FragLites - minimal, halogenated fragments displaying pharmacophore doublets. An efficient approach to druggability assessment and hit generation.

Daniel J. Wood[1], J. Daniel Lopez-Fernandez[2], Leanne E. Knight[2], Islam Al-Khawaldeh[2], Conghao Gai[2], Shengyin Lin[2], Mathew P. Martin[1], Duncan C. Miller[2], Céline Cano[2], Jane A. Endicott[1], Ian R. Hardcastle[2], Martin E. M. Noble[1*] and Michael J. Waring[2*].

[1]Northern Institute for Cancer Research, Paul O'Gorman Building, Medical School, Framlington Place, Newcastle University, Newcastle upon Tyne, NE2 4HH, U.K.; [2]Northern Institute for Cancer Research, Chemistry, School of Natural and Environmental Sciences, Bedson Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, U.K.

ABSTRACT: Identifying ligand binding sites on proteins is a critical step in target-based drug discovery. Current approaches to this require resource intensive screening of large libraries of lead-like or fragment molecules. Here we describe an efficient and effective experimental approach to mapping interaction sites using a set of halogenated compounds expressing paired hydrogen-bonding motifs, termed FragLites. The FragLites identify productive drug-like interactions, which are identified sensitively and unambiguously by X-ray crystallography, exploiting the anomalous scattering of the halogen substituent. This mapping of protein interaction surfaces provides an assessment of druggability and can identify efficient start points for the de novo design of hit molecules incorporating the interacting motifs. The approach is illustrated by mapping cyclin-dependent kinase 2, which successfully identifies orthosteric and allosteric sites. The hits were rapidly elaborated to develop efficient lead-like molecules. Hence, the approach provides a new method of identifying ligand sites, assessing tractability and discovering new leads.

Introduction

The identification of ligand binding sites is a critical initial step in assessing whether a protein target is amenable to modulation by a chemical probe and, ultimately, a drug molecule.[1] Binding of a small molecule to a specific site preferentially occurs where the protein offers a poorly solvated, hydrophobic pocket.[2] However, for a small molecule to be specific in its interactions (as is required for a chemical probe[3]) as well as possess the necessary physical properties to be functional in a biological environment (as is required for a drug), it must also contain polar groups that may contribute to solubility, modulate pharmacokinetics, and form geometrically constrained, and hence selective, interactions with the protein.[4,5,6] This combination of requirements makes it difficult to predict the number and location of drug-like compound binding sites on a protein from the 3D structure of the protein alone. The active sites of proteins, which are refined by evolution to bind to other macromolecules, co-factors or substrates, can generally be recognised from their shape and physicochemical properties.[7] Consequently, such sites can, in some cases, be targeted in drug or probe discovery campaigns. Allosteric sites or sites of regulatory protein-protein interaction, by contrast, are harder to identify and harder to exploit.[8,9] Since the existence of potential binding sites, and their propensity to bind drug-like molecules, cannot be known prior to commencing investigations, a significant amount of effort is wasted in attempting to identify leads against intractable targets.

Considerable attempts to address this issue have been made using both computational and experimental methods. Computational methods can provide useful indications of the presence of binding sites[10] but have, in general, not been accurate enough to be used with confidence as a tractability filter.[9,11] Experimental methods are generally based on screening compounds. In such a setting, the application of screening by NMR[12,13] or X-ray[14,15,16] based methods is attractive, since these techniques can detect the binding of ligands that have relatively low affinity for their target (K_d values up to 10 mM).[17] X-ray crystallography, in particular, provides an information rich assessment of binding sites, revealing not only the location of the binding site, but also the specific protein-ligand interactions. Although crystal structures of compounds with very small ligands are known, compounds usually need to have a threshold level of affinity to achieve the degree of occupancy to be detected with confidence. This typically requires compounds of a certain size (molecular mass greater than 200 Da) as well as complementarity of shape and pharmacophoric interactions.[18,19] Smaller compounds are likely to bind with lower affinity, hence with lower occupancy, and be hard to distinguish from ordered solvent in electron density maps. The number of drug-like molecules within the higher mass range is enormous,[20,21] consequently libraries that sufficiently cover the corresponding chemical space are both large (containing thousands of compounds[22]) and hard to design to populate chemical space systematically. It would be possible to cover chemical space far more comprehensively with smaller compounds, but such compounds may lack the affinity and shape recognisability required for their binding to be unambiguously confirmed by X-ray crystallography.

A further complication that arises when experimentally assessing target tractability is that the observation of molecules bound to a protein pocket does not necessarily establish that pocket to be amenable to binding a small molecule that has properties compatible with development into a drug or chemical probe. For this to be the case, some of the molecule's binding affinity must derive from polar interactions, in particular through the formation of hydrogen bonds that are energetically more favourable than those it might otherwise make with solvent.[23] The penalty of desolvation upon protein binding for functional groups capable to forming such interactions can be such that they contribute little or nothing to the overall binding energy.[2] This penalty can be minimised if more than one polar interaction can be made by groups on a ligand that are in close proximity to one another, such that the combined solvation sphere of the groups is reduced relative to that of the two groups in isolation. Protein – ligand pairs that are capable of making such cooperative interactions are likely to have the combination of high affinity and relatively low lipophilicity required for a drug molecule or chemical probe.[22]

Results

We hypothesised that such potential interacting sites could be found by designing a library of molecules of minimal size and complexity, that comprises compounds displaying a combination of two functionalities capable of forming hydrogen bonding interactions in proximity but with different spatial orientations, termed pharmacophore doublets (Fig. 1). In addition to providing potential for polar protein-ligand interactions, the doublet of polar groups would be expected to confer a degree of aqueous solubility to library members, allowing them to be used at high concentrations in crystallographic and other assay conditions - a requirement because molecules of this size are expected to have low affinity for their targets. Presenting these doublets on an aromatic scaffold would provide potential for lipophilic interactions and to conform to the typical architecture of drug-like molecules. Incorporating a heavy halogen atom (bromine or iodine) into each compound would further allow unambiguous identification of the position and orientation of the fragments by detection of a signal in X-ray crystallography that arises from the anomalous dispersion of the halogen atom.[24,25,26] Confidence in this concept was provided by a recent paper in which bromo- and iodopyrazole were shown to bind to numerous sites on three proteins, with ligand binding occurring at known active sites and at previously undescribed allosteric sites.[27]

