Fragment-based drug discovery: opportunities for organic synthesis

Jeffrey D. St. Denis, Richard J. Hall, Christopher W. Murray, Tom D. Heightman and David C. Rees

This Review describes the increasing demand for organic synthesis to facilitate fragment-based drug discovery (FBDD), focusing on polar, unprotected fragments. In FBDD, X-ray crystal structures are used to design target molecules for synthesis with new groups added onto a fragment via specific growth vectors. This requires challenging synthesis which slows down drug discovery, and some fragments are not progressed into optimisation due to synthetic intractability. We have evaluated the output from Astex’s fragment screenings for a number of programs, including urokinase-type plasminogen activator, hematopoietic prostaglandin D2 synthase, and hepatitis C virus NS3 protease-helicase, and identified fragments that were not elaborated due, in part, to a lack of commercially available analogues and/or suitable synthetic methodology. This represents an opportunity for the development of new synthetic research to enable rapid access to novel chemical space and fragment optimisation.

Introduction

In view of the increasing interest in and success of fragment-based drug discovery (FBDD), this Review describes the current chemistry challenges within the field: the design of new fragments and the elaboration of weakly binding fragments into nM leads guided by X-ray crystal structures. Ideally, synthetic elaboration of fragment hits in three-dimensions from many different growth vectors is experimentally worked out prior to fragment screening. This will increase the chance of success during fragment-to-lead optimisation.

As the field of small molecule drug discovery has advanced, the demand for molecular complexity has increased in line with ambitions to modulate the functions of increasingly complex human protein systems. This evolution has resulted in calls to the synthetic organic chemistry community for advances in synthesis methodology to keep pace with the demands of modern drug design in an attempt to avoid situations where desired molecules cannot be synthesised or, more commonly, are avoided in favour of...
designs that are more accessible. Such calls are being met by recent advances in broadly impactful organic synthesis methodology including transition metal catalysed couplings, electrocatalysis, photocatalysis, and C–H activation together with new technologies such as high-throughput experimentation (HTE), flow chemistry, and artificial intelligence.

**Fragment-based drug discovery**

Over the past 20 years fragment-based drug discovery (FBDD) has become widely used in pharma, biotech and academic institutes to identify over 40 compounds in clinical trials and 4 launched drugs pexidartinib, vemurafenib, erdafitinib, and venetoclax. In FBDD, the binding of low molecular weight fragments to their target protein is typically characterised by high resolution X-ray crystallography. This is critical not only in understanding and optimising the interactions underlying fragment binding, but in providing insights for the progression into high affinity leads.

The fragment-to-lead process involves the bespoke design of small molecules with 3D shape and electrostatics complementary to the target protein. In order to engage in additional interactions with the protein, substituents need to be added to the starting fragment at precise positions referred to as growth vectors (Fig. 1). Clearly, the direction and synthetic tractability of growth vector elaboration are critical in defining the suitability of a fragment for further development, together with other properties relevant to drug discovery. Furthermore, fragment elaboration may also reveal cryptic subpockets that result from residue movement to accommodate binding of the ligand. In addition to these factors, commercial availability of closely related analogues with exemplified growth vectors, heterocycle core modifications, or well-established scaffold modifications are important to consider when selecting which fragment to consider as a suitable starting point for a fragment-to-lead program.

For fragment optimisation, access to close analogues of the fragment hit determines the speed with which fragments can be evaluated and prioritised. This involves the analysis of protein fragment X-ray crystal structures to identify suitable growth vectors for the introduction of common functional groups onto the fragment and should be limited to a HAC (heavy atom count) ≤ 6 (Fig. 2). Frequently, closely related analogues can be obtained from commercial suppliers to determine structure activity relationships (SAR). However, this process can be a limiting factor for unexemplified fragments given the timelines of many commercial drug discovery projects. Recently, several groups have explored synthetic methods to address this issue, such as the concept of ‘poised fragments’ by Brennan and colleagues which utilise pre-functionalised fragments and the Spring group with derivatization of DOS-derived fragments.

