Fragment library design, synthesis and expansion

Ray, Peter C.; Kiczun, Michael; Huggett, Margaret; Lim, Andrew; Prati, Federica; Gilbert, Ian; Wyatt, Paul G.

Published in:
Drug Discovery Today

DOI:
10.1016/j.drudis.2016.10.005

Publication date:
2017

Document Version
Peer reviewed version

Fragment Library Design, Synthesis & Expansion: Nurturing a Synthesis and Training Platform

Peter C. Ray,* Michael Kiczun, Margaret Huggett, Andrew Lim, Federica Prati, Ian H. Gilbert,* and Paul G. Wyatt.*

Drug Discovery Unit, Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dow Street, Dundee, DD1 5EH, Scotland, United Kingdom

Key Words: fragment-based drug discovery; compound library; physicochemical properties; scaffold; drug discovery.

Teaser: We describe an approach to develop diverse and novel fragment libraries, which can also be used as a training platform for medicinal chemists, at the undergraduate and graduate levels.

* Tel + 44 1382 386 231

ABSTRACT: The availability of suitable diverse fragment- and lead-oriented screening compounds is key for the identification of suitable chemical start points for drug discovery programs. The physicochemical properties of molecules are critical in determining the success of small molecules in clinical development, yet reports suggest that pharmaceutical and academic sectors often produce molecules with poor drug-like properties. We present a platform to design novel, high quality and diverse fragment and lead-oriented libraries with appropriate physicochemical properties in a cost-efficient manner. This approach has the potential to assist the way libraries are constructed by significantly addressing the historical uneven exploration of chemical space for drug discovery. Additionally, this platform can teach both undergraduates and graduates compound library design.

INTRODUCTION

Fragment-based drug discovery (FBDD) and high throughput screening are important approaches to find chemical start points for drug discovery programmes. These libraries rely on synthesis as well as commercial acquisition of new compounds. The majority of this extensive effort, mainly carried out over the last 10-15 years, remains unpublished, as novel proprietary compounds provide a competitive edge for the respective pharmaceutical companies. Hence, there is no full understanding of the proportion of chemical space [1,2] covered by current compound libraries and screened in biological assays. It is estimated that there are more than 10^{60} possible organic compounds that fulfill Lipinski's rules [1,3]. However, a framework analysis of the CAS Registry suggests that chemists are more likely to use a particular framework to make a compound, the more often that framework has been used in the past [4]. This results in the proliferation of certain frameworks and limits the exploration of novel chemical space.

Bemis and Murcko analyzed drug molecules according to ring, linker, framework and side chain atoms so that the information could be employed for the synthesis of new scaffolds with biological rationale [5,6]. Their analysis suggests that the scaffold and side chain diversity associated with known drugs is relatively low. More recently, analysis of drug space (until end of 2012) by Taylor et al., describing the rings and

molecular scaffolds [7], showed that there were only 351 ring systems and 1197 frameworks. Further, only six new ring systems enter drug space each year and 28% of new drugs contain a new ring system. This is a relatively small number compared, for example, to the predicted number of small aromatic rings [8]. Increasing the diversity and novelty of compound libraries is likely to be important in probing the drug-like chemical space that addresses biological space [2], to tackle both existing and emerging drug targets (for example protein-protein interactions).

For many drug discovery organizations, particularly small- and medium-sized enterprises (SMEs) and the public sector, access to chemical matter is largely reliant on commercially available fragment and lead-like libraries. However, a recent analysis has suggested that "existing synthetic methodology is unintentionally predisposed to producing molecules with poorer drug-like properties and that this is likely to have ramifications to the early hit- and lead-finding phases of the drug discovery process" [9], particularly in addressing emerging target classes. In contrast, academic synthetic chemists are developing new synthetic methodology, which would be a powerful way to increase the novelty and diversity of our chemical libraries, if leveraged for compound library synthesis. This could be particularly beneficial for SMEs and academic drug discovery groups, as well as the pharmaceutical industry (Figure 1) [10].

Fragment-based drug discovery (FBDD) is an important approach to identify new leads for drug discovery [11-14]. It has the advantage of being able to address an area of chemical space with many fewer compounds than conventional lead-like or drug-like libraries. During an attempt to increase the chemical diversity of our fragment library, we realized that relatively simple derivatization ("capping") of commercially available building blocks could expand into new areas of chemical space that are not commercially available, often with very different chemical and physicochemical properties (Figure 2), [15]. These compounds could subsequently be rapidly expanded into lead-like space for hits, using the same or similar chemistry. We therefore decided to establish a platform to design and generate compounds to address a wider range of fragment-space. This platform encompasses a set of criteria for compound design, and procedures for compound preparation and carrying out quality assurance.

FBDD typically uses structural knowledge for fragment optimization using a range of strategies including merging, linking or growing [12]. Fragments typically bind to "hot spots" within the ligand binding site of the target [16] and options for optimization are prioritized according to synthetic appeal, opportunity to access relevant areas of the binding site, binding affinity and ligand efficiency. The prioritized fragments are then typically elaborated through addition of a suitable functional group, or "chemical handle", to attach and elaborate the new substituent to identify further interactions

[11]. Ideally the chemical handle provides additional interactions within the target to improve affinity as well aid both fragment growth and linking.

Screening fragments with “built-in” handles has the potential for additional interactions with the target (protein) and faster elaboration of a fragment. However care has to be taken as increasing complexity in a fragment decreases the probability of it achieving optimal ligand - protein interactions [17]. Conversely, too little complexity can lead to interesting interactions being missed [18]. Therefore, a careful balance is required between the “built-in” handles and the complexity of the fragment. Subsequent introduction of a chemical handle to a fragment may alter the binding conformation. This was demonstrated by Shoichet et al., who showed that deconstruction of a larger potent β-lactamase inhibitor into small fragments with “minimal complexity” does not necessarily recapitulate its binding to the enzyme [19]. However, fragments with additional functional-group complexity could recapitulate the larger potent β-lactamases inhibitor binding. Conversely, smaller fragments could identify more ligand efficient binding modes to the “hot spot”, which after alternative optimization strategies, could potentially lead to development of compounds with better drug-like properties e.g. lower molecular weight.

A case study in support of the use of fragments with pendent functional groups suitable for rapid elaboration comes from Nazaré et al. [20] who demonstrated the

super-additivity effect of linking two fragments (derived from deconstructing a potent factor Xa inhibitor) containing amide and sulfonamide functional groups respectively (Figure 3).

To make maximum use of the fragment library, an ideal scenario would be to have a linkage between the scaffolds in a fragment library and a lead-like library. This would have two advantages: firstly, a hit in a fragment screen could be rapidly expanded into lead-like space using analogues of the scaffold in the lead-like library, or using known robust chemistry to grow the fragments. Secondly, if there is a hit on screening a lead-like library, this could be rapidly "de-constructed" into fragments to probe the key receptor-ligand interactions. Given the rapid increase in the number of possible compounds as the heavy atom count in a molecule increases, there will be a limit to the examples of a fragment-scaffold that can be in a lead-like library. Therefore, it will be important to have chemistry suitable for fragment-scaffold elaboration.

RESULTS AND DISCUSSION

Design Process. A key component of our library platform was to identify under-represented areas of commercial chemical space and create diversity based on functional group manipulation. A library design team, composed of experienced medicinal chemists, investigated a number of procedures for synthesis of fragments

with differing synthetic complexity, based on the analysis of fragment scaffolds in relation to published biologically relevant data. Our workflow is shown in Figure 4.

We had three start points to develop novel fragments (Figure 5): (1) in-house assembly of diverse monomer (appendage) sets, which could be functionalized into both fragment and lead-like libraries, (2) commercial and in-house project intermediates, and (3) novel chemical scaffolds. Designs were filtered to ensure that compounds selected for synthesis filled in “gaps” in chemical space in our current fragment library.

Our focus here was on producing fragment and lead-like libraries to identify non-covalent, reversible inhibitors. Functional groups known to be chemically reactive or toxic were removed (so called “Structural Alerts”), as were compounds reported to be pan assay interference compounds (PAINS) [21]. Supporting information includes the Drug Discovery Unit revised in-house [22] and modified Eli Lilly-published [23] structural alerts and the PAINS alerts. After applying the structural alerts, we then set about defining descriptors and selection criteria for non-covalent reversible fragment libraries, monomer sets and lead-like libraries (Table 1). We report the final version we now use, which is the result of several rounds of iteration and optimization of the library selection parameters; so several of the early libraries were designed with slightly different criteria.

Although a number of key reports have evaluated the impact of aromatic ring count [24-26] and Fsp^3 on solubility and compound developability within drug-like chemical space [27,28], in practice, we did not explicitly factor in the aromatic ring count, as the other parameters took care of this. As historical fragment libraries (including our own) sample limited shape diversity [15], which in practice may impact opportunities to introduce shape diversity [29], we introduced routine calculations of the principal moments of inertia (PMI). Computational tools of evaluating novelty and diversity included ECFP4 fingerprint analysis, principal component analysis (PCA) and commercial availability versus established libraries (see Table 1).

Monomer Set Selection. Diverse monomer appendage sets were compiled to support our early hit- and lead-finding phases. The process for diverse monomer appendage set selection and purchase has evolved with the selection of each monomer set and was often dovetailed to the availability from the internal inventory. For example, selection of approximately 60 diverse primary and secondary amines included the following:

1. Extracting amines from the eMolecules database, salt stripping and filtering based on commercial availability from a set of suppliers (Aldrich, Enamine, Fluorochem, Tyger, Acros, Chembridge, Key Organics, ChemDiv, Otava and Maybridge, Combi-Blocks and Frontier)

2. Filtering compounds based on the properties described in Table 1 for monomer sets, which included in-house curated structural alerts.
3. Clustering of the filtered compounds using the ECFP4 fingerprint and binning by molecular weight for visual selection by a focused group of experienced medicinal chemists.
4. Purposely skewing the selection towards the lower molecular weight range within each cluster, to provide a greater diversity of cores of varying molecular weight and physicochemical properties.
5. A final set selection by further splitting into bins to ensure maximal coverage of both fragment and lead-like chemical space: 30 monomers were selected with MW ≤ 120, 20 with 120 < MW < 160 and 10 with 160 < MW ≤ 200.

We selected subsequent monomer sets in a similar manner and have thus far included carboxylic acids, sulfonyl chlorides and aldehydes. Learning from initial experience resulted in the monomers being enumerated after the initial filtering of the eMolecules search to afford the capped products and then examples taken from each cluster by chemist's eye selection. Cost, commercial availability and specific project requirements were factored into the final design of monomer sets.

Monomer Set Capping Fragment Library Design. Since the monomer sets were diverse, as defined by fingerprint analysis and visual inspection, we considered that "minimal functional group transformation" would give a diverse selection of functional group containing fragments (see Figure 6). Based on the case study by Nazaré et al., [20] the initial project focused on acetylation and mesylation of the amine monomer set and relevant inventory amines (see Scheme 1).

We employed the following protocol to identify the most suitable fragments for synthesis and selected examples of the acetylated and mesylated amine monomer set are shown in Charts 1 and 2:

1. Enumeration to give the acetylated and mesylated products.
2. The virtual products were filtered according to the fragment properties in Table 1.
3. Structural alert filters (PAINS, Eli Lilly and in-house) were applied [21-23].
4. Filtered compounds were clustered using the ECFP4 fingerprint (0.5 Tanimoto).
5. Commercial availability of the exact compound within the eMolecules database was checked. In general, we avoided re-making commercially available compounds. However sometimes when the compound was part of an array, it was cost-effective to include it.

