Fragment library design
the evolution of fragment-based lead discovery

With the growing need to streamline the drug discovery process, screening against fragment libraries rather than drug-like molecules has become increasingly adopted as an integral part of many drug discovery programmes. However, success depends on the quality of the fragment library, and many factors dictate quality. This review will look at how recent research has influenced the paradigms underlining fragment library design, and its evolution from infancy to its current status today.

In drug discovery programmes, Fragment Based Hit Generation (FBHG) overcomes many of the potential issues and inaccuracies encountered with the traditional approach of high throughput screening (HTS). For example, traditional screening libraries are commonly populated with molecules designed to adhere to Lipinski's rule of five; however, they tend to be large and lipophilic, making them difficult to develop into potent compounds without compromising their ADME properties. Fragment libraries, on the other hand, can be filtered against a number of physiochemical properties, for instance to remove reactive groups and assure solubility.

FBHG burst on the scene in 1996 when Abbott introduced the use of nuclear magnetic resonance (NMR) to guide structure-activity relationships (SAR) between compound and target. This approach, termed 'SAR by NMR'[1] coalesced the ideas of many existing concepts into an easily understandable process. Thus, FBHG started as an inherently NMR-based technique, and while crystallography played a significant role in early FBHG, NMR has traditionally remained the gold standard. In 2001, Chris Lepre laid out the key concepts for assembling a fragment library for NMR, which became the prevailing thought for fragment libraries[2]. He explained that they should include a large number of diverse compounds with high aqueous solubility, and be synthetically tractable for building into compounds.

The early days of FBHG

Fragment screening gained major support following studies by Hann and colleagues[3], who used a simple model to study the interactions between ligands and receptors of varying complexities, calculating the probabilities of binding. It was observed that as ligands become more complex the chance of observing a valuable interaction for a randomly chosen ligand falls dramatically (**Figure 1**). The key to FBHG is that minimally complex molecules are

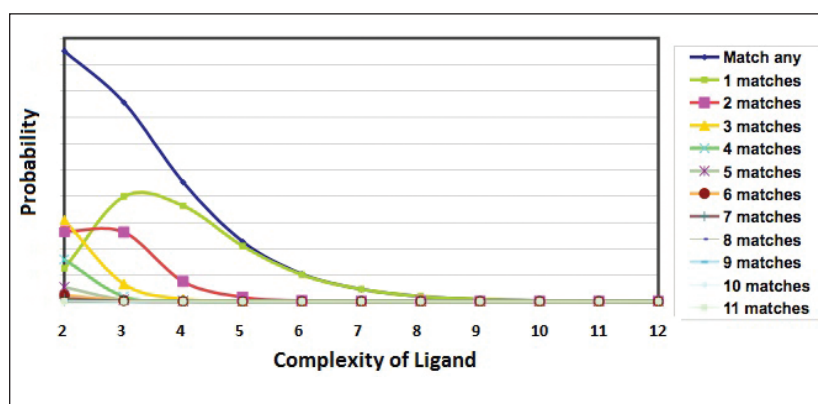**By Dr Edward Zartler, Dr Chris Swain and Simon Pearce**

Figure 1
Probabilities of ligands of
varying complexity matching a
binding site of complexity[3]. As
the ligand/receptor match
becomes more complex, the
probability of any given
molecule matching falls to
zero. The flip side of the coin
being that less complex ligands
will have lower affinity,
requiring more sensitive
methods of detection

more likely to exhibit favourable interactions, albeit weak ones, necessitating sensitive techniques for detection. Building on this idea Pfizer devised an approach designed to minimise the probability of unfavourable interactions, examining compounds based on 'ligand efficiency' metrics. These metrics are intended to compare molecules based on the observed affinity relative to the number of heavy atoms, molecular weight, or calculated lipophilicity[4].

The next major development in the fragment library zeitgeist was the proposal of a 'Rule of 3' (Ro3) for fragments[5], similar to Lipinski's Rule of 5[6]. Congreve and colleagues proposed this 'rule' based upon structural analysis of fragments found to bind to a variety of kinase and protease targets. The Ro3 states that on average, fragment hits tend to exhibit a molecular weight <300 Da, three or less hydrogen bond donors , less than three hydrogen bond acceptors, and finally a cLogP ≤3. Results also suggest that NROT (≤3) and PSA (≤60) are also favourable. It has now come to pass that the Ro3 has become a law that must not be violated, as evidenced by the number of commercially available libraries that are marketed as 'Rule of 3' compliant. Fragment-based screening has been used for a number of drug discovery programmes and a large number of publications describe the results of fragment-based screening, in total identifying more than 350 fragments against more than 60 different molecular targets. The majority of these published fragments obey the Ro3, while the average calculated LogP is perhaps slightly high, the calculated LogD is reduced because almost 40% of the ligands would be predicted to be ionised at physiological pH.

