ELSEVIER

Hit discovery and hit-to-lead approaches

György M. Keserű[1] and Gergely M. Makara[2]

[1] CADD&HTS Unit, Gedeon Richter Ltd, 19-21 Gyömrői út, Budapest, H-1103, Hungary
[2] Merck Research Laboratories, Merck & Co, RY80Y-325, 126 E. Lincoln Ave, Rahway, New Jersey, 07065, USA

Hit discovery technologies range from traditional high-throughput screening to affinity selection of large libraries, fragment-based techniques and computer-aided *de novo* design, many of which have been extensively reviewed. Development of quality leads using hit confirmation and hit-to-lead approaches present their own challenges, depending on the hit discovery method used to identify the initial hits. In this paper, we summarize common industry practices adopted to tackle hit-to-lead challenges and review how the advantages and drawbacks of different hit discovery techniques could affect the various issues hit-to-lead groups face.

It has been shown that marketed drugs are very frequently highly similar to the leads from which they were derived [1]. Thus, both the quality and the quantity of lead classes available to medicinal chemists are primary drivers for discovering best-in-class medicines. This makes lead generation a crucial step in the drug discovery process. Over the past decade, high-throughput screening (HTS) of corporate compound decks has become the major paradigm for hit or lead discovery in big Pharma. The role of HTS in hit discovery has been recently complemented by several fragment-based screening technologies, which require different hit-to-lead processes. Affinity screening approaches have also emerged as orthogonal methods for early lead discovery [2,3]. For affinity-based techniques, the readout is typically a qualitative or quantitative signal based on the physical interaction between macromolecule (RNA, DNA or protein) and the small molecule partner. With the emerging principle of fragment-based hit generation [4–6], hit discovery techniques can be further subdivided based on the nature of the molecules used to interrogate targets: drug-like compounds that might be tailored to conform to the filters related to the Rule of Five [7], or fragment-like molecules for which the principle of Rule of Three [8] has been put forth. In addition, both compound classes have been the subject of *in silico* lead discovery techniques.

Corresponding author: Makara, G.M. (gergely_makara@merck.com)

All hit discovery approaches have – often orthogonal – shortcomings prompting the frequent use of multiple techniques for hit confirmation (Table 1). Challenges facing traditional HTS technologies include high false-positive rates, the need for reporter assays and the limitation in throughput imposed by testing compounds individually [9]. False negatives in HTS probably occur frequently but, owing to the nature of false negativity, the true incidence is difficult to establish [10]. Affinity methods address some of these issues and possess several advantages over other techniques. However, they are not without their own particular problems in that they do not deliver a functional readout and, therefore, hits require biochemical, cellular or *in vivo* validation. Affinity methods are often carried out using mixtures of compounds to increase throughput and, thus, could also require a deconvolution process to identify single actives. Typical primary detection methods of binding in affinity screenings are NMR spectroscopy, X-ray crystallography, mass spectroscopy (MS) and surface plasmon resonance (SPR).

Sampling of therapeutically relevant chemistry space is the key to the success of any hit discovery program. As a result of this, and in view of the comments made previously, decision making is often biased by hit-rate expectations because the size of chemistry space, the potency of initial hits and the number of compounds screened are clearly interrelated. Better sampling is likely to increase hit-rate and improve the potency of some of the primary hits, whereas the chemistry space of smaller fragment-like mole-

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2006.06.016

TABLE 1

Hit discovery and confirmation techniques

Primary assay	Molecule class	Detection	Hit discovery advantages	Sampling	Hit confirmation challenges	Hit discovery techniques
Biochemical	Drug-like	Mostly fluorescence	Preliminary SAR[a]	Very poor	Data noise, false positives, false negatives	Affinity confirmation, counter screening (proteins, artifacts), preference for cluster hits
	Fragment-like	Mostly fluorescence	Minimal resource need	Better	False positives, weak potency	Multiple assays, counter screening (artifacts), preference for structural data
Affinity	Drug-like	MS	Large mixture capability	Poor	Deconvolution, false positives by MS, false negatives	Biochemical confirmation in discrete format, mixture optimization
	Fragment-like	NMR	Structural information	Better	Weak potency, false positives, no SAR	Biochemical confirmation, structure guided optimization, site-directed screening
	Fragment-like	X-ray	Structural information	Better	Weak potency, false negatives, no SAR	Biochemical confirmation, structure guided optimization
In silico	Drug-like	Scoring	Chemistry-less filtering	Better	Need experimental confirmation and preferably structural data for model	Biochemical confirmation, model refinement by enrichment studies
	Fragment-like	Scoring	Chemistry-less filtering	Good	Need experimental confirmation and preferably structural data for model	Biochemical confirmation, model refinement by enrichment studies

[a] Assuming preference for cluster hits.

cules is exponentially smaller than that of drug-like molecules [11]. In the lead-discovery environment, false positives can result from multiple mechanisms, including: nonspecific hydrophobic binding, poor solubility (protein or substrate precipitation), the tendency of small molecules to form aggregates, reactive functional groups, low purity, incorrect structural assignment or compound concentration, interference with the assay or its readout, and experimental errors. False negatives can be the result of poor solubility, chemical instability, low purity, lower than expected compound concentration, interference with the assay or its readout, and experimental errors.

The generation of viable chemical leads requires at least two components: (i) a clear and rational definition of the requirements of the lead in terms of physical, chemical, biological and pharmacological properties; and (ii) a straightforward strategy to achieve these goals. A generic lead profile can be used for defining the most important characteristics of lead compounds, such as potency, selectivity, physicochemical parameters, and *in vitro* ADME features like microsomal and/or hepatocyte clearance and permeability. The exact composition of the necessary requirements for a lead compound are highly dependent on the therapeutic area or need, and are usually extended by some less exact needs (e.g. synthetic accessibility, optimization potential and patentability) that can be of equal or even greater importance. The consequences associated with this multidimensional hit-to-lead process are the requirement for (i) dedicated hit-to-lead or exploratory medicinal chemistry groups with significant chemistry automation, who are able to provide fast validation of the promising hits, (ii) extensive hit profiling activity including cell-based assays and secondary screening with acceptable throughput at the early phase of the project, and (iii) integrated *in vitro* and *in vivo* ADME characterization that might include predictive toxicology. Even with optimal resources, hit-to-lead optimization is a challenging and often resource-intensive phase of lead discovery that requires strictly value-based management tools and high-throughput processes for all its components. Hit-to-lead processes are anticipated to have high failure rates compared with lead optimization programs, but the early and relatively less costly elimination of undesirable or intractable lead classes is of significant value before extensive medicinal chemistry efforts are initiated.