In this Review, we describe a retrospective comparison of fragments from screening campaigns at Astex Pharmaceuticals. Through this retrospective analysis we have developed a working definition of fragment sociability: an unsociable fragment is one that has limited (if any) synthetic methodology to enable growth vector elaboration and few...
commercially available close analogues. In contrast, a sociable fragment is one supported with robust synthetic methodology that enables every growth vector to be elaborated and a significant number of commercially available close analogues. To illustrate the concept of sociability we identified two X-ray hits from each of the selected programs (uPA, H-PGDS, and HCV NS3 protease-helicase); an unsociable fragment that was down prioritized due to synthetic intractability and a sociable fragment that progressed to a lead compound (Table 1).

**Results and discussion**

**Urokinase-type plasminogen activator (uPA)**

Urokinase-type plasminogen activator (uPA) is a trypsin-like serine protease that catalyses the conversion of plasminogen to plasmin through amide bond hydrolysis. Plasmin is responsible for a number of proteolytic processes that degrade components of the extracellular matrix enabling cellular migration. As such, uPA is linked to metastasis in cancer and therefore a target for therapeutic intervention.

We screened our fragment library against uPA and detected 105 X-ray hits bound in the catalytic site of the enzyme. This enabled the team to identify the essential binding pharmacophore as the charged ammonium or amidine acting as a formal hydrogen bond donor (Fig. 3A). The pharmacophore interacts with the backbone carbonyls of Gly219 and Ser190 as well as forming a salt bridge with the side chain of Asp189. Among the fragments binding to the active site of the enzyme we identified the fragments 1 and 2 (Fig. 3A and C). Additionally, both fragments make beneficial hydrophobic interactions with the protein.

Fragment 1 was a ligand efficient starting point (LE > 0.51) for fragment-to-lead program. The X-ray structure of fragment 1 bound to uPA identified several potential growth vectors from the saturated ring. Fortuitously, during the initial screening campaign Abbott Laboratories (now Abbvie) reported on the
and led to the lead compound amide bond formation. This facilitated rapid SAR gathering with an IC50 = 0.07 μM (LE = 0.31) (Fig. 3B and C). Although commercially available from several suppliers, the exact synthetic route of fragment 1 is not reported and close analogues that were not commercially available, therefore early SAR screening identified a strong polar interaction within the lipophilic pocket.58

This was an unexpected result as H-PGDS is known to bind to-lead. This fragment does not undergo a formal hydrogen bond donor or acceptor interaction but has a beneficial π-π-stacking interaction with Trp104 (Fig. 4B). As with all fragment 1 was not selected for optimisation against uPA.

In contrast to fragment 2, the orally active drug mexiletine was identified as a more attractive fragment hit for follow-up.67 While the clogP of 2 was quite high, this fragment has clear developability and substantial chemistry reported in the literature, there was inherent confidence in this fragment as a suitable starting point. Thorough exploration of the active site pocket from the 4-position of the fragment was enabled by chemistries such as Suzuki–Miyaura cross-coupling and amide bond formation. This facilitated rapid SAR gathering and led to the lead compound 7, a potent inhibitor of uPA with an IC50 = 0.07 μM (LE = 0.31) (Fig. 3B and C).64

Hematopoietic prostaglandin D2 synthase (H-PGDS)

H-PGDS is an enzyme responsible for the isomerisation of prostaglandin H2 to prostaglandin D2 (PGD2).68 The biological effects of PGD2 include vasodilation, bronchoconstriction, inhibition of platelet aggregation among others. As such, H-PGDS has been indicated as a target for allergic rhinitis and other inflammatory disorders.69

Fragment screening against H-PGDS identified 76 fragments binding to the catalytic site of the protein.70 These fragments contain a similar chemical architecture that compliments the binding pocket with a polar heterocycle linked via a carbon–carbon bond to an aromatic ring. This is illustrated in both fragments 3 and 4 (Fig. 4A and B). Additionally, the pyrazole moiety of 4 binds to the protein through a donor–acceptor interaction with Asp96 and Tyr152. This was an unexpected result as H-PGDS is known to bind lipophilic aryl pharmacophores, whereas our fragment screening identified a strong polar interaction within the lipophilic pocket.58

Similar to compound 2, fragment 3 is an approved oral drug ((+/-)-tetramisole, (+)-dexamisole, and (-)-levamisole),57 and therefore represents an attractive starting point for hit-to-lead. This fragment does not undergo a formal hydrogen bond donor or acceptor interaction but has a beneficial π-π-stacking interaction with Trp104 and excellent shape complementarity with the protein (Fig. 4A). However, the route to 3 is linear and does not lend itself to the rapid generation of analogues, in particular from the sp3-growth vectors of the core bicycle. Thus, limited access to close analogues and a linear synthetic route rendered compound 3 an unsociable fragment. Coupled with the sub-optimal ligand efficiency (LE ~ 0.29) this hit was down prioritised in favour of 4.