6. Tanimoto similarity of each compound to the current fragment library was calculated; however this was treated as a guide and all structures were inspected by a medicinal chemist before preparation.
7. PCA (ECFP4) and PMI plots were generated for enumerated products to better understand the chemical diversity.

In our platform, designed compounds were synthesized through in-house chemistry or through outsourcing. The fragment functional group capping initiative, as part of an academic-based drug discovery unit, provided an excellent opportunity to nurture undergraduate students on how to design and synthesize fragments. Undergraduate project students were paired with a mentor and provided with a set of enumerated fragments for a particular monomer set coupling reaction. Guiding of students by the mentor through the selection process, provided valuable training in the use of modern *in silico* prediction (StarDrop™ (www.optibrium.com)) and visualization (Vortex) software. The reagent sets were then selected and the students guided to identify robust and safe synthetic routes for parallel synthesis. The optimal synthetic route was identified with a trial set of compounds and the library was then prepared using parallel techniques, starting with 6 compounds and progressing towards 24 compound arrays. As well as using the monomer sets, this approach was extended to

commercially available building blocks and appropriate in-house project intermediates. This platform also has the advantage from a training context of being tunable, ranging from straightforward functional group interconversion to more complex scaffold synthesis.

We extended the approach to a number of other types of chemistry, including urea formation (Figure 6) and cross-coupling of the monomer sets. This was rolled out initially to final (fourth) year project students, working in our laboratories and then subsequently to third year undergraduate students. The students gained a significant training in the parallel synthesis and purification of small polar compounds and excitement stimulated by production of novel compounds.

Synthesis of Multiple Diverse Scaffolds from Common Intermediates. The next step was to identify flexible platforms for the synthesis of multiple unexplored diverse fragment scaffolds from readily available key intermediates. A pilot platform was based around the evaluation of fused bicyclic fragment frameworks described by Bemis and Murcko [5,6], in particular 6,5- and 6,6-fused bicyclic scaffolds, with the requirements of a saturated ring and sp^3 vectors with diverse functional groups capable of forming varied pharmacophoric points. To better understand the existing landscape for such scaffolds, we conducted database searches and refined the output based on

commercial availability as well as property selection filters based on our guide criteria for fragment selection within Table 1. A selection of the output in Figure 7 shows an uneven distribution of the explored potential chemical space for a selected set of different vectors as well as decreased representation because of increased structural complexity through the addition of chirality and heteroatoms into the saturated six-membered ring. From the authors' experience, the fact that a compound is offered commercially does not necessarily mean it is readily available at a suitable cost, as compound suppliers often assign a synthesis time based on literature routes and a minimum scale. In our opinion, the output in Figure 7 provides a balanced approach for identifying compounds that may be available and are likely to be accessed via literature synthetic routes.

To sample wide chemical space, we opted to identify common intermediates that could be prepared on a large scale and readily expanded into multiple structurally complex and diverse unexplored scaffolds. This approach led us to evaluate cyclic ketone building blocks, based on validated literature chemistry to make tetrahydroquinazoline, tetrahydroindazole and tetrahydrobenzothiazole 6,5- and 6,6-fused bicyclic scaffold derivatives. These derivatives were targeted from three simple commercially available ketone starting materials, ethyl 4-oxocyclohexanecarboxylate (**21**: X = CH_2CO_2Et), *N*-Boc-4-piperidone (**22**: X = NBoc), and *N*-4-Boc-

aminocyclohexanone (**23**: X = CHNHBoc), which were subsequently converted into diverse fused ring systems using divergent synthetic protocols (Scheme 2). These included α-substitution with *N*,*N*-dimethylformamide dimethyl acetal (DMF–DMA) to provide the corresponding enones **24–26** (**24**: X = $CHCO_2Et$, **25**: X = NBoc, and **26**: X = CHNHBoc) and α-monobromination with Br_2 to afford α-bromo ketones **27–29** (**27**: X = $CHCO_2Et$, **28**: X = NBoc, and **29**: X = CHNHBoc).

Acid-catalyzed cyclization of enone **24** with *S*-methylisothiourea afforded the 2-(methylthio)-tetrahydroquinazoline **30**, which was subsequently oxidized with *m*-chloroperoxybenzoic acid (*m*-CPBA) to the corresponding sulfone **31**. The reactive 2-(methylsulfonyl) group of **31** was treated with ammonia, dimethylamine, and methylamine to access amines **32** ($R_1 = R_2 = H$), **33** ($R_1 = R_2 = Me$), and **34** ($R_1 = Me$, $R_2 = H$) respectively.

Cyclization of enones **25** and **26** with the appropriate guanidine salts (Scheme 2), guanidine carbonate, 1,1-dimethylguanidine sulfate, methylguanidine hydrochloride provided a corresponding set of minimal amino-derived fragments **35** (X = NBoc, $R_1 = R_2 = H$), **36** (X = CHNBoc, $R_1 = R_2 = H$), **37** (X = NBoc, $R_1 = R_2 = Me$), **38** (X = CHNBoc, $R_1 = R_2 = Me$), **39** (X = NBoc, $R_1 = Me$, $R_2 = H$), and **40** (X = CHNBoc, $R_1 = Me$, $R_2 = H$). Treatment of enones **24–26** with hydrazine hydrate afforded indazoles **41–43** (**41**: X = $CHCO_2Et$, **42**: X = NBoc, and **43**: X = CHNHBoc), whereas coupling of α-bromo ketones **27–29** with

thioacetamide led to thiazoles **44–46** (**44**: X = $CHCO_2Et$, **45**: X = NBoc, and **46**: X = CHNHBoc) (Scheme 2). Further reactions from these common intermediates could also be envisaged, such as reaction of enones **24–26** with hydroxylamine, cycloaddition reactions of the enones **24–26**, or Robinson annulation of **21–23**, followed by subsequent chemistry or building diverse ring systems from the ketone intermediates **21–29** (Scheme 2).

However for this study, having created a platform of diverse scaffolds, we next turned our attention on how to best represent potential larger lead-like arrays, by increasing functional group diversity and applying "capping" chemistry to generate more diverse fragments or to move them into lead-like space. To further diversify the functional groups of the 15 scaffolds in a parallel fashion, esters **32–34**, **41**, and **44** were reduced using lithium aluminium hydride to afford the corresponding alcohols **47–51** or hydrolyzed with NaOH to give acids **52–56**. Acids **53**, **54**, and **56** were further coupled with methylamine or dimethylamine to afford the primary methyl amides **57–59**, or the corresponding secondary dimethyl amides **60–62** respectively (see Scheme 3).

Further examples are included in the Supporting Information. Of course many other derivatizations can be applied to the scaffolds in Scheme 3, such as reductive amination, ether formation, conversion of the carboxylate to a five-membered heterocycle, and so forth.

The above functional group modifications provided a more even sampling of potential pharmacophores for these 6,5- and 6,6-fused bicyclic scaffolds. These initial scaffolds are limited to a specific vector; however, we envisage that the methodology can be extended to different vectors, saturated ring sizes as well as further addition of heteroatoms into the saturated ring system.

Evaluation of the Compounds. PCA (ECFP4) analysis of these diverse fragments showed that they were significantly different from the historical fragment library (Figure 8a). This was considered largely to be as a result of the synthetic design to incorporate a heteroaromatic ring fused to a sp^3-saturated ring with a chiral vector of varied functional groups. PMI analysis was used to evaluate the 3D diversity of the fragments, due their relatively limited numbers of conformations due to their small size (Figure 8b) [30]. These fragments do not probe the more sphere-like region of the PMI plot, as the vectors explored provided a degree of rod- and disc-shaped character, but it is key to emphasize that these fragments are not flat and highly conjugated due to their design strategy (Figure 8c). To ensure that the new fragments were distinct from our historical fragment library, an ECFP4 fingerprint with a Tanimoto cutoff ≤ 0.6 was generally employed, although some compounds with higher Tanimoto similarity were included following visual analysis. Commercial availability (i.e., novelty) was determined

by searching eMolecules (Figure 8f); revealing that many compounds were not commercially available, enabling exploration of novel fragment chemical space.

Evaluation of the commercial availability, shape and diversity for a range of the fragment sets synthesized to date are shown below, together with highlights of selected molecules (Figure 9).

Plating, Quality Control, Solubility, and Stability. Compound quality and handling are key to the successful deployment of any screening library. This is particularly true in the area of fragment-based lead discovery where fragments may be routinely screened at concentrations up to 1 mM. When screening at such a high concentrations small amounts of impurity, can have a significant impact on the false positive rate and can lead to the wasteful deployment of valuable resources. To minimize wasted time and effort spent on following up artifacts, we established the following practices for quality control and compound handling.

Upon synthesis, fragments were routinely analyzed from solid by ^{1}H NMR and LC–MS to confirm identity and ensure appropriate purity (> 95%) before registration in our compound management database. At this point the weighed sample was submitted to our compound-handling group and the data captured in the database used to drive solubilization protocols for preparation of stock solutions and samples for further characterization. Further characterization involved the collection and analysis of one-

dimensional NMR spectra of an approximately 2 mM solution of the compound in phosphate buffer with suppression of the water signal by excitation sculpting [31]. Analysis of this data allowed further confirmation of identity, purity (> 95%) and allowed us to generate an estimate of solubility based on a comparison of the intensity of the compound signal with the DMSO-d_5 peak. Water–LOGSY was also acquired for each compound to assess the levels of self-aggregation [32]. Further physicochemical characterization included kinetic solubility, by assesses the compound solubility in the range of 250 µM to 1 mM, and CHI logD (pH7.4) data on selected fragment. As expected, based on the *in silico* design predictions, the kinetic solubility was overall very good for the majority of the compounds, with only a handful with kinetic solubility between 250 µM to 1 mM (Figure 10a). Since fragment and lead optimization based drug discovery programs commonly use ligand efficiency metrics based on *in silico* predictions of logP and logD, we also obtain a good understanding of the correlation between measured and *in silico* predicted values for a scaffold at an early stage. The coefficient of determination (R^2) for a plot of the *in silico* predicted logD (Stardrop™) versus measured CHI-logD (pH7.4)[33] for fragments with a suitable chromophore, showed a moderate correlation (Figure 10b) [34]. This data is helpful when making choices between fragment hits and facilitates better in-house CHI-logD predictions for hit expansion and lead orientated synthesis.

To ensure that compound quality is maintained throughout the lifetime of the library, a strict compound-handling regime has been implemented. Fragments are solubilized and held as a 200 mM DMSO solution in master plates. These master plates are used to create multiple daughter plates that are sealed and stored in low-humidity, inert atmosphere at –20 °C until required. Once in use, each daughter set is stored at room temperature and used for a maximum of six months. This important part of our compound management workflow minimizes the significant degradation that may be encountered when subjecting screening sets to multiple freeze–thaw cycles [35]. An important part of any fragment screening protocol would be confirmation of activity with fresh material that has been properly checked for identity and purity.

CONCLUSION

A successful academic-based platform has been set up to nurture both undergraduates and graduates to understand the importance of physicochemical properties in the design and synthesis of novel polar fragments. This powerful, yet general design and synthesis platform provides a much needed source of novel high-quality diverse fragments to complement our internal library enhancement efforts. To date this platform has delivered 356 diverse compounds within our selected criteria. Screening of these compounds has already led to the identification of fragment hits and

the results will be presented at a future date. We intend to apply the lessons learned from current and future screens to feedback into the design and selection process. As such, we continue to expand our platforms to synthesize diverse scaffolds from common bulk intermediates, develop diverse monomer sets to support internal drug discovery programs as well as generate novel fragment and lead-like arrays through the training of undergraduates and graduates. Expansion of such platforms within the academic sector could significantly address the historical uneven exploration of chemical space and improve access to high quality and diverse fragment and lead-like chemical matter. Such an effort should significantly improve the probability of success for both academic- and industrial-based translational research and we are open to collaborate with parties interested in adopting or supporting such platforms. Whilst preparing this manuscript, we were encouraged to see the recent publication of a conceptually similar approach by Marsden *et. al.* towards the synthesis of diverse scaffolds that can be elaborated into novel lead-like chemical space [36].