The increasing interest in the area of FBHG has coincided with a surge in the number of screening technologies available for drug development programmes. In addition to NMR and x-ray crystal-

lography we now see a diverse range of innovative technologies, from biosensors to microscale thermophoresis. This in turn has facilitated an increase in the range of possible targets. While originally, biological targets were limited to proteases/kinases, the chemical space of fragment-based screening applications has evolved to include GPCRs, ion-channels and protein-protein interactions. In light of the targets prosecuted today, it would seem that such hard rules are outdated and need to be evaluated on a target-by-target basis.

Recent advances in fragment library design

In the past two years several reviews have appeared discussing the current thought on fragment library design[7-9]. These papers demonstrate the evolution in thought behind fragment library design and highlight how much library design has changed over time. Libraries today are designed to have at least some inherent SAR, enabling medicinal chemists to rapidly generate hypotheses without the need for extensive follow-up. However, fragment libraries are relatively small (500-2,000 molecules) and thus there is limited opportunity to incorporate SAR unless the library members are chosen such that related analogues are commercially or readily available. In this case it is then possible to rapidly evaluate SAR using the so-called SAR-by-catalogue approach. Many vendors have pre-built follow-up libraries for purchase depending on the active fragments found in the initial screen. Schulz and colleagues propose a secondary library of larger molecules (as opposed to fragments), for matching fragments to existing full size compounds. This is an interesting approach to SAR, which previously would typically make only incremental changes to advance potency step-by-step. Alternatively, some vendor libraries are compiled of chemistry-enabled fragments, functionalised with a selection of the reactive groups for synthesis. These fragments are in a sense building blocks for creating a compound of multiple discrete fragments. It has also become apparent that many library designers are beginning to favour a minimum threshold for the size of fragments, as a recent online survey found that most FBHG practitioners limit the size of their fragments to between 16-20 heavy atoms (224-280 Da)[10]. Even more stringent libraries limit size from seven to 10 heavy atoms (~100-140 Da), driven by more sensitive methods of detection and prioritising favourable interactions over binding affinity. For example, although a seven-heavy atom fragment with 5mM affinity still has an acceptable binding

efficiency per atom, most screening methods do not robustly detect such weakly binding fragments, and thus such very small fragments rarely lead to screening hits.

Deciding factors in the quality of a fragment library

Diversity

One of the key aspects of molecule library design is to attempt to cover as much of chemical space as possible with the minimum number of molecules, thereby maximising library diversity. However, the initial offerings from commercial vendors were simple fragments from their existing catalogue using the Ro3. While this enabled them to offer a large selection of compounds, it rapidly became clear that many of the fragments had issues with reactivity and solubility, and that the collections contained many very similar compounds, ie high SAR and low diversity. As awareness of the importance of diversity increased, methods for measuring molecular similarity began to be implemented, and this is now often measured by comparing molecular fingerprints. These can be simple bitstrings where each bit represents the presence or absence of a structural feature. While this approach works fine for conventional drug-sized molecules, unfortunately with small fragments most of the bits are set to zero and a comparison of the sparsely-populated fingerprints is less useful. More discriminating tools for comparing fragments would certainly be useful, and there have been several attempts to quantify the diversity in fragment libraries, employing analysis of functional groups, topological and pharmacophore-based descriptors. Despite these shortcomings, the fundamental approach to library construction has changed, and Maybridge was one of the first vendors to offer smaller, more manageable screening libraries, while maintaining a larger collection for follow up SAR exploration. However, novelty in the starting fragment is probably much less important since there are many opportunities for the ligand to evolve as it is extended towards the final candidate.
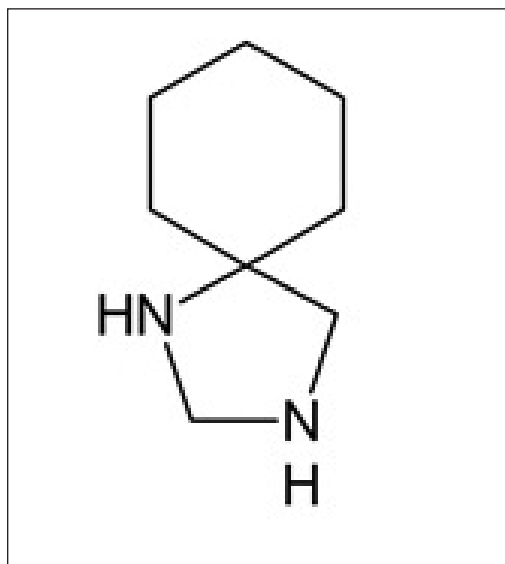
Figure 2: Archetypal 3D Fragment. 3D fragments, also known as scaffolds, may be replacing the standard 2D fragment space

