ELSEVIER

Integration of fragment screening and library design

Gregg Siegal[1,2], Eiso AB[2] and Jan Schultz[3]

[1] Leiden Institute of Chemistry, Leiden University, The Netherlands
[2] ZoBio, Leiden, The Netherlands
[3] Pyxis Discovery, Delft, The Netherlands

With more than 10 years of practical experience and theoretical analysis, fragment-based drug discovery (FBDD) has entered the mainstream of the pharmaceutical and biotech industries. An array of biophysical techniques has been used to detect the weak interaction between a fragment and the target. Each technique presents its own requirements regarding the fragment collection and the target; therefore, in order to optimize the potential of FBDD, the nature of the target should be a driving factor for simultaneous development of both the library and the screening technology. A roadmap is now available to guide fragment-to-lead evolution when structural information is available. The next challenge is to apply FBDD to targets for which high-resolution structural information is not available.

Introduction – advantages of fragment screening

The fragment-based approach to drug discovery (FBDD) has been in use for more than 10 years now [1]. During this time, significant energy has been invested in defining and optimizing collections of compounds ('fragments') and developing an array of techniques to screen the collections for ligands. This investment is beginning to pay off, as compounds derived (primarily) from FBDD are entering the clinic [2]. The underpinning logic of FBDD is simple; leads derived from high-throughput screens (HTS) of large corporate compound collections were achieving limited success in the clinic. Work by a number of groups (Lipinski [3], Hann and Oprea [4], and Wenlock *et al.* [5]) has led to a better understanding of the reasons for this failure and has had profound influence on the FBDD field. In summary, drug molecules should generally conform to the 'Rule of 5' as formulated by Lipinski *et al.* [6] to increase the likelihood of their having good PK (pharmacokinetic) and ADME (adsorption, distribution, metabolism, and excretion) properties. Working backwards, the best way to meet the Lipinski condition is to start with collections of small (less than 300 Da), highly soluble fragments that fit the 'Rule of 3' [7] (Figure 1).

In addition to maximizing the chance of remaining within Lipinski's limits, FBDD provides two additional benefits that can be of great importance. First, even a small (few thousands) collection of fragments covers a much greater proportion of all of the possible compounds that could exist, termed 'chemical space', than large (million) corporate compound collections for HTS [8]. Better coverage of chemical space derives from the fact that the number of possible compounds that can in principle be constructed is a very steep exponential function of the number of atoms from which they are constituted. The divergence between the number of compounds present in a collection and the number theoretically possible is minimized by limiting the size of molecules in the collection. At the same time, diversity can be maximized since fragment collections are often sampled combinatorially [9,1]. Greater chemical space coverage and the combinatorial approach provide a larger 'palette' of ligands to select from when evolving a hit to a lead. The increased choice allows one to avoid chemotypes which may be covered by existing intellectual property (IP), providing an important but little mentioned second benefit of FBDD. However, the small fragments typically interact so weakly with the target (1–2500 μM dissociation constant, K_D) that standard HTS technologies cannot readily be used to detect binding. Hence, FBDD depends heavily on advanced biophysical approaches to detect ligand binding.

For any screening campaign intended to identify small molecule modulators, the quality of the compound library will strongly

Corresponding author: Siegal, G. (g.siegal@chem.leidenuniv.nl)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2007.08.005

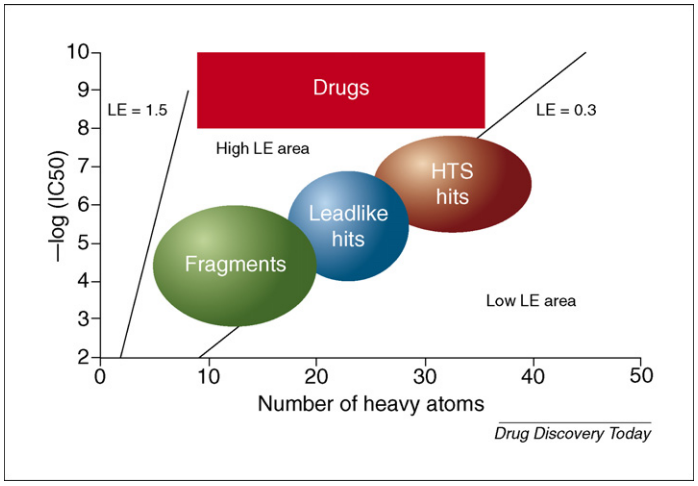

FIGURE 1

Hit-to-lead evolution for different starting points. The size of a compound (as measured by the number of nonhydrogen atoms) is compared to the affinity for the target. The Ligand Efficiency (LE) achievable in practice is shown with its slope of about 0.3 kcal/heavy atom, as well as the maximum theoretically available LE with a slope of about 1.5 kcal/heavy atom. The space defined by the 'Rule of 5' is shown in red. The lower right-hand corner of the 'Rule of 5' space represents a compound of approximately 500 Da with a potency of 10 nM. For a significant fraction of HTS hits, it is difficult to end up in the desired space, hence the move toward 'Leadlike' compounds which are smaller and more tractable. Fragments however, can be much smaller and therefore readily evolved toward 'Rule of 5' compliant drugs (figure based on ideas and data in Rees *et al.* [46], Hajduk [12], and Abad-Zapatero and Metz [47]).

