ELSEVIER

feature

Crystallographic screening using ultra-low-molecular-weight ligands to guide drug design

Marc O'Reilly, Anne Cleasby, Thomas G. Davies, Richard J. Hall, R. Frederick Ludlow, Christopher W. Murray, Dominic Tisi and Harren Jhoti, harren.jhoti@astx.com

We present a novel crystallographic screening methodology (MiniFrags) that employs high-concentration aqueous soaks with a chemically diverse and ultra-low-molecular-weight library (heavy atom count 5–7) to identify ligand-binding hot and warm spots on proteins. We propose that MiniFrag screening represents a highly effective method for guiding optimisation of fragment-derived lead compounds or chemical tools and that the high screening hit rates reflect enhanced sampling of chemical space.

Introduction

Conventional fragment-based drug discovery (FBDD) using X-ray crystallographic screening utilises protein crystal soaking at moderate ligand concentrations (50–100 mM) to identify fragment hits [ligands with non-hydrogen heavy atom count (HAC) <20] bound to a target [1,2]. These fragments are then evolved into drug-like molecules via iterative cycles of medicinal chemistry and structure determination of protein–ligand complexes. Fragment binding is characterised by energetically optimal, and thus highly ligand-efficient [3], interactions at ligand-binding hot spots [4,5] on a protein's surface. Furthermore, fragment soaking interrogates the protein surface in a site-agnostic manner, so detection of binding in unprecedented pockets, with potential utility for drug discovery, is not uncommon [6].

A key benefit of fragment screening relates to the efficiency of chemical space sampling [7]. Molecular complexity and thus the size of molecular space increases with HAC and the chance of observing good complementarity with a protein decreases with the chemical complexity of a ligand [8]. It is therefore argued that fragments should probe chemical space more efficiently than conventional HTS libraries. A logical evolution of the fragment-based approach might be to explore the effects of continued reductions in ligand size [9], perhaps to a level that approaches discrete functional group probes, and well below those generally employed by FBDD practitioners. The caveat is that smaller ligands tend to bind weakly, meaning that hit detection becomes progressively more challenging as HAC decreases; X-ray crystallography is well placed to detect hits with ultra-low affinity but requires exposure of crystals to increasing concentrations of compound to reliably identify these binding events.

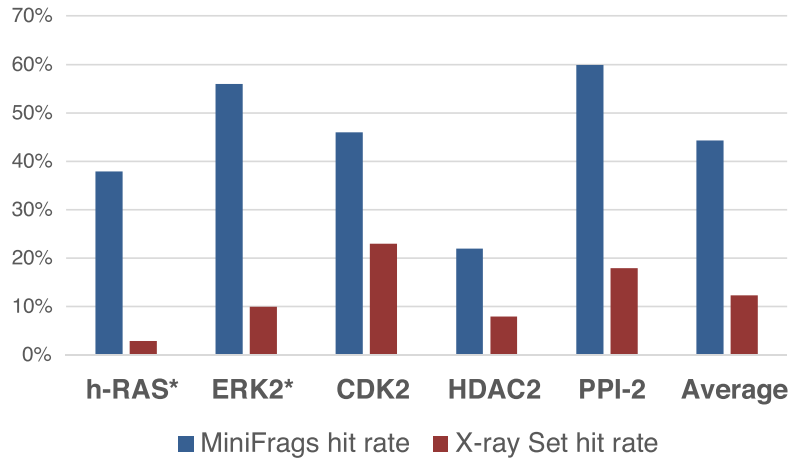
The early work by Ringe and co-workers [10,11] and others [12,13] explored the use of functional group mapping using the multiple solvent crystal structures (MSCS) methodology, which relies upon exposing protein crystals to very high concentrations (1–20 M) of organic solvents. Although the MSCS approach is effective at determining energetically favourable interaction points, few protein crystal systems can tolerate the high organic loads employed during soaking, and the nonaqueous conditions are physiologically less relevant for proteins and can promote protein structure perturbations. Thus, the challenge is to define a crystal-soaking methodology that retains the high sensitivity required to rigorously interrogate a protein's molecular recognition landscape but that can be applied straightforwardly to a variety of systems. Ultimately, the aim is to deliver hit matter that can be used by medicinal chemists and modellers to inform the

