This description of the fragment library and approach of AstraZeneca to fragment-based lead generation shows that 2D and 3D fragments provide complementary hits to explore binding pockets, and that both can deliver 3D lead series.

An improved model for fragment-based lead generation at AstraZeneca

Nathan Fuller[1,5], Loredana Spadola[2,5], Scott Cowen[1], Joe Patel[1], Heike Schönherr[1], Qing Cao[1], Andrew McKenzie[1,6], Fredrik Edfeldt[3], Al Rabow[4] and Robert Goodnow[1]

[1] AstraZeneca, Discovery Sciences, Innovative Medicines and Early Development Biotech Unit, 35 Gatehouse Drive, Waltham, MA 02451, USA
[2] AstraZeneca, Discovery Sciences, Innovative Medicines and Early Development Biotech Unit, Mereside, Alderley Park, Macclesfield SK10 4TG, UK
[3] AstraZeneca, Discovery Sciences, Innovative Medicines and Early Development Biotech Unit, Pepparedsleden 1, Mölndal 431 83, Sweden
[4] AstraZeneca Oncology, Innovative Medicines and Early Development Biotech Unit, AstraZeneca, Mereside, Alderley Park, Macclesfield SK10 4TG, UK

Modest success rates in fragment-based lead generation (FBLG) projects at AstraZeneca (AZ) prompted operational changes to improve performance. In this review, we summarize these changes, emphasizing the construction and composition of the AZ fragment library, screening practices and working model. We describe the profiles of the screening method for specific fragment subsets and statistically assess our ability to follow up on fragment hits through near-neighbor selection. Performance analysis of our second-generation fragment library (FL2) in screening campaigns illustrates the complementary nature of flat and 3D fragments in exploring protein-binding pockets and highlights our ability to deliver fragment hits using multiple screening techniques for various target classes. The new model has had profound impact on the successful delivery of lead series to drug discovery projects.

Introduction

To remain successful, lead generation teams in pursuit of quality chemical starting points in drug discovery must enhance and diversify the ability to identify chemical equity against important targets. Whereas a high-throughput screening (HTS) approach typically results in lead compounds similar to those identified in the screen, a fragment-based lead generation (FBLG) approach provides structurally simple starting points that evolve via a more creative or innovative path, leading to

Corresponding authors: Fuller, N. (nathan@rodintherapeutics.com), Spadola, L. (loredana.spadola@astrazeneca.com)
[5] Current address: Rodin Therapeutics, 400 Technology Square, 10th Floor, Cambridge, MA 02139, USA.
[6] Current address: Moderna Therapeutics, 200 Technology Square, Cambridge, MA 02139, USA.

Nathan Fuller joined the FBLG team at the Chemistry Innovation Center of AZ in 2013, where he has applied the principles of medicinal chemistry and fragment-based drug discovery to projects in a variety of target classes and therapeutic areas. He received his PhD in 2005 from the University of North Carolina at Chapel Hill working under James Morken. After conducting postdoctoral research with Stephen F. Martin at the University of Texas at Austin, he began his career as a medicinal chemist in 2007 at Wyeth Research. In 2009, he joined Satori Pharmaceuticals, where he worked on gamma-secretase modulators as therapeutics for Alzheimer's disease. He is currently the Director of Chemistry at Rodin Therapeutics, Cambridge, MA, USA.

Loredana Spadola is a computational chemist in the Discovery Sciences Innovative Medicines and Early Development Biotech Unit at AstraZeneca (AZ). She received her PhD from the University of Naples 'Federico II' in 2002 and joined AstraZeneca shortly after. Since the start of her career, she has worked in fragment-based lead generation (FBLG) in several disease areas, including respiratory and inflammation, infection and, most recently, oncology. She has been responsible for the development and maintenance of the AZ global fragment library since 2006. Her interests include molecular similarity, structural chemistry and library design.

Joe Patel is a FBLG specialist who joined the Structure & Biophysics Group at AZ in 2010, where he has continued to champion the use of FBLG across therapeutic areas, particularly the utility of X-ray cocktail screening to characterize ligand-binding pockets. Before working at AZ, he received his PhD in 2001 from the University of Cambridge under the advisement of Tom Blundell. He began his career as a structural biologist in 2001 at Astex Pharmaceuticals, a biotechnology company focused on using fragment-based methodology, and contributed to multiple programs that are currently in late-stage clinical studies.

http://dx.doi.org/10.1016/j.drudis.2016.04.023

series that are under-represented or nonexistent in the current screening deck. This attribute makes FBLG well suited to unprecedented target classes, providing chemical equity where no prior art exists. Thus, FBLG not only adds both novelty and flexibility to overall lead generation capabilities, but also increases the overall probability of success in lead generation.

Since the concepts of FBLG were first introduced [1,2], this strategy has become widely practiced and acknowledged as a powerful approach to creating high-quality drug-like molecules in the quest to invent new pharmaceutical therapeutics [3–5]. In 2007, AZ reported on the application of FBLG to its internal drug discovery projects [6]. Included was a discussion on the creation of the first dedicated fragment library used for screening, referred to in the previous publication as 'GFSL05,' but which we term here 'Fragment Library 1' (FL1). That retrospective analysis, along with subsequent internal evaluations on performance of FBLG within AZ over the ensuing years, prompted us to modify and improve the strategy behind the application of FBLG. As a result, we have implemented new technologies, improved practices and utilized learnings from practical experience to develop a highly standardized approach that we consider to be a model for FBLG Best Practice at AZ.

Recognizing a need for change: FBLG at AZ from 2002 to 2014

In an evaluation of internal lead generation capabilities, AZ conducted an analysis of all fragment screens conducted between 2002 and 2011. Efforts were considered successful if screening and subsequent optimization delivered a lead-like series with progressable structure–activity relationship (SAR). During the period from 2002 to 2008, the success rate of 63 FBLG campaigns at AstraZeneca (AZ) was a meager 10% (Fig. 1). Some key determinants for success were found to be robust crystallography systems along with commitment from a project team to engage in chemistry on fragment hits. A fragment approach had often been utilized as a rescue attempt for particularly challenging targets, thus decreasing the probability of success. Learning from this experience, workflows were modified and upfront investment was made in determining the tractability of crystallographic systems for FBLG to provide confidence that fragments could be optimized using structure-based design. This resulted in a marked improvement in performance, as in the period from 2009 to 2011, when successful delivery of lead series improved to 37% over the course of 19 fragment-screening campaigns. The decision to only progress programs with robust structural systems was a significant factor. Importantly, there were no failures because of the lack of further chemistry optimization from the project team. Still, there was room for improvement, and operational and organizational changes were made with this goal in mind. The fragment portfolio during 2009–2011 began to shrink compared with previous years, as the overall global drug discovery portfolio decreased. Still, nearly half of the projects were closed because of failed target validation or strategic portfolio reasons. Failures for these reasons have decreased in recent years with more rigorous target selection within the AZ 5R framework [7].