influence the chance of finding a good starting point for further optimization. It is clear that this is true irrespective of the screening approach. The importance of quality became apparent during the 1990s with HTS of historical collections where the emphasis placed on screening sheer numbers of compounds yielded many false positives due to impure samples, reactive compounds, and aggregation of compounds in the screening collection [10]. Experience gathered during the intervening years has helped to better define the requirements for building a good compound collection that can form the basis of any new drug discovery and development program [4,11]. Below we try to summarize the demands of screening and some design principles useful for constructing a fragment library that will produce hits that can be readily evolved. We then provide an overview of current and possibly upcoming screening approaches and attempt to define the advantages and disadvantages of each.

Fragment library design

Choosing the upper and lower size limits

Two recent papers provide critical information that allows logical limits to be placed on both the maximal and minimal size of compounds that should be included in a fragment collection. In the first, a retrospective analysis of 18 different drug leads was performed by Hajduk [12]. By plotting the relationship between affinity and molecular mass (M_r) for the optimal fragment obtained at each stage of hit evolution, Hajduk determined an upper limit for this function of about 64 Da in increased mass per unit of pK_D (the $\log_{10}$ of the dissociation constant). This line has been plotted in Figure 1 (LE = 0.3) assuming that an 'acceptable' drug has maximal M_r of 500 and 10 nM affinity. Since fragments typically do not bind tighter than 10 μM, they should not be larger than 20 heavy atoms or about 300 Da. Furthermore, if during medicinal chemistry efforts the expected improvement in affinity per additional mass unit is not achieved, the compound is probably interacting with the target suboptimally. Colleagues at Abbott have suggested related rules of thumb for initial selection of hits for evolution [13].

The hit evolution process itself is based on rational modification of small molecules, using 3D structural information from target–ligand complexes. While abundant tales of fragments changing orientation on the target during hit evolution can be heard at FBDD conferences, a fascinating publication has now documented a fragment changing binding sites on the target. In experimentally deconstructing a moderate affinity compound, Shoichet and colleagues [14] found that one fragment, when stripped to a minimal binding core, bound a different site on the target than it did as part of the parent compound. This data suggests a lower limit of approximately 150 Da to minimize the chance that a fragment reorients on the target upon elaboration. Smaller, less complex fragments will only contain single rings with small substituents which therefore have a greater likelihood of binding in multiple orientations. Here we use the term complexity as defined by Xu [15] where it scales with the number of atoms and bonds. Indeed, the elegant theoretical approach of Hann *et al.* [16] has shown a relationship between the complexity of compounds in a library and the expected hit rate, which we find experimentally in our own screens (Figure 2). Schuffenhauer *et al.* reached the same conclusion in their analysis of actives identified at Novartis [17].

Taken together, these studies suggest that there are both upper and lower limits to the M_r of compounds that should be included in a fragment library. In addition to M_r, there are a number of other physicochemical properties that should be taken into account when designing a fragment library. Furthermore, the nature of

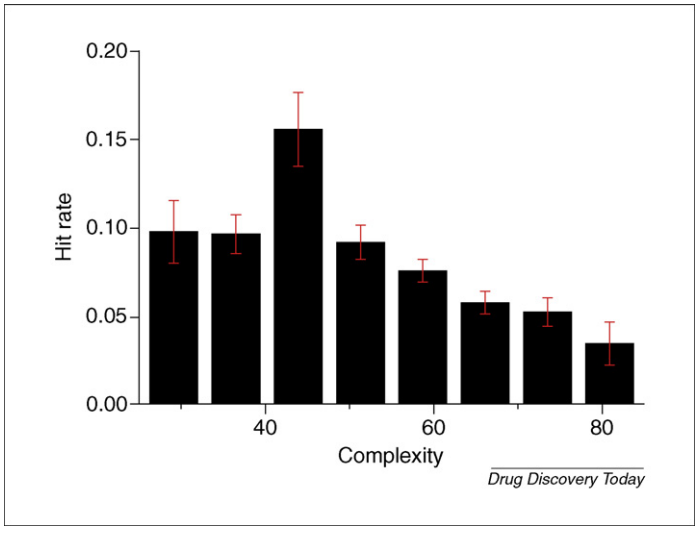

FIGURE 2

Plot of fragment complexity versus hit rate. Combined results of screens of five different targets against the ZoBio/Pyxis discovery fragment library are presented. The relationship between complexity and hit rate predicted by Hann *et al.* [16] is clearly seen. The complexity is defined as in Xu [15], where a larger number represents a more complex compound. A similar profile is observed if the M_r is plotted versus hit rate.

the intended screening method and even the types of targets to be screened need also be carefully considered.