Accordingly, we designed a set of compounds that individually display all the combinations of a hydrogen bond donor-acceptor or acceptor-acceptor doublet with 1 to 5 bond connectivity between them. We also incorporated some redundancy into the set, populating the likely more productive combinations such as 1,2 donor-acceptors and 1,2 acceptor-acceptors multiple times. This allows assessment of reproducibility and incorporates variety in the orientation of the pharmacophore doublet with respect to the halogen. This resulted in the selection of a set of 31 compounds that populate all combinations of pharmacophore doublets (Fig. 1). Where not commercially available, these molecules were synthesised using standard chemistry (see Supporting Information). We term these molecules "FragLites" [defined as small (≤13 heavy atoms) compounds bearing a pharmacophore doublet and a heavy halogen atom] due to their minimal size, maximal simplicity and high visibility in X-ray crystallography.

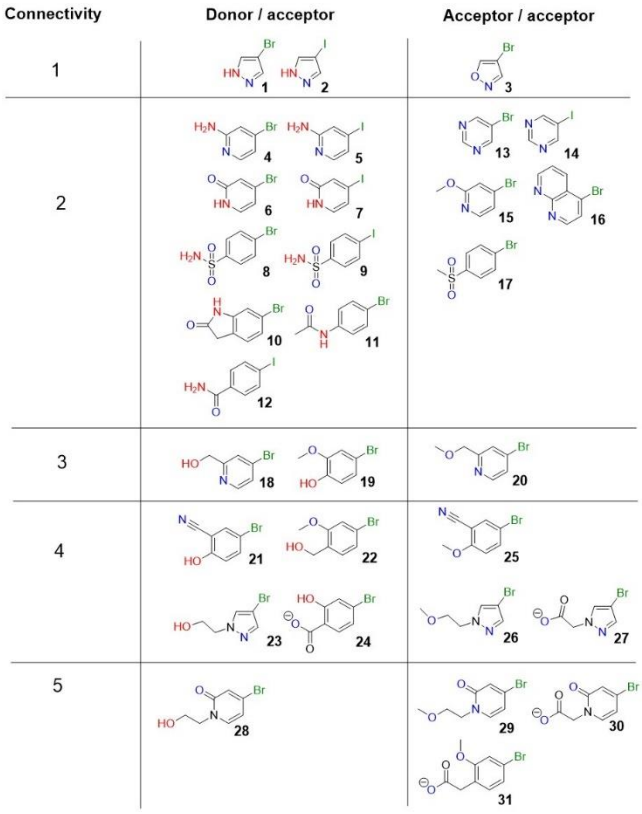


Figure 1 – Design of the FragLite set and their matching to pharmacophore doublets of donors (red) and acceptors (blue) along with their halogen tag (green), grouped by the connectivity (number refers to the number of bonds separating the doublet.

To test our hypothesis, each FragLite was soaked into an exemplar protein. The well diffracting system of cyclin-dependent kinase 2 (CDK2) was selected for this initial assessment due to its precedent for binding small molecules at the orthosteric site,[28] isolated reports of allosteric small molecule binding sites,[16,29] the well characterised allosteric binding of cyclin modulators and the potential biological significance of the discovery of new allosteric sites. Of the 31 FragLites that were soaked, 9 were observed to bind to CDK2 in 6 fully distinct sites (compounds **1**, **2**, **4**, **6**, **7**, **13**, **14**, **16** and **31**, Fig. 2). These include the orthosteric site, a known allosteric site[29] and 4 previously unidentified allosteric sites. FragLites **21**, **22**, **24** and **30** compromised the crystal and so no diffraction could be collected. The remaining compounds gave rise to no identifiable signal in normal or anomalous scattering electron density maps (Fig.2bc). This represents a significant overall hit rate, supporting the hypothesis that the FragLites have general utility as universal fragments.

Unambiguous detection of the low affinity binding events was achieved by looking for a peak more than 5 standard deviations above the mean value in an anomalous Fourier – a Fourier map calculated from phased X-ray diffraction data that presents features at sites occupied by anomalously scattering atoms. Whereas all atoms display a small amount of anomalous scattering, the magnitude of the scattering broadly increases with atomic number, being detectable for atoms with an atomic number greater than 8, depending on the wavelength of X-rays used and the signal to noise radio of the data collected.

To evaluate the enhancement in sensitivity provided by the anomalous signal, we compared the hit rate achieved using anomalous scattering with that achieved using only normal scattering. For this purpose, we applied the Pan-Dataset Density Analysis (PanDDA) algorithm, a state-of-the art statistical approach for detecting weak binding events in electron density maps.[30] The sensitivity of PanDDA derives from building a model for the variance of pixels in "apo" electron density maps so that significant deviations can be identified. Subsequently, the PanDDA algorithm exploits these variances to optimally subtract a proportion of the confounding ground state, thereby revealing a deconvoluted image of the ligand-bound state.[31] PanDDA analysis was performed on 31 ligand bound crystal datasets, bootstrapped from a variance model made from 43 apo-CDK2 crystal datasets. Through this analysis, PanDDA was able to identify 10 of the 16 binding events identified through anomalous scattering. Binding events caught by anomalous scattering but missed in the PanDDA analysis involved the unique sites observed for **6** and **14** and four of the six binding events found for **31**. In addition to highlighting sites that were not picked up by PanDDA, the anomalous signal from FragLite molecules contributed to the unambiguous determination of binding pose for several fragments, including compound **1** (Fig. 2b). While the deconvoluted PanDDA map was suggestive of a bound molecule, additional consideration of the associated anomalous peaks allowed a clear identification of two alternate binding modes for the fragment.