https://doi.org/10.1016/j.drudis.2019.03.009

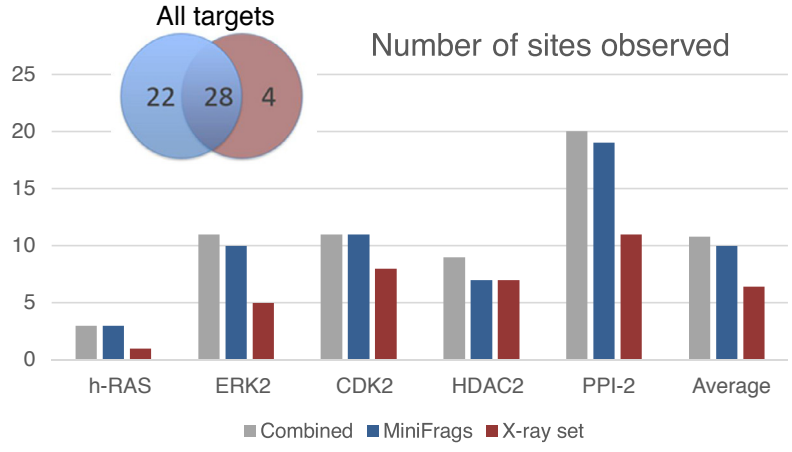
(a)

		Average properties							
Library	N	HAC	MW	clogP	PSA	NDon	NAcc	Nrot	Fsp^3
MiniFrags	81	6.4	94	-0.4	43	1.9	1.5	0.4	0.5
X-ray set	440	10.6	150	0.6	50	1.7	2.0	0.8	0.3

(b)



(c)



Drug Discovery Today

FIGURE 1

MiniFrag statistics. **(a)** Summary of average properties of the MiniFrag and historical X-ray sets used for X-ray screening. Standard deviations are provided in Table S2 (see supplementary material online). N is the number of ligands in each set, HAC the average number of non-hydrogen heavy atoms, MW is average molecular weight (Da) and clogP is the calculated logarithm of the partition coefficient between n-octanol and water. PSA is the calculated polar surface area ($Å^2$). NDon, NAcc, Nrot are the number of H-bond donors, H-bond acceptors and number of rotatable bonds, respectively. Fsp^3 is the fraction of sp^3 hybridised carbons (see MiniFrags section in supplementary material online). 80 of the 81 MiniFrags are HAC = (5–7) and one is HAC = 8. Chemical structures and Smiles strings for all of the ligands within the Astex MiniFrag set are provided in Fig. S9 (see supplementary material online). **(b)** Output summary from the MiniFrag and historical X-ray set screens. A hit is defined as a MiniFrag that binds to a target protein at least once. Targets marked with a * were screened using a historical X-ray set containing 352 fragments, rather than the 440 used for the remaining targets. PPI-2 is a small GTPase target. For all targets, MiniFrags gave substantially higher hit rates than the X-ray set (average MiniFrag and X-ray set hit rates were 44% and 12%, respectively). The difference in hit rates for all targets are highly statistically significant ($P < 0.00003$), although it should be noted that this intercomparison of hit rates is not completely rigorous owing to small differences in the experimental parameters. MiniFrag screening (all targets) identified a total of 68 unique hits and 316 hits total. Table S3 (see supplementary material online) gives a detailed breakdown on how many hits were observed for each MiniFrag against ERK2 and whether the binding was influenced by symmetry copies of the protein. The X-ray set produced a total of 317 hits of which 198 were unique. **(c)** Number of sites identified from screening with MiniFrags (blue bars) or the historical X-ray set (red bars). A site is defined as a location containing two, or more, spatially overlapping hits derived from the unified set of screening results. The grey bars show the combined, total number of non-spatially overlapping, unique sites identified for each target. In all instances MiniFrags identified a greater fraction of the available sites, but there were instances where MiniFrags failed to detect sites that had been highlighted in the X-ray set screen (as can be seen from the Venn diagram above). The Venn diagram refers to the five targets screened and shows the number of sites solely identified by MiniFrags, solely identified from the X-ray set and those sampled by MiniFrags and the X-ray set. Equivalent Venn diagrams for each target are presented in Fig. S1 (see supplementary material online).