Notes

The authors declare a collaboration with Key Organics (http://www.keyorganics.net/) to make fragments (which have passed the plating, quality control, solubility and stability

criteria) available for purchase, with the view of investing proceeds into additional design and synthesis platforms.

ACKNOWLEDGMENTS

We would like to acknowledge the following experienced synthetic medicinal chemistry mentors: Susan Davis, Caroline Wilson, Neil Norcross, Lauren Webster, Claire Mackenzie and David Foley. In particular, we would like to acknowledge Susan Davis, who mentored a large number of undergraduate students. We would also like to acknowledge the following undergraduate students and hope they enjoyed and will benefit from the training they received: Lucy Holcroft, Karen Wilson, Robert Broome, Sophie Gilbert, Joel Gilbert, Jemma Huxton, James Martin, Rachel Gilroy, Jennifer Auer, Kaoana Oleinik, Hannah Leech, Roddy McNeill, Samuel Reed, Timmy Tarr, Helena Vinas Gaza, Felix Breyer, Antrea Aristotelous, Christopher Hastings and Daniel Bode. We would also like to acknowledge the support of Art Crossman and Linda Morris, in helping us to include a fragment design and synthesis module within the undergraduate laboratory practical. We would also like to thank Dan Fletcher and David Robinson for the fragment library NMR analysis and purity determinations. We would like to thank Jenifer Riley and Lucy Ellis for determining the experimental solubility and CHIlogD. We would also like to acknowledge the contributions Marina Modric, through an EU Leonardo da Vinci work experience grant. We would also like to acknowledge helpful

discussions with Berry Matijssen and Lars Sandberg. We would also like to acknowledge helpful discussions with Andrew Morley from the 3D Fragment Consortium, which led to the adoption of similar activities within the 3D Fragment Consortium. This work is part-funded by the MSD Scottish Life Sciences Fund (AL). As part of an ongoing contribution to Scottish Life Sciences, Merck Sharp & Dohme (MSD) Ltd (known as Merck & Co., Inc., Kenilworth, NJ, USA, in the United States and Canada), has given substantial monetary funding to the Scottish Funding Council (SFC) for distribution via the Scottish Universities Life Sciences Alliance (SULSA) to develop and deliver a high quality drug discovery research and training programme. All aspects of the programme have been geared towards attaining the highest value in terms of scientific discovery, training and impact. The opinions expressed in this research are those of the authors and do not necessarily represent those of MSD, nor its affiliates.

ABBREVIATIONS

Ar, aromatic; CHI, chromatographic hydrophobicity index; DIPEA, *N*,*N*-diisopropylethylamine; ECFP4, extended-connectivity fingerprints; FG, functional group; Fsp^3, sp^3 fraction; HAC, heavy atom count; RotB, rotatable bonds, Water-LOGSY, water-ligand observed via gradient spectroscopy.

REFERENCES

1 Dobson, C.M. (2004) Chemical Space and Biology. *Nature* 432, 824-828

2 Lipinski, C. and Hopkins, A. (2004) Navigating chemical space for biology and medicine. *Nature* 432, 855-861

3 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

4 Lipkus, A.H. et al. (2008) Structural diversity of organic chemistry. A scaffold analysis of the CAS Registry. *J. Org. Chem.* 73, 4443-4451

5 Bemis, G.W. and Murcko, M.A. (1996) The properties of known drugs .1. Molecular frameworks. *J. Med. Chem.* 39, 2887-2893

6 Bemis, G.W. and Murcko, M.A. (1999) Properties of known drugs. 2. Side chains. *J. Med. Chem.* 42, 5095-5099

7 Taylor, R.D. et al. (2014) Rings in drugs. *J. Med. Chem.* 57, 5845-5859

8 Pitt, W.R. et al. (2009) Heteroaromatic Rings of the Future. *J. Med. Chem.* 52, 2952-2963

9 Nadin, A. et al. (2012) Lead-Oriented Synthesis: A New Opportunity for Synthetic Chemistry. *Angew. Chem. Int. Edit.* 51, 1114-1122

10 Laird, T. (2004) Does Industry Get Its Synthetic Methodology on the Cheap? *Org. Process Res. Dev.* 9, 1

11 Scott, D.E. et al. (2012) Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* 51, 4990-5003

12 Congreve, M. et al. (2008) Recent developments in fragment-based drug discovery. *J. Med. Chem.* 51, 3661-3680

13 Fuller, N. et al. (2016) An improved model for fragment-based lead generation at AstraZeneca. *Drug Discov. Today*

14 Keseru, G.M. et al. (2016) Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *J. Med. Chem.*

15 Morley, A.D. et al. (2013) Fragment-based hit identification: thinking in 3D. *Drug Discov. Today* 18, 1221-1227

16 Radoux, C.J. et al. (2016) Identifying Interactions that Determine Fragment Binding at Protein Hotspots. *J. Med. Chem.* 59, 4314-4325

17 Hann, M.M. et al. (2001) Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* , 856-864

18 Van Molle, I. et al. (2012) Dissecting fragment-based lead discovery at the von Hippel-Lindau protein:hypoxia inducible factor 1alpha protein-protein interface. *Chem. Biol.* 19, 1300-1312

19 Babaoglu, K. and Shoichet, B.K. (2006) Deconstructing fragment-based inhibitor discovery. *Nat. Chem. Biol.* 2, 720-723

20 Nazaré, M. et al. (2012) Fragment Deconstruction of Small, Potent Factor Xa Inhibitors: Exploring the Superadditivity Energetics of Fragment Linking in Protein–Ligand Complexes. *Angew. Chem. Int. Ed. Engl.* 51, 905-911

21 Baell, J.B. and Holloway, G.A. (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* 53, 2719-2740

22 Brenk, R. et al. (2008) Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem* 3, 435-444

23 Bruns, R.F. and Watson, I.A. (2012) Rules of identifying potentially reactive or promiscuous compounds. *J. Med. Chem.* 55, 9763-9772

24 Ritchie, T.J. and MacDonald, S.J.F. (2009) The impact of aromatic ring count on compound developability - are too many aromatic rings a liability in drug design? *Drug Discov. Today* 14, 1011-1020

25 Ritchie, T.J. et al. (2011) The impact of aromatic ring count on compound developability: further insights by examiing carbo- and hetero-aromatic and -aliphatic ring types. *Drug Discov. Today* 16, 164-171

26 Hill, A.P. and Young, R.J. (2010) Getting physical in drug discovery: a contemporary perspective on solubility and hydrophobicity. *Drug Discov. Today* 15, 648-655

27 Lovering, F. et al. (2009) Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem.* 52, 6752-6756

28 Yang, Y. et al. (2012) Beyond size, ionization state, and lipophilicity: influence of molecular topology on absorption, distribution, metabolism, excretion, and toxicity for druglike compounds. *J. Med. Chem.* 55, 3667-3677

29 Hall, R.J. et al. (2014) Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.* 116, 82-91

30 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

31 Hwang, T.L. and Shaka, A.J. (1995) Water suppression that works - excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *J. Magn. Reson., Series A* 112, 275-279

32 Dalvit, C. (2009) NMR Methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Discov. Today* 14, 1051-1057

33 Valko, K. (2004) Application of high-performance liquid chromatography based measurements of lipophilicity to model biological distribution. *J. Chromatogr.* 1037, 299-310

34 Young, R.J. et al. (2011) Getting physical in drug discovery II: the impact of chromatographic hydrophobicity measurements and aromaticity. *Drug Discov. Today* 16, 822-830

35 Oldenburg, K. et al. (2005) High throughput sonication: evaluation for compound solubilization. *Comb. Chem. High Throughput Screen.* 8, 499-512

36 Foley, D.J. et al. (2015) A systematic approach to diverse, lead-like scaffolds from α,α-disubstituted amino acids. *Chem. Commun.* 51, 11174-11177

Figure 1. The importance of synthetic methodology to provide fragment, lead-like and drug-oriented libraries to initiate and drive translational drug discovery.

Figure 2. The effect of "capping" a carboxylic acid can give rise to a number of different pharmacophores.

Figure 3. Nazaré et al.[20] case study showing the superadditive effect of linking amide and sulfonamide functionalized fragments to afford a potent factor Xa inhibitor.

Figure 4. Typical workflow. The precise order of steps was sometimes varied.

Figure 5. Strategy for compiling fragment and lead-like libraries.

Figure 6. Example monomer sets and minimal functional group transformations.

Figure 7. Database substructure searches for selected 6,5- and 6,6-fused bicyclic scaffolds with a functional group, based on commercial availability as well as property selection filter (MW ≤ 250, LogP ≤ 2.5, HBD ≤ 3, HBA ≤ 6, PSA ≤ 90 Å, RotB ≤ 3).

Figure 8. (a) PCA of synthesized diverse scaffolds (green) vs original fragment library (grey). (b) PMI plot of synthesized diverse scaffolds (green) vs original fragment library (grey). (c) Fsp3 of synthesized diverse scaffolds. (d) Intralibrary Tanimoto similarity of synthesized diverse scaffolds using the ECFP4 fingerprint. (e) Fsp^3 of original fragment library. (f) Commercial availability of synthesized diverse scaffolds (eMolecules).

Figure 9. (a) PCA of capped monomer sets (green) vs original fragment library (grey). (b) PMI plot of capped monomer sets (green) vs original fragment library (grey). (c) Fsp^3 of capped monomer sets. (d) Intralibrary Tanimoto similarity of capped monomer sets using the ECFP4 fingerprint. (e) Fsp^3 of original fragment library. (f) Commercial availability of synthesized capped monomer sets (eMolecules).

Figure 10. (a) Kinetic solubility versus *in silico* predicted logD. (b) *in silico* predicted logD versus measured CHI-logD (pH7.4), showing the line of unity.

Scheme 1. Acetylation and mesylation of the amine monomer set and inventory amines: Reagents and conditions: (a) CH_3COCl, DIPEA, DCM, rt; (b) CH_3SO_2Cl, DIPEA, DCM, rt.

Scheme 2. Synthesis of 6,5- and 6,6-fused bicyclic scaffold derivatives. Reagents and conditions: (a) DMF–DMA, Et_3N, microwave irradiation at 130 °C for **24**, or DMF–DMA, toluene, 100 °C for **25** and **26**; (b) Br_2, $AlCl_3$, MeCN (**27**) or EtOAc (**28** and **29**), 0 °C → rt; (c) *S*-methylisothiourea, 4 M dioxane HCl solution, DMSO, 130 °C; (d) *m*-CPBA, DCM, rt; (e) 33% aqueous NH_3 solution, dioxane, microwave irradiation at 160 °C for **32**, or dimethylamine and methylamine, THF, microwave irradiation at 160 °C for **33** and **34**; (f) guanidine carbonate, KOAc, EtOH, microwave irradiation at 100 °C for **35** and **36**, or 1,1-dimethylguanidine sulfate, and methylguanidine hydrochloride, Cs_2CO_3, DMSO, 80 °C for 37–40; (g) NH_2NH_2, H_2O, EtOH, 40 °C (**41**), or rt (**42** and **43**); (h) thioacetamide,

DMF, 80 °C for **44**, or thioacetamide, EtOH, microwave irradiation at 100 °C for **45** and **46**.