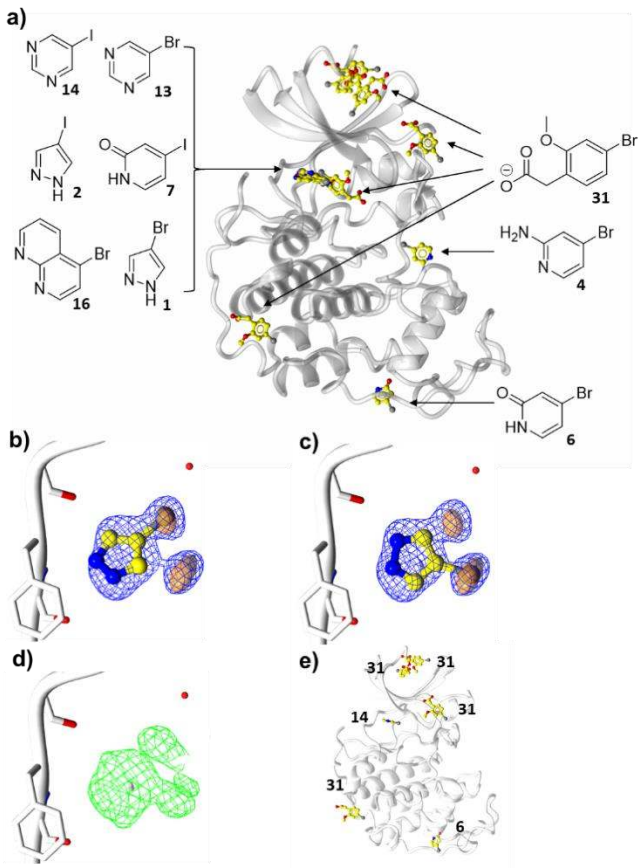


Figure 2 – Binding sites of CDK2 revealed by FragLites. a) Overview of binding events to CDK2 (white ribbon), depicting the 6 distinct sites identified by the FragLite library (yellow), PDB codes: **1** 6Q3C, **2** 6Q3B, **4** 6Q3F, **6** 6Q49, **7** 6Q48, b 6Q4B, 14 6Q4A, 16 6Q4C, 31 6Q4D, b) The anomalous signal generated resulting from the binding of **1** allows for unambiguous fragment hit identification and aids orientation of the two binding modes (b & c). Anomalous LLG map (orange surface contoured at 3σ), Weighted map from refinement (blue mesh contoured at 1σ). d) Evaluation of the dataset through PanDDA, shows native PanDDA event map (green contoured at 1σ) allowing identification of both binding modes of **1**. e) Six additional binding events identified through anomalous signal, not identified by PanDDA analysis, **6**,**14** and four copies of **31**.

The orthosteric (ATP binding) pocket bound seven different FragLites, with six (**1**, **2, 7**, **13**, **14** and **16**) making one or more hydrogen bonding contacts with the "hinge" region and 2-methoxy-4-bromophenyl)acetic acid **31** binding in a distal pocket. It is well established that targeting the ATP-competitive site of CDK2 with drug-like small molecules is feasible, with most such molecules forming hydrogen bonding interactions with the hinge region. Hence, the observation that this region binds the FragLite molecules with the highest frequency establishes that the mapping exercise can successfully identify the most tractable site, and therefore contribute to ranking and/or prioritising sites for further investigation.

3-Bromopyrazole **1** bound in two subtly different modes, the first with the pyrazole 2-nitrogen accepting a hydrogen bond from the backbone of Leu83 and the other showed the 2-nitrogen accepting a H-bond from Leu83 and the NH forming a H-bond with the backbone carbonyl of Glu81 (Fig 3a); 3-iodopyrazole **2** showed the same interactions (Supplementary Fig. 1). 4-Iodopyridone **7** also formed the donor acceptor paired interaction with the NH and carbonyl of Leu83 (Fig. 3b) and both 5-bromopyrimdine **13** and 5-iodopyrmidine **14** formed a hydrogen bond to the NH and a CH hydrogen bond from the pyrimidine 6-hydrogen to the backbone carbonyl of Glu81 (Fig. 3c and Supplementary Fig. 2). Bromonaphthyridine **16** bound to the hinge via a halogen mediated interaction.

In addition to the direct interactions with the hinge, (2-methoxy-4-bromophenyl)acetic acid **31** bound in the ATP-pocket, slightly offset from the hinge, with the carboxylate interacting with the backbone NH and sidechain hydroxyl of Thr14 and the amino group of Lys129, the benzene ring also formed a π-cation interaction with Lys33 (Fig. 3d).

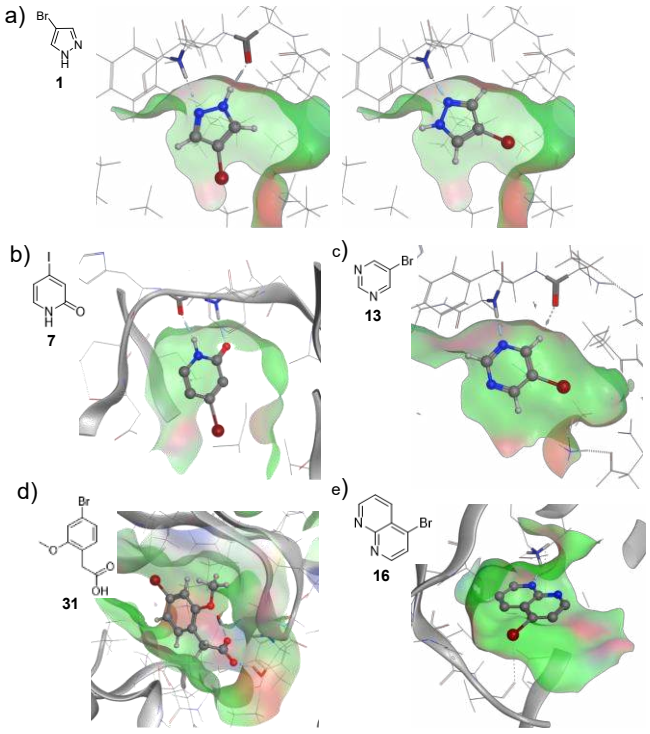


Figure 3 – Interactions of FragLites in the ATP pocket. a) 3-bromopyrazole **1**, b) 4-iodopyridone **7**, PDB code 6Q48, c) 5-bromopyrimidine **13**, PDB code 6Q4B, d) (2-methoxy-4-bromophenyl)acetic acid **31**, PDB code 6Q4D, e) 4-bromonaphthyridine **16**, PDB code 6Q4C. Protein atoms involved in direct interactions are highlighted in bold. The surface colour indicates regions of hydrophobicity (green), polarity (red) and solvent exposure (blue).