Fragment 4 is a weak but ligand efficient inhibitor of H-PGDS and the aromatic substituent of 4 forms a π-π-stacking interaction with Trp104 (Fig. 4B). As with all sociable fragments, there were a substantial number of commercially available analogues that facilitated rapid follow up of fragment 4 from the 3-position of the pyrazole and from the 4-position of the other aromatic ring. Once the commercially available compounds were exhausted, reliable synthetic chemistry enabled rapid progress to the lead.
compound 8 which transitioned into lead optimization (Fig. 4C).58

**Hepatitis C virus NS3 protease-helicase (HCV NS3 protease-helicase)**

The hepatitis C virus genome encodes ten viral proteins that ensure the propagation of the viral particle.71,72 Of these proteins, the NS3 protein is a bifunctional enzyme that contains an N-terminal serine protease domain and a C-terminal helicase domain that are closely associated in the full length protein.73 While there had been extensive studies on the isolated protease domain74 and early reports on the helicase domain,75 we performed the first fragment screen on the full-length protein.

The fragment screen was performed against the HCV NS3 full-length genotype 1b holozyme.76 The output from the screen identified a novel binding site at the interface of the protease-helicase domains.77 It is a relatively lipophilic binding site with acidic amino acid residues (Asp527, Asp79, Glu628) lining the entrance to the tunnel. This was reflected in the prevalence of lipophilic fragment hits, such as compounds 5 and 6 (Fig. 5A and C).

Compound 5, the commercially available efaroxan,59 contains a semi-saturated bicycle with a quaternary stereocentre that interacts with the entrance to the tunnel of the protein (Fig. 5A). This was an essential growth vector to interact with the acidic residues around the entrance of the pocket. Another key growth vector that needs to be synthetically enabled originates from the 5-position of the aromatic ring. While 5 itself is commercially available and there is literature precedent for this compound, there is a limited number of commercially available analogues and the synthetic route is a lengthy linear process.78,79 This necessitates the installation of the desired substitution or a synthetic handle at an early stage of the synthesis which limits throughput and lengthens the design-synthesise-test paradigm of medicinal chemistry, making this an unsociable fragment.80

The HCV NS3 fragment co-complexes of 6 enabled the effective deployment of structure-based drug design (SBDD) to identify viable growth vectors to target specific residues and sub-pockets within the protein to rapidly improve affinity. Examining this structural data, a systematic exploration of the SAR in the tunnel site was carried out whilst maintaining the benzylamine as the binding pharmacophore. This process was facilitated by the large number of commercially available analogues with this motif. Additionally, the facile introduction of the benzylc stereocentre and ease of amine substitution enabled the efficient growth of the ligand around the tunnel entrance. Finally, exploration of the entrance to the tunnel from the benzylc amine growth vector culminated in the identification of the lead molecule 9, a potent allosteric inhibitor (IC50 = 0.1 μM, LE = 0.38) of the HCV NS3 protease-helicase (Fig. 5B and C).77 When the aforementioned factors are considered, it is clear that compound 6 can be classified as a sociable fragment.

**Method for the identification of unsocial fragments**

The fragment network, recently described by Hall and co-workers, is based on a graph database, which is a data structure that is common in social media applications. In social media a node in the network represents a person and each edge in the network represents a friendship between two people. A person with many friendships can be thought of as sociable. In the fragment network a node in the network represents a fragment molecule and an edge represents a relationship between two fragments (based on their similarity). By analogy to social media we denote a fragment with many edges to be a sociable fragment.81

To identify unsocial fragments, we utilized the fragment network on all fragments in the current version of our core screening library (1651 compounds). For each fragment, we interrogated the corresponding node in the fragment network and focused on edges between this node and neighbouring nodes that have a higher heavy atom count, indicating commercially available analogues that are growth vector enabled. By grouping the connecting edges by positional growth vector and comparing the ratio of observed positions at which a fragment is grown to the maximum theoretical number of growth points, we could estimate how many growth points are synthetically accessible. We then ranked the compounds in our fragment library according to this ratio of observed growth points. The number of protein targets against which each fragment had been observed as an X-ray hit was also used to rank the least sociable fragments.