The method of choice for primary hit generation is often determined by the available infrastructure, resources, prior screening experience and predicted hit rate in the context of therapeutic needs (e.g. anticipated mechanistic limitations, off target selectivity, and pharmacokinetic and pharmacodynamic issues). The next, and equally crucial, phases in lead discovery, hit confirmation and the hit-to-lead processes, need to be tailored to deal with the various challenges encountered in the selected hit discovery processes as summarized in the preceding paragraph. Only after proper follow up can a confirmed hit list be compiled for planning a potential hit-to-lead optimization campaign. Because blind screening will remain the predominant method for orphan or validated targets with no known leads, we review common industry practices for small-molecule hit generation and follow up, and elaborate on common relationships between the hit discovery method of choice and the follow-up issues presented.

Hit identification

Hits and leads can be derived from many sources. Novel targets are typically interrogated in a high-throughput campaign if a suitable assay can be developed. Before – or simultaneously with – that process, small focused compound sets are frequently assayed in a medium-throughput fashion to identify molecules for pharmacological target validation. The latter can become the primary source of leads for well-understood target families because targeted focused libraries are readily available internally in most pharmaceutical companies. Fragment-based lead discovery can also be applied, either as an alternative to HTS, or as an additional lead

source when HTS hits are inadequate to initiate a medicinal chemistry program. For novel targets, focused- and fragment-based screening would also become the method of choice, if no suitable high-throughput assay can be developed for HTS. Novel ideas generated by information on small molecules reported in the literature are potentially the fastest entry to new leads for validated targets. In addition, *in silico* methods remain an attractive option for prioritizing structures for focused screening [12,13] or fragment-based lead discovery [14,15].

In vitro biochemical and cellular assays have long been used for HTS of approximately a million drug-like molecules in miniaturized formats [16]. Typically, single-point experiments are performed with small molecules at concentrations of 1–50 μM with a 30–50% activity cutoff to identify wells for confirmation. Confirmation is performed either by single-point repeats or by titration in the original high-throughput or medium-throughput format, or in both. This HTS workflow is customary in the industry but is hampered by several drawbacks. One can only sample a tiny proportion of drug-like chemistry space (estimated to contain $\geq 10^{60}$ molecules) with approximately a million compounds. Material improvement of the sampling rate is unlikely under the current paradigm [11,17]. False-positive and false-negative rates in HTS are relatively high owing to the contributing factors detailed earlier in this review. Fragment-like compounds can also be screened in biochemical assays. Several factors are easier to control in fragment-based lead discovery because the number of samples is significantly smaller than that used for HTS. Compounds can be prefiltered for solubility, purity and reactive functional group considerations, and structural assignment and compound concentration can be confirmed before adding new entries to the screening deck. Owing to the expectation of weak potency of the primary hits, it is necessary to perform biochemical assays of the low molecular weight compounds at relatively high concentrations [18]. Major issues plaguing this approach include aggregate formation [19], promiscuous binding [20,21] and assay interference. An attractive alternative approach – substrate activity screening (SAS) – has also been proposed, in which the accumulating cleavage product of *N*-acyl aminocoumarin libraries against proteases can be monitored [22]. The underlying idea can undoubtedly be extended for other enzymatic protein families, but it is likely to remain limited for receptors or protein–protein interactions.

Affinity screening techniques can accommodate small molecule mixture formats and, therefore, can increase throughput (and sampling) for drug- or fragment-like molecules. The mixture capability is the result of the assay readout being specific to the interacting small molecules as opposed to, for example, the substrate [23,24]. Affinity methods used to screen drug-like molecules use MS readouts and usually eliminate false positives caused by nonspecific hydrophobic binding, poor solubility, reactive functional groups, low purity or assay interference. However, the occurrence of false negatives remains an issue. The most widely applied affinity techniques used in the primary screening of fragment-like molecules include NMR [25], MS [26] and X-ray [27,28] methods, which are primarily affected by false positives (NMR and MS) [6], false negatives (X-ray and MS) [29], limitations related to target size and ligand off rates [6], tendency to form crystals, or need for mutations or labeling.

Hit confirmation

Hits can be derived from many sources, but ultimately an actual sample of the presumed hit structure must demonstrate activity in a primary biological assay. However, it is crucial to establish that the observed signal is caused by a desirable mechanism. In the past decade, the significant experience gathered from multiple HTS campaigns has facilitated improvement and refinement of the process. Particle count measurements for varying sample concentration, 2D fluorescence intensity distribution analysis for fluorescence interference [30] or color quenching corrections in scintillation proximity assays (SPAs) have been proposed to eliminate some artifacts [31]. A useful practice for identifying promiscuous or frequent hitters is to monitor hit rates across all screening campaigns of individual samples [32]. It has been proposed that one of the common mechanisms for promiscuous binding might be the tendency of some small molecules to form aggregates [19]. Such compounds have been pinpointed by repeated biochemical screenings at different protein concentrations or with additives such as detergents [19]. Alternatively, dynamic light-scattering readings can be used to look for large aggregate particles in biological buffer systems [33]. Counterscreens have also been used to eliminate false positives caused by assay interference or aggregate formation. To this end, actives in the primary biochemical screen are assayed against another member of the target family under identical assay conditions. Ideally, the only difference between the primary and counter assay is the target, whereas other reagents and parameters, such as concentration, are kept unchanged. If the observed activity is the same in both assays the small molecule is likely to be either promiscuous or a false positive. This technique is also expected to eliminate false positives caused by aggregate formation and small molecule precipitation. By contrast, follow-up assays using different assay readouts (at several fluorescence wavelengths or with radiolabeled ligands, etc.) provide less information content and might be limited to the elimination of molecules that interfere with the assay readout of the primary screen. A few methods have been suggested for filtering out undesirable mechanisms, such as oxidation or alkylation caused by reactive moieties. For many targets or therapeutic areas, elimination of often unobvious offenders is crucial for avoiding costly lead optimization on chemical series that would not become drugs. Reactivity towards reactive thiol groups, such as glutathione or a cysteine in La antigen protein, has been proposed to signal such mechanisms. The former is quantified by fluorogenic quenching [34], whereas the latter can be applied using ALARM ('a La assay to detect reactive molecules') NMR or ALARM MS [35]. In addition, microcalorimetry and affinity-based techniques, mostly SPR, have been used [36,37] to establish binding for HTS positives and to weed out false positives using mechanisms such as protein precipitation. Representative members of the remaining hits, which survive all the postscreening filters, are usually resynthesized to verify their structural assignment and then reassayed to confirm that the observed activity in HTS was not caused by an impurity.