Scheme 3. Capping chemistry to increase diversity. **Reagents and conditions**: (a) $LiAlH_4$, THF, rt; (b) NaOH, MeOH, reflux; (c) methylamine (**57–59**) or dimethylamine (**60–61**), DIPEA, PPA, DCM, rt; (d) TFA, DCM, rt; (e) CH_3COCl (**64** and **67**) or CH_3SO_2Cl (**65** and **68**), DIPEA, DCM, rt.

Chart 1. Selected fragments from acetylation of the amine monomer set

Chart 2. Selected fragments from mesylation of amine monomer set

Table 1. Descriptors and guide selection criteria employed for non-covalent reversible fragment libraries, monomer sets and lead-like libraries

Peter Ray

Peter Ray graduated from University of Cape Town then completed an industrial MSc fellowship working on estradiol analogues. Moved to the UK to complete his PhD at the University of Liverpool before moving to the University of Nottingham for a lecturer position. Joined Amura Therapeutics working on Cathepsin inhibitors, then Organon (which became Schering Plough-MSD) working as a project leader in multiple therapy areas. Moved to UCB as a Group Leader and is currently the Tuberculosis Project lead at the University of Dundee Drug Discovery Unit where he has helped evolve the fragment and library expansion programs.

Ian Gilbert

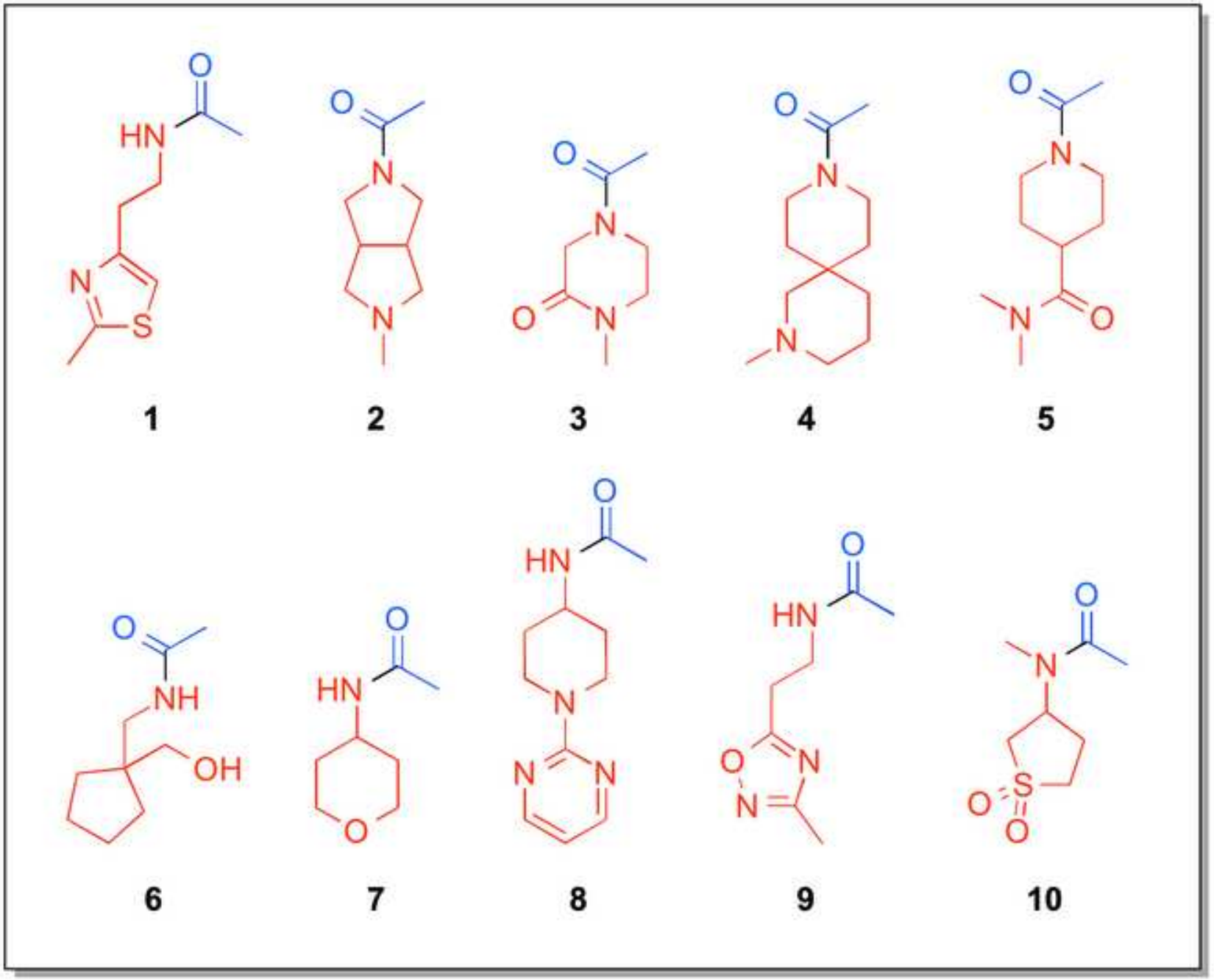
Ian is Head of Medicinal Chemistry in the Drug Discovery Unit, which he helped to set up on moving to Dundee in 2005. Ian started his research career with a PhD in synthetic chemistry, working with Andrew Holmes at the University of Cambridge (1990). This was followed by a post-doctoral fellowship with Parke-Davis Research, before spending a year teaching chemistry at the University of Zambia. On returning to the UK, Ian worked on a chemical biology project at the University of Cambridge, before establishing his own independent medicinal chemistry research group at the Welsh School of Pharmacy in Cardiff University, focusing on drug discovery for neglected tropical diseases.

Paul Wyatt

Professor Paul Wyatt's role as Head of Drug Discovery Unit, University of Dundee, UK is to develop translational research at Dundee, by bringing together the DDU's expertise of Drug Discovery in the Pharma/Biotech sector with basic academic research to de-risk novel targets for drug discovery and deliver new treatments for diseases including, TB, malaria, cancer and rheumatoid arthritis.

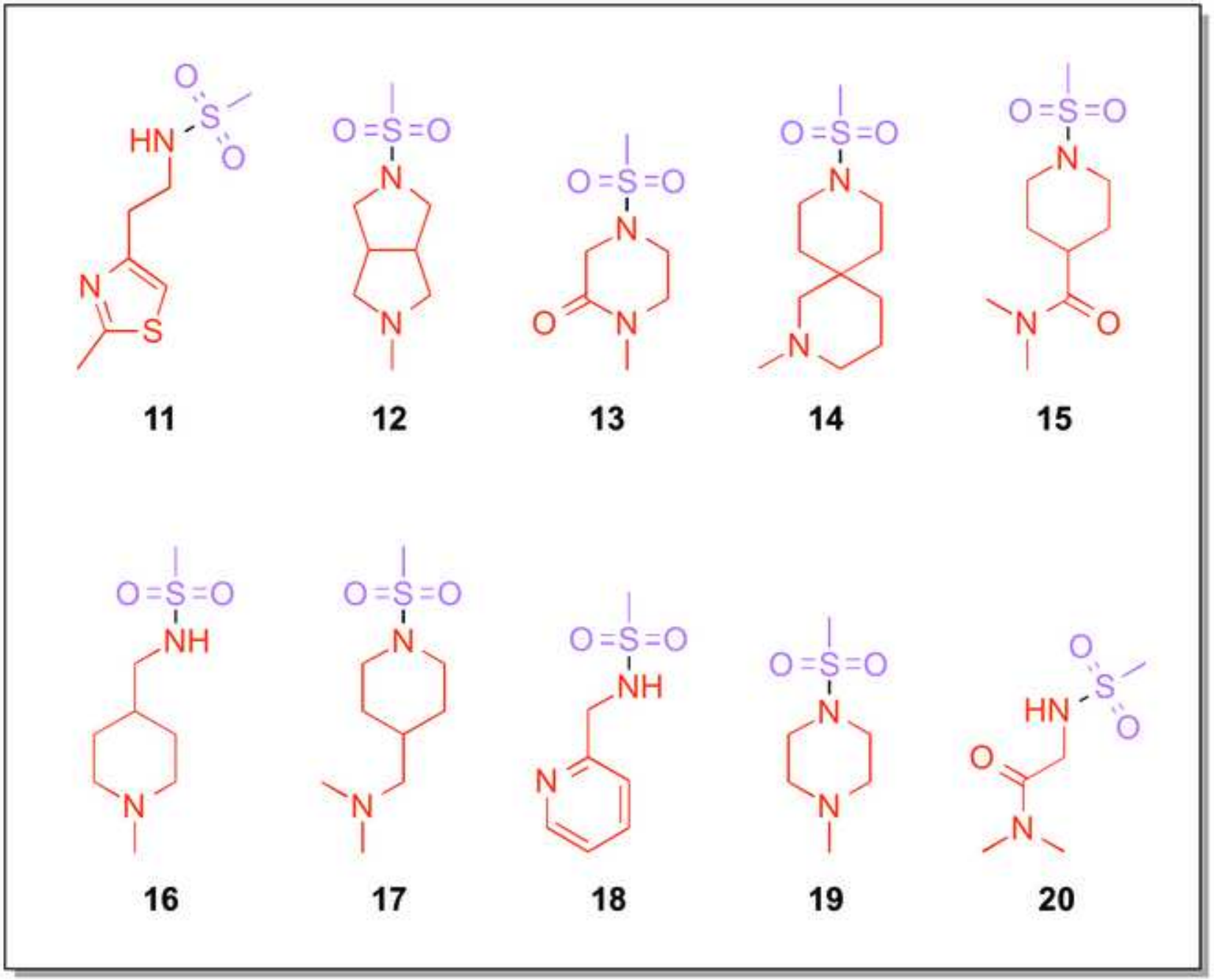
Previously Paul worked in the BioPharma industry for 23 years; playing a significant part in seven compounds entering pre-clinical development. Paul obtained his BSc and PhD in Chemistry from the University of Birmingham. Paul is an author on over 50 papers.