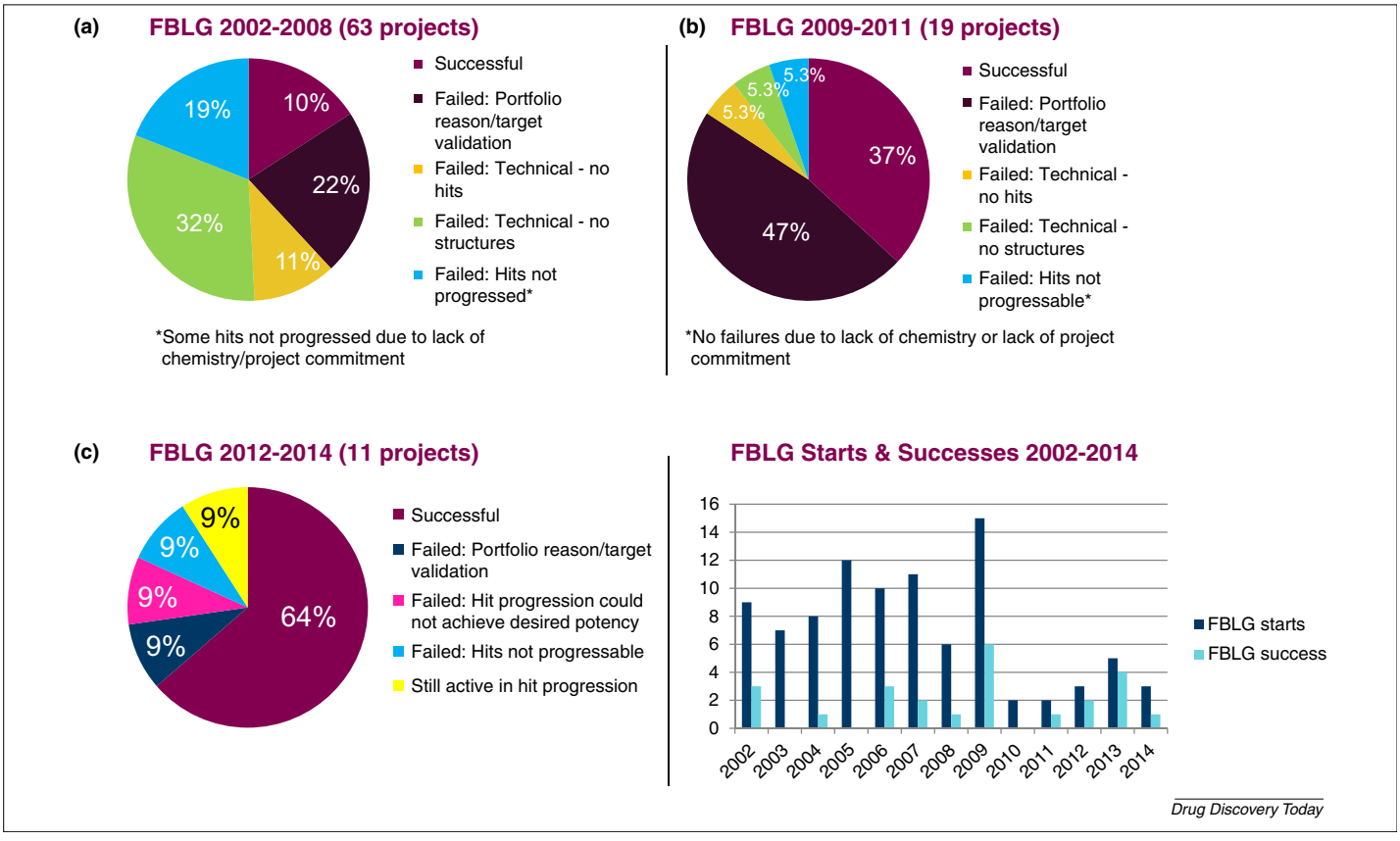

FIGURE 1

Success rates in the delivery of lead series to drug discovery projects from fragment screening and reasons for fragment-based lead generation (FBLG) failures at AstraZeneca (AZ) between **(a)** 2002 and 2008, **(b)** 2009 and 2011 and **(c)** 2012 and 2014.

Historically, and at the time of the 2007 report [6], FBLG activities were widely dispersed throughout the research and development (R&D) organization at AZ, a situation that limited the potential impact that fragment-based approaches might have on project success rates. To improve the rate of success in fragment chemistry efforts, the need for a dedicated team of medicinal chemists was identified. This specialized group would build expertise in the practice of evolving low-affinity fragment hits (Kds in the range of 100 μM to mM) into lead series. In 2012, a dedicated chemistry team was established to support all fragment-based efforts across the portfolio of AZ and tasked with generating quality lead-like series from any fragment-screening output. This ensured that no projects would fail because of the lack of chemistry support. As a result, since the inception of the fragment chemistry team in 2012, lead series delivery has been achieved on 64% of FBLG campaigns, a significant improvement over the previous 10-year period. The following description of the composition of the AZ fragment library and analysis of the AZ screening campaigns describes how FBLG teams at AZ have established an approach that has led to the observed improvement in project delivery and is well positioned for continued success.

Contemporary FBLG practices at AZ

Currently at AZ, FBLG is considered a viable hit-finding strategy for all water-soluble targets. As an integral part of the lead generation toolbox of AZ, it is considered as a complement or alternative to other hit-finding methodologies, such as HTS or encoded library screening [8]. Fragment campaigns primarily involve structurally enabled targets, but with the dedicated fragment chemistry team and strengthened biophysical capabilities, we are willing to use an FBLG approach for high-interest targets in the absence of structure when a robust suite of biophysical techniques provides confidence through orthogonal hit validation and clear SAR. From 2002 to 2011, most fragment screens were performed as either nuclear magnetic resonance (NMR) screens or high-concentration biochemical screens (HCS). Although NMR remains a preferred method of fragment-binding detection in both the ligand and protein observed configurations, technological and throughput advances in surface plasmon resonance (SPR)-based systems have elevated this screening approach as an alternative primary screening option. We consider NMR to be the most sensitive screening method. 2D NMR is the gold standard fragment assay, but is limited to targets of modest size. 1D NMR assays can be established for almost all targets and hits can be detected at concentrations well below the Kd. The weaknesses of both 1D and 2D NMR are the low throughput and high protein consumption. The main strength of SPR is the powerful combination of high throughput with accurate Kd determinations, making SPR particularly effective when screening near neighbors. That said, SPR is more prone to assay artifacts and is sensitive to poorly behaving compounds. The use of X-ray cocktail screening is another recent addition to the FBLG approach, but generating robust crystal systems capable of supporting this method remains challenging and it is difficult to predict the likelihood of success at project outset. The obvious benefit of X-ray screening is that the fragment-binding pose of any hit is available to enable rapid fragment elaboration and chemistry design without the need for extensive near-neighbor screening

In addition to a dedicated team of medicinal chemists focused on fragment series elaboration and optimization, an FBLG project team comprises a structural biologist, biophysicist and computational chemist, and might also include an *in vitro* assay screening scientist. We believe that it is advantageous to have computational chemists who are experienced in FBLG embedded in the project team to apply state-of-the-art methods based on specific project needs and lead generation plans, including but not limited to: screening result analysis and clustering, 2D fingerprint and 3D shape-based similarity searching, SAR analysis, pharmacophore modeling, virtual screening, and docking and other structure-based design technologies.

Updating the AZ fragment library

The AZ fragment library has undergone considerable transformation in the years since FL1 was first described [6,9]. This was driven in part by a shift in focus from using high-throughput HCS of fragments to lower-throughput but more information-rich techniques of X-ray, NMR and SPR as the principal primary screening methods. When using HCS, there had been wide variation in the size and composition of the screening libraries on a case-by-case basis and the concentration at which the fragments were screened. HCS also carried with it the limitations of higher rates of false negatives and false positives because of protein aggregation and assay interference. With our change in screening philosophy, it became important to adapt the ~20,000-member generic FL1 to these new practices so that it was fit for purpose and our approach was more standardized.

To reconfigure the fragment library, we first removed any compounds prone to decomposition and assay interference, and next eliminated compounds that did not meet our new criteria for aqueous solubility of >500 μM as measured using nephelometry [10]. Our focus on X-ray, NMR and SPR as principal screening methods made high solubility a crucial parameter for members of the fragment library. Next, a team of medicinal chemists flagged for removal reactive fragments, structures deemed unattractive from a follow-up chemistry perspective, and undesirable chemical functionalities (e.g. Michael acceptors, thioureas, aldehydes, among others) contained in FL1. To rebuild the fragment library composition to a larger size, we initiated a program to design and synthesize novel fragment libraries to expand and enhance the collection. Compound designs were inspired by new synthetic methods emerging in the literature, novel building blocks designed and synthesized internally, identification of chemical scaffolds and functional groups under-represented in the collection, and analysis of gaps in chemical shape space as measured by principal moments of inertia (PMI) [11] and shape fingerprints [12]. In addition, the AZ compound collection was mined to identify under-represented chemotypes and ring systems. These combined efforts resulted in the second-generation fragment-screening library, FL2, which has 15,000 members. In general, these compounds comply with the fragment 'Rule of Three' [13], but this principle was utilized as a loose guideline rather than a rule, with pharmacophore representation and structural diversity being given priority. Refreshing the library also provided the opportunity to introduce proprietary fragments into the collection, and these now represent approximately 25% of FL2.

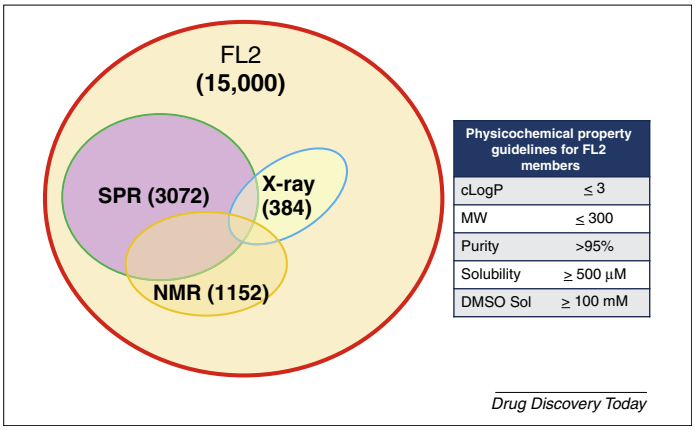

FIGURE 2

Graphical depiction of the composition of the AstraZeneca fragment library FL2. *Abbreviations*: NMR, nuclear magnetic resonance; SPR, surface plasmon resonance.

With the composition of the fragment library updated and enhanced, FL2 was organized into technology-appropriate subsets, sized to be commensurate with the principal screening techniques: X-ray, NMR and SPR (Fig. 2). Emphasis was placed on structural diversity, pharmacophore representation, physicochemical properties and good coverage of chemical space. There is an intentional degree of overlap between the three screening subsets to ensure that some compounds are screened with two different methods to assist in hit confirmation and engender confidence in a screening technique for a particular target. The design principles used in building the three screening subsets were slightly different based on the technique, and are discussed below. The property profiles of the complete FL2, the fragment-screening subsets, and the hits corresponding to each library are shown in Fig. 3.