Small fragment-like compounds evaluated in traditional biochemical screens have been reported: 'needle' (an alternative name for fragment-based) screening [18] was followed up with several studies to handle false positives, including multiple assay formats, SPR binding and structural studies (NMR and X-ray).

Biochemical screens have also been used to prioritize fragments for crystallography, thereby enabling the evaluation of larger fragment pools while keeping the required X-ray resources low [28]. Issues related to aggregate formation, or fluorescence interference, become more frequent with increasing compound concentration but can be dealt with in a manner similar to that detailed in the preceding paragraph. It has also been noted by multiple groups that structural information has been found to be invaluable in aiding the progression of weak fragment hits to more advanced leads [18,20]. SAS detects accumulating substrate turnover and can therefore circumvent issues related to aggregates, solubility and nonspecific binding observed in simple reporter assays. SAS has recently been demonstrated for proteases, but similar techniques can conceivably also be adapted for the substrate-binding site of other enzymes.

An important aspect of selecting chemical series for follow-up is their binding mechanism. For some targets, active site ligands are preferred, whereas for other protein families, an allosteric mechanism might be considered advantageous. Typically, selected members of hit clusters are evaluated in competition studies in biochemical or affinity-based [38] settings, although final determination can often only be made by structural studies [39].

Hit-to-lead techniques

Once the list of confirmed hits has been compiled with supporting data, the selection process begins for prioritizing chemical series for hit-to-lead follow up. The key considerations might be radically different for hits derived from the different hit discovery techniques. Potency *per se*, one of the primary drivers for prioritization in the dawn of the HTS era, has been seen as being of lesser importance, following the realization that gaining potency relatively rapidly is not a major bottleneck. For hits from HTS, selectivity, chemical tractability, binding mechanism, pharmacokinetic properties and freedom of operation (patentability) are usually viewed as more significant than the actual potency of the confirmed hits. In addition, a strong preference for cluster hits over singletons in HTS screens is common. However, for fragment-based hits, there is generally little value in judging selectivity or clustering for weak hits originating from small and diverse fragment libraries. The availability of structural information on binding mode and opportunities to rapidly and easily produce structural analogues might be highly valued, whereas pharmacokinetic or intellectual property (IP) evaluation often makes little sense at such an early stage.

The hit-to-lead optimization paradigms are also dependent on the screening technology used for hit discovery. We have here collated many of the published success stories of hit-to-lead optimization and analyzed them in relation to the technologies they used (Table 2). The hit-to-lead process for drug-like hits might involve techniques such as hit evolution, (bio)isosteric replacements and hit fragmentation, or any combination of these (Figure 1). Hit evolution is one of the most frequently used techniques for hit optimization when analogues of the original hits are synthesized with different substitution patterns and tested. These focused libraries are usually created by solid- or solution-phase parallel synthesis, coupled with high throughput purification to facilitate compound output for screening. Alternatively, crude or purified compounds, synthesized in mixtures, can be screened under competitive conditions with affinity screening methods to identify analogs with improved K_d [38]. Initial SAR data produced by these libraries can then be used to drive exploratory medicinal chemistry efforts, producing compounds with an improved lead-like profile. (Bio)isosteric replacements represent a case of hit evolution when hits are transformed using bioisosteric rules. Unlike chemical similarity, biological similarity depends on the structure of the target and, as such, prevents the application of bioisosterism as a general method. Nevertheless, bioisosteric replacements, if applicable, are useful for improving the hit profile while maintaining potency. Hit fragmentation, most often applied when the initial HTS hits are large molecules, is the structural decomposition of HTS hits that leads to the identification of promising fragments or minimum pharmacophores. Although it has been demonstrated that HTS hits with significant molecular complexity left reduced chemistry space for subsequent hit-to-lead optimization, the structural information encoded in the large complex molecules can be used for fragmentation. Fragments produced from this approach can be subjected to fragment-based hit-to-lead technologies. Identification of the minimal core fragments can then serve as a new starting point for fragment expansion. Fragmentation can be followed by the combination of fragments identified even in different hits (fragment linking or merging). Retrosynthetic fragmentation can facilitate developing dynamic combinatorial libraries for fragment self-assembly.

The examples collected in Table 2 suggest that successful hit-to-lead optimization of HTS hits resulted in drug-like leads, regardless of the lead-likeness of the original hit. Problems in the lead-likeness of the hits are basically attributed to molecular weight and logP, reflecting the fact that HTS hits are often large lipophilic molecules. Rapid resolution of such issues has been largely aided by the considerable expertise accumulated in these areas by the medicinal chemistry community. The chemical similarity between hits and leads derived from them is somewhat larger for HTS hits (average 0.61) than that for fragment-based hits (average 0.56). This is probably a consequence of the larger molecular weight of hits, as well as saddle modifications needed to tweak the initial hits during the hit evolution process. However, the value of this conventional approach is clearly demonstrated by improvements of ligand efficiency (LE), defined as free energy of binding per heavy atom: $LE = \Delta G/N_{\text{non-hydrogen atoms}}$, for the HTS hits during the lead optimization (LE gain in 87% of cases, average HTS hit-to-lead $\Delta LE = 0.07$, Table 2).