Figure
Click here to download high resolution image



Figure

Click here to download high resolution image



Figure

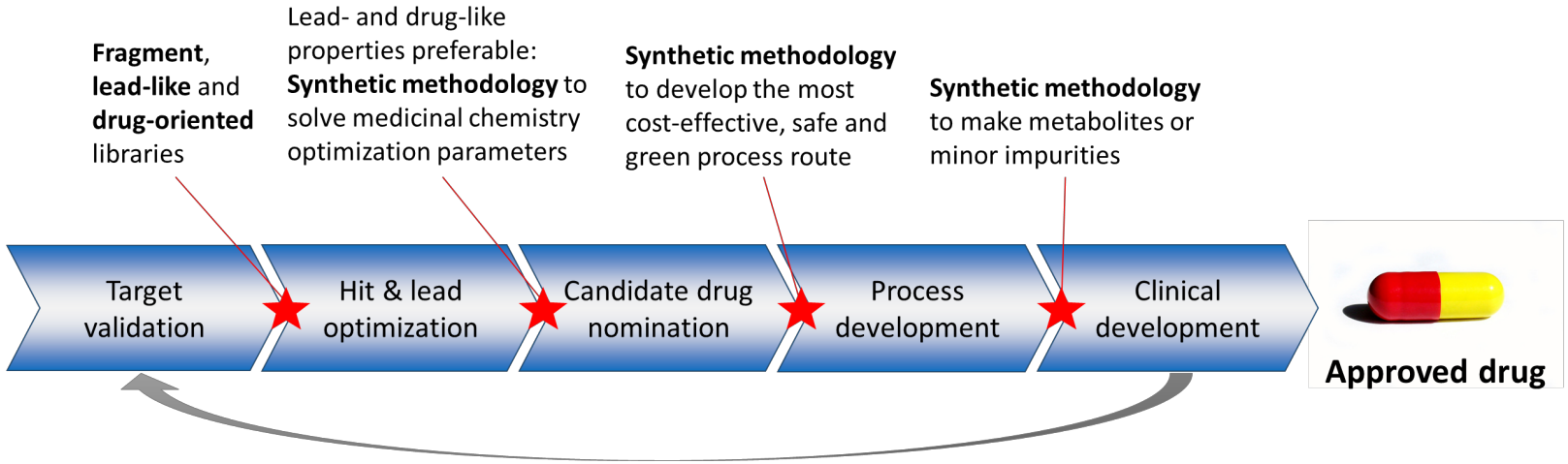

Fragment, **lead-like** and **drug-oriented** libraries
Lead- and drug-like properties preferable: **Synthetic methodology** to solve medicinal chemistry optimization parameters
Synthetic methodology to develop the most cost-effective, safe and green process route
Synthetic methodology to make metabolites or minor impurities
Target validation
Hit & lead optimization
Candidate drug nomination
Process development
Clinical development
Approved drug

Figure
Click here to download high resolution image

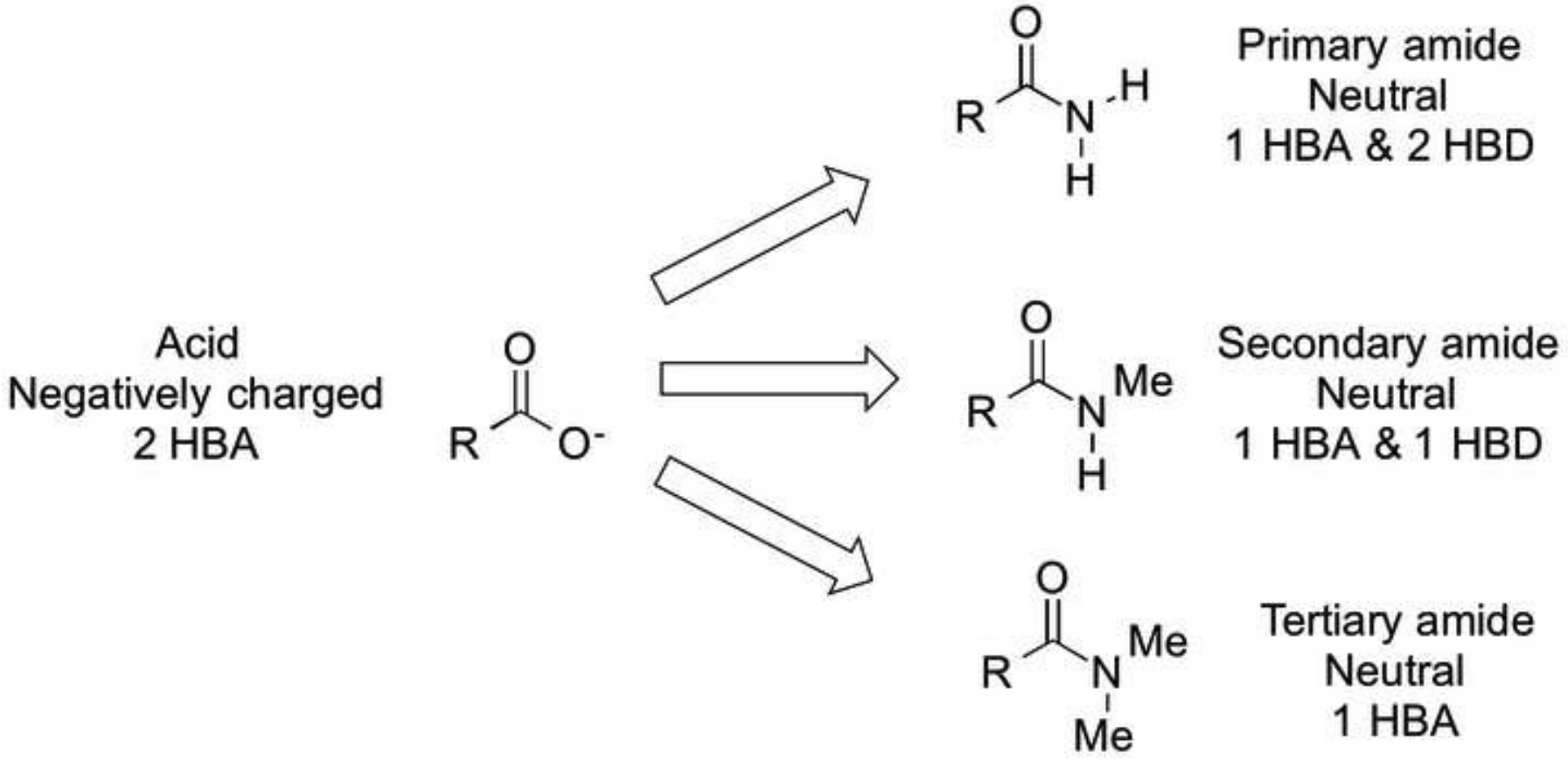
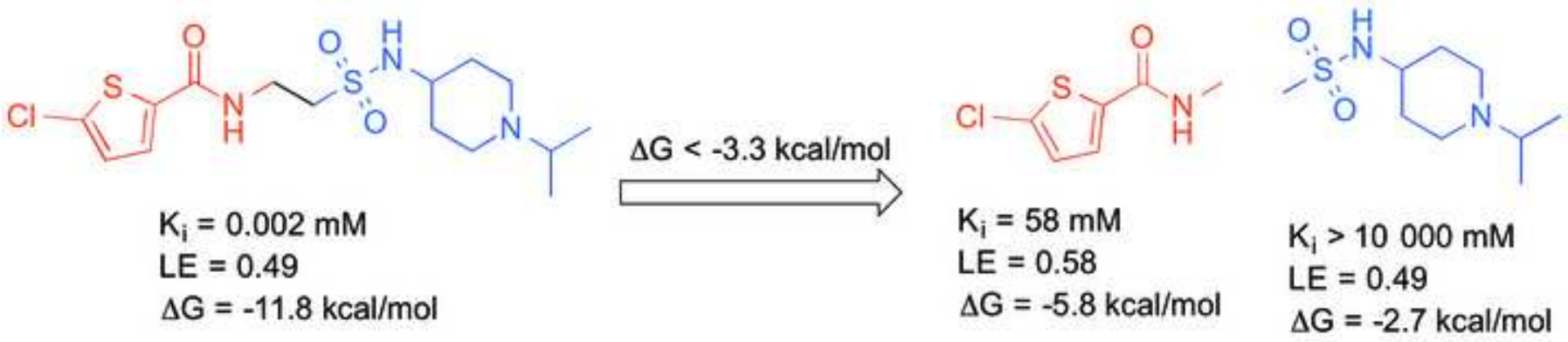
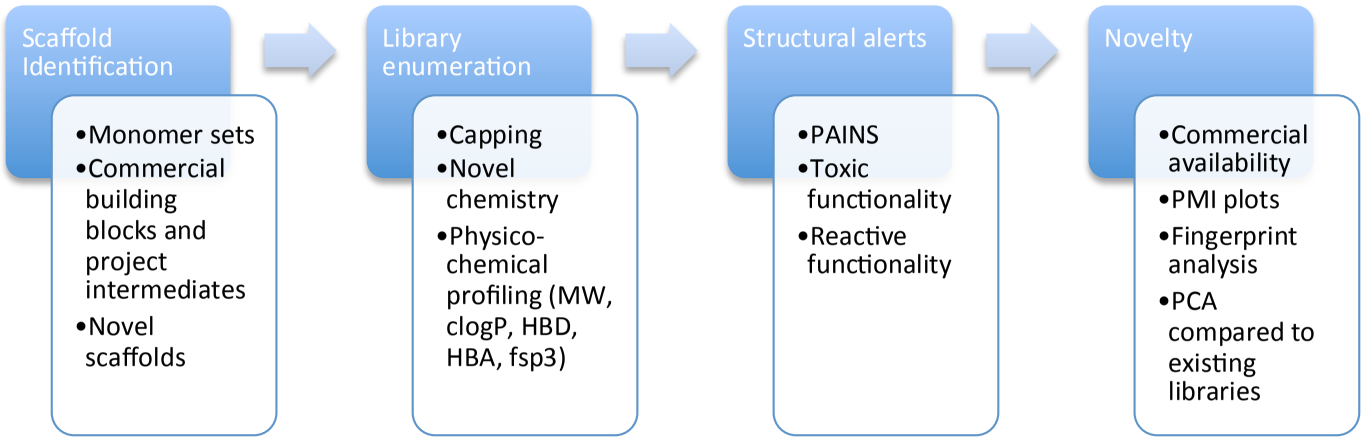


Figure
Click here to download high resolution image



Figure


Scaffold Identification
•Monomer sets
•Commercial building blocks and project intermediates
•Novel scaffolds
Library enumeration
•Capping
•Novel chemistry
•Physico-chemical profiling (MW, clogP, HBD, HBA, fsp3)
Structural alerts
•PAINS
•Toxic functionality
•Reactive functionality
Novelty
•Commercial availability
•PMI plots
•Fingerprint analysis
•PCA compared to existing libraries

Figure

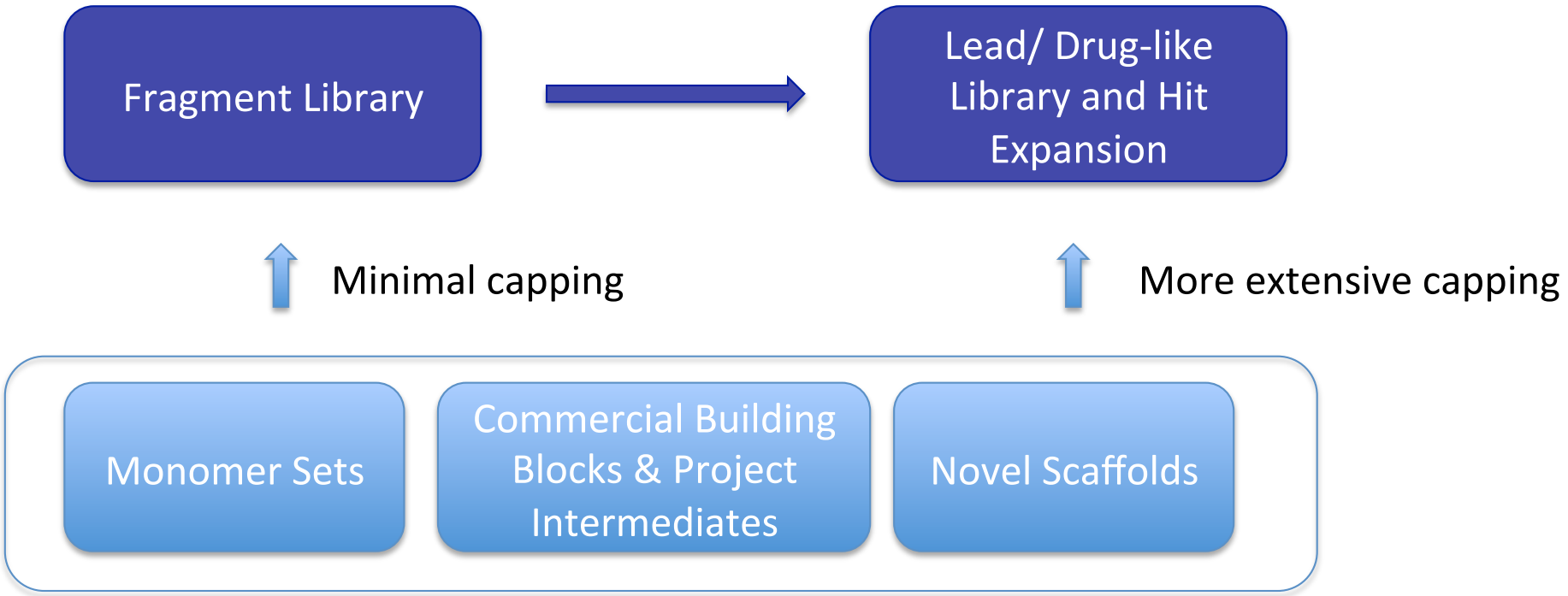

Fragment Library
Lead/ Drug-like Library and Hit Expansion
Minimal capping
More extensive capping
Monomer Sets
Commercial Building Blocks & Project Intermediates
Novel Scaffolds