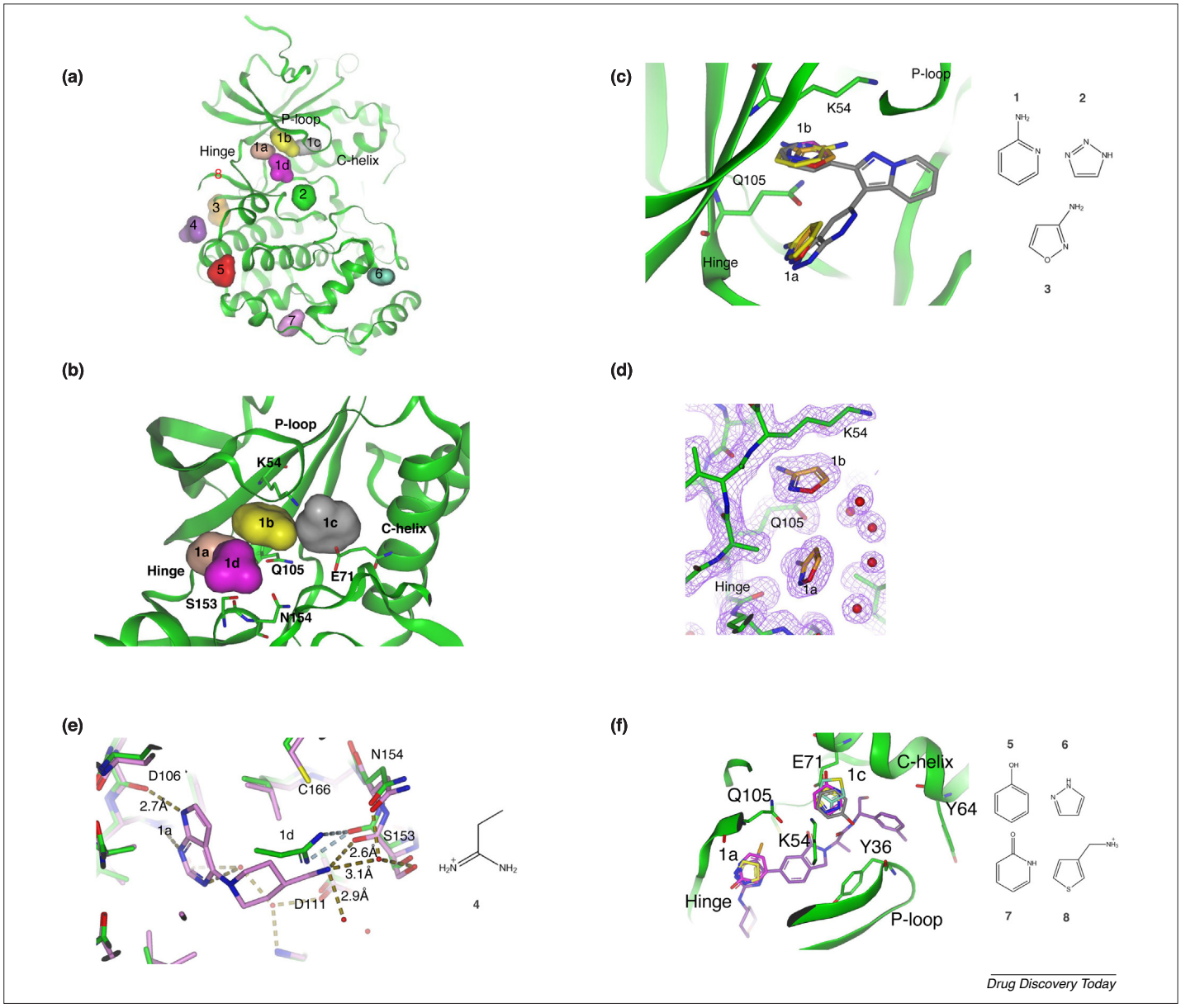


FIGURE 2

ERK2 MiniFrag sites. **(a)** The 11 ligand-binding sites identified on ERK2 (green ribbon representation). Site 1 corresponds to the ATP-binding cleft and has been subdivided into subsites 1a–d. Sites (2–7) are distal to the active site. MiniFrags were observed bound in sites (1–7), but site 8 (red number) was only detected with X-ray set fragments. The ERK2 substrate recruitment sites are referred to as the D- and F-sites [14]. Sites 3 and 4 are within the D-site and site 6 is in the F-site. Sites 4 and 6 are proximal to lattice symmetry contacts. The hits within the biologically relevant, but challenging, D- and F-sites could indicate SAR relevant to the development of chemical probes interrogating these sites. **(b)** ERK2 (green) active site (site 1) with key kinase structural features labelled. Also highlighted are the four MiniFrag binding subsites identified within the active site (subsites 1a–d). 32 MiniFrags were observed bound at the kinase hinge (subsite 1a) and nine MiniFrags were distributed across the three additional subsites (b–d). Q105 is the gatekeeper residue. **(c)** Overlap of MiniFrags in subsite 1b with the phenyl group of ERK2 inhibitor FR180204 [15] (grey: see Fig. S8 in supplementary material online). The phenyl group of FR180204 is accommodated within a pocket formed by subtle sidechain movements of the amino acid residues Q105 (gatekeeper) and K54 (see also Fig. 2b). The three heterocyclic MiniFrags [aminopyridine: yellow (1); triazole: magenta (2); aminooxazole: orange (3), chemical structures shown to the right of the figure] that were observed bound in subsite 1b appear to mimic the interactions formed by FR180204. Binding of compounds **1** and **2** appears to be mediated exclusively via lipophilic and aromatic interactions and **1** and **2** overlap with the phenyl group of FR180204. Compound **3** forms lipophilic/aromatic interactions and H-bonds with the main-chain carbonyl of D106 and the sidechain of Q105. The MiniFrags that bind within subsite 1b also simultaneously bind at the hinge (subsite 1a). Sidechain movements from K54 and Q105 are required to accommodate MiniFrag binding within the subsite 1b cryptic pocket. **(d)** Final refined electron density ($2mF_o$-DF_c contoured at 1σ) for MiniFrag 3 from Fig. 2c that simultaneously binds to subsites 1a and 1b. ERK2 is coloured green, the MiniFrag is orange and water molecules are shown as red spheres. **(e)** Subsite 1d overlay of amidine MiniFrag [green (4) – chemical structure of the protonated state is shown to right of the