The methods in the preceding paragraph applied in hit fragmentation are also used as primary hit-to-lead technologies for fragments identified by fragment-based screenings (Figure 2). Like drug-like hits, hit-to-lead optimization of fragments is often a combination of site-directed screening or fragment expansion, and fragment linking or merging, although in practice these technologies tend to be used together. High-throughput parallel expansion of the initial fragment hits has been found over the years in Abbot to be the most effective way to progress fragment hits obtained in NMR screening [40]. It was also noted that although library design was typically aided by structural data, serendipitous findings arising from the addition of random diversity to the focused library often resulted in discoveries of novel and unpredictable interactions [40,41]. Throughout the process of fragment expansion, new functionalities, structural moieties or hydrophobic surfaces are built in to the original fragment picked

TABLE 2

Hit evolution to leads

Target	Screening strategy	Hit activity[a]	Hit LE	Lead-likeness[b]	Primary H2L Strategy	Lead activity	Lead LE	Chemical Similarity[c]	Drug-likeness[b]	Ref.
JNK1	HTS	0.1 μM IC_{50}	0.36	MW 364.4	Isosteric replacement	16 nM IC_{50}	0.38	0.50	OK	[42]
mGluR5	HTS	0.37 μM IC_{50}	0.55	OK	Isosteric replacement	5 nM IC_{50}	0.82	0.41	OK	[43,44]
Tryptase	HTS	1.26 μM IC_{50}	0.29	MW 411.5, logP 3.66	Isosteric replacement	100 nM IC_{50}	0.32	0.56	OK	[45]
NR2B	HTS	0.1 μM IC_{50}	0.41	MW 321.4, logP 4.28	Isosteric replacement	2 nM IC_{50}	0.43	0.57	OK	[46,47]
BACE1	HTS, MS	25 μM IC_{50}	0.17	MW 508.7, logP 3.99	Isosteric replacement, hit evolution	11 nM IC_{50}	0.27	0.62	MW = 578.8	[48,49]
CXCR2	HTS	4.6 μM IC_{50}	0.39	logP 3.44	Hit evolution	28 nM IC_{50}	0.49	0.89	OK	[50]
MC4	HTS	2.75 μM IC_{50}[d]	0.35	MW 334.8, logP 4.55	Hit evolution	46 nM K_i	0.42	0.81	OK	[51]
JNK3	HTS	0.3 μM IC_{50}	0.45	MW 365.6, logP 3.45	Hit evolution	80 nM IC_{50}	0.36	0.73	OK	[52]
Adenosine A2	HTS	0.15 μM K_i	0.42	OK[e]	Hit evolution	27 nM K_i	0.42	0.68	OK	[53]
JNK1	HTS	1 μM IC_{50}	0.41	logP 4.22	Hit evolution	20 nM IC_{50}	0.44	0.68	OK	[42]
CDK2	HTS	1.5 μM IC_{50}	0.53	OK	Hit evolution	37 nM IC_{50}	0.46	0.42	OK	[54]
MCHr1	HTS	1.68 μM IC_{50}	0.26	logP 4.39	Hit evolution	3 nM IC_{50}	0.46	0.52	OK	[55]
P2X7	HTS	0.12 μM IC_{50}	0.26	MW 540.3, logP 6.31	Hit fragmentation	40 nM IC_{50}	0.44	0.44	OK	[56]
HCV NS5B	HTS	12 μM IC_{50}	0.18	MW 484.6, logP 5.78	Hit fragmentation	4.3 μM IC_{50}	0.3	0.64	OK	[57]
IKK2	HTS	1.6 μM IC_{50}	0.57	OK	Hit fragmentation	25 nM IC_{50}	0.61	0.71	OK	[58]
c-Src	Biochemical	40 μM IC_{50}	0.32	logP 3.80	Fragment linking	64 nM IC_{50}	0.32	0.69	OK	[59]
U1061A RNA	MS	>100 μM D_d	N/A	OK	Fragment linking	6.5 μM D_d	0.17	0.67	MW = 558.6	[60]
PTP1B	NMR	293 μM K_i	0.21	MW 313.3	Fragment linking	22 nM K_i	0.2	0.51	MW = 713.8 HBA = 6	[61]
PTP1B	NMR	800 μM D_d	0.29	OK	Fragment linking	6.9 μM K_i	0.24	0.56	OK	[62]
FKPB	NMR	2 μM D_d	0.3	MW 365.4	Fragment linking	19 nM D_d	0.24	0.79	MW = 620.7	[63]
MMP3	NMR	20 μM D_d	0.43	logP 3.24	Fragment linking	25 nM D_d	0.49	0.31	OK	[64]
HPV E2	NMR	1.9 mM D_d	0.2	logP 3.52	Fragment linking	10 μM IC_{50}	0.31	0.56	logP = 5.35	[65]
Bcl-xL	NMR	180 μM IC_{50}	0.24	MW 320.2, logP 5.00	Fragment linking	10 μM IC_{50}	0.16	0.71	MW = 611.7 logP = 8.10	[66]
Bcl-2	NMR	300 μM D_d	0.3	OK	Fragment linking	36 nM K_i	0.27	0.24	MW = 551.6 logP = 5.11	[67]
LFA-1/ICAM-1	NMR	80 μM D_d	0.26	MW 303.3, HBA 4	Fragment linking	40 nM K_i	0.33	0.56	OK	[68]
Glycogen phosphorylase	XRA	130 nM IC_{50}	0.34	MW 379.5	Fragment linking	6 nM IC_{50}	0.33	0.57	OK	[69]
Urokinase	XRA	56 μM K_i	0.49	OK	Fragment linking	370 nM K_i	0.49	0.51	OK	[70]
Gelatinase B	Biochemical	2.7 mM IC_{50}[d]	0.32	OK	Fragment expansion	13 μM IC_{50}	0.25	0.60	OK	[71]
DNA gyrase	Biochemical	41 mM MNEC[f]	0.13	logP 3.45	Fragment expansion	62 nM MNEC[f]	0.28	0.26	logP = 6.39	[18]
PDE IV	Biochemical	60 μM IC_{50}	0.49	OK	Fragment expansion	33 nM IC_{50}	0.49	0.62	OK	[28]
HPV E1 helicase	Biochemical	2 μM IC_{50}	0.41	OK	Fragment expansion	4.3 nM IC_{50}	0.34	0.62	OK	[72]
Thymidilate synthase	MS	1.1 mM K_i	0.23	OK	Fragment expansion	330 nM K_i	0.3	0.68	HBA = 6	[73]
ErmAM	NMR	1 mM D_d	0.41	OK	Fragment expansion	8 μM K_i	0.3	0.46	OK	[74]