Figure
Click here to download high resolution image

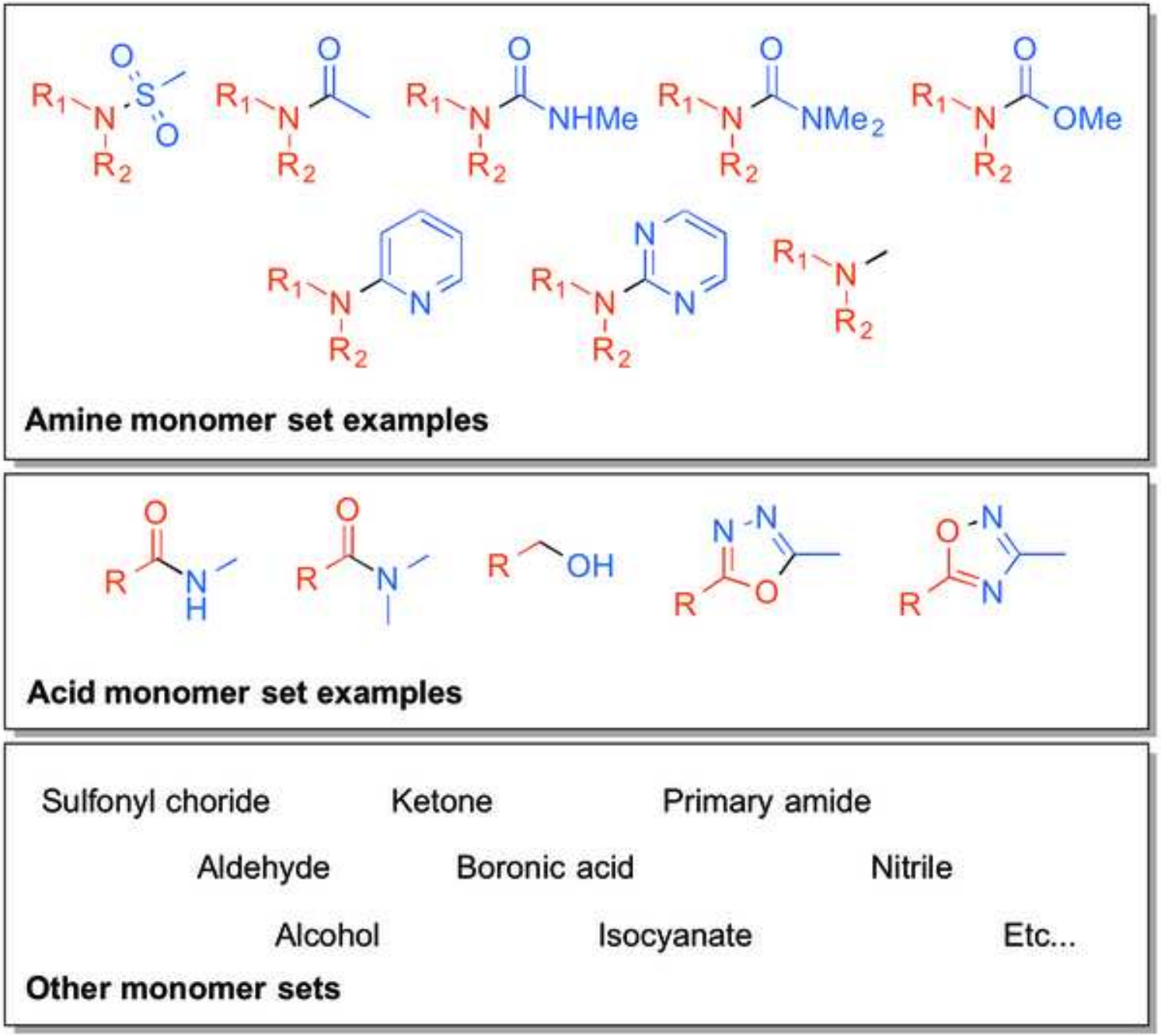
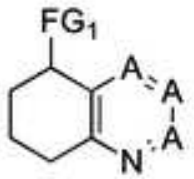
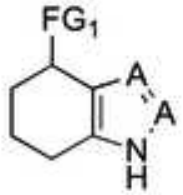


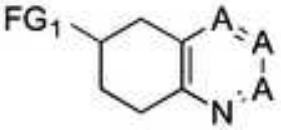
Figure
Click here to download high resolution image



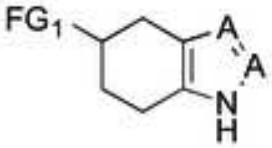
$FG_1 = NH_2$: 145
$FG_1 = CO_2H$: 9
$FG_1 = H$: 2150



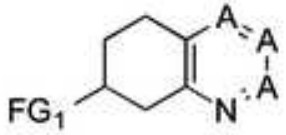
$FG_1 = NH_2$: 153
$FG_1 = CO_2H$: 61
$FG_1 = H$: 1494



