

Letting the target determine your compound acquisition strategy

With compound libraries likely to remain the cornerstone of hit and lead discovery for the foreseeable future, this paper seeks to summarise current and future approaches to optimal library design and compound acquisition, using examples and experimental evidence of the success, limitations and weaknesses of target-led approaches in comparison with diversity-screening alternatives.

The global pharmaceutical industry is facing considerable challenges, such as rising research costs, drug failures and decreasing returns on investment. In addition, while the human genome project has revealed a considerable amount of valuable information, it has served also to exacerbate these challenges, particularly with respect to the use of the knowledge gained to create better drugs. Such pressures have exposed the need for increased productivity in the pharmaceutical industry, and this means a greater focus on activities that reduce the ballooning downstream costs resulting from high clinical failure rates. Fast, flexible and cost-effective strategies are needed to meet the demands of creating and sustaining a pipeline of high-value leads with improved prospects for clinical success. How to effectively and efficiently identify targets that are amenable to therapeutic intervention and how to design, test and select candidate compounds are thus key activities in modern drug discovery. Increasingly, challenging tasks, such as the identification of drug-like small molecule modulators of protein function and the translation of these into high-value leads, are now seen as central to 21st century drug discovery. The decisions made at these early stages have extensive consequences for success later in pre-clinical and, even more importantly, in clinical development.

The majority of pharmaceutical and biotech

companies today tend essentially to follow the same well-tested processes, from the cloning and expression of human receptors and enzymes in formats that allow high-throughput, automated screening through to fuelling these screens with compounds. As an approach, high-throughput screening (HTS) has become an embedded technology, in concert with the vast range of potential targets that have emerged from genomics and the newer field of epigenetics. Furthermore, over the last decade, HTS using large so-called 'diverse' compound libraries has been widely adopted by pharmaceutical and biotechnology companies in the pursuit of novel compounds as potential drug candidates. However, this approach, where massive numbers of diverse compounds from earlier combinatorial libraries are screened, has not, by general consent, always lived up to expectations of 10 years ago. This is due, at least in part, to the fact that no collection of a few million compounds can ever represent even the limited property space of known drug-like compounds, let alone the colossal numbers of possible screening entities that could exist. Increasingly, such tactics are being replaced by more limited campaigns where knowledge of the target type and the desired property profile of the hits are used to limit the size of the screening collection, or 'library'. Of course, the corollary is that the design of these libraries is critical to achieving the

By Dr John Harris

Compound Management

required improvements in drug discovery efficiency – in effect, one has traded a certain degree of rare serendipity for greater certainty that the molecules one chooses to invest in have the ‘legs’ to go the distance in development.

From a more or less standing start a decade ago, compound library design is now a very sophisticated art, and this short commentary seeks to examine current and future trends in optimum library design and compound acquisition, especially drawing on experience of the successes, limitations and weaknesses of target-led approaches in comparison with diversity-screening alternatives.

What screens? And what do you find or miss?

Screening assays, by their very nature reductionist in concept, are typically performed using either isolated (purified or cloned) targets, such as enzymes or receptors, or instead use cell-based assays where the target is anticipated to be in a more ‘relevant’ biological context. Isolated target assays offer the ideal format for reproducible and robust screening because they measure a functional consequence of ligand activity at the target protein as directly as possible. Moreover, isolated targets are not constrained by the limiting behaviour of a cellular ‘host’ (for example, through cytotoxicity) and generally they possess a higher tolerance towards universal organic solvents, such as DMSO. Cell-based assays, being at least one step closer to the relevant pathophysiology, have potentially much higher information content than isolated protein assays. As a consequence, the technique of high-content screening, where changes in multiple cell parameters are measured at the same time, has emerged. The challenge is how to convert the resulting massive data stream into classified and interpretable information (structure-activity relationships, or SAR) that can be used by the medicinal chemist and biologist to drive the optimisation of the lead molecules. Of course, the use of cell-based primary assays is not new – most cell-surface targets, such as cell-surface receptors and ion channels, cannot be adequately configured using an isolated protein assay. However, the use of cellular primary screening assays to study intracellular targets is a newer and growing paradigm within the industry.

Nevertheless, in the majority of current drug discovery programmes, it remains most common to use isolated target assays as primary screens, and to use lower-throughput cell-based assays as secondary screens in an attempt to verify the observed ‘primary’ activity in a more physiological environment. Historically it is also the case that certain

targets are more druggable using HTS than others. G-protein coupled receptors (GPCRs), ion channels and proteases are among the most exploited target classes and drugs against these targets produce the highest sales of prescription drugs. Kinases are another major druggable class that often afford excellent lead compounds from HTS. If, as is critical with voltage-dependent ion channels and, as is beginning to be fully appreciated, with protein kinases, different states of the protein can be usefully targeted, critical questions need to be asked about whether the commonly used HTS assays are configured to detect modulators of these other states. If they are not, which is currently the case with most kinase HTS systems used by pharma, it means that screening efficiency can be very poor indeed and many useful hits will be missed.

What compounds to screen?