Solubility

Since fragments would be predicted to have only modest affinity for the molecular target, screening has to be carried out at relatively high concentrations. For this reason solubility is a critical property, which is as true today as it was in 2001. While there are several algorithms to predict solubility, many vendors have taken the approach that all fragments should have measured aqueous solubility. This crucial quality control component ensures that every fragment will not be hindered by individual solubility. In particular, Maybridge's quality control approach to ensuring aqueous solubility was demonstrated by the development of a robust solubility assay, described by Myszka and Paul[11]. The experimental data also clearly show a correlation between increasing molecular weight and decreasing solubility, as would be expected. Because solubility is such a critical requirement, the presence of ionisable groups within the fragments becomes an important consideration. Based upon their analysis that ~25% of marketed drugs contain an ionisable group, Pfizer has taken steps to ensure that ionisable groups were fairly represented in their final library[9].

Selectivity

For high-throughput screening there is often a desire to exclude compounds that display off-target activity such as HERG inhibition or CYP450 interactions. However, unlike HTS approaches, since the fragment might be expected to represent only a very small portion of the final candidate there is not the same need to evaluate selectivity in fragment libraries. Interestingly, in results published from fragment screening programmes, out of 350 fragments against 60 different targets, it became apparent that several molecules were active against multiple different targets. While there may be concerns about the potential for promiscuous fragments to be involved in non-specific interactions, it is probable that these structures represent clearly defined binding motifs.

Impurities

Another factor to be considered is that because screening is done at high concentrations, any impurities are also present at much higher concentrations than would be found in conventional HTS. Thus, routine and appropriate quality control should be part of the curation of any fragment library. For example, the Thermo Fisher Scientific Maybridge collection is tested by a variety of analytical methods to confirm the identity and concentration of any impurity, both initially and during storage in solution.

Reactive groups

The fragments should not contain reactive groups or groups that are rapidly converted to reactive groups. While the presence of Fluorine is essential for screening by F-NMR technology, there are other factors that might need to be considered, depending on the screening technology. For example when screening mixtures using x-ray crystallography there is a distinct advantage in choosing mixtures of fragments that are readily distinguishable when comparing the density maps. Fluorescence-based technologies may also be compromised if the fragments are quenchers or fluorescent in themselves.

Focused libraries – quality over quantity

One important aspect of fragment library construction that is often overlooked is the ability to create focused libraries. In these cases, a pharmacophore model is built from a structure or homology model, perhaps with a bound inhibitor, or knowledge of the natural substrate. This model can then be used to select fitting fragments and the subsequent use of highly targeted fragments has been demonstrated to increase the hit rate almost 10-fold. Typically, these focused libraries tend to be small, around 200-300 fragments. However, despite their small size, and because of their high hit rate, a fair amount of SAR can be built in, resulting in very robust SAR hypotheses. One

References

1 Shuker, SB, Hajduk, PJ, Meadows, RP, Fesik, SW (1996). Discovering high-affinity ligands for proteins: SAR by NMR. Science 29; 274(5292), 1531-4.
2 Lepre, CA (2001). Library design for NMR-based screening. Drug Discov. Today, 6(3), 133-140.
3 Hann, MM, Leach, AR, Harper, G (2001). Molecular complexity and its impact on the probability of finding leads for drug discovery. J. Chem. Inf. Comput. Sci., 41(3), 856-64.
4 Hopkins, AL et al (2004). Ligand efficiency: a useful metric for lead selection. Drug Discov. Today, 9, 430-431.
5 Congreve, Carr, Murray and Jhoti (2003). A 'Rule of Three' for fragment-based lead discovery? Drug Discov. Today, 8, 876-877.
6 Lipinski, CA, Lombardo, F, Dominy, BW and Feeney, PJ (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Del. Rev., 46, 3-26.
7 Schulz, MN, Landström, J, Bright, K and Hubbard, RE (2011). Design of a fragment library that maximally represents available chemical space. J. Comput. Aided Mol. Des., 25(7), 611-20.
8 Na, J and Hu, QY (2011). Design of Screening Collections for Successful Fragment-Based Lead Discovery. In Chemical Library Design (Zhou, J. ZX, ed), Vol. 685, pp. 219-240. Humana Press Inc, Totowa.
9 Lau, WF et al (2011). Design of a multi-purpose fragment screening library using molecular complexity and orthogonal diversity metrics. J. Comput. Aided Mol. Des., 25(7), 621-36.
10 Erlanson, D (2012). 'Poll Results: How big are your fragments?' http://practicalfragments.blogspot.co.uk/2012/06/poll-results-how-big-are-your-fragments.html.