Library design principles

When building a fragment library, the water solubility of the fragments will be of paramount importance, since they will be screened at high concentration (0.2–1.0 mM) in aqueous buffer. In one of the first publications concerning fragment library design, only about 50% of the selected fragments possessed sufficient solubility (1 mM) to be screened [18]. In more recent publications, better results for the water solubility of fragment libraries have been reported [19,20]. The prediction of water solubility remains a challenge because one has to take into consideration both the crystal and solution state of the compound. Moreover, in our own analysis, we have not been able to find a simple correlation between the number of hydrogen bond (H-bond) donors/acceptors and water solubility. Since computational methods for better prediction of water solubility are still under development, one must experimentally determine the solubility of a given fragment. However, by applying cut-off values based on experience, for properties that can be better predicted, such as cLogP and the number of hydrogen bond donors and acceptors, which have a profound influence on water solubility, the fraction of water-soluble fragments can be increased considerably. In our own efforts, about 90% of compounds that were selected were soluble as singletons at 500 μM in phosphate-buffered saline and 5% DMSO.

There is now more or less consensus that fragments should be 'Rule of 3' compliant. The 'Rule of 3', as proposed by Astex Therapeutics [7], is defined as: $M_r < 300$ Da, cLogP < 3, H-bond donors < 3, H-bond acceptors < 3, number of rotatable bonds < 3, and total polar surface area (TPSA) < 60 Å^2. In the light of previous considerations, we would also propose a lower limit of 150 Da for a general use, diversity library. Further, each fragment should preferably possess at least one chemical 'handle' that can be used for further elaboration. At the same time, this chemical handle should not be so reactive that it can form a covalent bond with a (target) protein. Therefore, selection filters should be applied, in order to avoid the inclusion of reactive compounds and known toxicophores in a fragment collection. Taking all these requirements into account, a database of 2.6 million commercially available compounds from 14 vendors was analyzed by removing duplicates, applying 60 selection filters [21] and applying the 'Rule of 3', which yielded 70,000 unique fragments (2.7%). However, it is important to realize that this analysis only addressed the relevant 'property space' of the fragments. In order to include fragments in the library with a good chance of binding to a (protein) target, a further selection strategy needs to be employed. One possible selection strategy to build a generic fragment library is to analyze current drugs, since they have already passed toxicity and ADME studies. Several approaches have been reported to 'fragment' active/drug molecules into smaller parts such as SHAPES [22], RECAP [23], and MS [24]. In most cases, databases such as CMC,[1] MDDR,[2] and WDI[3] with drugs or molecules in development are analyzed in order to identify interesting structural elements that can be used for library design. When applying this strategy, one must be aware of the fact that this is a retrospective analysis and that this approach is then biased toward what is already known. Another strategy is to look at the chemical tractability of the fragments. In this approach, the presence of a linker functionality is required and building blocks such as used in combinatorial chemistry would be suitable fragments [23,17]. A further elaboration of this strategy is to use masked linker groups, in which the building block in its native form is transformed into a screening fragment using the simplest possible reaction partner with minimal molecular weight [17]. Fragment libraries that target specific protein families (e.g. kinases) can also be developed, but we focus here on more generic, diversity-oriented collections. Finally, the size of the actual fragment library is important. The size will be determined, to some extent, by the throughput that can be accomplished and this varies according to the method. Information about the size of fragment libraries found in the literature and some of the companies involved in fragment-based screening has been summarized in Table 1.

In designing a fragment collection for our target-immobilized NMR screening (TINS, see 'Screening Technologies' for a description) technology, we took into consideration that TINS is meant to be applied to a broader range of targets, including membrane proteins, than solution-based methods. Accordingly, we are experimenting with incorporating new classes of compounds, all of which obey the 'Rule of 3'.[4] Specifically, we have chosen compounds according to four themes that include (1) diversity using the scaffold-based classification approach (SCA) [15], (2) amino acid derivatives, (3) scaffolds found in natural products, and (4) shape diversity. The members of each theme were selected from the pool of 70,000 fragments and represent about 25% of the total library.

Maximizing the diversity of a given library (within a biologically relevant chemical space) will ensure that active molecules from a given screen represent different chemotypes. This will help in prioritizing actives with respect to patentability. Therefore, the first 500 fragments of the collection were selected for maximal chemical diversity, using the SCA approach [15]. In this method, molecules are broken up into their constituent ring systems and side chains. Four structural descriptors are calculated to determine the 'complexity' of a scaffold. The descriptors are the maximum number of the smallest set of smallest rings, the number of heavy atoms, the number of bonds, and the sum of heavy atomic numbers. Subsequently, the ratio of 'ring atoms' to 'side chain atoms' is determined, resulting in a 'cyclicity' score for each molecule. Next, realizing that protein–protein interactions are based on amino acid side chain chemistry, it was decided that using structural motifs from the 20 natural occurring amino acids connected to small rings should yield an interesting generic, but at the same time biased, set of molecules. Since drugs and natural products have historically been a good source of biologically active compounds, this was used as a third approach. SCA was used to

[1] Comprehensive Medicinal Chemistry; available from Elsevier MDL, San Ramon, CA, USA.

[2] MDL Drug Data Report; available from Elsevier MDL, San Ramon, CA, USA.

[3] World Drug Index; available from Thomson Scientific, Alexandria, VA, USA.

[4] In order to include some desirable compounds in the amino acid theme, we did not apply an absolute cutoff to the number of rotatable bonds and the number of H-bond acceptors.

