Lead Discovery for Hotspots –FBDD Informed by Smart Fragment Libraries and Expanded Chemical Space

Fragment based drug discovery (**FBDD**) is a strategy for identifying target-specific leads by screening libraries composed of small, low molecular weight molecules and subject them to cycles of synthetic elaborations. Fragment screening libraries are generally compiled using ligand based physicochemical property criteria such as molecular weight, lipophilicity, number of H-bond acceptors and donors etc.[1] When reviewing fragment starting points, promiscuous binding due to the compactness of the fragments was often identified as an issue. However, a recent analysis of active fragments from 35 campaigns on 20 targets at Novartis indicated that 63% of the screened fragments had never been observed as hits,[2] while privileged fragments appeared as high-value library members active on more than one target.

Fragment screening libraries – the value of information content

Aiming to create a library that is an entry to large, yet unbiased and diverse, chemical space the **Comprehensive Fragment Library (CFL)** was developed by BioBlocks. The CFL is a set of small, rigid, pharmacologically relevant, fragment-like virtual molecules derived from a starting set of over 3 million potentially synthesizable fragments which all conform to strict fragment definitions[1]. With over 580k virtual compounds and 830k diastereomers, each characterized by their target interaction patterns, the library covers a vast and diverse area of fragment chemical space. Each virtual fragment contains a single methyl group acting as an anchor point, called a 'handle', representing a vector for synthetic expansion. The library is 3D enabled, *i.e.* all molecules in the library are clustered based on their 3D conformations derived by Cresset software.[8] This allows us to explore the chemical space around a particular fragment based on both two-dimensional and three-dimensional properties.

A representative subset of the CFL has been generated as a physical library, the available **CFL screening library,** currently 160 compounds. Half of the fragments are novel, exclusive to BioBlocks as it proved necessary to supplement available compounds with *de novo* synthetic compounds to provide library members with rare pharmacophore patterns. Each physical compound represents >1000 additional fragments with related core structures, 3D structures, and interaction patterns[6].

The **SpotXplorer library** was designed to capitalize on the structurally conserved nature of binding hotspots[3] and the conservation of pharmacophores found in certain target classes[4]. As the number of distinct structure-based pharmacophores found in fragment-binding crystal structures is limited, a small, efficient fragment library is able to recognize this diversity.[5] Indeed, we found that a non-redundant set of 425 pharmacophores (with 2-4 H-bond donors/acceptors, ionic centers, aromatic rings, *etc.* each) represents the full scope of binding patterns found in >3000 experimental protein-fragment complexes.

A **pilot SpotXplorer library (labeled SX0)** was designed to optimize coverage for the experimental binding pharmacophores with commercial compounds. A compiled set of only 96 compounds displayed 76 % of the 2-point and 94 % of the 3-point pharmacophores. This SX0 library provided diverse fragment hits against well-known drug targets such as GPCRs and proteases (Fig. 2). The active hits included 70% of the known pharmacophores for these targets based on their reported ligands in ChEMBL. The library also produced hits against

newly identified target proteins, such as the SETD2 methyltransferase and two viral proteins of SARS-CoV-2: the main protease 3CLPro, and the NSP3 macrodomain.[5]

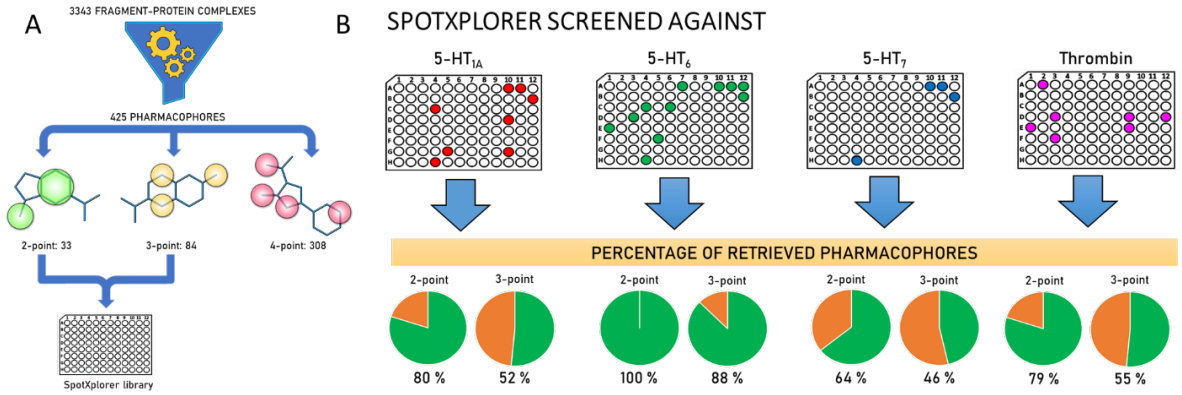

Figure 2 A) 425 pharmacophores represent the variety of binding patterns found in >3000 publicly available fragment-protein complexes. Due to the limited complexity of fragments, we optimized the SpotXplorer library for the coverage of 2- and 3-point pharmacophores. B) The SX0 library provides diverse fragment hits (highlighted in the relevant plate positions) against a panel of three GPCRs and thrombin. The hits represent the majority of binding patterns observed in the known fragment-sized ligands of these targets (4-11 SX0 hits *vs.* 43-81 fragments reported in ChEMBL).

Combining SpotXplorer pharmacophores with tractable, unbiased CFL chemical space

An improved strategy for lead finding and development utilizes both screening libraries presented in this paper. The approach consists of initial screening with the **SpotXplorer library** utilizing its efficient pharmacophore coverage to identify the key binding pharmacophores that are mirrored in the physical CFL inventory and the power of the CFL's built-in follow-up strategy to find analogues for those hits.