An additional application of fragment screening at AZ is 'ligandability' screening to evaluate the probably tractability of novel targets and target classes to standard lead-generation approaches, especially where no chemical equity has been previously disclosed in the literature. The ligandability fragment set is typically screened using NMR and/or SPR. On the basis of the hit rate, affinity and diversity of hits, the target is assigned a ligandability score of low, medium, or high. The results of ligandability screens are used to inform the lead-generation strategy for projects. AZ has demonstrated the value of this strategy for many years and has previously reported on it [14].

X-ray screening subset

The X-ray screening subset comprises 384 compounds that are screened in soaking experiments. As a small set of compounds, it is designed to sample chemical space efficiently and, thus, contains small fragments with heavy atom count (HAC) <17, and the lowest average HAC of any subset (13). To make our selections for this subset, we first identified all the fragments with HAC <17 that had ever been successfully crystallized in a protein target at AZ. The molecules were then clustered using ECFI fingerprints [15] and representatives were chosen from each cluster to achieve maximum structural diversity. The compounds are organized in 96 cocktails of four, selected to maximize the shape diversity of the compounds using the principles of shape fingerprints [12].

NMR screening subset

The NMR screening subset comprises 1152 compounds and is screened by either 1D- or 2D-NMR. The compounds were selected to maximize the structural diversity and are screened in cocktails of between four and 12 compounds. Included is the 768-member 'ligandability set', the library often screened during the target identification phase to establish the ability of a protein target to bind small molecules with reasonable affinity ('ligandability') [2].

SPR screening subset

The SPR screening subset contains 3072 compounds and also includes the 768-member 'ligandability set,' providing a good degree of overlap with the NMR subset. Members of the SPR subset have higher median HAC to account for the lower sensitivity for binding detection in SPR techniques. The library was designed with small clusters of close structural analogs around most fragments (three to five compounds per cluster). This is beneficial for facilitating hit confirmation by building confidence in a hit when similar structures are also identified as hits, highlights emerging SAR, and can also help to recognize any false positives. All members of the SPR subset are subjected to a nonspecific binding clean screen to confirm that no members of the library interfere with the screening technique by nonspecifically binding to the control protein or chip surface [16].

In our continuous efforts to explore and evaluate new and emerging approaches to FBLG, we recently constructed a library of covalent fragments that is directed toward cysteine proteases and other targets with reactive residues in the binding pocket for which there is literature precedent and clear benefit to using covalent modulation [17,18]. In addition to structural diversity, the library features a range of covalent warheads exhibiting a wide degree of electrophilicity. A glutathione reactivity assay was used to ensure compounds were within an appropriate reactivity range to engage a specific nucleophilic residue without causing uncontrolled side reactions ($10 < t_{1/2} < 3000$ min). This library of 800 compounds stands alone from FL2 and is not included in standard fragment screens. Given that the covalent fragment library has been established only recently, it is not included in the fragment-screening analysis in this article. Screening data will be collected on this library and the results and analysis reported in future publications.

Ability to follow up on fragment hits

An important factor in the success of any FBLG campaign is the ability to rapidly follow up on primary hits by screening close structural relatives (near neighbors) that are available from either the internal compound collection or commercial sources to explore the SAR. Evaluating a series of minor changes around initial fragment hits provides a better understanding of the binding pocket and allows for optimization of the fragment core scaffold before growing or merging activities. Although the highly curated FL2 comprises 15,000 fragments, a major asset in our ability to successfully prosecute FBLG campaigns is the AZ global compound collection, which contains over 750,000 additional molecules that can be classified as fragments from which similar compounds can be selected for screening. For the purposes of the following analysis, we refer to this lower-molecular-weight portion of the corporate compound collection as 'AZ-Frag.' To assess our ability to

follow up on fragment hits with related chemical equity, we analyzed the number of near neighbors available to all fragment components of the screening subsets (X-ray, NMR and SPR) in different compound collections, FL2, AZ-Frag and fragments from the commercial Sigma-Aldrich database (http://www.sigmaaldrich.com/united-states.html; SA-Frag) using two different methods (Fig. 4). Fragments from SA-Frag were considered to be any molecule in the Sigma-Aldrich database with molecular weight less than 300 Da.

The first method calculated the fingerprint-based Tanimoto distance between the query molecule and the molecules in different libraries. We used the in-house Foyfi fingerprint method [9]

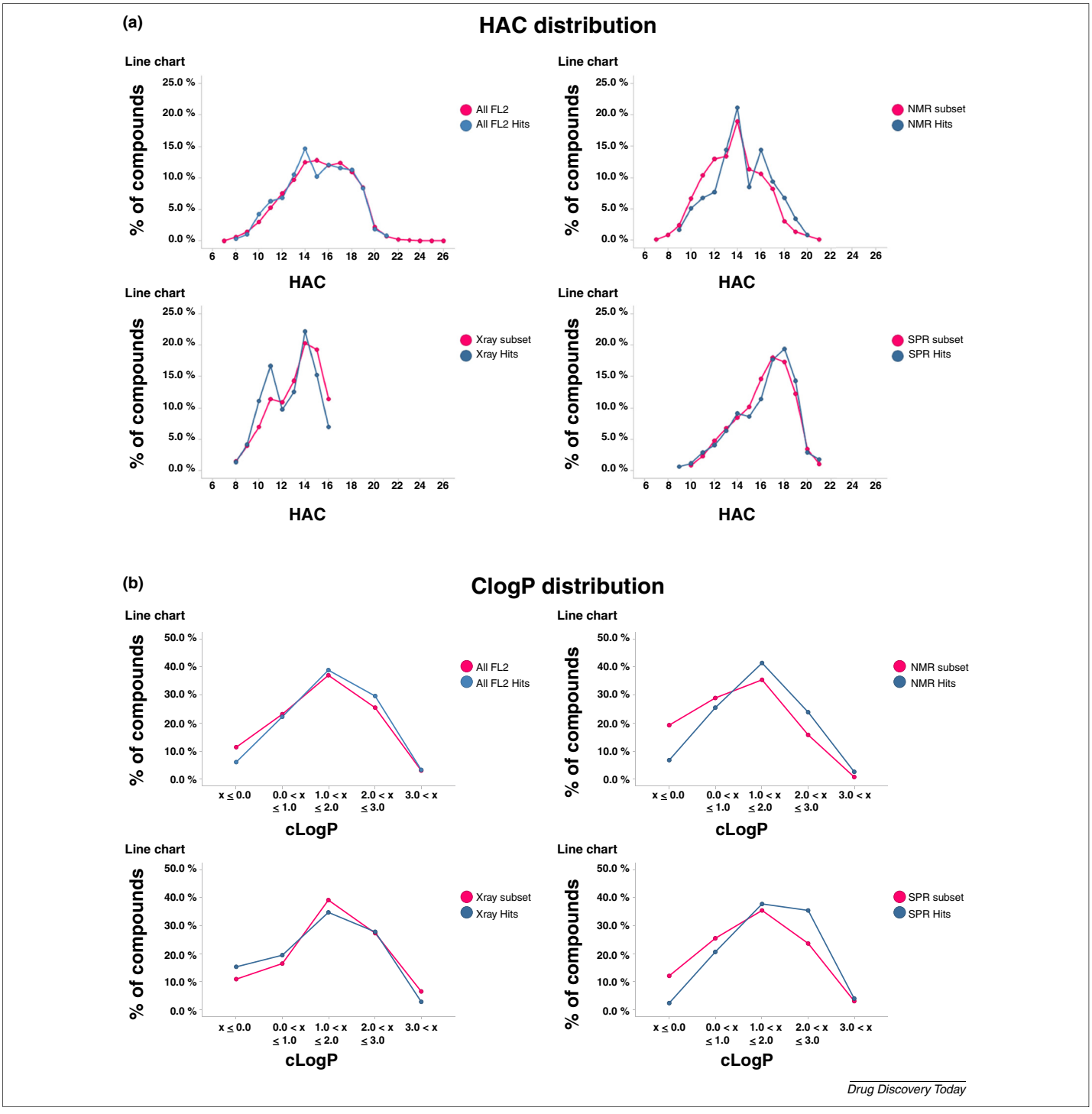

FIGURE 3

Property comparison between AstraZeneca fragment-screening hits and library subsets with values normalized by taking the percentage of compounds at each data point in relation to the number of hits from each library or subset and comparing them with the percentage of overall compounds at that data point in the overall library or subset. **(a)** Heavy atom count (HAC) distribution. **(b)** *c* Log *P* distribution. **(c)** Plane of best fit (PBF) distribution. **(d)** Principle moment of inertia (PMI) distribution.