Of the five allosteric sites, 2-amino-4-bromopyridine **4** bound in the lower region of the C-terminal lobe on the C-helix side (Fig. 2). Binding interactions were made by the donation of hydrogen bonds from both hydrogens of the amino group to the backbone amide of Lys178 and Tyr180, along with π-cation interactions between the pyridine ring and Arg126 and Arg150 (Fig. 4a). A second allosteric site at the very bottom of the C-lobe was identified by 4-bromopyridone **6**, which formed a hydrogen bond between the carbonyl and the backbone NH of Tyr269 as well as a π-cation interaction with Arg245 (Fig. 4b).

The remaining allosteric sites were all identified by (2-methoxy-4-bromophenyl)acetic acid **31**, which bound in three different orientations at a region in the palm of N-terminal lobe, a single location below the hinge region and a peripheral site close to the cyclin binding region (Fig. 2). Each of these binding events were dominated by interactions of the carboxylate group. In the N-lobe palm, one copy formed interactions with the amino groups of Lys34, Lys75 and Tyr77 (Fig. 4c), a second copy formed an interaction with the amino group of Lys6 (Fig. 4d) and the third copy formed hydrogen bonds to the backbone NH groups of Glu2, Asn3 and Phe4 (Fig. 4e). At the bottom left of the C-lobe, making a single hydrogen bond contact to the backbone NH of Leu101 in a relatively solvent exposed region (Fig. 4f).

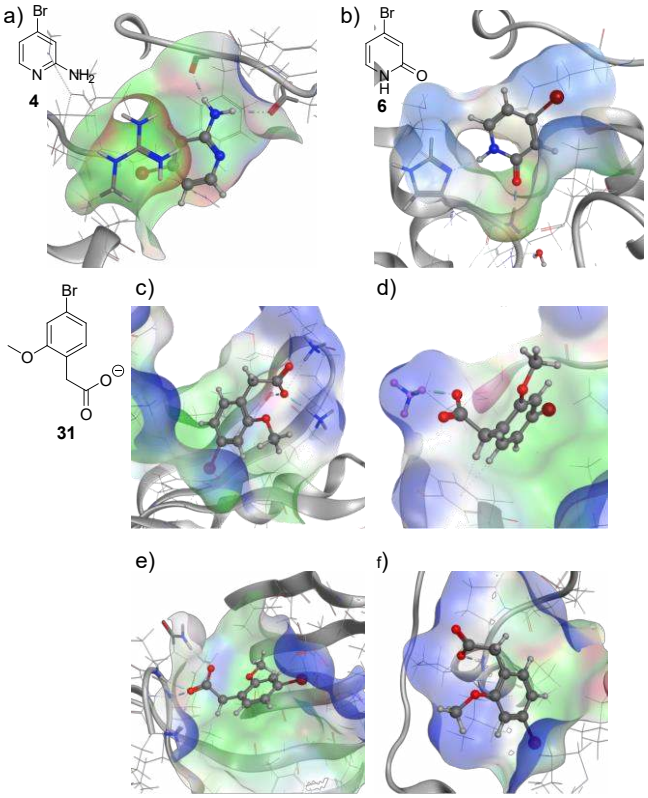


Figure 4 – Interactions of the FragLites in the identified allosteric sites. a) 2-amino-4-bromopyridine **4**, PDB code 6Q3F, b) 4-bromopyridone **6**, PDB code 6Q49, c-f) (2-methoxy-4-bromophenyl)acetic acid **31**, PDB code 6Q4D. Protein atoms involved in direct interactions are highlighted in bold. The surface colour indicates regions of hydrophobicity (green), polarity (red) and solvent exposure (blue).

Together, the mapping of the ATP pocket identifies several productive hydrogen bonding interactions that reinforce the conclusion that this location is the most druggable site. Moreover, their proximity suggests that it may be possible to use the results of these experiments to design active leads by combining the FragLite signatures into merged fragments. For example, three hydrogen bonding contacts at the hinge were identified by a combination of 3-bromopyrazole and 4-iodopyridone, overlaying these structures suggested molecules that may exploit all three of these contacts simultaneously via a contiguous array of hydrogen bond donor - acceptor - donor, each separated by two bonds (Fig. 5a). Accordingly, we conceived of fragments such as 2,6-diaminopyridine **32**, and 6-aminopurine **33** that would satisfy this combined pharmacophore.

Upon soaking into CDK2, both **32** and **33** were observed to make all three of the postulated hydrogen bond contacts in a manner that directly overlaid with that suggested by the FragLite map (Fig. 5b and c). Moreover, compound **33** showed increased affinity relative to the original hit; **33** showed a K_d of 350 μM (ligand efficiency 0.44) by isothermal calorimetry (ITC). In contrast, original hits **1** and **8** showed no detectable binding by ITC at 5 mM, consistent with the premise that such small fragments have low affinity (Supplementary Fig. 3). The affinity of **32** could not be determined due to poor solubility. These results show that the FragLite map can be used to effectively design highly ligand efficient fragment-like leads in a rational manner without extensive screening; the ligand

efficiency (LE)[32] value for **36** is 0.44. Moreover, the lead structures are reminiscent of substructures of known CDK2 ligands, which have been optimised into potent compounds.[33]