TABLE 1

Summary of screening methods and associated fragment libraries

Method	Sensitivity limit	Hit rate[a]	Pros	Cons	Library requirement	Library size (organization)
NMR protein detected (SAR by NMR)	1 mM K_D	Up to 1%	Low false positive rate 3D structure information	Large quantity of isotope labelled protein, M_r of target <40 kDa, low hit rate	Soluble in large mixes (>10)	17,000 (Abbott)[b] 20,000 (Combinature)
NMR ligand detected (STD/WATERLOGSY)	2 mM K_D	Up to 5%	Simple to setup, moderate to low target requirement, good throughput	Higher false positive, best with targets at least 20 kDa, cannot detect tight binders	No limitations	1200–1500 (Vernalis [19])
NMR competition binding	2–5 mM K_D	Up to 1%	Very low false positive	Focus on one binding site, must have ligand *a priori*	^{19}F labelled for finding competitor ligand	Up to 100,000 but only for enzymes (Nerviano Med. Sci. [31])
NMR TINS	10 mM K_D	Up to 10%	Efficient use of protein, screen protein in all states, low false positives, membrane proteins, uses reference protein	Must immobilize protein	Soluble at 500 μM each in mixes	2000 (ZoBio [29]/Pyxis Discovery)
SPR protein	0.5–1 mM K_D	0.5–1%	Highly efficient use of protein, kinetic data	Must immobilize compounds, protein must be soluble	Immobilized chemical array, fragment like	10,000 (Graffinity)
SPR compound	5 mM (K_i)	3–5%[c]	Highly efficient use of protein, kinetic data and K_D, uses reference protein	Must immobilize protein on Au surface	Fragment like, minimal M_r of compound 160 Da	2000 (Roche)
High concentration screening	NF	Up to 2%	Upfront knowledge of biochemical activity	Requires biochemical function (usually enzyme), false negatives, and false positives	125–350 Da	20,000 (Plexxikon [25])
Fluorescence (Thermofluor)	nM–μM K_D (possibly to 1 mM) [41]	NF	Enables search for novel protein modulators such as stabilizers	Sensitivity to weak binding. Requires up to 10 mg of pure, soluble protein	Lead like	300,000 (Johnson & Johnson [42])
Frontal affinity chromatography–mass spectroscopy (FAC–MS)	10 μM K_D – direct mode; 500 μM K_D – indicator mode [36]	NF	High-throughput, competition mode	Immobilize protein sensitivity, DMSO concentration/compound solubility, required known ligand as 'indicator'	Lead like, compound must be visible in MS	Potentially large
Affinity capillary electrophoresis (ACE)	nM–μM K_D	0.3%	High throughput	Target must be fluorescently labelled	Lead like	44,000 (Johnson & Johnson [38])
Crystallography	Up to 1 mM K_D	0.5–10%	3D structure info	Must have relatively high-resolution crystals, false negatives (occluded binding site), false positives, extremely resource intensive	Fragment like, up to 1000 compounds, SGX library contains 50% compounds with an aromatic bromine atom for crystallography and for the follow-up chemistry	850 (Astex [43]) 1000 (SGX [44])
Tethering	1–5 mM	0.10%	Sensitivity	Must generate surface exposed cysteine mutants of target	Compound must have SH group	7000 (Sunesis[45])

Note: NF – not found.

[a] Hit rate refers to a diversity library under 'standard' conditions for the method.
[b] Philip Hajduk, personal communication.
[c] Walter Huber, personal communication.

TABLE 2

Summary of physicochemical properties of the ZoBio/Pyxis fragment library

Property	Average SCA diversity approach	Average amino acid derivatives	Average drugs and natural products	Average shape diversity approach	Average for 2000 library
M_r	212.9	237.3	216.1	203.7	217.5
ClogP	1.7	0.4	1.3	1.2	1.2
TPSA[a] (Å^2)	47.2	59.7	51.2	46.4	52.6
Rotatable bonds	1.5	3.0	1.5	1.6	1.9
H-acceptors	2.2	3.0	2.8	2.3	2.7
H-donors	1.2	1.9	1.0	1.3	1.3

[a] Total polar surface area.

map the scaffolds of all orally available drugs and commercially available natural products. The pool of 70,000 fragments was searched to identify those that contained these scaffolds. Finally, an attempt was made to maximize the 'shape' diversity of the entire collection by comparing the Shape Fingerprints[5] of the pool of 70,000 fragments and those fragments already selected. The 500 fragments that best optimized the overall shape diversity of our collection were selected. Table 2 provides an overview of the calculated physicochemical properties for the library and per theme. An interesting question that we wish to answer by analysis of screening results is whether a given class of targets might exhibit selectivity toward one theme. This will be valuable information for the design of future fragment libraries.