figure] and a larger tool compound (pink: see Fig. S8 in supplementary material online) structures. The MiniFrag binds proximal to the ERK2 acidic patch (localised around amino acid residues S153, D111, N154), forming two H-bonds with the carbonyl of S153 (Figs. S2 and S3, see supplementary material online), and appears to highlight a propensity for ERK2 to bind basic groups at this site. The tool compound engages the kinase hinge but the primary amine of the piperidine methylamine substituent directly engages the carbonyl of S153 and forms a water-mediated H-bond with the sidechain of S153 (Fig. S4, see supplementary material online). **(f)** Overlay of the MiniFrag cluster in subsite 1c, towards the rear of the ATP-binding pocket, with the protein structure of ERK2 (green) complexed with a late-stage lead [31] (lilac) (PDB code 6g9n; Fig. S8, see supplementary material online). The subsite 1c cluster comprises four MiniFrags [phenol: grey (5); pyrazole:

development of chemical biology tools and guide structure-based design.

Here, we present a crystal-soaking methodology that employs high-concentration (1 M) aqueous soaks of a dedicated library of ultra-low-molecular-weight compounds (typical HAC 5–7) called MiniFrags. The MiniFrag library was specifically constructed to sample chemical space in the ultra-low-molecular-weight regime and comprises 81 chemically diverse and highly soluble ligands, the average properties of which are summarised in Fig. 1a (chemical structures and SMILES strings for all ligands within the Astex MiniFrag set are provided in Fig. S9, see supplementary material online).

MiniFrag screening

We selected five high-resolution crystal systems (≤2 Å) on which to perform a series of proof-of-concept MiniFrag experiments (Crystal systems, see supplementary material online). The proteins included two kinases, two protein–protein interaction (PPI) targets and one metalloenzyme, all of which had been prosecuted previously using a crystallographic screen of our historical X-ray set (Fig. 1a shows the average properties of this X-ray set). The MiniFrags were screened as individual compounds at 1 M, in the absence of any organic solvent, and the X-ray set was screened at 50–100 mM ligand and up to 10% dimethylsulfoxide solvent. Comparative outputs from the X-ray set and MiniFrag screens are summarised in Fig. 1b,c. In all instances the MiniFrag screens displayed higher hit rates (44% average, maximum 60%) and identified more ligand binding sites (10.0 average) than the conventional X-ray set screens (average 12%, maximum 23% and 6.2 sites average).

