Fragment-based lead discovery: leads by design

Robin A.E. Carr, Miles Congreve, Christopher W. Murray and David C. Rees

Fragment-based lead discovery (also referred to as needles, shapes, binding elements, seed templates or scaffolds) is a new lead discovery approach in which much lower molecular weight (120–250Da) compounds are screened relative to HTS campaigns. Fragment-based hits are typically weak inhibitors (10μM–mM), and therefore need to be screened at higher concentration using very sensitive biophysical detection techniques such as protein crystallography and NMR as the primary screening techniques, rather than bioassays. Compared with HTS hits, these fragments are simpler, less functionalized compounds with correspondingly lower affinity. However, fragment hits typically possess high 'ligand efficiency' (binding affinity per heavy atom) and so are highly suitable for optimization into clinical candidates with good drug-like properties.

The problems with drug discovery productivity during the 1990s catalyzed significant investment by the pharmaceutical industry into major new high-throughput technologies such, as combinatorial chemistry and HTS. These techniques are now widely adopted and are central to most organizations' lead-generation approaches. HTS typically involves screening approximately a million relatively complex drug-sized compounds, with identification of the most potent hits as the primary objective. Combinatorial chemistry has been central to increasing the numbers of such compounds available for screening. Although these approaches have undoubtedly identified many high-value hits, the limitations of screening drug-sized compounds are starting to become apparent. Hit rates are often low and many of the hits fail to progress into optimization [1,2]. For those that do progress, their optimization into potent compounds tends to actually reduce their initial drug likeness and therefore reduce the developability of the final optimized compounds [3–5]. Fragment-based discovery [6,7] is targeted at addressing the issue of hit rate and also the ability to optimize hits into compounds possessing drug-like physical properties. The term 'fragment' is used here to describe a low molecular weight compound (~120–250Da) that is suitable for screening at high concentrations.

The hit rate from screening fragments is typically much higher than observed with HTS, as there is an inverse relationship between the molecular complexity of compounds screened and the probability of a compound possessing good complimentarity with the target protein [3]. Additionally, a library of small fragments represents a much higher proportion of the available 'chemical space' for low molecular weight compounds than a large library of drug-sized molecules does for higher molecular weight compounds, because the number of possible molecules rises exponentially as molecular weight increases [8]. A consequence of the higher hit rate for fragments is that fewer compounds need to be screened (typically <1000) to identify multiple hits, enabling fragment-based lead discovery to exploit a range of high information content screening techniques such as

Robin A. E. Carr*
Miles Congreve
Christopher W. Murray
David C. Rees
Astex Technology,
436 Cambridge Science Park,
Milton Road,
Cambridge CB4 0QA,
UK
*e-mail:
r.carr@astex-technology.com

1359-6446/05/$ – see front matter PII: S1359-6446(05)03511-7

TABLE 1

Comparison of methods used for fragment screening arranged in order of decreasing throughput and increasing information content

Approach	Typical throughput per screen (compounds)	Quality of information about ligand binding mode	Resource and instrumentation requirements	Protein structure required	Key technical considerations	Representative references
HTS	100–1000K	None	Specialised infrastructure required	No	Not suitable for fragments	
High concentration bioassay	10–50K	None	Very straightforward method	No	High false-positive rates can often hinder interpretation of data	[17,20,31]
Surface plasmon resonance	10–50K	None	Straightforward method, but requires costly instrumentation	No	Protein or compounds must be immobilized; false positives possible	[32]
Affinity mass spectrometry	10–50K	None	Straightforward method, but requires costly instrumentation	No	Limited applications reported	[33,34]
Covalent attachment and MS	10–50K	None	Specialized infrastructure required	No	Requires cysteine residue close to active site	[22,35]
Dynamic combinatorial chemistry and LC/MS	1–10K	None	Straightforward method	No	Limited range of chemistry is suitable	[23,24]
Ligand-detected NMR (1D/2D)	1–10K	Can distinguish active site vs. non-active site binders	Straightforward methods using ^{1}H or ^{19}F, but requires costly instrumentation. Well-suited to screening of mixtures	No	Protein typically >20kDa in size; moderate protein requirements	[15,36,37]
Protein-detected NMR (2D)	1–10K	Information on principle interactions between ligand and protein	Requirement for labeled protein and (usually) $^{1}H/^{15}N$ NMR resonance assignments for amide groups. Requires costly instrumentation	Usually	Protein typically <30kDa in size; high protein requirements	[21,38]
X-ray crystallography	500–1000	Detailed binding mode elucidated	Specialized infrastructure required	Yes	Limited to ~35% drug targets where structure can be solved; moderate protein requirements	[18,39]

X-ray crystallography, NMR and biophysical methods, as well as bioassay techniques. Fragment-screening approaches need to possess high sensitivity in order to detect low-affinity hits in the millimolar or high micromolar range. In this context it is worth noting that, when some types of compounds are screened at high concentration (20–400μM), aggregation can occur leading to non-selective inhibition of many enzymes. This has been referred to as promiscuous aggregating inhibition [9] but such 'false positives' are more easily identified and are therefore much less problematic when biophysical techniques are used for the screening instead of conventional bioassays. Table 1 outlines the principle methods used for screening of fragments, compared with HTS.

Integration of protein X-ray crystallography into the subsequent screening cascade allows detailed structural ligand-binding information to be obtained and enables highly efficient hit validation and optimization. Figure 1 shows the differences between a typical HTS hit and a fragment hit and outlines the reasons why fragment optimization is often more straightforward.

Ligand efficiency and chemical tractability

The concept of ligand efficiency can be used to assess the quality of initial screening hits and also to monitor the quality of leads as they are being optimized. Hopkins *et al.* [10] have defined ligand efficiency (LE) simply as:

$$LE = -\Delta G/HAC \approx -RT\ln(IC_{50})/HAC \quad \text{[Equation 1]}$$

where ΔG is the free energy of binding of the ligand for a specific protein, HAC is the number of heavy atoms in the ligand and the IC_{50} represents the measured potency of the ligand for the protein. The origin of this concept can be traced back to pioneering work of Kuntz, Kollman and colleagues, who showed that, for strong-binding ligands of up to 15 non-hydrogen atoms, the free energy of binding is approximately 1.5 kcal/mol for each non-hydrogen atom [11].

Figure 2 illustrates the concept of ligand efficiency and how it relates to the 'chemical tractability' of a hit. The figure shows graphically a broad generalization of the range of molecular weights and potencies for HTS hits and fragments, superimposed on typical requirements for