Screening technologies

An array of methods has been elaborated to detect the weak interactions typically found between the members of a fragment collection and the target to be screened, almost all of which are biophysical approaches, aimed at directly detecting binding (Table 1). Naturally, one can screen fragment libraries for biologically active compounds, however due to the lack of potency an *in vitro* enzymatic assay is normally required [25]. For this reason we focus on more generally applicable biophysical screening methodologies and in particular, on the interface of screening techniques and compound collections. For a recent review focused exclusively on screening technologies see [26]. An important advantage of biophysical screening methods is that one knows from the outset that hits interact directly with the target in a reversible manner. Although NMR is not sensitive on a mass basis, it is exquisitely sensitive to intermolecular interactions. Unsurprisingly then, NMR formed the basis of the first FBDD screening approach, SAR by NMR [1]. Here the protein is isotopically labelled and compounds are added in mixes of up to 30 at a time, with each compound present at approximately 400 μM. The advantages of this patent-protected approach, which 10 years later remains a robust solution, include: very low false positive rates and 3D structural information on the ligand-binding site when the sequential assignment of the target is available. SAR by NMR has been used to develop a large number of high-affinity inhibitors, some of which have moved into the clinic [2]. Disadvantages of SAR by NMR include the fact that it is very resource-intensive, requiring at least 200 mg of labelled protein, which should be less than about 40 kDa and needs to be soluble in the presence of up to 12 mM total concentration of fragments. These requirements severely restrict the range of proteins to which SAR by NMR can be applied. In addition, the low hit rate necessitates a rather large library (typically more than 15,000 compounds).

In an effort to expand the range of targets to which FBDD could be applied, a number of NMR-based methods were introduced that detected binding via changes in the spectrum of the ligand. Collectively, this approach has the advantage that the target need not be isotopically labelled and the concentration can be lowered by 10–100-fold over the SAR by NMR approach. Since many of these methods detect changes in the NMR spectrum resulting from the increase in mass when a compound is bound to the target, they actually benefit from using large proteins. Furthermore, ligand-based approaches generally have 5–10-fold higher hit rates and, therefore, the library size can be reduced, although in some cases this may come at the expense of increased false positives. Amongst the variety of ligand-based methods, the saturation transfer difference (STD) method of Meyer [27,28] and the WATERLOGSY [28] method are the most commonly used in practice. Both methods rely on the selective transfer of magnetization to the compound upon target binding and are straightforward to implement. Clearly, the target must be soluble and not aggregate in order for these methods to give reliable results; however, the target should also be relatively large (>20 kDa) to achieve maximal sensitivity. Recently, we have also developed a ligand-based approach called TINS for target immobilized NMR screening [29,30]. TINS uses a single sample of immobilized protein to screen the entire compound library and, as such, makes extremely efficient use of the target, while exhibiting very low false positive rates, due to the use of a reference sample. Target immobilization also allows unstable forms of targets, such as those that tend to aggregate or self-destruct, as in uninhibited proteases, to be screened (Figure 3). Finally, all ligand-detected NMR experiments are compatible with competition mode screening in which displacement of a known ligand is used as the readout. Competition screening lowers the false positive rate and allows one to focus on a particular binding site on the target. The use of [^{19}F]-labelled 'spy molecules' has particular advantages of sensitivity and throughput in competition screening [31]. As a result of this sensitivity, exceptionally small amounts of an enzymatically active target can be used for screening.

The use of surface plasmon resonance (SPR) as a ligand-screening technique has gained increased attention due to significant technical advances. As with NMR, SPR can be implemented using modes observing either the ligands or the target. Biacore instru-

[5] Molecular Operating Environment; available from the Chemical Computing Group, Montreal, Quebec, Canada.