For the purposes of this work, only commercially available fragments available in eMolecules® were considered in the analysis.82 This analysis resulted in the identification of 30 putative unsociable fragments.
The fragment network is our preferred methodology for the rapid assessment of sociability for a large number of fragments, however, sociability can also be assessed by a collection of substructure searches and analysis of individual fragments in standard searching tools such as SciFinder®. To further confirm the initial analysis performed by the fragment network, we performed a manual assessment of the 30 compounds for commercial availability of close analogues on SciFinder®. We subsequently went further and manually assessed the literature for synthetic methods to access the target molecule and analogues thereof within 4-synthetic steps. Robustness of the chemistry with a focus on commonly utilised medicinal chemistry reactions was also considered to target suitable quantities of material. This manual process led to the identification of 12 compounds within our fragment library that we consider as unsociable fragments (Scheme 1).

Synthetic routes could be envisioned for these unsocial fragments, but they may not be consistent with the time constraints of typical drug discovery projects with the pressure of pressing unmet medical needs. Some of these fragments could become more synthetically tractable if minor changes were made to the structure (e.g. introduction of a carbonyl at C-4 of 1 or removal of the –Et from 5) or selecting different growth vectors. However, in general these changes disrupt the protein binding interactions and therefore are not suitable from a SBDD perspective. If there was methodology to selectively functionalize these unsociable fragments at each and every carbon growth vector with the common functional groups added during fragment elaboration (Fig. 2) while maintaining the binding pharmacophore, we would consider this a valuable development in FBDD and organic chemistry as a whole.

The Astex fragment library has been subjected to constant analysis to improve performance. These analyses have resulted in evolution of the library over several generations taking into account these learnings. Two of these key factors – synthetic tractability and synthetic vectors – have influenced the design and chemical space occupied by the library and are important aspects influencing which fragments that are included in the library. This correlates to the low rate of unsociability for the fragments contained within the Astex library. However, if more esoteric fragments were socialized then they would warrant inclusion into our library and provide an opportunity to identify novel starting points for drug discovery.

Identification of false positives – the eye of an organic chemist

Computational analysis (such as our fragment network) can result in sociable fragments being identified as unsociable (apparent unsociable fragment), so it is important to engage a skilled organic chemist prior to assessing if a compound is or is not sociable. As the implications of an apparently unsociable fragment can down-prioritise an otherwise valuable fragment in favour of other chemical matter. For example, this is observed for simple bond disconnections or single-step functional group transformations and will be discussed in more detail below.

One class of false positive is a double scaffold in which two ring systems are joined by a simple linker or single bond (Scheme 2). Such molecules do not score well in the fragment network analysis because the growth points on each ring system are not well represented in commercial databases; yet analogues would be facile to synthesise as the apparent unsociable fragment yields two highly social compounds upon disconnection. These functionalised reagents represent good starting points for diversification through robust chemical transformations.

![Scheme 1](image1)

**Scheme 1** 12 fragments contained within our fragment library are considered unsociable fragments. These are examples of fragments that require organic methodology development to become sociable.

![Scheme 2](image2)

**Scheme 2** Examples of apparently unsociable fragments and the single bond transformation that yields functionalised sociable reagents that enable rapid analogue synthesis via robust organic methods.
A second type of false positive manifests when a simple functional group modification can result in an otherwise social fragment being misidentified. As a representative example, the dihydrobenzothiazine-dioxide (10) is poorly socialised when analysed by the fragment network. However, ‘simplification’ to the reduced dihydrobenzothiazine (11) results in identification of several commercially available analogues. It is important to note that synthetic elaboration of every carbon position of 11 is exemplified thus enabling access to every growth vector of 10 through a single-synthetic transformation step. As such, complex molecules should be simplified to the core scaffold by a single bond-forming or bond breaking chemical transformation to identify near neighbours (Scheme 3). We expect that the recent advances in AI to make a significant impact in this area of fragment sociability and growth vector elaboration in the near future.