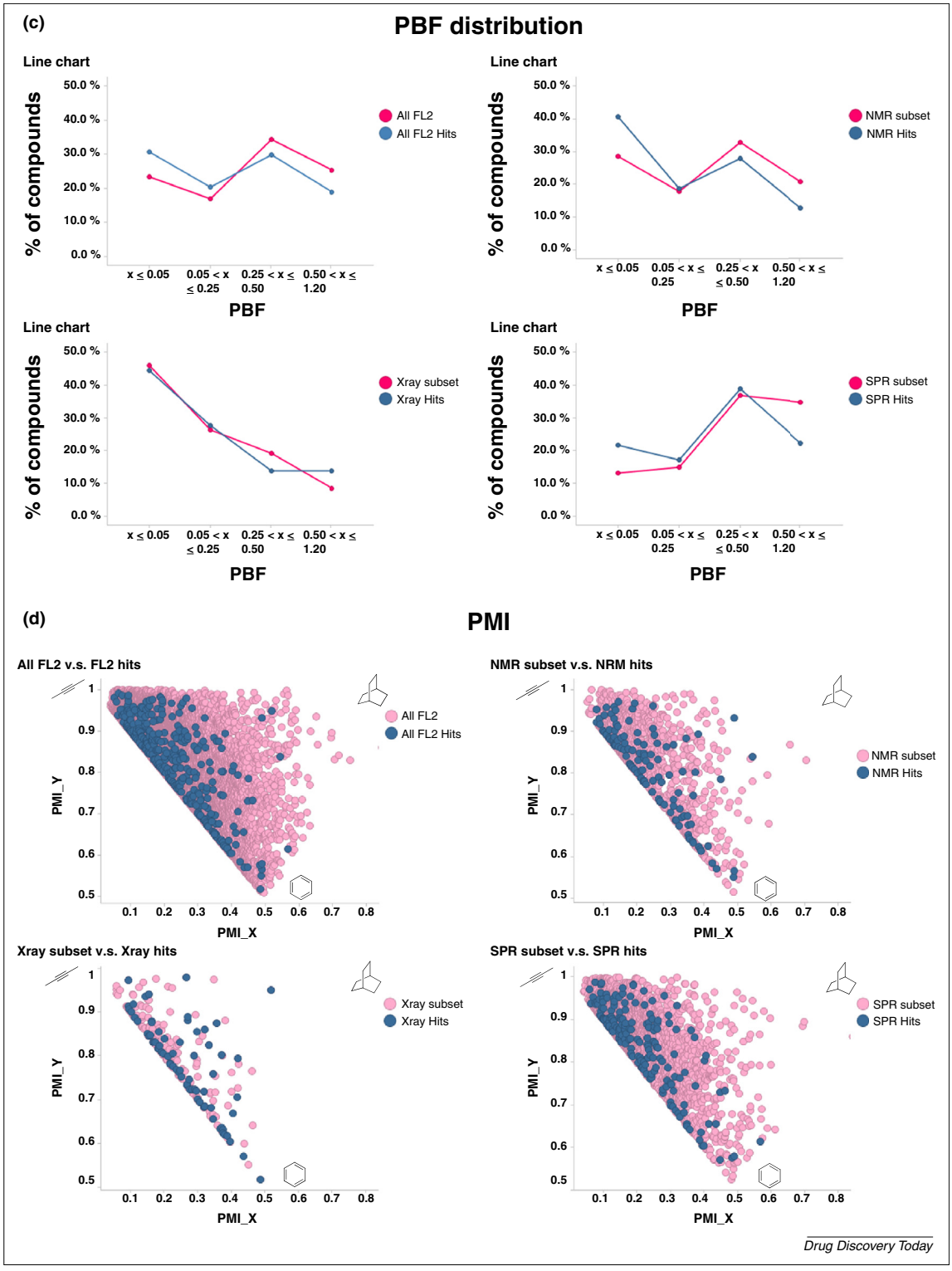

FIGURE 3 (*Continued*).

and considered two compounds with Tanimoto distance below 0.3 to be similar. This threshold was selected after visual analysis of molecules retrieved at different Tanimoto distances up to 0.4.

The second method used to identify near neighbors from a database was based on matched molecular pairs (MMP) [9]. MMP are compounds that differ only by a single atom or R-group, where an R-group is defined as a maximum eight heavy atoms separated by a single bond from the rest of the molecule. We recently developed an in-house tool, MATCH, that exhaustively retrieves all MMP of a given compound available in the AZ compound collection. To ensure that the identified MMP were relevant analogs to the fragment hits, we restricted the database search to fragment-like compounds with a maximum HAC of 22 and $c\,\mathrm{Log}\,P \leq 3$, thus avoiding compounds that had a very high molecular weight compared with the initial fragment, and were not relevant to our objectives.

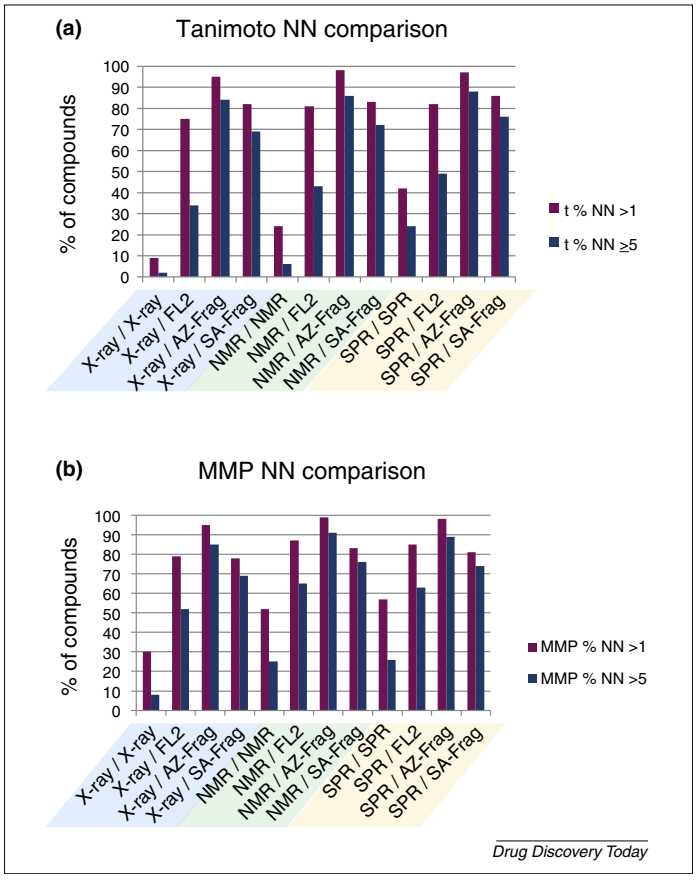

FIGURE 4

Ability to follow up on fragment hits as judged by the number of near-neighbor (NN) structures in AstraZeneca compound libraries or commercial sources. **(a)** Tanimoto analysis: percentage of compounds in screening subsets that have at least one and ≥5 NN for the fragment library-screening subsets within different sets (FL2, AZ-Frag and SigmaAldrich-Frag) using fingerprint similarity searches (t, tanimoto distance = 0.3). **(b)**. Matched molecular pair (MMP) analysis: percentage of compounds in screening subsets that have at least one to five NN for the fragment library-screening subsets within different sets (FL2, AZ-Frag and SigmaAldrich-Frag) using MMP similarity searches.

In our similarity analysis, comparison of a library with itself shows the lower percentages of near neighbors within a library, an indication of the structural diversity within each of the respective libraries. As expected based on the design principles used in constructing the library subsets, these percentages were lowest for the X-ray set (reflecting maximum diversity); the value was intermediate for the NMR subset, and highest for the SPR subset. There were fewer near neighbors for all subset libraries in SA-Frag compared with the in-house fragment database AZ-Frag, reflecting the high percentage of fragments proprietary to AZ. These results highlight the ability of FBLG project teams at AZ to rapidly identify and screen close structural analogs of fragment hits, enabled by software tools such as MATCH. Extensive elucidation of SAR early on in a project without the need to involve synthetic chemistry resources can greatly accelerate the hit-to-lead process.

Analysis of fragment hit properties from FL2 in screening campaigns

Since completing the reconfiguration of our global fragment-screening library, the strategic subsets of FL2 have been screened against several targets: five targets were screened via X-ray crystallography, seven targets via SPR and seven targets via 1D- or 2D-NMR. The data from each fragment-screening campaign were collected, and the properties of fragment hits were compared with those of their library of origin. For the purposes of our analysis, a 'fragment hit' is a confirmed hit. Primary screening hits were considered confirmed if they resulted from X-ray screening or 2D-NMR screening. Hits from 1D-NMR and SPR were considered confirmed only after positive validation by an orthogonal technique. For the physicochemical property analysis of our fragment hits, we focused on HAC, *c* Log *P*, plane of best fit (PBF) [19] and PMI [11]. We used PBF and PMI plots to explore the 3D shape and chemical space coverage of the fragments in the library and to compare them to the profile of hits generated. The PBF and PMI were calculated as described in the literature [11,19].