We will now use ERK2 as a representative example of MiniFrag screening output. There is a substantial amount of ERK2 structural data in the literature [14] and the ERK2 crystal system is robust and well characterised. The MiniFrag and X-ray set screening with ERK2 identified a total of 11 ligand-binding-site clusters (a cluster is defined as a minimum of two spatially overlapping ligands), with ten and five sites being sampled by the MiniFrag and X-ray sets, respectively (Figure 2a). Not unexpectedly, numerous MiniFrag hits were observed in the kinase active site, with hits partitioning across the hinge, and three additional subsites (Fig. 2b). MiniFrag hits in subsite 1b are shown in Fig. 2c,d and they overlay well with, and mimic, the phenyl in the ERK inhibitor FR180204 [15]. Fig. 2e shows an overlay of a MiniFrag hit in subsite 1d with the basic amino group of a larger tool compound, again indicating how MiniFrags appear to recapitulate key interactions of larger molecules. Both the above subsites were not detected in the original X-ray set screen and illustrate how the MiniFrag output could have been used to drive ligand design by introducing functionality to target these areas. The MiniFrags within subsite 1c (Fig. 2f) appear to have identified a pocket that has not been fully exploited for inhibitor design and might be used prospectively to design ligands that occupy this novel pocket. Despite their exceptionally small size, MiniFrags can induce conformational changes. For example, several sidechain movements are observed in subsites 1b and 1c (Fig. 2c and f), and Fig. S7 (see supplementary material online) exemplifies how MiniFrag hits can induce more substantial main-chain movements. The fact that MiniFrag hits were observed at diverse sites on ERK2, including the known substrate recruitment ERK2 D- and F-sites (Fig. 2a), suggests that it might be feasible to use MiniFrags to guide the development of chemical probes aimed at elucidating the potential functionality and allostery of ligand-binding sites on a protein [6].

The concept that compounds with reduced molecular complexity provide efficient sampling of chemical space is central to FBDD [16,17]. As such, our fragment library has progressively evolved to lower average molecular weight, driven largely by analysis of the impact of compound size on historical screening hit rates [7]. In our experience, X-ray crystallography, using soaks at moderate (50–100 mM) concentrations, has proved highly sensitive for the detection of fragments from our historical X-ray set.

We have now extended this concept using 1 M MiniFrag soaks. The concept of exploiting the binding of ultra-low-molecular-weight molecules has similarities with solvent mapping approaches such as MSCS, but we have shown that the more extreme concentrations utilised by MSCS are not required for detection of hits from our MiniFrag set. We therefore believe this is a more widely applicable approach, and that this work is the first time that results have been presented from screening a dedicated fragment library specifically constructed to sample chemical space in this ultra-low-molecular-weight regime.

As proposed by Hann *et al.* [8], smaller compounds are expected to give higher screening hit rates, provided the technique for detecting binding has sufficient sensitivity. Consistent with this, we observed significant increases in hit rates across multiple target classes, and a greater range of binding sites for MiniFrags, compared with those observed from screening our standard fragment library. We do, however, note that the elevated hit rates observed with MiniFrags are likely to be the product of the higher concentrations used for the MiniFrag soaks and more-efficient sampling of chemical space. Interestingly, there are seven fragments common to the MiniFrag set and X-ray set and, for each of these compounds, higher hit rates were observed in the MiniFrag screens compared with the historical screens. The increased hit rates are therefore at least partly attributable to the higher soak concentrations employed in the MiniFrag screens (1 M vs 100 mM) and this is consistent with previous studies by English *et al.* [12].

As exemplified by ERK2, our primary interest has not been the increase in hit rate but rather the ability of MiniFrags to identify additional, weaker points of interaction proximal to more-readily detected fragment binding hot spots. We recently highlighted that these warm spots, despite supporting weaker binding, could nevertheless be productively engaged during medicinal chemistry campaigns to provide substantial increases in affinity [18]. The theoretical prediction of protein–ligand energetics is challenging; but recapitulating patterns of protein–ligand interactions known empirically to increase affinity is a useful approach in structure-based design. Fragment screening is in theory well placed to provide this information but, using standard protocols, it is unusual to observe hits that provide an energetic mapping

yellow (6); pyridinone: magenta (7) and thiophen-3-ylmethanamine: cyan (8) – chemical structures are shown to the right of the figure] (Figs. S5 and S6, see supplementary material online). The four subsite 1c MiniFrags sit within a pocket, at the rear of the kinase active site, that is proximal to the sidechain of E71 and the C-helix. Although the late-stage lead shown accesses the rear of the active site, we were not able to identify examples of ERK2 inhibitors that explicitly access the pocket occupied by the MiniFrag cluster (subsite 1c). We suggest that MiniFrag binding at this site indicates that there could be additional potency available to ligands that can productively access this region of the ERK2 structure. It should be noted that the pyrazole and pyridinone MiniFrags also bind to the ERK2 hinge (subsite 1a).

of a binding site in its entirety. We therefore regard the MiniFrag screen as an experimental approach to help guide more-efficient fragment optimisation through the identification of favourable interactions that are too weak to be detected using standard fragment screening protocols, but which help to characterise target binding sites more fully.