As **CFL** and **SXL** fragments encode pharmacophore and chemical profiles in different ways, we have established a correspondence between the two definitions and utilized it to interrogate the virtual CFL starting from either library. The virtual CFL space is then available to explore the wider chemical space around the confirmed hits, thus nominating further compounds for hit-to-lead follow-up. By creating this connection between libraries, we are not only connecting ligand-based and structure-based fragment spaces, but also enabling a powerful and focused lead finding strategy.

Hit follow up – Fragment Analog Workflow

The CFL screening library has been screened against a wide range of targets including SGK1, HPGDS and NSD2[7]. Follow-up of these hits invariably led to low µM activity compounds with high ligand efficiency suitable for lead development. Application of the hit follow-up protocol for SGK1 hits has resulted in a novel, potent lead series[7,9].

Following up on **CFL fragment(s)** to find more hit-to-lead candidates involves making fingerprint-based searches in the virtual **CFL** space. These can result in a large number of **CFL** fragments (Fig. 1B), with varying degrees of similarity to the originally identified fragment(s). The follow-up process, carried out in multiple cycles, investigates fragment analogues, possibly revealing new, unexpected chemotypes and scaffolds.

Established CFL protocols enable follow-up from any fragment hit set (Fig. 1).

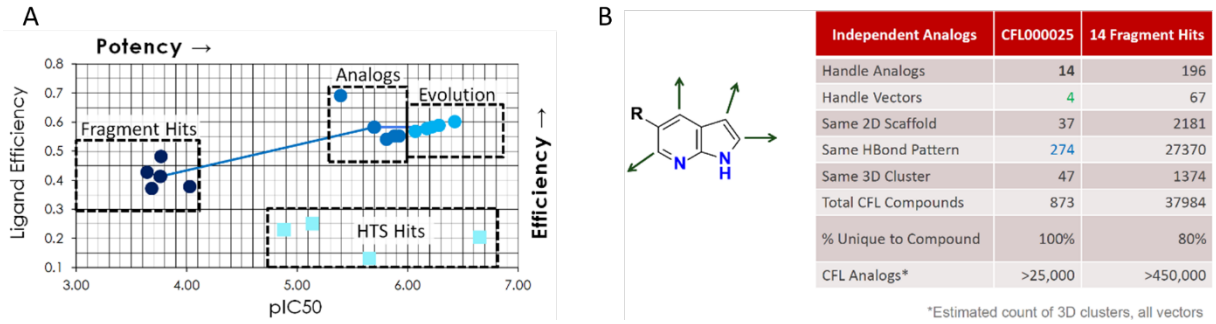

Figure 1 A) Project data summarizing the CFL's standard follow-up protocol improvement in potency starting from >100 µM fragment hits. B) The distribution of analogue fragments in the CFL for the representative fragment hit CFL000025.

For an active set of analogue fragments, the best handle orientations can be identified to allow the expansion of the set to larger molecules with enhanced interaction to their binding pocket, and expanding the pharmacophore set while maintaining the features associated with activity. Further exploration of analogues by expansion of the chemical space is also possible with Syntheverse™, a proprietary tool of BioBlocks Inc. One such approach involves iterating over all handle positions, allowing the fragment to be expanded in all possible directions. With Syntheverse™, expansion of fragments can be performed by finding all possible reactions the compound can take part in and using a curated list of commercially available set of reagents to generate a set of synthetically possible reaction products.

In summary, the combination of **SXL** and **CFL** libraries leads to a highly data driven fragment-based approach for finding chemotypes, scaffolds and synthetic pathways for proprietary lead identification against challenging targets. By using the structure-based **SpotXplorer** screening library to identify initial hits with pharmacophore patterns, then the vast **CFL** defined virtual space to follow-up on the hits, one can achieve a focused, streamlined lead discovery process without the need to screen large, unconnected screening libraries or biased fragment collections.

References

1 G. M. Keserű, D. A. Erlanson, G. G. Ferenczy, M. M. Hann, C. W. Murray and S. D. Pickett, *J. Med. Chem.*, 2016, **59**, 8189–8206.

2 P. S. Kutchukian, A. M. Wassermann, M. K. Lindvall, S. K. Wright, J. Ottl, J. Jacob, C. Scheufler, A. Marzinzik, N. Brooijmans and M. Glick, *J. Biomol. Screen.*, 2015, **20**, 588–596.

3 C. Ehrt, T. Brinkjost and O. Koch, *J. Med. Chem.*, 2016, **59**, 4121–4151.

4 T. Fehlmann and M. C. Hutter, *J. Chem. Inf. Model.*, 2019, **59**, 1314–1323.

5 D. Bajusz, W. S. Wade, G. Satała, A. J. Bojarski, J. Ilaš, J. Ebner, F. Grebien, H. Papp, F. Jakab, A. Douangamath, D. Fearon, F. von Delft, M. Schuller, I. Ahel, A. Wakefield, S. Vajda, J. Gerencsér, P. Pallai and G. M. Keserű, *Nat. Commun.*, 2021, **12**, 3201.

6 P. V. Pallai, W. S. Wade, *US Patent 9946847 B2*, 2018.

7 P. V. Pallai, W. S. Wade, *Leap-To-Lead Summary [White Paper]*, 2019, https://www.bioblocks.com/s/BioBlocks-Leap-to-Lead-White-Paper-2p-09262019.pdf

8 W. S. Wade, K. Chang, P. V. Pallai, P. Tosco, J. Zapf, L. Linguardo and G. Alton, *Design and Synthesis of the Comprehensive Fragment Library*. Poster presented at: Cresset UGM, 2017 Jun 29-30.

9 S. T. Meyer, W. S. Wade, J. W. Zapf, J. Gerencsér, B. Gyimóthy, *WO Patent 2020023393 A1,* 2020.