Figure 3 illustrates that the profile of all fragment hits compared with all of FL2 for HAC, *c* Log *P*, PBF and PMI was very similar. In addition, the hits from the X-ray, NMR and SPR subsets generally reflected the property distribution of their library of origin. For example, the average HAC was lowest for X-ray hits (13) and highest for SPR hits (16), as would be expected based on their composition. From the plots of *c* Log *P* (Fig. 3b), the NMR and SPR fragment hits tended to be slightly more lipophilic than their respective subsets. Others previously reported that hits tended to be more hydrophobic than the average fragment [20,21]. Overall, the general lipophilicity profile for our fragment hits was similar to the average *c* Log *P* of the screening sets. One of the stated advantages of fragment screening is the ability to identify chemical starting points for projects that are smaller and less lipophilic than would typically be identified through use of HTS. This benefit is obvious when comparing the distribution of HAC and *c* Log *P* of fragment hits from FL2 with the HAC and *c* Log *P* plots for all HTS hits generated at AZ from 2010 to July 2015 (Fig. 5).

In terms of PBF, we see that the hits were slightly less 3D than their respective libraries. This was true for FL2 as well as the NMR and SPR subsets (Fig. 3c). Recent literature has debated the merits of the 3D character in fragment-based drug discovery in addressing difficult targets, in particular those for which protein–protein interactions (PPIs) have a fundamental role [22]. Some feel that the 3D shape will provide added benefit when targeting PPIs with fragment libraries [23]. Others have reported that there is little difference in the 3D shape between fragment hits bound to PPIs and those bound to more traditional targets [24,25]. Multiple groups are conducting focused studies to explore trends in 3D fragment screening in more detail [26,27]. Although we have a limited data set, we were intrigued to explore the shapeliness of our fragment hits for all targets as well as for PPI targets. In a recent paper [25], scientists at Astex stated that they considered PBF the most meaningful 3D measure for their fragment library, and used PBF >0.05 to indicate the 3D character in their molecules. In calculating both PBF and PMI, we used 3D coordinates of molecules as generated by Corina [28], with the exception of the hits from the X-ray fragment library, for which the PBF was calculated based on the actual bound conformation observed in the crystal structure. In our own analysis of the conformation of compounds with PBF in the interval 0–0.3, our perspective was that most compounds with PBF ≤0.25 could not be considered 3D. Thus,

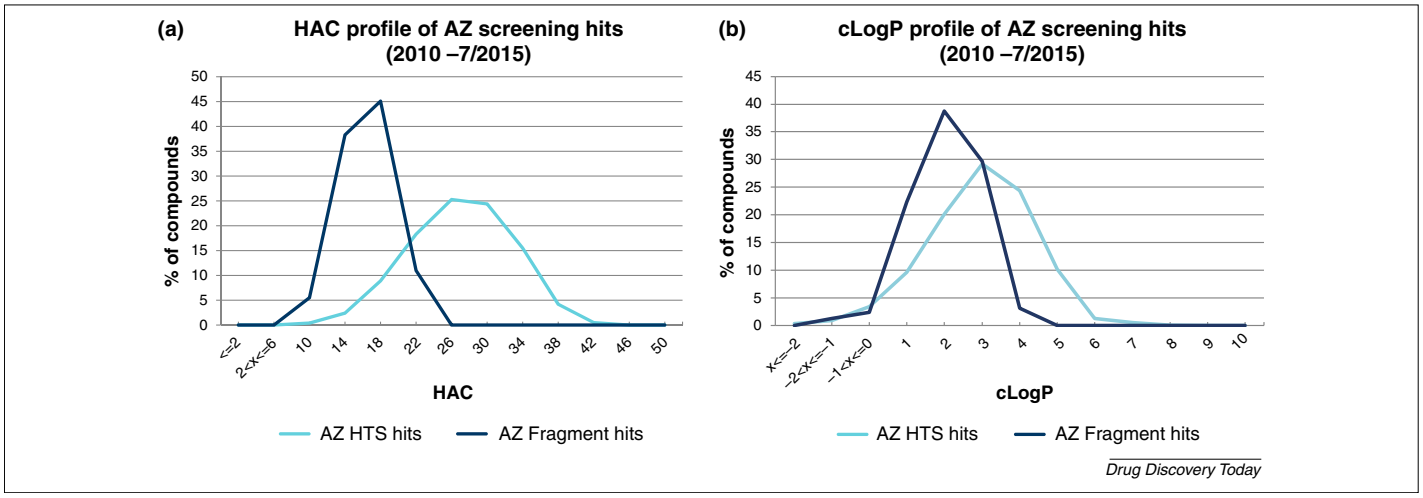

FIGURE 5

Comparison of the **(a)** heavy atom count (HAC) and **(b)** *c* Log *P* profiles between AstraZeneca high-throughput screening (HTS) hits (light-blue lines) and fragment-screening hits (dark-blue lines) from FL2 in the years 2010–2015.

we set our threshold for characterizing a compound as 3D at PBF >0.25.

Looking at the PBF profile of all fragment-screening hits from FL2, we observed that the hits showed slightly less 3D character than the entire library (Figs 3c and 6). While 60% of compounds comprising FL2 had PBF >0.25 and can be considered 3D by our metric, we see that only 49% of all fragment hits identified had PBF >0.25. As for PPI targets, with a limited data set from only three PPI campaigns, we see that 49% of the hits had PBF >0.25. Thus, in our internal fragment projects, the degree of nonplanarity of fragment hits for PPI targets was the same as for hits against all other target classes, albeit based on a small sample set. This agrees with other recent assessments of 3D character of PPI fragment hits: that there is no clear trend toward greater 3D character for this target class [24,25]. On the basis of the slightly more 3D nature of FL2 and the NMR and SPR subsets than the corresponding hits, one might suggest that our library is slightly over-represented in terms of 3D fragments. However, based on nearly half of our fragment hits exhibiting 3D character across all target classes, we feel a strong argument is made for maintaining a good balance of 2D and 3D molecular topology in any fragment-screening collection.

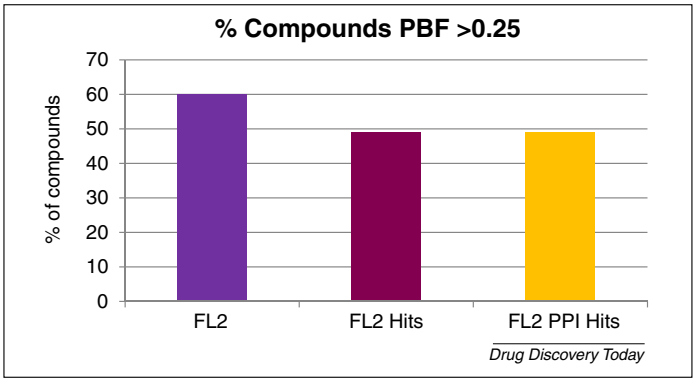

FIGURE 6

Percentage of three-dimensionality for FL2, all fragment hits and fragment hits for protein–protein interaction (PPI) targets.

Furthermore, 3D fragments are an excellent complement to flat fragments, because they provide additional and alternative vectors for exploring the binding pocket of interest.

The concept of chemical space coverage and the advantages of sampling chemical space with fragments has been discussed in detail elsewhere [25,29,30]. Using PMI as a measure of shape diversity and chemical space coverage, we see that FL2 was well represented in all regions of the chemical space plot (Fig. 3d). The PMI plots for each of the respective screening subsets showed reasonable coverage of chemical space, with the X-ray subset having the most limited coverage, trending toward more rod- and disk-like topology. This might be expected for compounds that have lower HAC. The PMI profile of the hits derived from each of the screening technologies generally resembled the distribution of the respective subset. Taken together, the PMI and PBF plots showed that our screening techniques are able to identify a balance of both flat and shapely molecules, and are able to generate hits that occupy distinct areas of chemical space.

While one might argue that the profile of hits would be expected to resemble the composition of the library from which they came, other FBLG practitioners have observed that this is not always the case [21,25]. We believe that the close resemblance of fragment hit profiles in terms of HAC, *c* Log *P*, PBF and PMI to the corresponding libraries highlights that our fragment library and screening subsets are appropriate for the screening techniques used and have excellent coverage of chemical space and physicochemical property balance. The fragment hits sample the full range of the library, validating the selection criteria used to build the screening subsets and highlighting the rigor put into constructing this new incarnation of the AZ fragment-screening library, FL2.