Conclusions

FBDD is a key hit-finding technology for drug discovery and has enabled the discovery of several approved drugs. However, as the field of FBDD has developed over the past 20 years it has revealed the need for further development in the field of organic synthesis to successfully functionalise specific growth vectors of polar, unprotected small molecules using medicinal chemistry relevant transformations. This Review is intended to inspire the development of organic methodology targeted at unsociable fragments in an effort to socialise them and thereby facilitate the development of novel medicines.

Conflicts of interest

The authors are employees of Astex Pharmaceuticals.

Acknowledgements

We would like to thank our Astex colleagues, in particular Gianni Chessari, Rachel Grainger, Christopher Johnson, David Twigg, and Andrew Woodhead for their insightful comments during the preparation of this manuscript.

Notes and references

1 (a) C. W. Murray and D. C. Rees, Angew. Chem., Int. Ed., 2016, 55, 488–492; (b) G. M. Keserü, D. A. Erlanson, G. G. Ferencyz, M. M. Hann, C. W. Murray and S. D. Pickett, J. Med. Chem., 2016, 59, 8189–8206.
2 P. C. Ray, M. Kiczun, M. Huggett, A. Lim, F. Prati, I. H. Gilbert and P. G. Wyatt, Drug Discovery Today, 2017, 22, 43–56.
3 P. O’Brien, T. D. Downes, S. P. Jones, H. F. Klein, M. C. Wheldon, M. Atobe, P. S. Bond, J. D. Firth, N. S. Chan, L. Waddelove, R. E. Hubbard, D. C. Blakemore, C. De Fusco, S. D. Roughley, L. R. Vidler, M. A. Whatton, A. J.-A. Woolford and G. L. Wrigley, Chem. – Eur. J., 2020, 26, 8969–8975.
4 N. S. Troelsen, E. Shanina, D. Gonzalez-Romero, D. Danková, I. S. A. Jensen, K. J. Šniady, F. Nami, H. Zhang, C. Rademacher, A. Cuenda, C. H. Gottfredsen and M. H. Clausen, Angew. Chem., 2020, 59, 2204–2210.
5 (a) C. N. Johnson, D. A. Erlanson, W. Jahnke, P. N. Mortenson and D. C. Rees, J. Med. Chem., 2018, 61, 1774–1784; (b) C. N. Johnson, D. A. Erlanson, C. W. Murray and D. C. Rees, J. Med. Chem., 2017, 60, 89–99; (c) P. N. Mortenson, D. A. Erlanson, I. J. P. de Esch, W. Jahnke and C. N. Johnson, J. Med. Chem., 2019, 62, 3857–3872.
6 D. A. Erlanson, I. J. P. de Esch, W. Jahnke, C. N. Johnson and P. M. Mortenson, J. Med. Chem., 2020, 63, 4430–4444.
7 O. O. Grygorenko, D. M. Volochynyk, S. V. Ryabukhin and D. B. Judd, Chem. – Eur. J., 2020, 26, 1196–1237.
8 M. D. Eastgate, M. A. Schmidt and K. R. Fandrick, Nat. Rev. Chem., 2017, 1, 0016.
9 D. C. Blakemore, L. Castro, I. Churcher, D. C. Rees, A. W. Thomas, D. M. Wilson and A. Wood, Nat. Chem., 2018, 10, 383–394.
10 K. Feng, R. E. Quevedo, J. T. Kohrt, M. S. Oderinde, U. Reilly and M. C. White, Nature, 2020, 580, 621–627.
11 R. Jana, T. P. Pathak and M. S. Sigman, Chem. Rev., 2011, 111, 1417–1492.
12 (a) A. J. Lennox and G. C. Lloyd-Jones, Chem. Soc. Rev., 2014, 43, 412–443; (b) J. D. St. Denis, C. F. Lee and A. K. Yudin, ACS Catal., 2015, 5, 5373–5379.