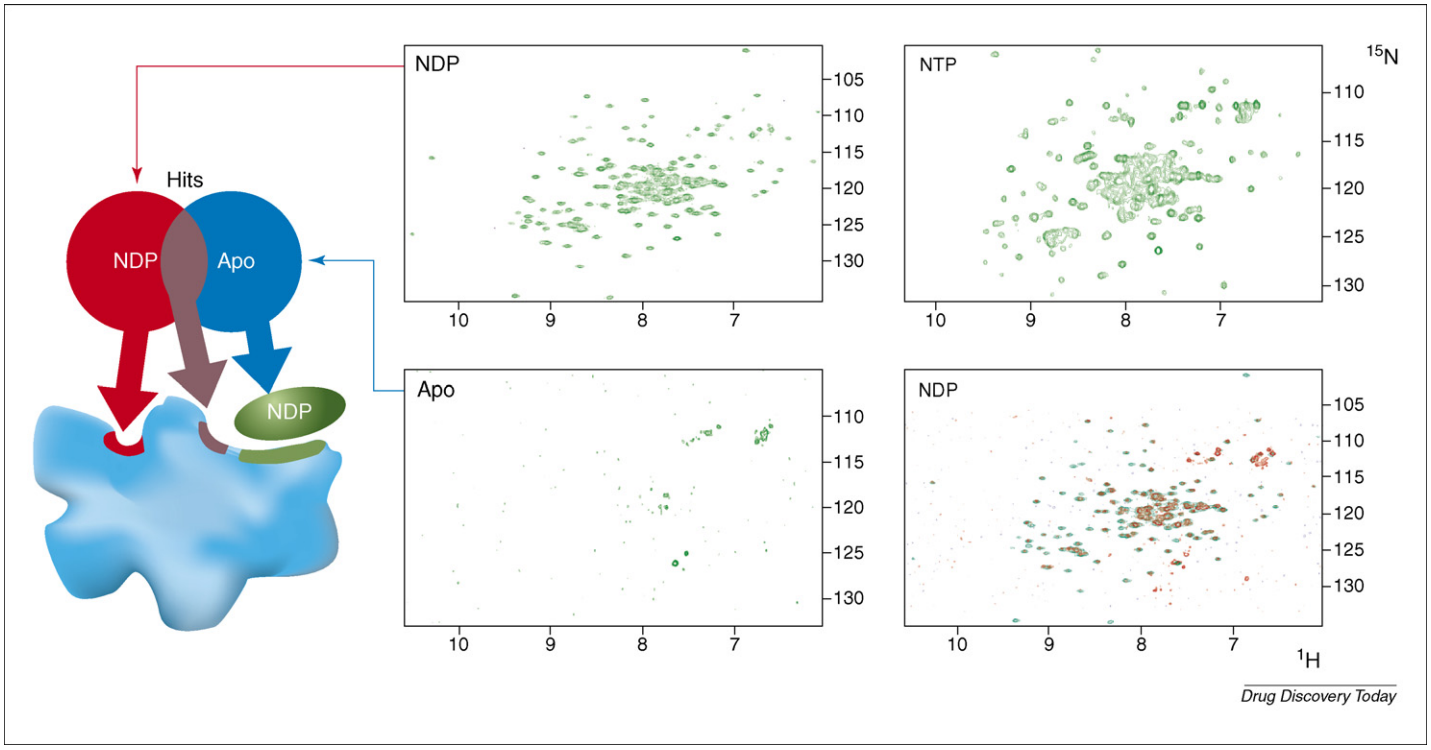

FIGURE 3

Illustration of one advantage of target immobilization. The four [^{15}N,^{1}H] Heteronuclear Single Quantum Coherence NMR spectra provide an overview of the solution state of the target. In the nucleotide diphosphate (NDP) bound state, the protein is monomeric and folded. The nucleotide triphosphate (NTP) bound state is monomeric but probably has floppy loops leading to weak aggregation. Removal of the nucleotide leads to large, soluble complexes that are essentially NMR invisible but can be reversed by subsequent addition of NDP (lower right, red spectrum reconstituted NDP–protein complex, green original). The aggregated state would lead to artefacts if screened by most solution techniques. Since 3D structures of all forms of the protein other than the NDP bound form would probably not be available, we chose to use differential screening to seek non-NDP competitive inhibitors. Ligands that bound exclusively to the apo protein could be classified as binding to the NDP site while those that bind exclusively to the complex with NDP likely bind a distal site induced by NDP binding. Hits common to both screens likely bind in a portion of the NDP site available for the γ phosphate of NTP. This type of analysis provides low-resolution structural information and allows one to choose which mechanism of inhibition one wishes to pursue.

ments flow single compounds over an immobilized target and detect binding by the increase of mass at the liquid–solid interface. Screening of fragment libraries of greater than 2000 members (150–350 Da) yields yes/no answers and must be followed by competition-binding experiments to confirm and rank affinities (W. Huber, personal communication). Graffinity (http://www.graffinity.com) has developed an instrument that uses SPR imaging to detect a soluble protein target binding to arrays of immobilized compounds. Having developed methods to reliably generate such chemical arrays, the method is routinely used to screen a 10,000-compound immobilized fragment library.

The use of high-throughput X-ray crystallography to screen fragment collections for target binding was pioneered by Astex Therapeutics (http://www.astex-therapeutics.com) and SGX Pharmaceuticals (http://www.sgxpharma.com). Clearly, the marriage of high-resolution structure and ligand screening is a powerful approach, but one that is also intensely resource demanding. Crystallography can be used to screen fragment-like libraries, but is typically limited to fewer than 1000 compounds [32]. The ligand-binding site of the target must also be accessible in the crystal (bearing in mind that the ligand-binding site may not be known *a priori*) and binding of the fragments should not disrupt the crystals. Supplementary biophysical forms of screening (NMR is common) are often employed to enhance the range of compounds that can be screened.

Tethering, through disulfide trapping of fragments to the target, has proven a powerful method of developing leads for soluble targets. The extremely high sensitivity of tethering, being able to detect binders with K_D greater than 5 mM, provides a significant advantage of the method. By modifying the first ligand to include the disulfide trap, second site ligands can be found which, when linked to the first, improve affinity and specificity [33]. An obvious limitation of tethering is the requirement that all compounds in the fragment library must contain an accessible SH group and additionally, a large number (up to 20) of cysteine mutants of the target must be produced.

Recently, three new biophysical methods have been proposed that show promise, but currently appear to be mostly applicable for the screening of lead-like libraries that contain larger compounds and bind the target with nM–μM affinity. The Thermofluor approach detects changes in the melting temperature of the target when a small molecule ligand binds to and stabilizes it. While it clearly can be used to screen libraries for ligands [34], a more innovative use is for deorphanization studies, or to find small molecules that can stabilize a particular protein. Frontal affinity chromatography (FAC) has also been proposed as a tool for ligand screening [35], albeit one that fits more into the HTS category. In FAC, the target is immobilized in a capillary tube and mixtures of compounds are applied. Binding is detected by an increase in the retention time of a ligand using mass spectroscopy.