Computational methods have also been used extensively to map the interaction landscape of protein surfaces in an approach analogous to the experimental one adopted here. These methods do not suffer from the practical difficulties associated with high concentration experimental screening (fragment solubility or aggregation and even the possibility of fragment-induced protein denaturation [19]) but they are inherently limited by the underlying accuracy of the methodology. Computational methods have been discussed in detail in a recent review [20] and include molecular dynamics approaches such as MCSS [21], SILCS [22], MDmix [23] and MixMD [24]; or knowledge-based approaches such as GRID [25], Superstar [26] and FTMAP [27]. The MiniFrag approach is complementary to these efforts, and in future should provide more experimental data to validate and improve computational methods.

Another approach to enhance the experimental detection of weak sites is through the incorporation of anomalous scatterers, such as bromine, into the fragments [28]. However, there is a danger that this will limit the number of interactions and extent of chemical space that can be successfully probed. Alternatively, the detection of low-occupancy binding sites can be enhanced using crystallographic techniques such as PanDDA [29]. It will be interesting to explore whether these methods can be used alongside MiniFrag screening to further improve applications to drug discovery.

The chemical simplicity of MiniFrags means that they effectively probe a protein's surface via a very discrete set of interactions and, as such, MiniFrag hits provide an elegant experimental exemplification of the minimal pharmacophore concept [1]. This concept suggests that efficient binding to a specified region of a protein requires a discrete pharmacophore (e.g., a positive charge or H-bond donor–acceptor motif) and that bound fragments usually contain this minimal pharmacophore. Our study begs the question as to whether there could be advantages in screening yet smaller compounds (HAC <5) that contain a minimal pharmacophore; and one might consider water as the ultimate example of such a fragment. However, in general, there could be additional requirements for binding such as aromaticity or lipophilicity that will not be captured by very small polar probes. It is known that lipophilic interactions as well as polar interactions are important to fragment binding [7,30] and we have shown that, although MiniFrags are more polar than normal fragments, they are still capable of sampling lipophilic hot spots and warm spots (Fig. 2c). Our results indicate that, although chemically simple, MiniFrags possess the required diversity of properties to form similar types of interaction to larger fragments, and even have the potential to induce conformational change.

Concluding remarks

In summary, we believe that the use of ultra-small fragments (typical HAC 5–7), called MiniFrags, for crystallographic screening represents a broadly applicable and powerful tool for identifying energetically favourable interaction points on proteins. We advocate using regular fragment screening to identify initial fragments bound to hot spots, and MiniFrag screening to identify proximal warm spots, which should then be targeted with specific functionalities during chemical elaboration of the initial fragments. MiniFrags provide an optimal balance between enhanced screening hit rates and the requirement for hits to have chemical relevance and utility for medicinal chemists and modellers engaged in drug discovery.

Acknowledgements

We would like to thank G. Chessari and D. Norton for help in selecting the MiniFrag set, M. Verdonk for discussion and statistical analysis of the MiniFrag screening output, P. Day, P. Pathuri, J. Dong and J. Reeks for work establishing the relevant crystal systems, G. Bunkóczi for assistance with data collection and analysis and D. Rees for comments and support.

Conflicts of interest

All the authors, with the exception of A. Cleasby, are employees of Astex Pharmaceuticals. A. Cleasby is a consultant for Astex Pharmaceuticals.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.drudis.2019.03.009.