13 R. Dorel, C. P. Grugel and A. M. Haydl, Angew. Chem., Int. Ed., 2019, 58, 17118–17129.
14 (a) D. A. Everson and D. J. Weix, J. Org. Chem., 2014, 79, 4793–4798; (b) C. F. Lee, A. Holownia, J. M. Bennett, J. M. Elkins, J. D. St. Denis, S. Adachi and A. K. Yudin, Angew. Chem., Int. Ed., 2017, 56, 6264–6267.
15 C. Kingston, M. D. Palkowitz, Y. Takahira, J. C. Vantourout, B. K. Peters, Y. Kawamata and P. S. Baran, Acc. Chem. Res., 2020, 53, 72–83.
41. D. Patel, J. D. Bauman and E. Arnold, *Prog. Biophys. Mol. Biol.*, 2014, **116**, 92–100.
42. I.-J. Chen and R. E. Hubbard, *J. Comput.-Aided Mol. Des.*, 2009, **23**, 603–620.
43. D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke and H. Jhoti, *Nat. Rev. Drug Discovery*, 2016, **15**, 605–619.
44. J. A. Johnson, C. A. Nicolaou, S. E. Kirberger, A. K. Pandey, H. Hu and W. C. K. Pomerantz, *ACS Med. Chem. Lett.*, 2019, **10**, 1648–1654.
45. A. R. Leach and M. M. Hann, *Curr. Opin. Chem. Biol.*, 2011, **15**, 489–496.
46. K. Ogata, T. Isomura, S. Kawata, H. Yamashita, H. Kubodera and S. J. Wodak, *Molecules*, 2010, **15**, 4382–4400.
47. C. W. Murray, M. L. Verdonk and D. C. Rees, *Trends Pharmacol. Sci.*, 2012, **33**, 224–232.
48. C. W. Murray, M. G. Carr, O. Callaghan, G. Chessari, M. Congreve, S. Cowan, J. E. Coyle, R. Downham, E. Figueroa, M. Frederickson, B. Graham, R. McNemarin, M. A. O’Brien, S. Patel, T. R. Phillips, G. Williams, A. J. Woodhead and A. J. Woolford, *J. Med. Chem.*, 2010, **53**, 5942–5955.
49. A. J. Woodhead, H. Angove, M. G. Carr, G. Chessari, M. Congreve, J. E. Coyle, J. Cosme, B. Graham, P. J. Day, R. Downham, L. Fazal, R. Feltell, E. Figueroa, M. Frederickson, J. Lewis, R. McNemarin, C. W. Murray, M. A. O’Brien, L. Parra, S. Patel, T. Phillips, D. C. Rees, S. Rich, D. M. Smith, G. Trewartha, M. Vinkovic, B. Williams and A. J. Woolford, *J. Med. Chem.*, 2010, **53**, 5956–5969.
50. C. J. Helal, M. Bundesmann, S. Hammond, M. Holmstrom, J. Klug-McLeod, B. A. Lefker, D. McLeod, C. Subramanyam, O. Zakaryants and S. Sakata, *ACS Med. Chem. Lett.*, 2019, **10**, 1104–1109.
51. F. W. Goldberg, J. G. Kettle, T. Kogej, M. W. D. Perry and N. P. Tomkinson, *Drug Discovery Today*, 2015, **20**, 11–17.
52. O. B. Cox, T. Krojer, P. Collins, O. Monteiro, R. Talon, A. Bradley, O. Fedorov, J. Amin, B. D. Marsden, J. Spencer, F. von Delft and P. E. Brennan, *Chem. Sci.*, 2016, **7**, 2322–2330.
53. D. G. Twigg, N. Kondo, S. L. Mitchell, W. R. Galloway, H. F. Sore, A. Madin and D. R. Spring, *Angew. Chem., Int. Ed.*, 2016, **55**, 12479–12483.
54. S. L. Kidd, E. Fowler, T. Reinhardt, T. Compton, N. Mateu, H. Newman, D. Bellini, R. Talon, J. McLoughlin, T. Krojer, A. Aimon, A. Bradley, M. Fairhead, P. Brear, L. Diaz-Saez, K. McAuley, H. F. Sore, A. Madin, D. H. O’Donovan, K. V. M. Huber, M. Hyvönen, F. von Delft, C. G. Dowson and D. R. Spring, *Chem. Sci.*, 2020, **11**, 10792–10801.