Comparison of pocket volume filling between 3D and 2D fragments

To further investigate the concept of 3D fragments, we were interested in assessing whether protein-bound 3D fragments explored the pocket space of targets more extensively than 2D fragments. It is tempting to assume that 3D molecules have the

TABLE 1

Percentage of binding pocket volume (Å^3) filled by 3D and 2D fragments in X-ray crystal structures from five AZ FBLG projects

Entry	Target class	Total pocket volume (Å^3) (no. of fragment X-ray structures)	% volume filled, 3D fragments (no. of molecules)	% volume filled, 2D fragments (no. of molecules)
1	PPI	822 (59)	88 (32)	69 (27)
2	Oxidoreductase/dehydrogenase	880 (125)	86 (43)	78 (82)
3	Nuclear hormone receptor	664 (23)	74 (6)	53 (17)
4	Kinase 1	521 (15)	80 (8)	83 (7)
5	Kinase 2	608 (18)	77 (7)	82 (11)

ability to interrogate additional pocket volume compared with flatter compounds [27], but to the best of our knowledge, there is no clear evidence to support this idea in the literature. Therefore, we undertook an analysis to explore the validity of this assumption. We examined five protein targets for which we had multiple fragment crystal structures (Table 1 and Fig. 7). The fragments were classified as 3D or 2D based on PBF values (3D defined as PBF >0.25) to allow for comparison of the total volume occupied in the pocket by these two sets calculated using a grid-based method [31,32].

For a PPI target, we found that the 3D molecules occupied 19% of the pocket volume unexplored by the 2D fragment X-ray structures (Table 1, entry 1). Of the remaining four targets included in the analysis, two showed a similar trend, with 3D fragments covering more space in the binding pocket than 2D fragments (Table 1, entries 2 and 3). Of note, for a dehydrogenase target (Entry 2), although there were nearly twice as many 2D fragment structures, the 3D fragments explored 8% more of the binding pocket volume. For the two kinase target examples (Table 1, entries 4 and 5), the 3D and 2D fragment structures explored roughly an equivalent percentage of the binding pocket. While we have only conducted the pocket volume-filling analysis on a small set of targets and feel that it is a fairly crude method, it is an intriguing way to compare 3D and 2D fragments and explore the concept of whether property profile of hits, particularly PBF, might correlate to the shape of the binding pocket. A more systematic study of this question might be useful in further evaluating the utility of 3D fragments in a fragment library. At this point, the analysis of the pocket volume agreed with our other data, suggesting that there are benefits to a fragment library having a balanced representation of topology for maximum exploration of target binding pocket. This is something that we will continue to collect statistics on to see whether any additional trends emerge, and will comment on this in due course.

Trends in fragment hit progression

To assess the progression of FBLG projects, a more thorough analysis was conducted for a range of targets that had successfully transitioned from fragment hits to lead series (Table 2 and Fig. 8). Several different target classes were included, representing known targets as well as unprecedented targets with no known small-molecule inhibitors. Each of our screening subsets was used as a primary hit-finding technique. While hit rates varied from very low (0.2%) to relatively high (8.8%), in each case chemical starting points were delivered that allowed for chemistry progression via elaboration and optimization. One parameter utilized in fragment

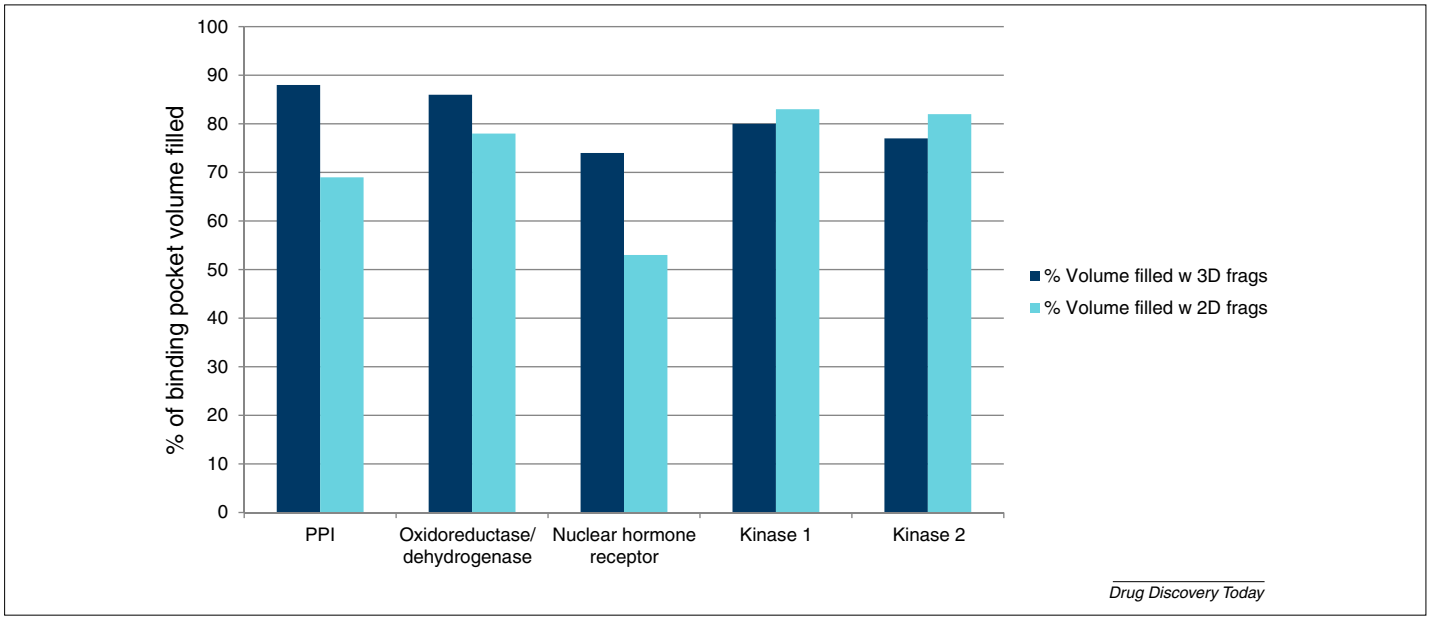

FIGURE 7

Graphical depiction comparing the percentage of the volume of binding pocket filled by 3D and 2D fragments in X-ray crystal structures for five targets at AZ.

TABLE 2

Summary of eight AZ FBLG projects that progressed from fragment hits to lead series, in terms of screening method, hit rate, fragment hit profile and representative lead compound profile derived from corresponding fragment hits

Entry	Target class	Screening method	No. of fragments screened (hit rate)	Fragment hit profile	Lead compound profile
1	PPI[a]	SPR	3072 (1.1%)	Kd = 690 μM; LE[b] = 0.25; LLE[c] = 1.0; PBF = 0.03	Kd = 0.0015 μM; LE = 0.34; LLE = 7.5; PBF = 0.77
2	PPI[a]	SPR	3072 (1.1%)	Kd = 830 μM; LE = 0.29; LLE = 0.60; PBF = 0.26	IC_{50} = 0.011 μM; LE = 0.37; LLE = 5.3; PBF = 0.51
3	Nuclear hormone receptor	X-ray	384 (4.7%)	Kd = 3230 μM; LE = 0.35; LLE = 0.9; PBF = 0	IC_{50} = 0.040 μM; LE = 0.43; LLE = 4.4; PBF = 0.64
4	Enoyl reductase (mInhA)	SPR; X-ray	3072 (1.5%); 384 (2.1%)	Kd = 1340 μM; LE = 0.37; LLE = 1.3; PBF = 0	IC_{50} = 0.0032 μM; LE = 0.44; LLE = 4.7; PBF = 0.73
5	Oxidoreductase/dehydrogenase[a] (PHGDH)	X-ray	384 (8.8%)	Kd = 470 μM; LE = 0.39; LLE = 1.5; PBF = 0	Kd = 0.18 μM; LE = 0.29; LLE = 5.5; PBF = 0.89
6	Protease[a]	NMR; WAC	3600 (1.5%); 5000 (0.2%)	IC_{50} = 97 μM; LE = 0.43; LLE = 1.2; PBF = 0.35	IC_{50} = 0.180 μM; LE = 0.38; LLE = 3.2; PBF = 0.72
7	Kinase (series 1)	NMR	1335 (7.3%)	Kd = 78 μM; LE = 0.40; LLE = 1.8; PBF = 0	IC_{50} = 0.079 μM; LE = 0.30; LLE = 2.2; PBF = 0.67
8	Kinase (series 2)	NMR	1335 (7.3%)	Kd = 50 μM; LE = 0.43; LLE = 1.0; PBF = 0	IC_{50} = 0.012 μM; LE = 0.41; LLE = 5.0; PBF = 0.30

[a] Unprecedented target (no known inhibitors reported).
[b] LE = 1.4 × pKd/HAC or 1.4 × pIC_{50}/HAC.
[c] Lipophilic LE (LLE) = pKd − *c* Log *P* or pIC_{50} − *c* Log *P*.

series selection and design is ligand efficiency (LE) [33]. In general, high levels of LE were maintained as hits were elaborated. Impressively, for difficult to drug, unprecedented PPIs, fragment hits with lower LEs were optimized into lead series with much improved LEs (Table 2, entries 1 and 2).