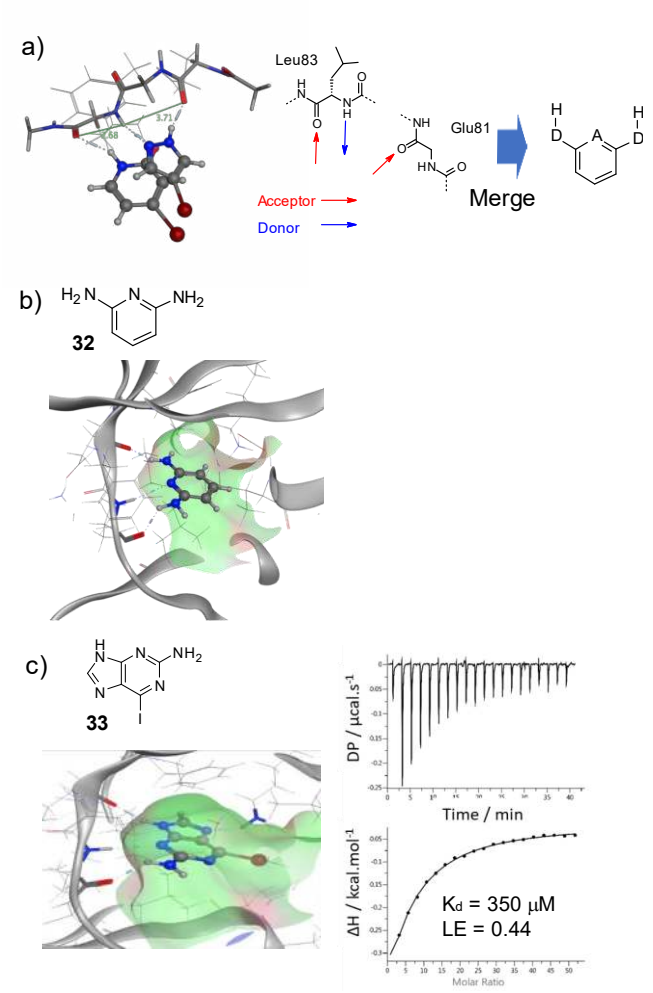
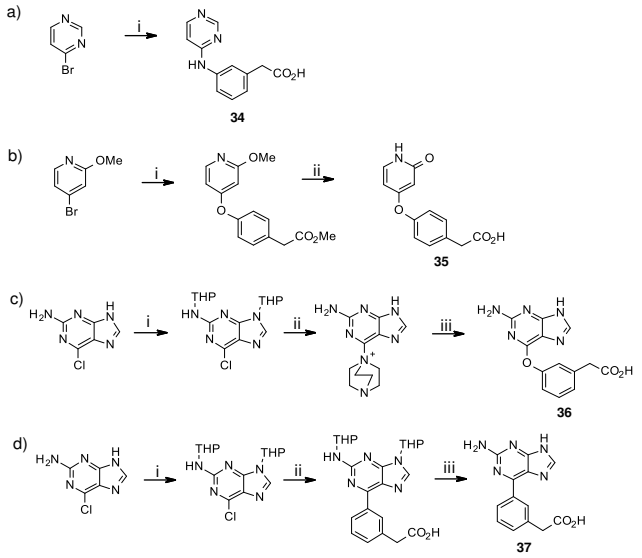


Figure 5 – Fragment merging from FragLites to produce fragment leads. a) 3-point pharmacophore derived from two FragLites binding to the hinge, b) 2,6-diaminopyridine **32** fulfilling the 3-point pharmacophore and making the predicted interactions, PDB code 6Q4F, c) 4-iodo-6-aminopurine **33** making the predicted interactions, PDB code 6Q4E, with a measurable a K_d by ITC.

The observation of (2-methoxy-4-bromophenyl)acetic acid **31** binding in the ATP pocket adjacent to the hinge provided further opportunities for increasing potency by fragment growing (Fig. 6a). The analysis of FragLite binding suggested it was the carboxylate of **31** that made the critical interactions and its orientation suggested that it could be linked to hinge binding Fragments via a direct bond or a single atom linker. Accordingly, compounds **34** and **35**, combining this template with the hinge binders **7** and **13** as well as biarylether **36** and biaryl **37** incorporating the merged aminopurine motif revealed in **31**, were designed.

Pyrimidine **34** was prepared in a single step S_NAr displacement of 4-bromopyrimidine with (3-aminophenyl)acetic acid (Scheme 1a). Pyridone **35** was prepared by copper catalyzed etherification of 4-bromo-2-methoxypyridine with methyl (4-hydroxyphenyl)acetate followed by global demethylation with pyridine hydrochloride (Scheme 1b).Compound **36** was synthesised from 6-chloro-9H-purin-2-amine by bis-dihydropyran protection followed by stepwise S_NAr substitutions of the chloro-substituent with DABCO, followed by methyl 2-(3-hydroxyphenyl)acetate and subsequent ester hydrolysis (Scheme 1c). Compound **37** was also prepared from bis-tetrahydropyranyl-protected 6-chloro-9H-purin-2-amine by Suzuki-Myaura coupling with ethyl 2-(3-(tetramethyldioxaborolanyl)phenyl)acetate followed by acidic deprotection (Scheme 1d).



Scheme 1 – Synthesis of merged FragLites. a) **34**, reagents and conditions: i) (3-aminophenyl)acetic acid, DMF, 80 °C; b) **35**, reagents and conditions: i) methyl (4-hydroxyphenyl)acetate (1.5 eq.), 3,4,7,8-tetramethylphenanthroline (cat.), CuI (cat.), Cs_2CO_3, toluene, 110 °C, 24h; ii) pyridine hydrochloride (5 eq.), 180 °C; c) **36**, reagents and conditions: i) dihydropyran (10 eq.), aq. 2M HCl (cat.), DCM, r.t., 16h; ii) DMSO, DABCO (2 eq.), r.t., 3h; iii) 1) methyl 2-(3-hydroxyphenyl)acetate (1.33 eq.), DBU (1.1 eq.), DMF, 65°C, 72h; 2) aq. 2M HCl, THF, r.t., 4h; d) **37**, reagents and conditions: i) dihydropyran (10 eq.), HCl (2M, cat.), DCM, r.t., 16h; ii) ethyl 2-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)acetate (2.5 eq.), K_2CO_3 (2.5 eq.), $Pd(PPh_3)_4$ (0.05 eq.), DMF:H_2O (1:1), 140 °C, 20', MW hv; iii) aq. 2M HCl, THF, r.t., 2h.