Although it clearly is an affinity method that detects reversible binding, FAC is more appropriate for screening collections of 'larger' compounds ('Rule of 5' compliant), due to its sensitivity to binding affinities in the range of nM–μM [36]. However, FAC in 'indicator' mode, has been used to screen a library derived from docking studies where a 'Rule of 3' compliant, 3 μM ligand for the protein kinase EphB2 was found [37]. A closely related method, affinity capillary electrophoresis (ACE) has been used to screen a 44,000-compound library against a bacterial target [38]. Again, while ACE does detect reversible, affinity-based interactions with the target, the hits from this study were much more 'drug-like' than 'fragment-like'.

Biological validation of screening hits

Ultimately, a fragment will have to become part of a biologically active lead molecule. However, little is known about the correlation between binding, as determined by a biophysical assay, and biological activity. At Abbott Labs, a set of 699 SAR by NMR hits against 13 targets was followed up with *in vitro* testing at 200 μM. Of these, 85% exhibited IC50 values less than 100 μM (P. Hajduk, personal communication and [39]). In our own experience, we find that 50% of TINS hits are biochemically active at a concentration of 500 μM or lower, a result consistent with the Abbott results, considering that TINS is more sensitive to weak interactions. In both cases, the remaining hits are likely to be biochemically active at higher concentration. Finally, for the ACE assay described above, 75% of the screening hits were active antibacterials against at least one strain [38].

Perspectives

It is becoming clear that the combination of biophysical screening of fragment libraries with structure-guided hit evolution is a powerful and general method of developing high affinity, high specificity lead molecules. As such, it has become widely adopted. It is still too early to test one of the fundamental promises of FBDD, increased success in the clinic. With a number of FBDD-derived compounds now in phase 2 clinical trials, it seems our first glimpse of the ability to fulfil this promise is imminent. The common denominator and crucial aspect in all of these campaigns is the availability of 3D structural information [2].

A large fraction of current drugs target proteins that are located within, or are tightly associated with, the plasma membrane [40] and yet, these proteins are drastically under-represented within the Protein Data Bank (the repository of 3D structures). Membrane proteins remain a significant challenge because they are difficult to produce, isolate, and solubilize. Target immobilization offers a promising approach to overcome the limitations of ligand screening. Anecdotal reports suggest GPCRs can be immobilized using the same technology, while maintaining a high level of functionality, although they have not yet been used in ligand screening. Additionally, the Thermofluor method has also been reported to have successfully detected ligand binding to a membrane protein [41]. So, it appears that the barrier to screening membrane proteins for ligands may be resolved in the near future. The lack of experience in screening fragments against membrane proteins may require a further refinement of the selection filters to optimize the libraries. Once hits have been found, evolution is typically guided by structural information on target–hit complexes. In the case of membrane proteins however, 3D structural information is usually unavailable. This situation suggests that more resources need to be applied to finding innovative solutions for the structural biology challenges posed by membrane proteins.

Acknowledgements

The authors would like to thank Philip Hajduk of Abbott Labs and Walter Huber of Roche for personal communication of information. We wish to apologize to those whom, for reasons of brevity, have not been cited in this review.

References

1 Shuker, S.B. *et al.* (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274, 1531–1534
2 Hajduk, P.J. and Greer, J. (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.* 6, 211–219
3 Lipinski, C.A. (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* 44, 235–249
4 Hann, M.M. and Oprea, T.I. (2004) Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.* 8, 255–263
5 Wenlock, M.C. *et al.* (2003) A comparison of physiochemical property profiles of development and marketed oral drugs. *J. Med. Chem.* 46, 1250–1256
6 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25
7 Congreve, M. *et al.* (2003) A rule of three for fragment-based lead discovery? *Drug Discov. Today* 8, 876–877
8 Lipinski, C.A. and Hopkins, A. (2004) Navigating chemical space for biology and medicine. *Nature* 432, 855–861
9 Leach, A.R. *et al.* (2006) Fragment screening: an introduction. In *Structure-based Drug Discovery: An Overview* (Hubbard, R.A., ed.), pp. 430–446, RSC Publishing
10 Milne, G.M. (2003) Pharmaceutical productivity – the imperative for new paradigms. *Annu. Rep. Med. Chem.* 38, 383–396
11 Jacoby, E. *et al.* (2005) Key aspects of the novartis compound collection enhancement project for the compilation of a comprehensive chemogenomics drug discovery screening collection. *Curr. Top. Med. Chem.* 5, 397–411
12 Hajduk, P.J. (2006) Fragment-based drug design: how big is too big? *J. Med. Chem.* 49, 6972–6976
13 Metz, J.T. *et al.* (2007) Enhancement of chemical rules for predicting compound reactivity towards protein thiol groups. *J. Comput. Aid. Mol. Des.* 21, 139–144
14 Babaoglu, K. and Shoichet, B.K. (2006) Deconstructing fragment-based inhibitor discovery. *Nat. Chem. Biol.* 2, 720–723
15 Xu, J. (2002) A new approach to finding natural chemical structure classes. *J. Med. Chem.* 45, 5311–5320
16 Hann, M.M. *et al.* (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comp. Sci.* 41, 856–864
17 Schuffenhauer, A. *et al.* (2005) Library design for fragment based screening. *Curr. Top. Med. Chem.* 5, 751–762
18 Lepre, C.A. (2001) Library design for NMR-based screening. *Drug Discov. Today* 6, 133–140
19 Baurin, N. *et al.* (2004) Design and characterization of libraries of molecular fragments for use in NMR screening against protein targets. *J. Chem. Inf. Comp. Sci.* 44, 2157–2166
20 Jacoby, E. *et al.* (2003) Design of small molecule libraries for NMR screening and other applications in drug discovery. *Curr. Top. Med. Chem.* 3, 11–23
21 Verheij, H.J. (2006) Leadlikeness and structural diversity of synthetic screening libraries. *Mol. Divers.* 10, 377–388
22 Fejzo, J. *et al.* (1999) The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.* 6, 755–769