It is well documented in the literature that compounds with high sp2 centers:sp3 centers ratios have been found to have higher rates of attrition in the clinic [34]. Some have suggested that it would be beneficial for the design of fragment libraries to move away from flat, aromatic components and contain more stereocenters and higher fraction of sp3 atoms (fsp3) to build in this structural complexity early on [23,26,27]. While we closely monitored LE metrics, LogD and *c* Log *P* during our elaboration and optimization process, we paid little attention to PBF and fsp3 in the initial stages of fragment SAR. Regardless, we found that compounds with high PBF can be derived from completely flat fragment hits (Table 2, entries 3–5, 7,8), as has also been observed by scientists at Astex [25,26]. Thus, while 3D character can be useful in fully interrogating a binding pocket, it is not required for a fragment hit to be highly nonplanar to result in a lead series with 3D character. In our experience, three-dimensionality is often introduced during optimization, as encouraged by the preferences of the protein target. These successful deliveries of attractive lead series provide confidence that FL2 can generate chemical starting points that can be progressed to viable lead series irrespective of target class.

Our interest in further expanding our repertoire of fragment-screening techniques is exhibited by entry 6 in Table 2. For this protease target, we conducted fragment screens with both NMR and weak-affinity chromatography (WAC) [35], and found good complementarity of the hits. We continue to evaluate new emerging technologies that can enhance our ability to screen a range of

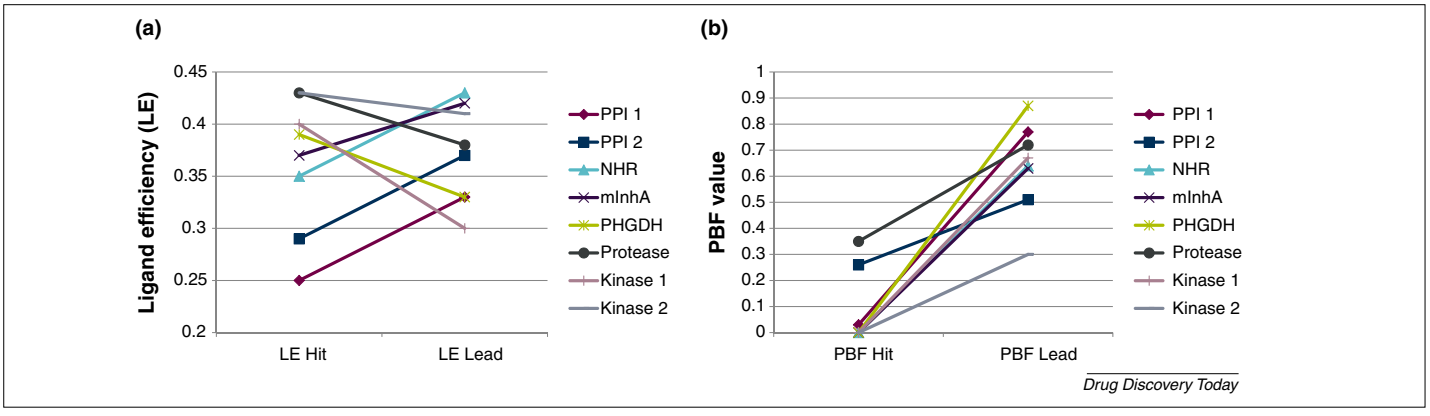

FIGURE 8

Graphical depiction of ligand efficiency (LE) **(a)** and plane of best fit progression **(b)** in going from fragment hits to lead compounds.

target classes to ensure that our capabilities remain at the forefront of the industry.

To illustrate the success of the new FBLG process, two case studies are discussed below (P. Madhavapeddi *et al.*, unpublished data).

Phosphoglycerate dehydrogenase

Phosphoglycerate dehydrogenase (PHGDH) is an NAD-dependent enzyme involved in serine biosynthesis and is postulated to be essential for tumor metabolism. It is strongly implicated in aggressive breast cancer [36]. At the time the project was initiated, there were no known PHGDH inhibitors and no small-molecule co-crystal structures were reported. A fragment screen was performed in advance of an HTS via crystal soaking with the X-ray screening subset, and provided multiple fragments bound in the adenine region of the NAD-binding site. Starting from the very simple but ligand efficient hit **1** (Fig. 9a), we were able to grow in multiple vectors to form key hydrogen bond interactions and improve potency, using SPR to drive SAR. Through successive rounds of optimization, the team was able to develop a series that showed potent enzyme inhibition. Starting from a completely planar fragment hit, a lead series with PBF >0.8 and high fsp3 was delivered. Notably, while HTS identified inhibitors with Kd of 1.5 μM and multiple series were pursued, none of these streams was able to produce compounds with Kd <1 μM and the fragment series became the lead project series.

mInhA

The ACP-enoyl-reductase mInhA catalyzes a crucial step in the synthesis of mycolic acids that form the protective waxy coating on *Mycobacterium tuberculosis* (TB) [37]. Isoniazid (INH) is a first-line TB prodrug that is activated within the bacterial cell wall by the catalase-peroxidase KatG. This activation generates a species that forms a covalent adduct with NADH, resulting in a NAD-competitive inhibitor [38]. The primary mechanism of resistance to isoniazide is attributed to deactivating mutations in KatG. The project strategy was to identify direct inhibitors of InhA that function through a KatG-independent mechanism, thus circumventing acquired resistance. In this case, substrate cost precluded a full HTS; thus, the team opted for a dual approach comprising a SPR-based fragment screen, and biochemical screen of a 26,000-compound subset to provide fragment and nonfragment hits. The fragment screen identified several interesting ligand efficient hits, including **4** (Fig. 9b). Crystallographic data from fragments and hits emerging from the biochemical screen enabled the design of highly potent and cell active inhibitors, such as **6**. Again, a fragment hit with low PBF gave rise to a lead series with high 3D character while maintaining excellent LE.

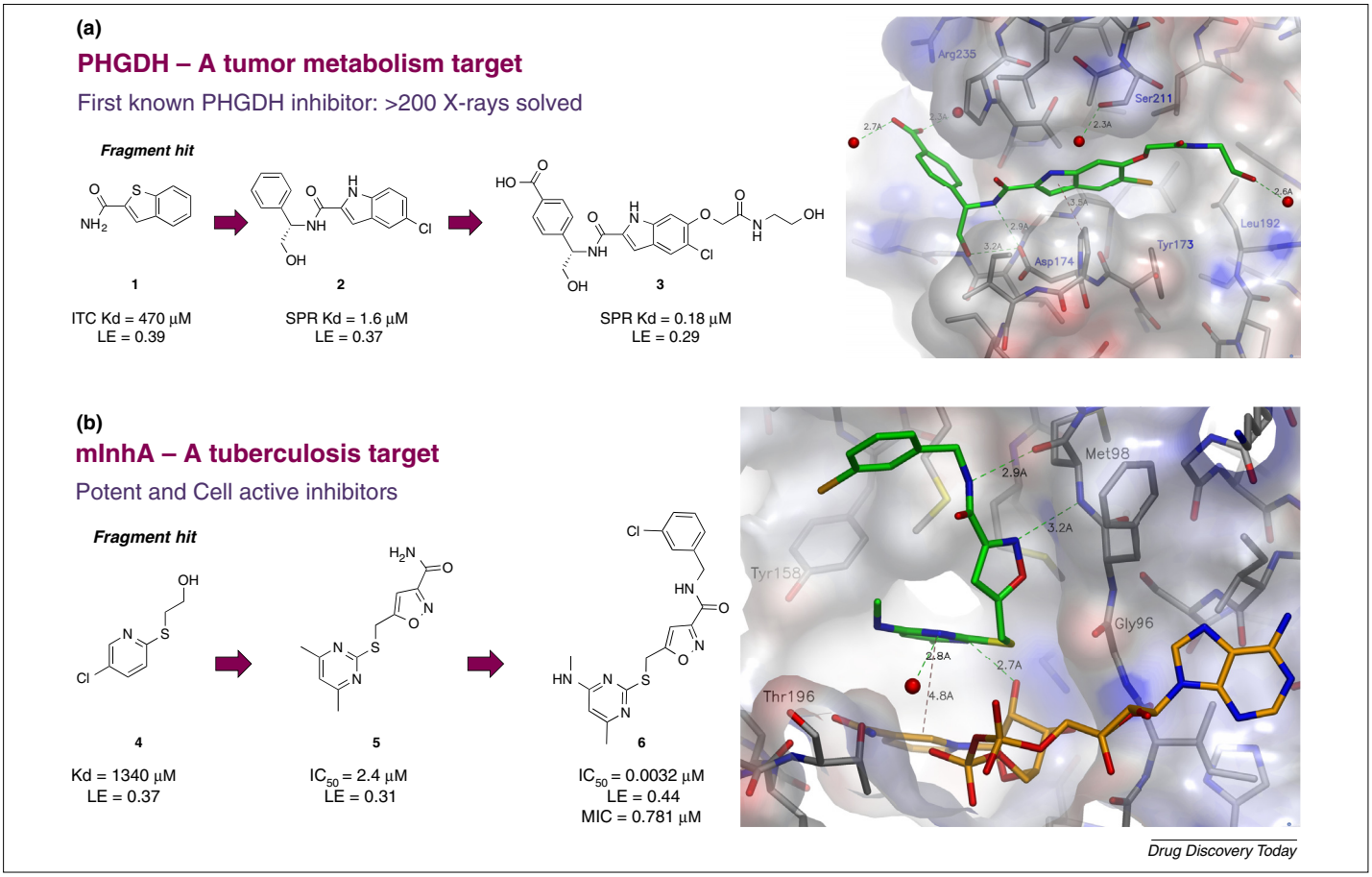

FIGURE 9

Case studies for two AstraZeneca fragment-based lead generation (FBLG) projects, phosphoglycerate dehydrogenase (PHGDH) **(a)** and mInhA **(b)**, showing the progression of fragment hit, to optimized fragment, to lead series. The co-crystal structures of the lead series compounds show the key interactions in the binding pocket.