Crystallisation of expanded structures pyrimidine **34** and pyridone **35** revealed their binding mode to be as predicted by the FragLite mapping. The pyrimidine of **34** and pyridone of **35** engaged the hinge region with the same interactions as **13** and **7** respectively and carboxylate **31** binding to Lys129 with good overlay with the original hits (Fig. 6b and c). Aminopurines **36** and **37** also bound in the manner predicted by the FragLite map, with the carboxylate and the hinge motifs of the combined fragments making the same interactions with the same orientations as the merged fragment **33** and original hit **31** (Figure 6d and e). Compound **36** did not show a binding isotherm by ITC, however, **37** demonstrated concentration dependent binding in ITC with a K_d of 50 µM, equivalent to a ligand efficiency of 0.30 (Fig. 6e). These results demonstrate that the FragLite mapping can be used to efficiently inform fragment growing strategies to derive lead molecules with promising binding affinity in short order without the synthesis of large numbers of analogues.

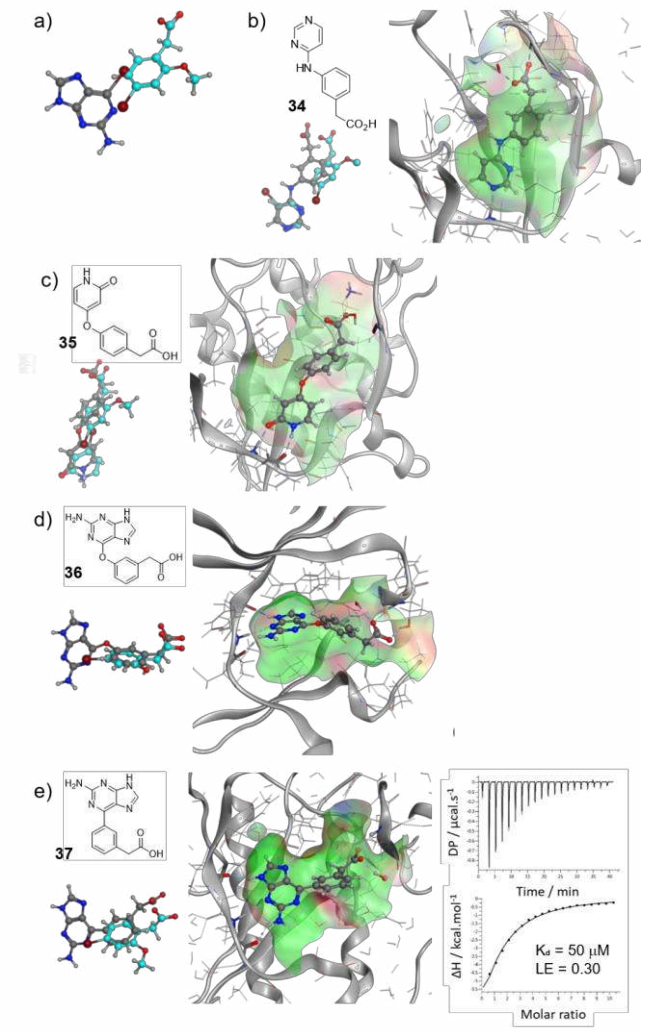


Figure 6 – Linking hinge binding compounds with carboxylate containing FragLite **31**. a) Overlay of **33** (grey carbon atoms), indicating orientation of the hinge motif, and **31** (cyan carbon atoms) as they bind to the CDK2 ATP pocket, b) Linked aminopyrimidine **34** binding to CDK2 and forming interactions with the hinge and Thr14, PDB code 6Q4J, overlay shows the alignment of **34** (grey) with the original hits **13** and **31** (cyan), c) Linked pyridone **35** binding to CDK2, PDB code 6Q4I, with **34** (grey) overlaid with **7** and **31** (cyan), d) Linked pyridone biarylether **36** (grey) binding with the acid making comparable interactions to **31** (cyan), PDB code 6Q4H, e) Linked biarylether **37** (grey) binding to CDK2making comparable interactions to **31** (cyan), PDB code 6Q4G, and demonstrating concentration dependent binding by ITC.

In addition to the hinge interactions, to illustrate that follow up hit development for the allosteric pockets was also possible, we studied further analogues of (2-methoxy-4-bromophenyl)acetic acid **31** to investigate their propensity to bind to the same allosteric pockets. This investigation revealed that amide substituted cinnamic acid **38** was able to bind to the site identified by **31** at the bottom of the C-lobe (Fig. 4f). In this case, the carboxylate does not make the same interactions in the analogue but the distal *p*-chlorophenyl ring overlays with the aromatic ring of **31**, in the lipophilic region of this relatively polar pocket (Fig. 7). This observation suggests that the FragLite mapping identifies tractable allosteric sites for which optimisable leads can be identified. It also suggests that the lipophilic scaffolds employed in the design of the FragLites, in addition to the polar hydrogen bonding pairs, are important for binding, and that the combination of the two is essential for identifying tractable sites.