References

1 Erlanson, D.A. *et al.* (2016) Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* 15, 605–619
2 Hartshorn, M.J. *et al.* (2005) Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* 48, 403–413
3 Hopkins, A.L. *et al.* (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discov.Today* 9, 430–431
4 Clackson, T. and Wells, J.A. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science* 267, 383–386
5 Hall, D.R. *et al.* (2015) Lessons from hot spot analysis for fragment-based drug discovery. *Trends Pharmacol. Sci.* 36, 724–736
6 Ludlow, R.F. *et al.* (2015) Detection of secondary binding sites in proteins using fragment screening. *Proc. Natl. Acad. Sci. U. S. A.* 112, 15910–15915
7 Hall, R.J. *et al.* (2014) Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.* 116, 82–91
8 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.* 41, 856–864
9 Behnen, J. *et al.* (2012) Experimental and computational active site mapping as a starting point to fragment-based lead discovery. *ChemMedChem* 7, 248–261
10 Mattos, C. and Ringe, D. (1996) Locating and characterizing binding sites on proteins. *Nat. Biotechnol.* 14, 595–599
11 Mattos, C. *et al.* (2006) Multiple solvent crystal structures: probing binding sites, plasticity and hydration. *J. Mol. Biol.* 357, 1471–1482
12 English, A.C. *et al.* (1999) Locating interaction sites on proteins: the crystal structure of thermolysin soaked in 2% to 100% isopropanol. *Proteins* 37, 628–640
13 English, A.C. *et al.* (2001) Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Eng.* 14, 47–59
14 Roskoski, R., Jr (2012) ERK1/2 MAP kinases: structure, function, and regulation. *Pharmacol. Res.* 66, 105–143
15 Ohori, M. *et al.* (2005) Identification of a selective ERK inhibitor and structural determination of the inhibitor-ERK2 complex. *Biochem. Biophys. Res. Commun.* 336, 357–363
16 Congreve, M. *et al.* (2003) A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today* 8, 876–877
17 Jhoti, H. *et al.* (2013) The 'rule of three' for fragment-based drug discovery: where are we now? *Nat Rev. Drug Discov.* 12, 644–645
18 Rathi, P.C. *et al.* (2017) Predicting Hot and warm spots for fragment binding. *J. Med. Chem.* 60, 4036–4046
19 Guvench, O. (2016) Computational functional group mapping for drug discovery. *Drug Discov. Today* 21, 1928–1931
20 Ghanakota, P. and Carlson, H.A. (2016) Driving structure-based drug discovery through cosolvent molecular dynamics. *J. Med. Chem.* 59, 10383–10399
21 Miranker, A. and Karplus, M. (1991) Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* 11, 29–34
22 Raman, E.P. *et al.* (2011) Reproducing crystal binding modes of ligand functional groups using site-identification by ligand competitive saturation (SILCS) simulations. *J. Chem. Inf. Model.* 51, 877–896
23 Alvarez-Garcia, D. and Barril, X. (2014) Molecular simulations with solvent competition quantify water displaceability and provide accurate interaction maps of protein binding sites. *J. Med. Chem.* 57, 8530–8539
24 Lexa, K.W. and Carlson, H.A. (2013) Improving protocols for protein mapping through proper comparison to crystallography data. *J. Chem. Inf. Model.* 53, 391–402

25 Goodford, P.J. (1985) A computational-procedure for determining energetically favorable binding-sites on biologically important macromolecules. *J. Med. Chem.* 28, 849–857
26 Verdonk, M.L. *et al.* (1999) SuperStar: a knowledge-based approach for identifying interaction sites in proteins. *J. Mol. Biol.* 289, 1093–1108
27 Dennis, S. *et al.* (2002) Computational mapping identifies the binding sites of organic solvents on proteins. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4290–4295
28 Bauman, J.D. *et al.* (2016) Rapid experimental SAD phasing and hot-spot identification with halogenated fragments. *IUCrJ* 3, 51–60
29 Pearce, N.M. *et al.* (2017) A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.* 8, 15123
30 Radoux, C.J. *et al.* (2016) Identifying interactions that determine fragment binding at protein hot spots. *J. Med. Chem.* 59, 4314–4325
31 Heightman, T.D. *et al.* (2018) Fragment-based discovery of a potent, orally bioavailable inhibitor that modulates the phosphorylation and catalytic activity of ERK1/2. *J. Med. Chem.* 61, 4978–4992

Marc O'Reilly
Anne Cleasby
Thomas G. Davies
Richard J. Hall
R. Frederick Ludlow
Christopher W. Murray
Dominic Tisi
Harren Jhoti*

Astex Pharmaceuticals,436 Cambridge Science Park, Milton Road, Cambridge, CB4 0QA, UK

**Corresponding author.*