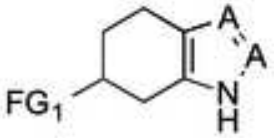
$FG_1 = NH_2$: 39
$FG_1 = CO_2H$: 73



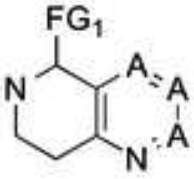
$FG_1 = NH_2$: 163
$FG_1 = CO_2H$: 96



$FG_1 = NH_2$: 10
$FG_1 = CO_2H$: 7

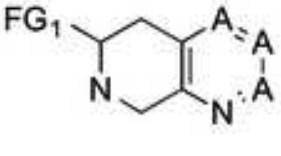


$FG_1 = NH_2$: 202
$FG_1 = CO_2H$: 153

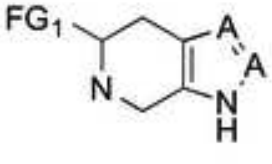


$FG_1 = CO_2H$: 8
$FG_1 = H$: 1605

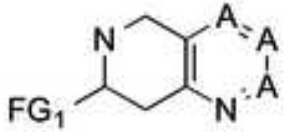
$FG_1 = CO_2H$: 10
$FG_1 = H$: 1370



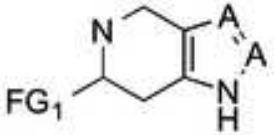
$FG_1 = CO_2H$: 1
$FG_1 = H$: 719



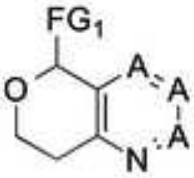
$FG_1 = CO_2H$: 40
$FG_1 = H$: 1000



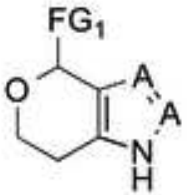
$FG_1 = CO_2H$: 2



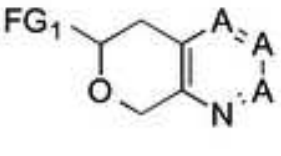
$FG_1 = CO_2H$: 24



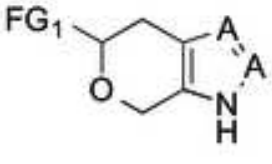
$FG_1 = CO_2H$: 0
$FG_1 = H$: 255



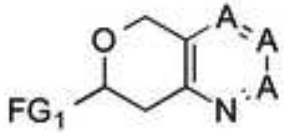
$FG_1 = CO_2H$: 0
$FG_1 = H$: 217



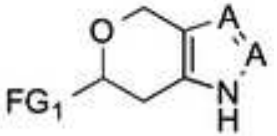
$FG_1 = CO_2H$: 0
$FG_1 = H$: 78



$FG_1 = CO_2H$: 0
$FG_1 = H$: 64

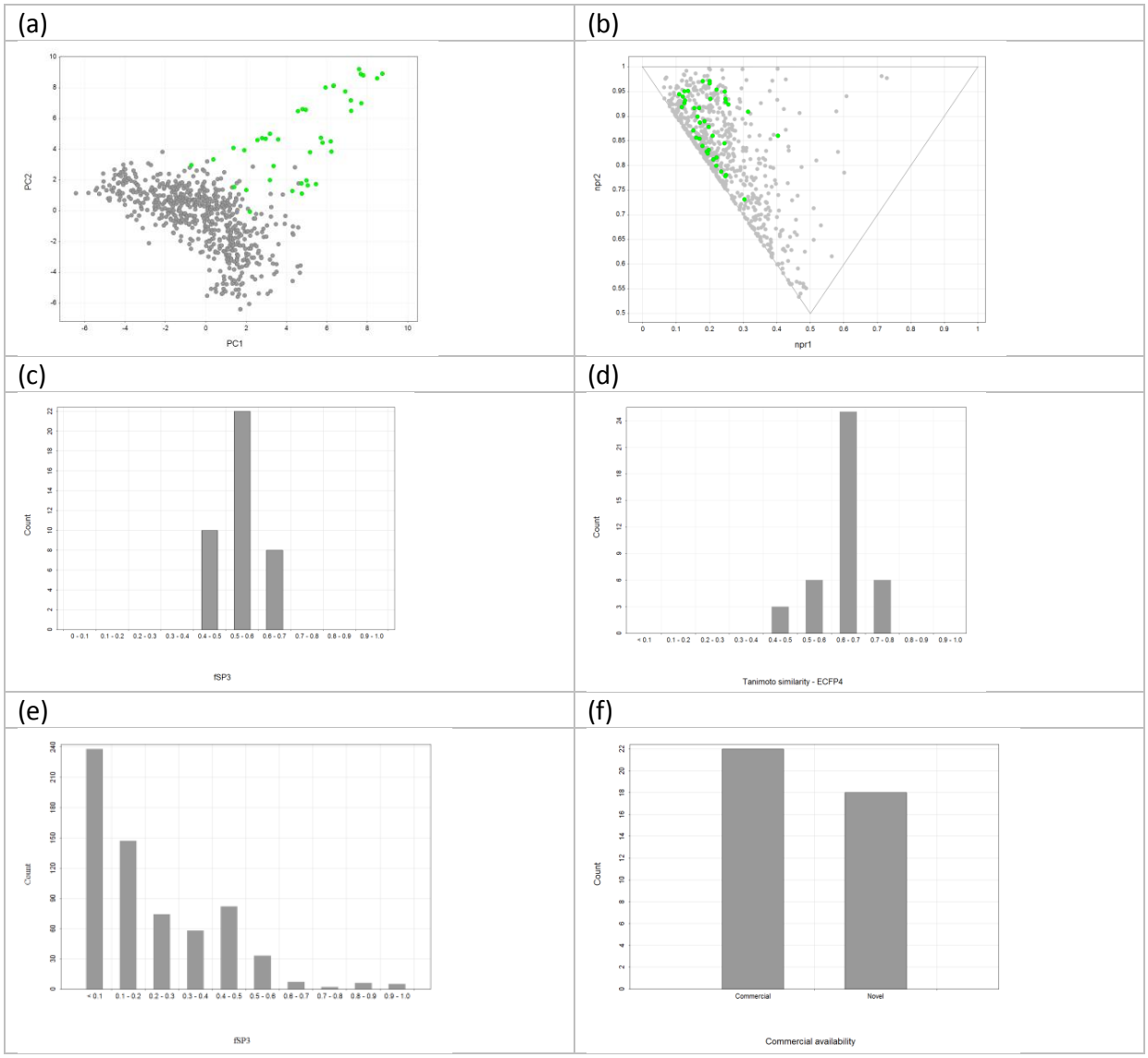


$FG_1 = CO_2H$: 0

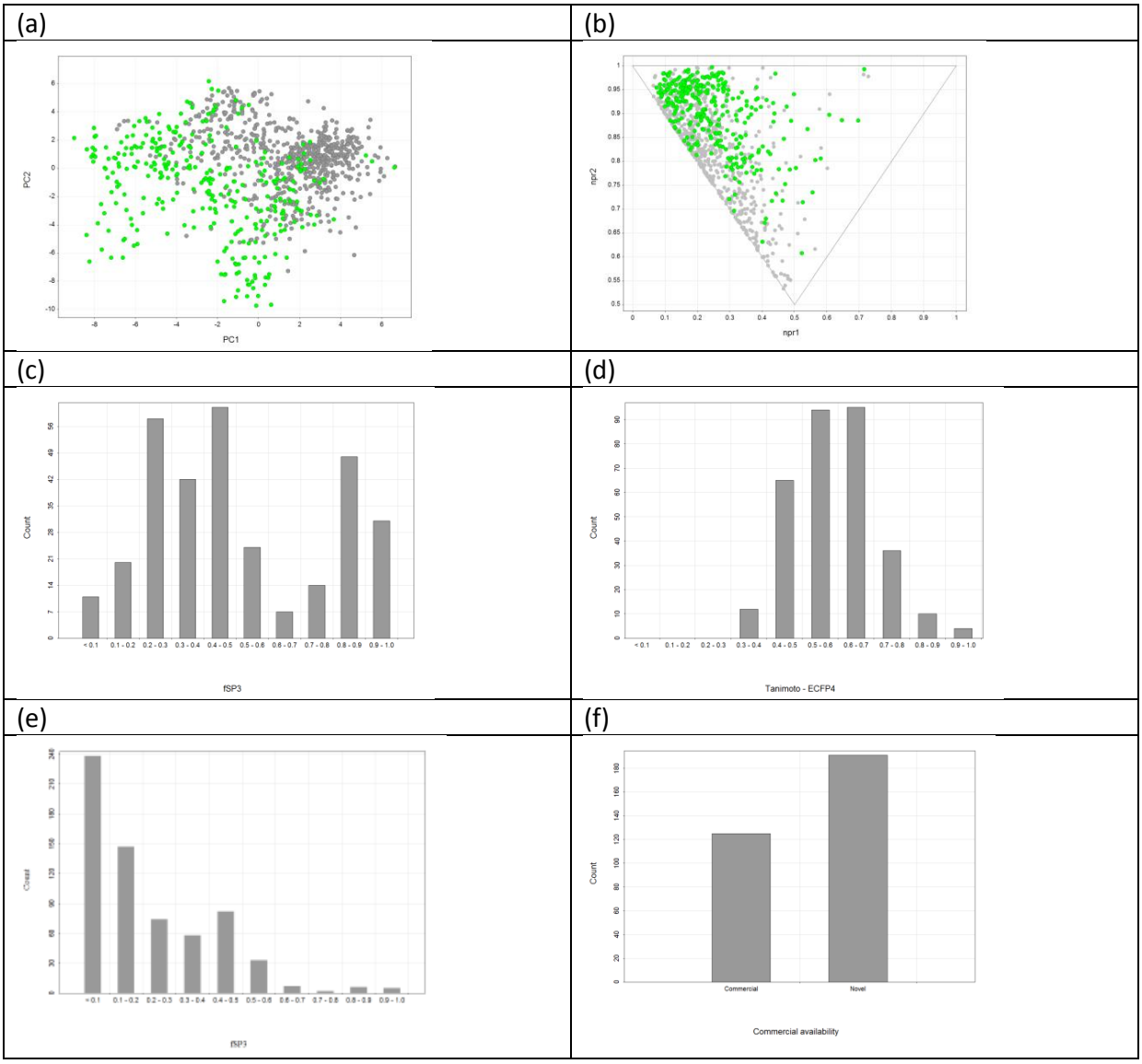


$FG_1 = CO_2H$: 0

Figure


(a)
PC2
PC1
(b)
npr2
npr1
(c)
Count
fSP3
(d)
Count
Tanimoto similarity - ECFP4
(e)
Count
fSP3
(f)
Count
Commercial
Novel
Commercial availability

Figure


(a)
PC2
PC1
(b)
npr2
npr1
(c)
Count
fSP3
(d)
Count
Tanimoto - ECFP4
(e)
fSP3
(f)
Count
Commercial
Novel
Commercial availability