23 Lewell, X.Q. *et al.* (1998) RECAP – retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comp. Sci.* 38, 511–522
24 Vieth, M. *et al.* (2004) Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.* 47, 224–232
25 Card, G.L. *et al.* (2005) A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design. *Nat. Biotechnol.* 23, 201–207
26 Bartoli, S. *et al.* (2006) The fragment-approach: an update. *Drug Discov. Today: Technol.* 3, 425–431
27 Mayer, M. and Meyer, B. (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chem. Int. Ed.* 38, 1784–1788
28 Dalvit, C. *et al.* (2000) Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J. Biomol. NMR* 18, 65–68
29 Marquardsen, T. *et al.* (2006) Development of a dual cell, flow-injection sample holder, and NMR probe for comparative ligand-binding studies. *J. Magn. Reson.* 182, 55–65
30 Vanwetswinkel, S. *et al.* (2005) TINS, target immobilized NMR screening: an efficient and sensitive method for ligand discovery. *Chem. Biol.* 12, 207–216
31 Dalvit, C. *et al.* (2005) Rapid NMR-based functional screening and IC50 measurements performed at unprecedentedly low enzyme concentration. *Drug Dev. Res.* 64, 105–113
32 Carr, R.A.E. *et al.* (2005) Fragment-based lead discovery: leads by design. *Drug Discov. Today* 10, 987–992
33 Choong, I.C. *et al.* (2002) Identification of potent and selective small-molecule inhibitors of caspase-3 through the use of extended tethering and structure-based drug design. *J. Med. Chem.* 45, 5005–5022
34 Cummings, M.D. *et al.* (2006) Universal screening methods and applications of ThermoFluor (R). *J. Biomol. Screen.* 11, 854–863
35 Ng, W. *et al.* (2007) Automated multiple ligand screening by frontal affinity chromatography–mass spectrometry (FAC–MS). *J. Biomol. Screen.* 12, 167–174
36 Slon-Usakiewicz, J.J. *et al.* (2005) Frontal affinity chromatography with MS detection (FAC–MS) in drug discovery. *Drug Discov. Today* 10, 409–416
37 Toledo-Sherman, L. *et al.* (2005) Frontal affinity chromatography with MS detection of EphB2 tyrosine kinase receptor. 2. Identification of small-molecule inhibitors via coupling with virtual screening. *J. Med. Chem.* 48, 3221–3230
38 Lewis, L.M. *et al.* (2004) Affinity capillary electrophoresis for the screening of novel antimicrobial targets. *J. Biomol. Screen.* 9, 303–308
39 Stamper, G.F. *et al.* (2006) Structure-based optimization of MurF inhibitors. *Chem. Biol. Drug Des.* 67, 58–65
40 Overington, J.P. *et al.* (2006) Opinion – how many drug targets are there? *Nat. Rev. Drug Discov.* 5, 993–996
41 Pantoliano, M.W. *et al.* (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screen.* 6, 429–440
42 Grasberger, B.L. *et al.* (2005) Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells. *J. Med. Chem.* 48, 909–912
43 Hartshorn, M.J. *et al.* (2005) Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* 48, 403–413
44 Blaney, J. *et al.* (2006) Fragment-based lead discovery and optimization using X-ray crystallography, computational chemistry, and high-throughput organic synthesis. In *Fragment-based Approaches in Drug Discovery* (Jahnke, W., Erlanson, D.A., Mannhold, R., Kubinyi, H., Folkers, G., eds), pp. 215–248, Wiley–VCH
45 Erlanson, D.A. *et al.* (2006) Tethering. In *Fragment-based Approaches in Drug Discovery* (Jahnke, W., Erlanson, D.A., Mannhold, R., Kubinyi, H., Folkers, G., eds), pp. 285–312, Wiley–VCH
46 Rees, D.C. *et al.* (2004) Fragment-based lead discovery. *Nat. Rev. Drug Discov.* 3, 660–672
47 Abad-Zapatero, C. and Metz, J.T. (2005) Ligand efficiency indices as guideposts for drug discovery. *Drug Discov. Today* 10, 464–469