The composition of an HTS screening collection is critical to the quality of lead generation. Although typically considered to consist of small organic molecules, such collections may also usefully include certain proteins (mAbs, toxins), iRNAs and known drugs and probes. Depending on the therapeutic area of interest, such a composite resource is most effective in drug discovery when it is used to extract information about what molecules to make, rather than expecting it to produce patentable leads in its own right.

It is clear that the target class and sub-family should be taken into account when seeking the most efficient screening strategy. These considerations will determine the optimum type of library or screening collections to be used. In classifying and clustering targets into families, the underlying hypothesis is that similar ligands should bind to similar targets, and thus the knowledge gained previously from one project should be useful in new screening campaigns within the same target family. This approach aims to group potential drug discovery targets into families based on their protein sequence and function, their structural similarity, and the relatedness of the SAR of their ligands. Following this principle, it has been argued that the need for diversity in a screening collection is inversely proportional to the accumulated knowledge of the individual target and, especially, of its family.

For individual organisations, their library collections should, ideally, also reflect the balance of the types of the target they are interested in, with focused libraries arguably being the most efficient at tractable lead generation where ‘family’ knowledge is greatest. The kinase inhibitor field is a case

in point, where many patentable leads and many of the current pre-clinical and clinical candidates have been derived from focused library screening campaigns. Two related types of focused library are commonly used; those that have been tailored specifically to the target family by *de novo* synthetic campaigns either in-house or from specialised commercial sources, and what have been termed ‘rear-view mirror (RVM)’ focused libraries³, where the screening subset is selected from pre-existing diverse libraries using *in silico* recognition methods. The former approach is preferable if the goal is to discover new patentable chemistry. Alternative techniques, involving the *in silico* screening of vast numbers of ‘virtual’ compounds, have also proved effective and they tend to be less expensive in terms of initial outlay but more expensive downstream as the output ‘designer’ molecules have to be synthesised in a one-off manner. Another fashionable and effective alternative approach, if somewhat recycled conceptually, is to screen small molecular fragments, typically much smaller than conventional drug molecules, either in high-concentration conventional bioassays or in physicochemical affinity-based assays. This latter so-called ‘fragment’ approach can be considered complementary to conventional library screening.

How does this translate into compound acquisition and library design?

In putting together a corporate screening library, ‘diverse’ solutions remain as common as ever, with enabling filtering tools becoming more sophisticated (good drug-likeness filters, *in silico* ADME and solubility assessment, toxicophore flagging, etc). Compound suppliers have responded with better quality control, relative ease of resupply and so forth. As a result, collections have tended to become smaller and the embedded chemistries more relevant, but progress towards true structural novelty has been much slower, with many suppliers’ collections overlapping substantially. This is particularly the case with *in silico* selection of RVM focused libraries from pre-existing collections. A common way around this problem is the design and construction in-house, or by outsourcing, of proprietary customised libraries – an intrinsically more expensive option. Many companies have pursued this route in an attempt to build a ‘representative’ library to cover their current and future target interests. The limitation, of course, is that such corporate ‘representative’ libraries can only be sparse in their coverage of drug space – the way to find a needle in a haystack is probably not to make the haystack bigger!

Focused libraries continue to grow in popularity – from a logistics point of view, they are smaller, therefore cheaper to store, maintain and screen, and very often much more efficient in hit-rate terms if they are well-designed in two key respects: to address both what is known about the target class (eg the scaffold ‘focus’) and what is NOT known (eg the ‘diversity’ of the scaffold decoration). Again, such libraries can be bought-in from the few specialist commercial suppliers, or commissioned exclusively. For such libraries, higher upfront costs are traded for overall cost-effectiveness downstream. While the rationale for *de novo* focused library design is clear, the practical aspects of developing and building such libraries is far from simple. Each target family poses different and complex problems of how to use and blend structural information and ligand information, and the best designs prioritise the biological requirement over the synthetic one – if the target requires certain chemotypes, the library must get as close to this as is possible rather than compromise with simpler chemistry.

What does a good corporate library look like?

What often distinguishes the leaders in drug discovery and development from the rest is the quality of their compound libraries and the ease of access that they have to the information within those libraries.

Such corporate collections will generally have a balance of both focused and diversity-oriented sub-libraries, and also a mix of novel (acquired and/or synthesised in-house) and generically-available compounds. They will also have a balance of compounds (synthesised or selected) based on known target structures and based on known target ligands¹. The ideal collection will give adequate (but not too large or promiscuous) hit-rates across a range of target types, and should also give the screener the maximum chance of identifying hit clusters rather than singletons, so that nascent SAR is already evident. Such clusters can be ‘designed-in’ by collating a variety of focused sub-libraries.

The chemistries embedded in the collection should give the medicinal chemist the best chance of developing new composition-of-matter intellectual property rights quickly. As discussed above, this is a problematic issue with RVM libraries in particular, as would be anticipated.

