase, diversity HTS must be able to cope with a range of chemical types. It must also be able to achieve higher than normal levels of accuracy and reproducibility because of the tendency for single sample, single point data within an HTS. Features such as long-term stability of signal, reproducibility and accuracy are all key, especially for screens on automated systems. An HTS-friendly screening format will also be substantially easier to carry out, in terms of manipulations required.

However, building a screen to these high standards has the significant benefit that data quality will be highly validated and, when applied to other lower throughput phases, the assay format is extremely robust (i.e. can cope with variability in sample types, operator performance and conditions). This generates value to the rest of the drug discovery pipeline in terms of time spent on repeating work and the capability to support subsequent hit progression and lead optimization in a more comprehensive and efficient manner.

Higher standards in the requirements placed on bioassay methodology and readouts is also beneficial to the efficiency of the rest of the research organization. Many of the current homogeneous formats widely used within the pharmaceutical industry, such as Scintillation Proximity Assay (SPA; Amersham Pharmacia Biotech) and Homogenous Time Resolved Fluorescence (HTRF; Packard Instrument Company, Meriden, CT, USA), made their initial entry into the industry by the HTS route and are now widely used for lower scale applications such as lead optimization and bioassay in support of *de novo* chemical synthesis. Indeed, the added value role of the HTS environment as a technology incubator and catalyst should not be underestimated: the requirement for speed and efficiency and reliability has driven the introduction of many of the key bioassay methodologies now taken for granted within other parts of the industry. Examples such as the concept of microtitre plate-based bioassay and reading devices, homogeneous screen formats and whole-cell reporter screen technologies⁸ have all been fuelled by the requirements of HTS. These developments are by no

means complete, as the maximum extent of HTS is a moving target. Current cutting-edge technologies, such as whole-plate imaging camera systems [e.g. the Viewlux™ (Perkin Elmer Life Sciences, Akron, OH, USA) and the LEADseeker miniaturized homogeneous radioactive format] are likely to find their way into the mainstream of the pharmaceutical industry as their potential applications and efficiencies become apparent.

Data handling for HTS

Diversity HTS generates a substantial data set that requires investment in high-volume database and data handling systems, for example OMMM™ (Oxford Molecular, Oxford, UK) and ActivityBase™ (ID Business Solutions, Guildford, UK). As well as the sheer volume of data, diversity HTS generates a high-content data set that is rich in potential for defining structure–activity relationships between the spectrum of chemical types that are represented. Hits are often simply defined by showing activity above a given predetermined threshold. The much greater value that can be derived from the whole data set by data mining techniques such as recursive partitioning⁷ and data visualization methodologies such as Spotfire (Spotfire, Cambridge, MA, USA) is only now being widely appreciated.

Why do diversity HTS?

A good implementation of diversity screening requires extensive investment in samples, screening and sample manipulation hardware, technology, simplified assay formats and extended assay-to-screen development to get the best from the strategy. A focussed strategy should be achievable much more simply with an assay, rather than a screen (Box 1), good computational expertise and access to samples. However, despite this, diversity HTS forms the mainstream of hit generation in many pharmaceutical companies. Why should this be?

Part of the answer could lie in the search for the novelty that this strategy offers. The use of computational and analytical tools is effective in generating biological activity in candidate molecules but the approaches rely on knowledge of a target. For many important classes of target, this knowledge is not necessarily available. Furthermore, in those instances where target knowledge is available, it is often non-proprietary and open to all investigators in the field. The promise of a diversity campaign is of a molecule with excellent IP (intellectual property), but at what cost?

In the real world, drugs get through the tortuous route to market with a wealth of different life-stories. Most owe their existence to good judgement or good timing

throughout their progression, but novel IP can provide a clear market advantage.

Partnership: the key to success

With current technology, neither diversity-based nor focussed strategies in isolation can fuel a consistently successful lead generation function. Diversity screening will find unexpected relationships between small molecules and proteins, but potentially at an unsustainable cost. Focussed screening will find novel hits, but the quantity of information required before this can occur is often not available for the majority of targets.

There are very useful complementarities that arise from partnering diversity and focussed screening approaches. For instance, the ability to quantify molecular diversity enables a much wider proportion of chemical space to be sampled as part of a diversity screen campaign. Even without target information, this methodology can be coupled to drive intelligent sample synthesis or acquisition strategies that aim to improve the coverage of chemical space that can be achieved. From the alternative perspective, where target-related computational methods can be applied, the availability of high-scale screening methodology enables useful progress to be made with much less accurate predictive tools; the degree of focus required can be reduced and this opens up the prospect of, for example, target class-specific focussed sample selection at the level of 10,000–50,000 samples.