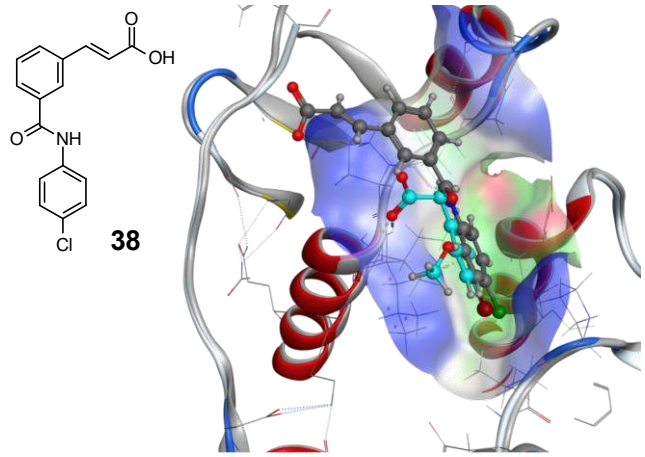


Figure 7 – Cinnamic acid fragment **38** binding to the allosteric pocket at the bottom of the C-lobe, PDB code 6Q4K.

Discussion

The key principle of fragment screening is to screen smaller compounds to cover more chemical space with fewer compounds than traditional lead-like libraries. The smaller the compounds, the fewer will be required to fully cover chemical space. FragLites provide the smallest possible constructs that can display hydrogen-bonding doublet on aromatic scaffolds. The hit rate for binding to CDK2 (8/31, 26%) is considerably higher than could be expected for the hit rate in a screen of fragments, which is typically around 5%.[19] This hit rate far exceeds that observed previously for fragment screening against CDK2 with regard to hit rate[34] and the number of identifiable sites.[29] This provides evidence that FragLites provide much more extensive coverage of chemical space with fewer compounds than is possible with compound libraries of larger molecules.

Our results demonstrate that FragLite mapping of a protein provides accurate experimental identification of productive ligand binding sites. The observation that the frequency of FragLites binding to the region established to be most druggable (the ATP pocket binds 6/31 compounds) and that the majority of these (5/31) make the well characterised hydrogen bonding interactions with the hinge gives confidence that this exercise can be used as an experimental druggability assessment. The identification of additional pockets, one of which we have shown, with very limited follow up to be amendable to binding traditional fragments, suggests that the analysis can also be used to identify novel allosteric pockets.

The points of interaction identified by the FragLites could be used prospectively to define a pharmacophore for virtual screening or to focus future library chemistry. For example, in our CDK example, further molecules could be selected satisfying the three-point hydrogen bonding pharmacophore identified at the CDK2 hinge coupled with the interaction made by the carboxylate. This is potential advantage of this approach in that it can identify productive interactions within a tractable pocket (and presumably not unproductive

ones) to focus future hit finding. Other approaches, such as computational druggability calculations are unable to do this with confidence.

Furthermore, the results of fragment merging leading to hit-like levels of affinity and high ligand efficiency demonstrate that the FragLite map can be used in hit generation directly, without recourse to further hit generation activities. We would not contend that this approach renders further screening unnecessary but could be applied if further screening approaches are unavailable or resource limited. The short order in which we were able to produce potent merged fragments demonstrates that this can be a very efficient way to find hit compounds.

A further extension of this work would be to investigate why certain residues within proteins are more able to form productive polar interactions such as hydrogen bonds when others are not. Why, for example, are the hydrogen bonding interactions with the peptide backbone in the hinge region identified when other backbone interactions are not? This particular case is well established with almost all kinase inhibitors requiring hydrogen bonding to the hinge region to achieve potency.[35]

The incorporation of the halogen atom is essential to this approach. Without the analysis of anomalous scattering, it would not be possible to determine the location and orientation of the hits. Another potential advantage of the incorporation of an anomalous scatter is their use in weaker diffracting crystal systems, to help resolve ambiguity of binding events. When reviewing the crystallographic binding event data, consideration was given to the generation of halide ions from the compound upon exposure to synchrotron radiation.[36,37,38] In some instances, we observed anomalous signal for a liberated halide ion but this was of no consequence for the analysis. The degree of radiation damage can be minimised by controlling the radiation energy.

The PanDDA analysis method for identification of low occupancy ligands identified the majority of those detected using the anomalous signal but failed to identify the binding events of two FragLites and several of the binding sites of FragLite **31**. Increasing ligand occupancy would require screening at much higher concentrations at which many of the desired compounds would be insoluble. Similarly, use of an anomalous signal may allow for application to weakly diffracting crystal systems, which are typically not suitable for a fragment crystallography campaign.

The FragLite mapping can be carried out very rapidly, due to the small number of compounds required. Our selected set can be soaked into protein crystals in a single experiment and X-ray diffraction data on the entire set collected within a day. Critical to the further development of this approach is the definition of the optimum set of compounds that cover chemical space most effectively with the minimal number of compounds. Our work shows that additional information can be obtained, at least for CDK2, by screening a larger compound set than that explored by Arnold.[27] It is noteworthy, however, that bromo- and iodopyrazole, the most promiscuous compounds in Arnold's experiment, have also proven to be the most useful here. We contend that this is likely due to that fact that the 1,2-donor-acceptor motif can form highly productive complementary hydrogen bonds with a protein target. This observation could be useful for future ligand design. This work also suggests that the promiscuous carboxylate **31** may offer an alternative "universal fragment" that could be used for SAD phasing as established by Arnold for bromopyrazole.

Despite these results, we believe that the definition of the optimal set of FragLites that strike the best balance between minimising the numbers of compounds to screen and maximising the coverage of chemical space will require further optimisation across a wide range of proteins and larger number of compounds. Our results suggest that some redundancy in the screening set is required. There are differences between the binding of analogous brominated and iodinated FragLites and not all the hits expressing the same pharmacophore doublet bind in the same manner. Perhaps most strikingly, it is surprising that 4-bromo-2-aminopyridine **4** was not detected binding to the hinge region and 4-iodo-2-aminopyridine **5** did not bind at all. These compounds express the same 1,3-donor-acceptor doublet as 4-iodopyridone **7** which made the critical pair of H-bonds in the hinge region. It is possible that the geometry of the bound pose required for **4** and **5** to bind to the hinge is incompatible with accommodation of the heavy halogen atom. Consistent with this theory, 4-chloropyridine did bind at the hinge in the manner that was predicted by the binding mode of **7** (Supplementary Fig. 4). This suggests that the ideal set should cover a range of differing geometries between the pharmacophore doublets and the halogen atom. Future work will explore these observations further to define the optimal FragLite set.