Drug Discovery

11 Myszka, D and Paul, J (2011). Exploring the horizons of small molecule drug discovery: the evolution and application of the ideal fragment library. Drug Disc. World, winter, 51.

advantage of some vendor libraries is the ability to cherry-pick individual compounds from their fragment collection in order to assemble these focused libraries. Furthermore, this flexibility decreases sample manipulation and allows for quicker turn-around on testing.

As befitting compounds originally designed to be detected by NMR (ie exhibiting well resolved chemical shifts), fragment libraries tend to be made up of planar, aromatic compounds. Many targets seem to prefer these types of fragments, eg kinases whose natural substrate is a planar, aromatic molecule. The number of hits containing sp2 carbons has led to a proposal that there is a need to have more '3D' nature to the fragments, and a number of companies have designed bespoke structures that explore 3D (or vector) space. However, it is apparent from typical 3D fragments (**Figure 2**) that they will have to be much larger than 2D fragments, and thus the libraries would have to be much larger to cover equivalent 2D fragment space. We suggest that the nomenclature 'scaffold' be used to refer to 3D fragments, in the sense that Plexxikon uses it. Such 3D structures are perhaps a double-edged sword: they do allow side chains at interesting vectors. However, one of the attractions of fragment-based screening has been the ability to mine catalogues quickly and cheaply to explore SAR, but with scaffolds this could become impossible without significant chemistry resources. The exception to this being targets with structural data, where 3D fragments do not require extensive SAR.

For many current targets, structural support is not, nor likely to be, available. So, do scaffolds have value as a complement to fragments, even though they may not, in themselves, directly interact with the target? Small fragments efficiently explore chemical space, but not vector space. Scaffolds on the other hand, explore vector space, but not necessarily chemical space. For example, imagine a furan being found as an active hit from a screen. Producing a SAR hypothesis from this would be straightforward for most chemists, but in the absence of structural data how would you efficiently explore vector space in the binding site? In our conceptualisation of how scaffolds inform fragment SAR, dihydrofuran would be the 3D scaffold that would subsequently be used to explore vector space.

Why explore vector space? If you have structure, it may be possible to readily define a framework that positions binding motifs precisely in 3D space. However, exploring vector space could provide important information if this structural information is unavailable. Structure is always valuable,

and Plexxikon showed that with structure and a good scaffold you can successfully fast-track from bench to bedside, but the challenge remains that many current targets are not going to yield structure, or yield it fast enough to impact hit discovery.

Conclusions

With the increasing pressures on the pharmaceutical industry as a whole, moving into the future of drug discovery necessitates ever more efficient methods for screening programmes. From its emergence in 1996, fragment-based hit generation has been used for a number of drug discovery programmes, and a large number of publications describe the results of fragment-based screening, in total identifying more than 350 fragments against more than 60 different molecular targets.

The success of this approach hinges on the quality of the fragment library and the ideas relating to this have undergone various shifts, initially with the Ro3 law with which all libraries now comply as standard. It is crucial that the quality of the fragment library must keep pace with advances in screening approaches such as biosensor technology. With the introduction of the 3D vector space concept, it will be interesting to see where further developments will lead.   DDW

Dr Edward Zartler founded Quantum Tessera Consulting in 2011 after more than a decade at Merck and Eli Lilly. Involved in the business development of ZoBio, fragment library design and individual project consulting, he is the co-editor of 'Fragment-based Drug Discovery: A Practical Approach' and the popular blog 'Practical Fragments'.

Dr Chris Swain founded Cambridge MedChem Consulting in May 2006 offering a range of consultancy services in drug discovery and medicinal chemistry. He recently joined the Scientific Advisory Board of Selcia Ltd for ongoing collaborations with small/medium and start-up pharmaceutical companies and academic groups, providing lead-finding oversight and lead optimisation input.

Simon Pearce joined Maybridge (now part of Thermo Fisher Scientific) in 1984, progressing to Chemistry Team Leader responsible for generating new building blocks and screening compounds for the Maybridge portfolio. Simon became Chemistry Manager and in 2009 he assumed the role of Global Product Manager for the Maybridge brand.