TABLE 2 (*Continued*)

Target	Screening strategy	Hit activity[a]	Hit LE	Lead-likeness[b]	Primary H2L Strategy	Lead activity	Lead LE	Chemical Similarity[c]	Drug-likeness[b]	Ref.
P38α	XRA	1 mM IC_{50}	0.29	OK	Fragment expansion	100 nM IC_{50}	0.31	0.28	OK	[75]
Carbonic anhydrase	Assembly	N/A	N/A	OK	Fragment assembly	84 nM IC_{50}	0.46	0.74	OK	[76]
Neuraminidase	Assembly	31 μM IC_{50}	0.41	HBD 4	Fragment assembly	85 nM IC_{50}	0.48	0.81	OK	[77]
CDK2	Assembly	N/A	N/A	OK	Fragment assembly	30 nM IC_{50}	0.45	0.58	OK	[78]
Caspase-3	Assembly	50 μM K_i	0.54	OK	Fragment assembly	2.8 μM K_i	0.26	0.46	HBA = 6	[79]

[a] Fragment activity belongs to the most active fragment was considered.
[b] Leadlikeness, drug likeness: MW, logP, H-bond acceptor (HBA), H-bond donor (HBD) were evaluated using rule-of-3 and rule-of-5, respectively.
[c] Tanimoto coefficients as calculated by sim2d (OpenEye) using MACCS keys.
[d] Estimated by logit transformation [80]: logit = log[(% activity ÷ 100) − (% activity)].
[e] Estimated logP owing to missing fragments in logP calculation.
[f] MNEC: maximum not effective concentration.

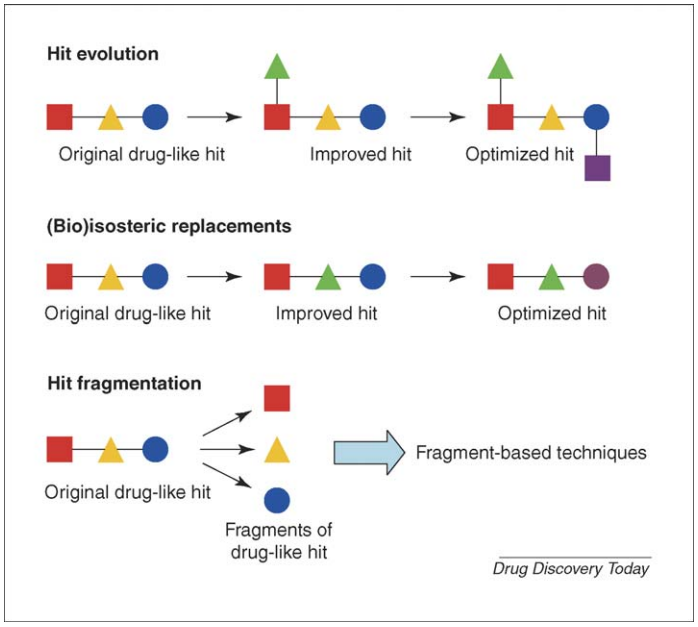

FIGURE 1

Schematic representation of hit-to-lead approaches for drug-like hits. The most common approach for evolving drug-like hits is hit evolution, a systematic SAR-driven analoging process. The strategy of (bio)isosteric displacement is mostly used for improving the pharmacokinetic or pharmacodynamic profile of the lead molecule. Hit fragmentation can be applied when the initial hit is a large molecule that cannot be significantly processed by hit evolution or (bio)isosteric displacement.

up in primary screening. As a result, the original fragment becomes more complex, with more potential target-interaction points, resulting in compounds with increased potency. Fragment expansion is well suited for the multidimensional optimization task of the hit-to-lead process because it easily allows simultaneous consideration of selectivity and ADME issues as potency improves. It has been noted that access to structural data has been crucial for the rapid evolution of fragments and keeping the number of compounds synthesized low. The whole process of fragment expansion is governed by human intervention and, ideally, supported by primary and secondary assays. Examples in Table 2 demonstrate that this tight control helps to preserve lead-like properties during optimization and often ends with a drug-like lead. Ligand efficiency is usually conserved, or decreases only slightly, during the hit-to-lead process with fragment expansion (average fragment expansion hit-to-lead $\Delta LE = 0$; Table 2), which is acceptable because the typical primary fragment hits are efficient binders. The basic premise of fragment linking or merging is that if two or more elements interact closely in the binding pocket with the protein target then, owing to the entropic component, connecting them can produce more-complex molecules with higher potency. However, linking or merging fragments is not a straightforward process even with structural information at hand, and it can significantly increase molecular complexity, although binding efficiency can be maintained. In fact, ligand efficiency increased only for 3 out of 12 cases (25%) shown in Table 2, with no change being the average (average fragment linking hit-to-lead $\Delta LE = 0$). The 'likeness' values, generated in Table 2 on the basis of Rule of Three for fragments and the Rule of Five for drug-like molecules, demonstrate that, unlike the similar comparison for HTS hits, if