Cost versus novelty

De novo focused designs possess novelty and IPR potential that commercial RVM libraries simply

Compound Management

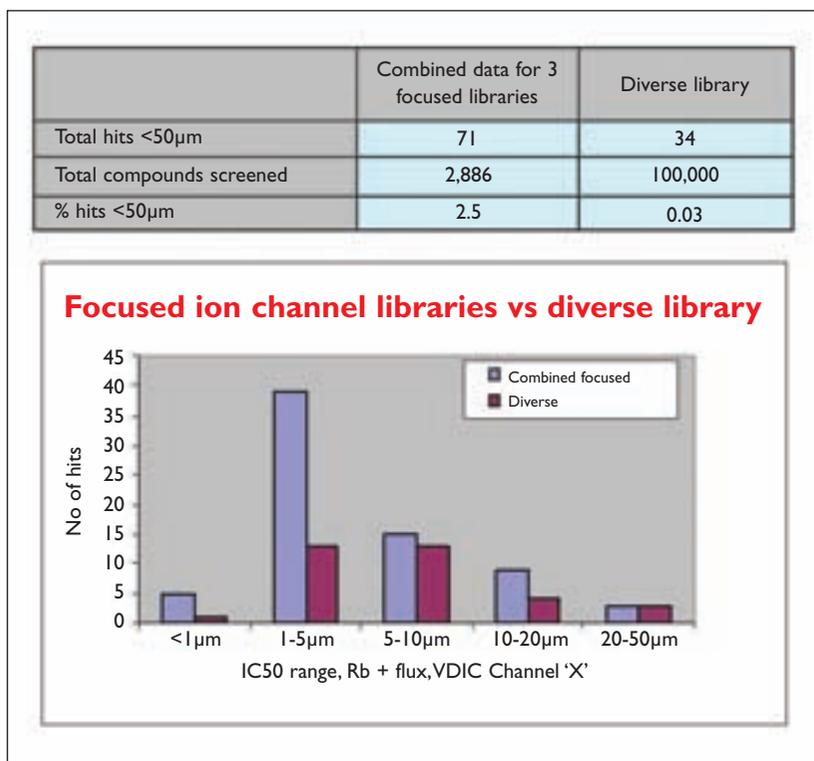


Figure 1
Performance comparison for
VGIC libraries²

cannot attain, but they are significantly more expensive to design and produce. The upfront cost can be offset by non-exclusive or semi-exclusive purchase deals. By comparison, the overall cost-effectiveness of exclusive libraries designed in-house in this way is very difficult to estimate, as so little information is released. However, where *de novo* focused libraries have been available commercially, great success has been achieved in the major protein target areas (for example, >50 patent applications and at least three known clinical candidates arising from BioFocus DPI's *de novo* focused kinase libraries²).

Where structure-based approaches are more tenuous due to a paucity of structural data, focused libraries based on various structural analogies have proven equally effective. In particular, focused libraries based on chemogenomic models of both GPCRs and even targets as tricky as voltage-gated ion channels (VGICs), have elicited much improved hit rates over diversity-based approaches (Figure 1), leading to several novel and highly selective VGIC leads.

The inherent novelty of *de novo* focused libraries can be enhanced by using new scaffold-hopping technologies, especially those which 'scramble' the connectivity of the starting scaffold. Two good examples are Cresset BioMolecular's FieldScreen™ technology (www.cresset-bmd.com) and BioSolveIT's 'ReCore'

approach (www.biosolveit.de). Again, libraries designed in this way are inherently more expensive in terms of initial outlay but the real payback is their enhanced hit-rate and hit quality potential.

The future of compound collections?

Compound libraries are likely to remain the cornerstone of hit and lead discovery for the foreseeable future, with fragment-based approaches and improved chemogenomics increasingly complementing their use. Cell-based (ie less reductionist) screening will increase, and the biochemical, pharmacodynamic and ADME/PK annotation of screening libraries will be in greater demand than at present, but this will have to be paid for upfront in order to achieve real cost and time-to-development efficiencies in the future.

In the near term, new design paradigms and focused libraries based on them will begin to supplant the traditional shotgun screening approaches in areas where small molecule discovery has traditionally been difficult, such as the general field of protein-protein interactions. In the longer term, experimental chemogenomics and theoretical biology will influence the design of new focused libraries that are designed to address and explore multiple targets and signalling pathways, rather than single protein targets. It is perhaps a contradiction in terms, but diversity is at its most powerful when it is focused – the future of novel focused libraries is indeed an exciting one. **DDW**

References

- 1 Sun, D et al. A Kinase-Focused Compound Collection: Compilation and Screening Strategy. *Chem. Biol. Drug Discovery* (2006) 67, 385-394.
- 2 Presented at 11th MipTec Conference, Basel, October 2008.
- 3 Christopher Hulme, presented at 237th ACS National Meeting, Salt Lake City March 22-26, 2009.

Dr John Harris was the co-founder and Chief Scientific Officer for BioFocus (now BioFocusDPI and a division of Galapagos NV) and now acts as an advisor and independent consultant to the company. Dr Harris started his industrial career in 1974 at Wellcome Laboratories as a medicinal chemist and during his two decades at the company went on to become the Head of Cardiovascular Research in the UK before setting up and leading the first 'combiChem' discovery unit within Wellcome in the early 1990s.