Figure

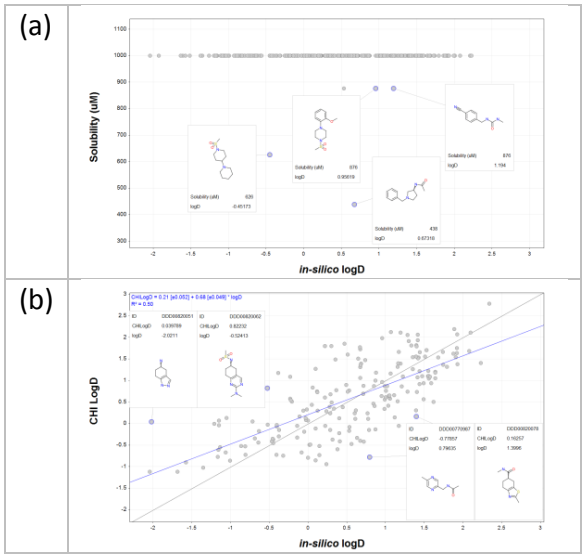


Figure
Click here to download high resolution image

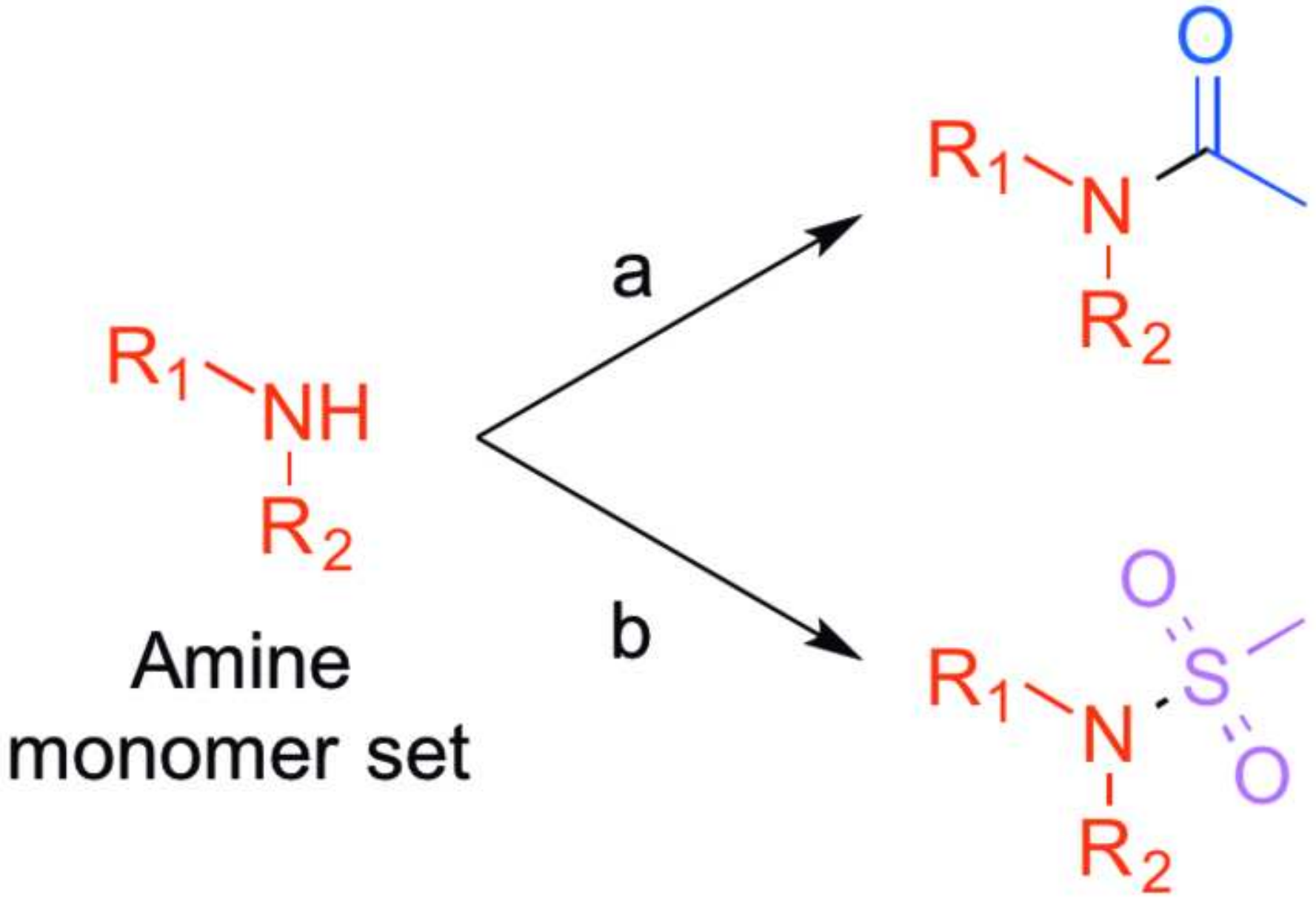


Figure
Click here to download high resolution image

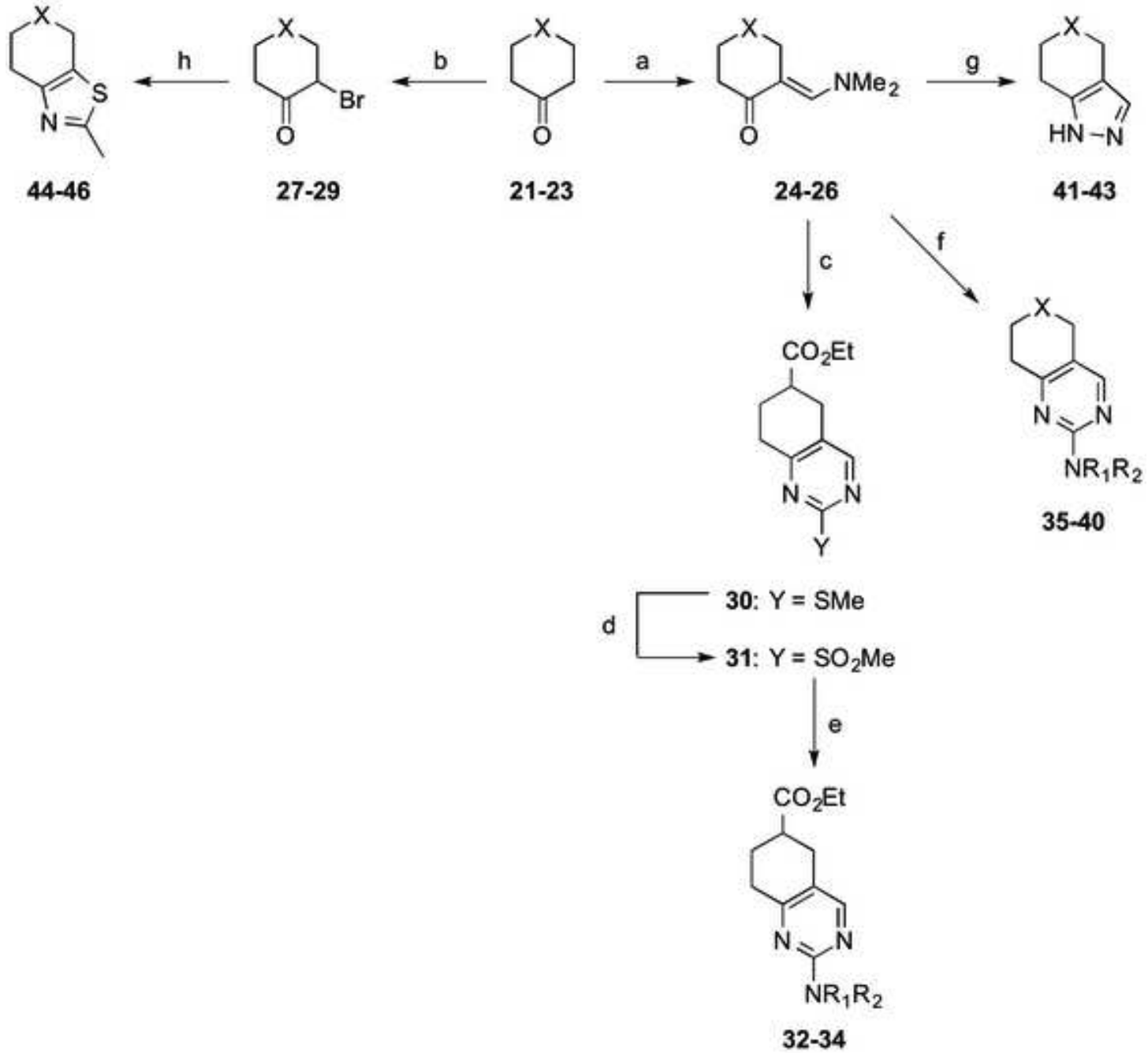
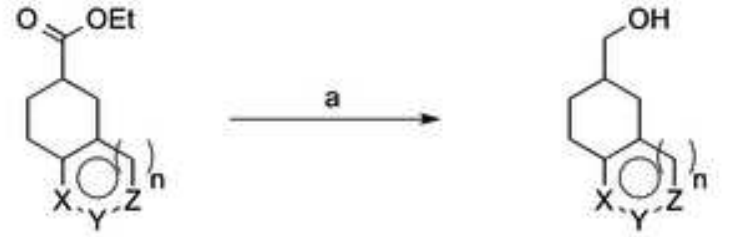
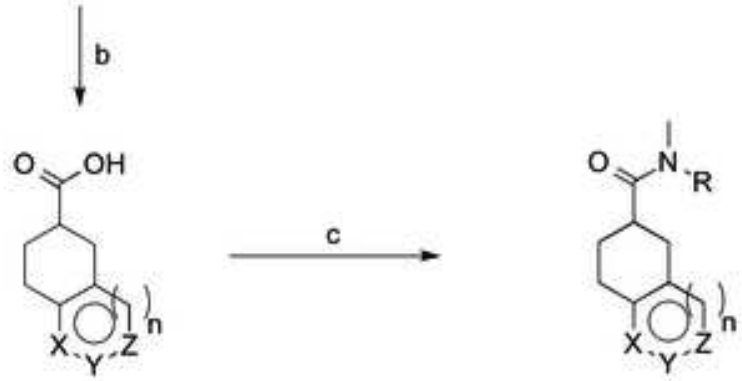


Figure
Click here to download high resolution image



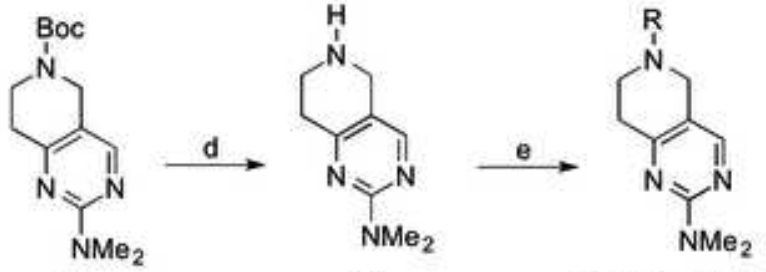
32 (n = 1, X = N, Y = CNH_2, Z = N)
33 (n = 1, X = N, Y = $CNMe_2$, Z = N)
34 (n = 1, X = N, Y = CNHMe, Z = N)
41 (n = 0, X = NH, Y = N, Z = C)
44 (n = 0, X = N, Y = CMe, Z = S)

47 (n = 1, X = N, Y = CNH_2, Z = N)
48 (n = 1, X = N, Y = $CNMe_2$, Z = N)
49 (n = 1, X = N, Y = CNHMe, Z = N)
50 (n = 0, X = NH, Y = N, Z = C)
51 (n = 0, X = N, Y = CMe, Z = S)

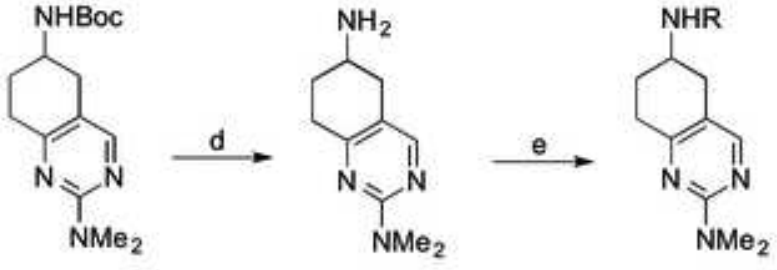


52 (n = 1, X = N, Y = CNH_2, Z = N)
53 (n = 1, X = N, Y = $CNMe_2$, Z = N)
54 (n = 1, X = N, Y = CNHMe, Z = N)
55 (n = 0, X = NH, Y = N, Z = C)
56 (n = 0, X = N, Y = CMe, Z = S)

57 (n = 1, X = N, Y = $CNMe_2$, Z = N, R = H)
58 (n = 1, X = N, Y = CNHMe, Z = N, R = H)
59 (n = 0, X = N, Y = CMe, Z = S, R = H)
60 (n = 1, X = N, Y = $CNMe_2$, Z = N, R = Me)
61 (n = 1, X = N, Y = CNHMe, Z = N, R = Me)
62 (n = 0, X = N, Y = CMe, Z = S, R = Me)



37 **63** **64** (R = COMe)
65 (R = SO_2Me)



38 **66** **67** (R = COMe)
68 (R = SO_2Me)

e-component

Click here to download e-component: Supporting Information.pdf

e-component

Click here to download e-component: DDU_structural_alerts_DDT.sdf

e-component

Click here to download e-component: Lilly_structural_alerts_DDT.sdf

e-component

Click here to download e-component: Pains_structural_alerts_v2_DDT.sdf

Table

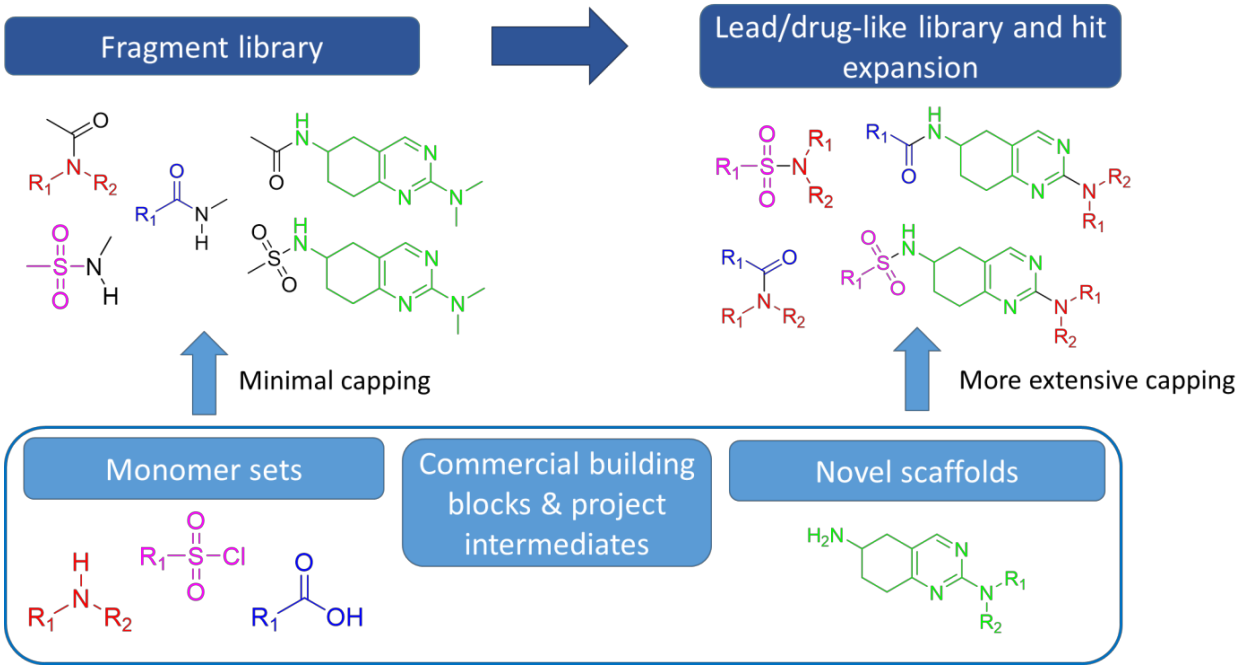
Table 1. Descriptors and guide selection criteria employed for non-covalent reversible fragment libraries, monomer sets and lead-like libraries

Fragment library	Monomer sets	Lead-like library
HAC 5–18 LogP ≤ 2.5 LogD ≤ 2.5 HBD ≤ 3 HBA ≤ 6 PSA ≤ 90 Å RotB ≤ 3 Ar rings ≤ 3	MW ≤ 200 LogP ≤ 2 HBD ≤ 3 RotB ≤ 3 Ar rings ≤ 3	HAC 14–26 LogP –1 to 3 Ar rings ≤ 3
Commercial availability, shape & diversity		
eMolecules sp^3 Content PMI (used for fragment library to impart understanding of shape where rotatable bonds ≤ 3) PCA (ECFP4) ECFP4 fingerprint Medicinal chemists eye (Vortex)		
Structural alerts		
PAINS, Eli Lilly and DDU integrated set		

Physicochemical data were calculated using StarDrop™ (www.optibrium.com). clogD values were calculated at pH 7.4

Graphical Abstract (for review)

Graphical Abstract



Highlights

- A platform for the production of novel, high quality fragment and lead-like libraries.
- Parameters for the design of monomer sets, fragments and lead-like libraries.
- Synthesis of multiple diverse scaffolds from common intermediates.
- Quality analysis of the fragments.