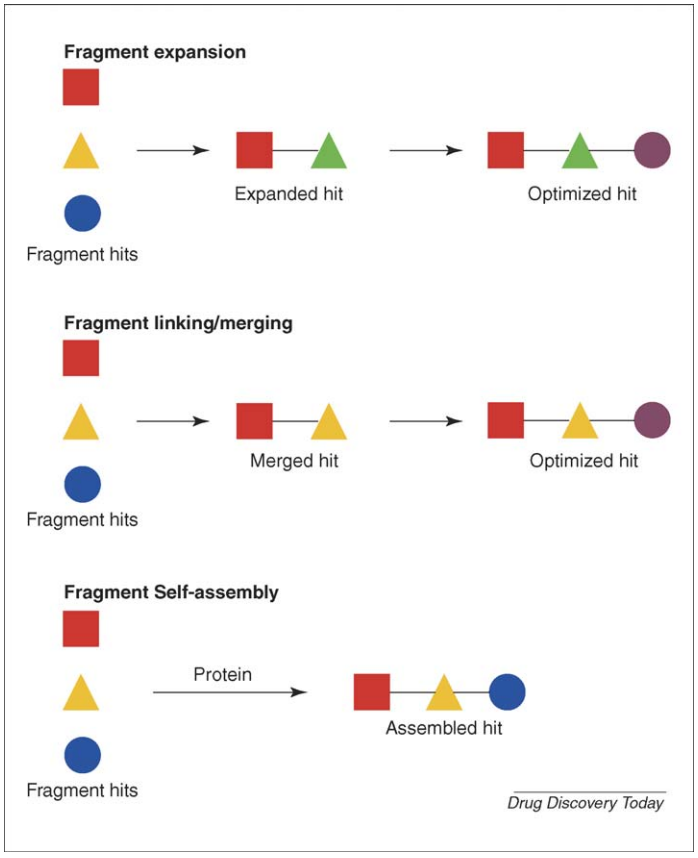

FIGURE 2

Schematic representation of hit-to-lead approaches for fragment hits. Fragment expansion, an analoging process similar to hit evolution, is the most successful technique used for evolving fragment hits. Fragment linking can be used when two or more fragments are believed to bind to the target in close proximity to one another. Fragment assembly is the active-site-driven assembly of chemically compatible fragment units.

the initial fragment is not lead-like then corresponding lead tends not to be drug-like either. On one hand, this means that there is little chance of maintaining favorable properties during hit-to-lead optimization using fragment linking exclusively. On the other hand, growing molecular weight and complexity caused by combining fragments of similar ligand efficiency keeps the ligand efficiency in check. Fragment self-assembly is a process where the target facilitates the self-assembly of fragments armed with compatible functionalities. Similar to other molecular self-assemblies, the catalytic effect of the protein template in generating new, more complex and, hopefully, more potent compounds is used. Although there are some examples reported in Table 2, the need for chemical complementarity between the fragments, as well as steric, electrostatic and hydrophobic complementarity for each fragment towards the target site, imposes severe challenges and limitations. Indeed, the evident drop in ligand efficiency can be attributed to the chemistry limitations described earlier (average fragment assembly hit-to-lead $\Delta LE = -0.1$; Table 2). Fragment self-assembly is therefore unlikely to become a general strategy for the hit-to-lead optimization of fragments.

Conclusions

After a decade of experience with various discovery technologies and optimization techniques, a universal lead-discovery process devoid of traps and artifacts remains elusive. Despite the ever-increasing understanding of cheminformatics and chemical diversity, and emerging techniques for parallel synthesis, assay signaling and miniaturization, the scientific community continues to struggle with finding solutions to the key conundrum of lead generation: a cost-effective optimal sampling of chemistry space and detection of a biological signal to provide an abundance of well-characterized lead classes for medicinal chemistry development or target validation. Instead, lead discovery groups are often forced to either prioritize too early or to spend significant resources on a relatively small number of unattractive chemical series. Fortunately, the solutions developed for the many issues faced by current technologies are slowly enhancing the fidelity and efficiency of the lead-discovery process. However, the addition of new technologies and checkpoints is also contributing to the astronomical and ever-rising cost of getting medicines to the market. Thus, assembling a successful chain of tools with the entire lead generation scheme in mind will become crucial for achieving systematic success in a rapidly changing pharmaceutical industry.

References

1 Proudfoot, J.R. (2002) Drugs, leads, and drug-likeness: an analysis of some recently launched drugs. *Bioorg. Med. Chem. Lett.* 12, 1647–1650
2 Makara, G.M. and Athanasopoulos, J. (2005) Ligand affinity binding in improving success-rates for lead generation. *Curr. Opin. Biotechnol.* 16, 666–673
3 Comes, K.M. and Schurdak, M.E. (2004) Affinity-based screening techniques for enhancing lead discovery. *Curr. Opin. Drug Discov. Devel.* 7, 411–416
4 Carr, R.A.E. *et al.* (2005) Fragment-based lead discovery: leads by design. *Drug Discov. Today* 10, 987–992
5 Erlanson, D.A. *et al.* (2004) Fragment-based drug discovery. *J. Med. Chem.* 47, 3463–3482
6 Zartler, E.R. and Shapiro, M.J. (2005) Fragonomics: fragment-based drug discovery. *Curr. Opin. Chem. Biol.* 9, 366–370
7 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug. Delivery Res.* 23, 3–25
8 Congreve, M. *et al.* (2003) A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today* 8, 876–877
9 Hann, M.M. and Oprea, T.I. (2004) Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.* 8, 255–263
10 Popa-Burke, I.G. *et al.* (2004) Streamlined system for purifying and quantifying a diverse library of compounds and the effect of compound concentration measurements on the accurate interpretation of biological assay results. *Anal. Chem.* 76, 7278–7287
11 Harper, G. *et al.* (2004) Design of a compound screening collection for use in high throughput screening. *Comb. Chem. High Throughput Screen.* 7, 63–70
12 Polgár, T. *et al.* Virtual Screening. In *Encyclopedia of Pharmaceutical Technology* (Swarbrick, J. and Boylan, J.C., eds) Taylor and Francis Group (in press)
13 Jain, A.N. (2004) Virtual screening in lead discovery and optimization. *Curr. Opin. Drug Discov. Devel.* 7, 396–403
14 Pickett, S.D. *et al.* (2003) Discovery of novel low molecular weight inhibitors of IMPDH via virtual needle screening. *Bioorg. Med. Chem. Lett.* 13, 1691–1694
15 Ward, R.A. *et al.* (2005) Structure-based virtual screening for low molecular weight chemical starting points for dipeptidyl peptidase IV inhibitors. *J. Med. Chem.* 48, 6991–6996
16 Golebiowski, A. *et al.* (2003) Lead compounds discovered from libraries: part 2. *Curr. Opin. Chem. Biol.* 7, 308–325
17 Mestres, J. and Veeneman, G.H. (2003) Identification of ''latent hits'' in compound screening collections. *J. Med. Chem.* 46, 3441–3444