Concluding remarks and outlook

In recent years, AZ has established a new internal model for FBLG. The impact of the operational changes on performance of FBLG at AZ has been profound. By combining dedicated chemistry resources with the established strengths in structural biology and computational chemistry and improving biophysical screening capabilities, AZ has placed emphasis on it becoming an industry leader in this field. The analysis of the fragment hits generated by FL2 has shown the property profile of the fragment hits to resemble the composition of the library from which the hits came, confirming that this fragment library is appropriately constructed and positions AZ for success in finding quality chemical starting points for drug projects. We have also seen that a high percentage of the fragment hits are considered 3D (49% with PBF >0.25). These data strongly advocate for including 3D fragments in the composition of any fragment library to complement more planar fragment hits in fully exploring binding pockets of interest. The skill of the AZ FBLG teams has resulted in the evolution of diverse fragment hits into quality lead series for a range of target classes, including unprecedented targets of low predicted druggability. We hope that sharing our experiences and findings will be helpful to those in the field of fragment-based drug discovery.

Going forward, we will continue to maintain rigor in the active improvement and crucial evaluation of our fragment library. We are exploring emerging tactics in FBLG and ensuring that our hit-finding approaches are sufficiently diversified to address all target classes, such as cysteine proteases, and unprecedented and lower tractability targets. A diverse array of screening techniques should provide more confidence in progressing fragment campaigns in the absence of structure. As we gather more data on the screening of FL2, we will be able to strengthen our understanding of trends in the physicochemical properties of the hits generated against different target classes and the impact of parameters, such as PBF, and adjust the composition of our library accordingly to create FL3. We look forward to reporting on our learnings and experiences with FL3, the covalent fragment set and additional PPI targets. The active management of our assets and technology uptake will enable AZ to stay at the leading edge of this exciting field of drug discovery.

Acknowledgements

The authors would like to thank several colleagues at AstraZeneca for their contributions: Richard Ward for his contributions to the design of the covalent fragment library; David Cosgrove for providing the code for the volume analysis; and Thierry Kogej for property profile statistics of HTS hits.

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. *et al.* (1997) Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. Chem. Soc.* 119, 5818–5827
3 Murray, C.W. *et al.* (2012) Experiences in fragment-based drug discovery. *Trends Pharmacol. Sci.* 33, 224–232
4 Murray, C.W. and Rees, D.C. (2009) The rise of fragment-based drug discovery. *Nat. Chem.* 1, 187–192
5 Baker, M. (2013) Fragment-based lead discovery grows up. *Nat. Rev. Drug Discov.* 12, 5–7
6 Albert, J.S. *et al.* (2007) An integrated approach to fragment-based lead generation: philosophy, strategy, and case studies from AstraZeneca's drug discovery programmes. *Curr. Top. Med. Chem.* 7, 1600–1629
7 Cook, D. *et al.* (2014) Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* 13, 419–431
8 Goodnow, Jr., R.A. (ed.), (2014) *A Handbook for DNA-Encoded Chemistry: Theory and Applications for Exploring Chemical Space and Drug Discovery*, John Wiley & Sons
9 Blomberg, N. *et al.* (2009) Design of compound libraries for fragment screening. *J. Comput. Aided Mol. Des.* 23, 513–525
10 Bevan, C.D. and Lloyd, R.S. (2000) A high-throughput screening method for the determination of aqueous drug solubility using laser nephelometry in microtiter plates. *Anal. Chem.* 72, 1781–1787
11 Sauer, W.H.B. and Schwarz, M.K. (2003) Molecular shape diversity of combinatorial libraries: a prerequisite for broad bioactivity. *J. Chem. Inf. Comput. Sci.* 43, 987–1003
12 Haigh, J.A. *et al.* (2005) Small molecule shape-fingerprints. *J. Chem. Inf. Model.* 45, 673–684
13 Congreve, M. *et al.* (2003) A 'Rule of Three' for fragment-based lead discovery? *Drug Discov. Today* 8, 876–877
14 Edfeldt, F.N.B. *et al.* (2011) Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discov. Today* 16, 284–287
15 Anon. (2008) *Fingerprint Theory Manual*. Daylight Chemical Information Systems Inc.
16 Giannetti, A.M. *et al.* (2008) Surface plasmon resonance based assay for the detection and characterization of promiscuous inhibitors. *J. Med. Chem.* 51, 574–580
17 Singh, J. *et al.* (2011) The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* 10, 307–317
18 Kalgutkar, A.S. and Dalvie, D.K. (2012) Drug discovery for a new generation of covalent drugs. *Exp. Opin. Drug Discov.* 7, 561–581
19 Firth, N.C. *et al.* (2012) Plane of best fit: a novel method to characterize the three-dimensionality of molecules. *J. Chem. Inf. Model.* 52, 2516–2525
20 Chen, I-J. and Hubbard, R.E. (2009) Lessons for fragment library design: analysis of output from multiple screening campaigns. *J. Comput. Aided Mol. Des.* 23, 603–620
21 Lau, W.F. *et al.* (2011) Design of a multi-purpose fragment screening library using molecular complexity and orthogonal diversity metrics. *J. Comput. Aided Mol. Des.* 25, 621–636
22 Fry, D.C. *et al.* (2013) Deconstruction of a Nutlin: dissecting the binding determinants of a potent protein–protein interaction inhibitor. *ACS Med. Chem. Lett.* 4, 660–665
23 Bower, J.F. and Pannifer, A. (2012) Using fragment-based technologies to target protein–protein interactions. *Curr. Pharm. Des.* 18, 4685–4696
24 Turnbull, A.P. *et al.* (2014) Fragment-based drug discovery and protein–protein interactions. *Res. Rep. Biochem.* 4, 13–26
25 Hall, R.J. *et al.* (2014) Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.* 116, 82–91
26 Morley, A.D. *et al.* (2013) Fragment-based hit identification: thinking in 3D. *Drug Discov. Today* 18, 1221–1227
27 Hung, A.W. *et al.* (2011) Route to three-dimensional fragments using diversity-oriented synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 108, 6799–6804
28 Sadowski, J. *et al.* (1994) Comparison of automatic three-dimensional model builders using 639 X-ray structures. *J. Chem. Inf. Model.* 34, 1000–1008
29 Hann, M.M. *et al.* (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Model.* 41, 856–864
30 Leach, A.R. and Hann, M.M. (2011) Molecular complexity and fragment-based drug discovery: ten years on. *Curr. Opin. Chem. Biol.* 15, 489–496
31 Gavezzotti, A. (1983) The calculation of molecular volumes and the use of volume analysis in the investigation of structured media and of solid-state organic reactivity. *J. Am. Chem. Soc.* 105, 5220–5225
32 Perkins, T.D.J. *et al.* (1995) Molecular surface-volume and property matching to superpose flexible dissimilar molecules. *J. Comput. Aided Mol. Des.* 9, 479–490
33 Hopkins, A.L. *et al.* (2004) Ligand efficiency: a useful metric for lead selection. *Drug. Discov. Today* 9, 430–431
34 Lovering, F. *et al.* (2009) Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* 52, 6752–6756
35 Duong-Thi, M-D. *et al.* (2011) Weak affinity chromatography as a new approach for fragment screening in drug discovery. *Anal. Biochem.* 414, 138–146
36 Possemato, R. *et al.* (2011) Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* 476, 346–353
37 Brennan, P.J. *et al.* (1970) The lipids of *Mycobacterium tuberculosis* BCG: fractionation, composition, turnover and the effects of isoniazid. *Iran. J. Med. Sci.* 3, 371–390
38 Rawat, R. *et al.* (2003) The isoniazid-NAD adduct is a slow, tight-binding inhibitor of InhA, the *Mycobacterium tuberculosis* enoyl reductase: adduct affinity and drug resistance. *Proc. Natl. Acad. Sci. U. S. A.* 100, 13881–13886