The potential to be able to test several hundred thousand data points for a screening campaign is also useful in formulating improved 'hybrid' screening strategies. The iterative, or sequential screening approach is an example of this and uses the strengths of each technology synergistically. In this case, a small diversity screen is run, not to find hits, but as a data acquisition exercise. A computational analysis is then performed to identify trends, and a further set of compounds selected for screening. This process continues until enough hits have been identified. This is a very exciting, if unproven, approach. In theory it has the best of both worlds, but it can also have the worst of both worlds – the diversity screen could fail to find any useful data, and the analysis and selection method could fail to interpret the data correctly. Another area of value is the potential to generate more accurate and informative data across a broader range of samples, for instance by duplicate primary testing or primary dose–response testing. The value of these data comes from its enhanced power to provide structure–activity relationship information directly from the primary data set, but this is only feasible in a screening environment

where high scale automated screening and sample handling are routine.

At GlaxoWellcome, both screening tactics have been carried out routinely. As a recent example of a comparative study, a large screening campaign was carried out on an enzyme target with known inhibitors. In diversity mode, in excess of 500,000 samples were tested and 517 hits found (>70% at 1 μM). In parallel, 6000 compounds were selected using a 3-D pharmacophore and run through a low-throughput screen. This pharmacophore, which was of reasonable complexity (five pharmacophore features), matched 250 of the 517 hits. It also provided the project rapidly with a novel lead series. After 18 months, the project is still optimizing two chemical series, one from the selected screen and one from the HTS.

This complementarity can even be seen in the clinic. HIV protease inhibitors are all from the rational, structure-based design school of drug discovery, whereas the non-nucleoside HIV reverse transcriptase inhibitors have almost all originated from diversity screening. Furthermore, as is now known, combination therapy is the best way to treat HIV infection. This is a good demonstration that the real battle is not between rival technologies, but against disease.

Conclusion

The history of recent drug discovery has swung from biology-directed chemical synthesis and through an era where large-scale zero-knowledge screening predominated. As the cost of carrying out large-scale diversity HTS continues to fall, driven by extreme miniaturization (for example the 3456-well NanoWell™ and the Ultra-High-Throughput Screen System⁹ (UHTSS™, Aurora Biosciences Corporation), a paradigm shift towards less centralized screening activities distributed more closely to end-user groups or individuals is a possible consequence. The power of the methodology therefore becomes more responsive to a more diverse range of individual require-

ments. This model is a logical extrapolation of current practice, taking account of the history of other technologies such as computing (the rise of the personal computer), DNA sequencing and other physicochemical analytical techniques. The era of personal or desktop screening cannot be far off but to get there will require a considerable degree of miniaturization to be applied to the whole screening system, not just to the sample carrier.

In parallel with these approaches, the power of computational predictive methodologies has now reached the stage where a third strategy can be applied to the discovery portfolio to aid in finding pharmaceutical interventions. Information-based screening, be it through the application of focussed, systems-based or the use of iterative screening, will become more efficient, selecting fewer false positives and, more importantly, fewer false negatives. This will be achieved not only through the development of better algorithms, but through established algorithms made accessible by an increase in computing power. The processor power in the average computer doubles every two years and, coupled with the advent of PC 'farms' and Beowulf clusters that give supercomputer performance with commodity components, this will enable the application of more accurate virtual screening methods. The other major contribution to the development of computational methods is the data generated over the years, particularly for target families such as kinases. In the next few years, predictive models for these targets might negate the need for random screening against them, as we will have learnt the essential features required for recognition.

In conclusion, diversity and focussed approaches are likely to continue to provide complementary benefits. The portfolio of techniques will together provide a still more effective and comprehensive approach to the effective identification of novel medicines for the multitude of novel drug targets that are being revealed through the Human Genome Mapping Project.

REFERENCES

- 1 Ashton, M.J. *et al.* (1996) New perspectives in lead generation. II: Evaluating molecular diversity. *Drug Discovery Today* 1, 71–78
- 2 Marriott, D.P. *et al.* (1999) Lead generation using pharmacophore mapping and three-dimensional database searching: application to muscarinic M3 receptor antagonists. *J. Med. Chem.* 42, 3210–3216
- 3 Walters, W.P. *et al.* (1998) Virtual screening – an overview. *Drug Discovery Today* 3, 160–178
- 4 Gillet, V.J. *et al.* (1998) Identification of biological activity profiles using substructural analysis and genetic algorithms. *J. Chem. Inf. Comput. Sci.* 38, 165–179
- 5 Ajay *et al.* (1999) Designing libraries with CNS activity. *J. Med. Chem.* 42, 4942–4951
- 6 Willett, P. *et al.* (1998) Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* 38, 983–996
- 7 Young, S.S. *et al.* (1997) Optimum utilization of a compound collection or chemical library for drug discovery. *J. Chem. Inf. Comput. Sci.* 37, 892–899
- 8 Rees, S. *et al.* (1999) Reporter gene systems for the study of G-protein coupled receptor signalling in mammalian cells. In *Signal Transduction: A Practical Approach* (Milligan, G., ed.), pp. 171–221, Oxford University Press
- 9 Mere, L. *et al.* (1999) Miniaturized FRET assays and microfluidics: key components for ultra-high throughput screening. *Drug Discovery Today* 4, 363–369