18 Boehm, H.-J. *et al.* (2000) Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. *J. Med. Chem.* 43, 2664–2674

19 McGovern, S.L. *et al.* (2002) A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J. Med. Chem.* 45, 1712–1722

20 Fejzo, J. *et al.* (1999) The SHAPES strategy: An NMR-based approach for lead generation for drug discovery. *Chem. Biol.* 6, 755–769

21 Hann, M.M. *et al.* (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* 41, 856–864

22 Wood, W.J.L. *et al.* (2005) Substrate activity screening: a fragment-based method for the rapid identification of nonpeptidic protease inhibitors. *J. Am. Chem. Soc.* 127, 15521–15527

23 Annis, A.D. *et al.* (2005) An affinity selection–mass spectrometry method for the identification of small molecule ligands from self-encoded combinatorial libraries: Discovery of a novel antagonist of *E. coli* dihydrofolate reductase. *Int. J. Mass. Spec.* 238, 77–78

24 Zehender, H. *et al.* (2004) SpeedScreen: The ''missing link'' between genomics and lead discovery. *J. Biomol. Screen.* 9, 498–505

25 Villar, H.O. *et al.* (2004) Using NMR for ligand discovery and optimization. *Curr. Opin. Chem. Biol.* 8, 387–391

26 Swayze, E.E. *et al.* (2002) SAR by MS: A Ligand Based Technique for Drug Lead Discovery Against Structured RNA Targets. *J. Med. Chem.* 45, 3816–3819

27 Hartshorn, M.J. *et al.* (2005) Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* 48, 403–413

28 Card, G.L. *et al.* (2005) A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design. *Nat. Biotechnol.* 23, 201–207

29 Sanders, W.J. *et al.* (2004) Discovery of potent inhibitors of dihydroneopterin aldolase using CrystaLEAD high-throughput X-ray crystallographic screening and structure-directed lead optimization. *J. Med. Chem.* 47, 1709–1718

30 Heyse, S. *et al.* (2005) Quantifying bioactivity on a large scale: quality assurance and analysis of multiparametric ultra-HTS data. *JALA* 10, 207–212

31 Park, Y.W. *et al.* (1999) Homogeneous proximity tyrosine kinase assays: Scintillation proximity assay versus homogeneous time-resolved fluorescence. *Anal. Biochem.* 269, 94–104

32 Priestle, J. *et al.* (2004) Molecular informatics as an enabling *in silico* technology platform for drug discovery. *Chimia* 58, 577–584

33 McGovern, S.L. (2003) A specific mechanism of nonspecific inhibition. *J. Med. Chem.* 46, 4265–4272

34 Epps, D.E. and Taylor, B.M. (2001) A Competitive fluorescence assay to measure the reactivity of compounds. *Anal. Biochem.* 295, 101–106

35 Huth, J.R. *et al.* (2005) ALARM NMR: A rapid and robust experimental method to detect reactive false positives in biochemical screens. *J. Am. Chem. Soc.* 127, 217–224

36 Davidson, W. *et al.* (2004) Discovery and characterization of a substrate selective p38alpha inhibitor. *Biochemistry* 43, 11658–11671

37 Cooper, M.A. (2004) Advances in membrane receptor screening and analysis. *J. Mol. Recognit.* 17, 286–315

38 Annis, D.A. *et al.* (2004) A general technique to rank protein-ligand binding affinities and determine allosteric versus direct binding site competition in compound mixtures. *J. Am. Chem. Soc.* 126, 15495–15503

39 Regan, J. *et al.* (2002) Pyrazole urea-based inhibitors of p38 MAP kinase: from lead compound to clinical candidate. *J. Med. Chem.* 45, 2994–3008

40 Hajduk, P. (2005) NMR-based screening in drug discovery and design. *230th ACS National Meeting*, 28 August – 1 September 2005, Washington, DC, USA (MEDI-462)

41 Petros, A.M. *et al.* (2006) Discovery of a potent inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis. *J. Med. Chem.* 49, 656–663

42 King, S. *et al.* (2005) Novel compounds. AstraZeneca AB. *Int. Appl.* WO 2003051277

43 Gasparini, F. *et al.* (1999) 2-Methyl-6-(phenylethynyl)-pyridine (MPEP), a potent, selective and systemically active mGlu5 receptor antagonist. *Neuropharmacology* 38, 1493–1503

44 Cosford, N.D. *et al.* (2003) 3-[(2-Methyl-1,3-thiazol-4-yl)ethynyl]-pyridine: a potent and highly selective metabotropic glutamate subtype 5 receptor antagonist with anxiolytic activity. *J. Med. Chem.* 46, 204–206

45 Pairaudeau, *et al.* (2005) Hit-to-Lead at AZ Charnwood. *IBC 'Hit to Lead Success Stories'*, 31 January – 1 February 2005, San Diego, CA, USA

46 Barta-Szalai, G. *et al.* (2004) Oxamides as novel NR2B selective NMDA receptor antagonists. *Bioorg. Med. Chem. Lett.* 14, 3953–3956

47 Borza, I. *et al.* (2003) Indole-2-carboxamides as novel NR2B selective NMDA receptor antagonists. *Bioorg. Med. Chem. Lett.* 13, 3859–3861

48 Coburn, C.A. *et al.* (2004) Identification of a small molecule nonpeptide active site β-secretase inhibitor that displays a nontraditional binding mode for aspartyl proteases. *J. Med. Chem.* 47, 6117–6119

49 Stachel, S.J. *et al.* (2004) Structure-based design of potent and selective cell-permeable inhibitors of human β-secretase (BACE-1). *J. Med. Chem.* 47, 6447–6450

50 Baxter, A. *et al.* (2003) Hit-to-lead studies: the discovery of potent, orally bioavailable triazolethiol CXCR2 receptor antagonists. *Bioorg. Med. Chem. Lett.* 13, 2625–2628

51 Vos, T.J. *et al.* (2004) Identification of 2-[2-[2-(5-bromo-2-methoxyphenyl)-ethyl]-3-fluorophenyl]-4,5-dihydro-1H-imidazole (ML00253764), a small molecule melanocortin 4 receptor antagonist that effectively reduces tumor-induced weight loss in a mouse model. *J. Med. Chem.* 47, 1602–1604

52 Gaillard, P. *et al.* (2005) Design and synthesis of the first generation of novel potent, selective, and *in vivo* active (benzothiazol-2-yl)acetonitrile inhibitors of the c-Jun N-terminal kinase. *J. Med. Chem.* 48, 4596–4607

53 Trottmann, G.H. *et al.* Hoffmann-La Roche. Amino-triazolopyridine derivatives. USP6355653

54 Pevarello, P. *et al.* (2005) 3-Aminopyrazole inhibitors of CDK2/cyclin A as antitumor agents. 2. Lead optimization. *J. Med. Chem.* 48, 2944–2956

55 Vasudevan, A. *et al.* (2004) Synthesis and evaluation of 2-amino-8-alkoxy quinolines as MCHr1 antagonists. Part 2. *Bioorg. Med. Chem. Lett.* 14, 4879–4882

56 Baxter, A. *et al.* (2004) Hit-to-lead studies: the discovery of potent, orally active, thiophenecarboxamide IKK-2 inhibitors. *Bioorg. Med. Chem. Lett.* 14, 2817–2822

57 Beaulieu, P.L. *et al.* (2004) Non-nucleoside inhibitors of the hepatitis C virus NS5B polymerase: discovery and preliminary SAR of benzimidazole derivatives. *Bioorg. Med. Chem. Lett.* 14, 119–124

58 Baxter, A. *et al.* (2003) Hit-to-Lead studies: the discovery of potent adamantane amide P2X7 receptor antagonists. *Bioorg. Med. Chem. Lett.* 13, 4047–4050

59 Maly, D.J. *et al.* (2000) Combinatorial target-guided ligand assembly: identification of potent subtype-selective c-Src inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 97, 2419–2424

60 Swayze, E.E. *et al.* (2002) SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. *J. Med. Chem.* 45, 3816–3819

61 Liu, G. *et al.* (2003) Selective protein tyrosine phosphatase 1B inhibitors: targeting the second phosphotyrosine binding site with non-carboxylic acid-containing ligands. *J. Med. Chem.* 46, 3437–3440

62 Szczepankiewicz, B.G. *et al.* (2003) Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy. *J. Am. Chem. Soc.* 125, 4087–4096

63 Shuker, S.B. *et al.* (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274, 1531–1534

64 Hajduk, P.J. *et al.* (1997) NMR-based discovery of lead inhibitors that block DNA binding of the human papillomavirus E2 protein. *J. Med. Chem.* 40, 3144–3150

65 Hajduk, P.J. *et al.* (1997) Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. Chem. Soc.* 119, 5818–5827

66 Jahnke, W. *et al.* (2003) Second-site NMR screening and linker design. *Curr. Top. Med. Chem.* 3, 69–80

67 Oltersdorf, T. *et al.* (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 435, 677–681

68 Liu, G. *et al.* (2001) Novel p-arylthio cinnamides as antagonists of leukocyte function-associated antigen-1/intracellular adhesion molecule-1 interaction. 2. Mechanism of inhibition and structure-based improvement of pharmaceutical properties. *J. Med. Chem.* 44, 1202–1210

69 Rath, V.L. *et al.* (2000) Human liver glycogen phosphorylase inhibitors bind at a new allosteric site. *Chem. Biol.* 7, 677–682

70 Nienaber, V.L. *et al.* (2000) Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotechnol.* 18, 1105–1108

71 Wang, X. *et al.* (2002) Design and synthesis of novel inhibitors of gelatinase B. *Bioorg. Med. Chem. Lett.* 12, 2201–2204

72 Faucher, A.M. *et al.* (2004) Discovery of small-molecule inhibitors of the ATPase activity of human papillomavirus E1 helicase. *J. Med. Chem.* 47, 18–21

73 Erlanson, D.A. *et al.* (2000) Site-directed ligand discovery. *Proc. Natl. Acad. Sci. U. S. A.* 97, 9367–9372

74 Hajduk, P.J. *et al.* (1999) Novel inhibitors of Erm methyltransferases from NMR and parallel synthesis. *J. Med. Chem.* 42, 3852–3859

75 Gill, A.L. *et al.* (2005) Identification of novel p38alpha MAP kinase inhibitors using fragment-based lead generation. *J. Med. Chem.* 48, 414–426

76 Nguyen, R. and Huc, I. (2001) Using an enzyme's active site to template inhibitors. *Angew. Chem. Int. Ed. Engl.* 40, 1774–1776

77 Hochgurtel, M. *et al.* (2002) Target-induced formation of neuraminidase inhibitors from *in vitro* virtual combinatorial libraries. *Proc. Natl. Acad. Sci. U. S. A.* 99, 3382–3387

78 Congreve, M.S. (2003) Detection of ligands from a dynamic combinatorial library by X-ray crystallography. *Angew. Chem. Int. Ed. Engl.* 42, 4479–4482

79 Erlanson, D.A. *et al.* (2003) In situ assembly of enzyme inhibitors using extended tethering. *Nat. Biotechnol.* 21, 308–314

80 Chu, K.C. (1980) The quantitative analysis of structure - activity relationships. In *Burger's Medicinal Chemistry. Part 1. The basis of medicinal chemistry* (4th ed.) (Wolff, M.E., ed.), pp. 393–418, Wiley