Conclusion

The definition of the set of 31 FragLites and mapping of CDK2 protein to identify the most tractable, orthosteric site and exploitable allosteric sites suggest that FragLites offer a new approach to assessment of druggability that provides a facile and informative hit-finding activity. Development of potent hit compounds directly from the FragLite map shows that the information can also be used for hit-finding in isolation. We believe that this has the potential to be a nuanced approach to hit finding that will supplement and inform screening of more traditional fragment and lead-like libraries.

Future work will explore the application of this approach to a wider range of proteins and with a wider range of designed FragLites. This will establish the general applicability of the approach and optimise the compound set for prospective screening. The results of these studies will be reported in due course.

Experimental Section

Protein Purification

Full length human CDK2 for crystallography was cloned into a pGEX6P1 vector (GE Healthcare) and grown in BL21(DE3)STAR Escherichia coli (E. coli) cells.[39] For ITC, full length human CDK2 was cloned into a modified pGEX3C plasmid[40] containing Saccharomyces cerevisiae Civ1 for activation loop T160 phosphorylation and grown in BL21(DE3)pLysS cells.

All E. coli were then grown at 37 ºC until an optical density of 0.4-0.6 before decreasing the temperature to 18 ºC with 0.2 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside; Sigma) induction for BL21(DE3)STAR and Rosetta(DE3)pLysS cells. BL21(DE3)pLysS cells were grown in auto-induction media (Formedium) and so didn't require induction with IPTG. All other constructs were grown in Luria Broth (LB; Invitrogen). All cells were incubated for a further 16-20 hours at 18 °C, before harvesting via centrifugation (4,000 xg, 30 mins, 4 ºC) and resuspension in 50 mM HEPES, 150 mM NaCl, 2 mM DTT (dithiothreitol), pH 7.4 and EDTA (ethylenediaminetetraacetic acid)-free protease inhibitor tablets (Roche). For lysis, cell suspensions were then supplemented with 2 µg/mL DNase I, 10 µg/mL RNase A, 25 µg/mL lysozyme and 5 mM $MgCl_2$ before pulsed sonication for 5 mins (20 s

on, 40 s off) on ice. Cell lysates were then recovered via centrifugation (40, 000 xg, 60 mins, 4 ºC) before 0.45 µm filtering and batch binding to glutathione sepharose 4B resin (GE Healthcare) for 60 mins, 4 ºC with rolling. Batch-bound lysate was then poured down a gravity flow column, with GST-CDK2 (all constructs are GST-fusions) eluted from the resin using 50 mM HEPES, 150 mM NaCl, 2 mM DTT, 10 mM reduced glutathione, pH 7.4. GST was then cleaved from CDK2 using 1:100 w/w 3C prescission protease overnight at 4 ºC with rolling. Protein was then separated from GST using gel filtration (26-60 HiLoad Superdex 75 preparative grade (GE Healthcare)) before subtractive GST using a 5 mL GSTrap fast flow column (GE Healthcare).

For the phosphorylated CDK2 construct (used for ITC experiments), protein was then concentrated appropriately using a Millipore spin concentrator and flash frozen using liquid nitrogen for storage at -80 ºC

For crystallisation, CDK2 was then buffer exchanged into 100 mM 4:1(Dibasic:monobasic) K2:Na phosphate, 2 mM DTT buffer and concentrated to 10 mg/mL prior to crystalisation.[41]

Crystallisation and structure determination

Crystals of CDK2 protein were prepared using the hanging drop vapour diffiusion method. Protein was mixed 1:1 with precipitant solution (6% PEG3350, 100 mM HEPES pH 7.4) over a reservoir of 50 mM HEPES pH 7.4, 25 mM K_2HPO_4, 25 mM NaH_2PO_4.

Crystals were then soaked for 48 hours in 50 mM compound prepared to 10 % DMSO in CDK2 precipitant solution (6% PEG3350, 100 mM HEPES pH 7.4). Crystals were then harvested in 30 % ethylene glycol prepared in CDK2 precipitant solution and flash-cooled in liquid nitrogen.

X-ray diffraction data were then recorded either at Diamond Light Source, Oxford, UK or in house using a rotating copper anode X-ray source. Structural data were determined using the CCP4 suite of programs.[42] Data were firstly processed using DIALS and XDS [43] within Xia2.[44] The structures were solved using molecular replacement using PHASER[45] and protein databank (PDB) code 1HCK[46] as a search model. Refinement was carried out using REFMAC5,[47] with model building performed and compound descriptions generated using COOT[48] and AceDRG,[49] respectively. Figures were prepared using the CCP4 molecular graphics programme CCP4mg[50] and MOE (Molecular Operating Environment, 2013.08; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2018). PanDDA (Pan-Dataset Density Analysis)[30] was undertaken using 31 FragLite datasets against a variance model derived from 42 apo CDK2 crystal datasets. Derived PanDDA maps were analysed using the panda.inspect command and results tabulated appropriately.

Isothermal Titration Calorimetry (ITC)

CDK2 was buffer exchanged using a 5 mL HiTrap desalting column (5 mL) (GE Healthcare) into ITC buffer (50 mM HEPES, 150 mM NaCl, 0.25 mM TCEP, pH 7.4) and protein concentration was then determined using a Nanodrop 2000 at an absorbance of 280 nm. with sequence derived extinction coefficients (http://web.expasy.org/protparam/). Protein was prepared in ITC buffer, with 1 % DMSO in the cell at 20 µM with compounds prepared at 5 mM or 1 mM in ITC buffer at 1 % DMSO final, in the syringe.

All ITC experiments were carried out at 30 °C using a Microcal PEAQ ITC instrument (Malvern) with 1× injection of 0.4 µL for 0.8s followed by 19× 2 µL, 4s injection with 120 s spacing between injections, using a reference power of 6 µcal/s. Data were then analysed using Malvern Microcal PEAQ-ITC analysis software using a one-set-of-sites model. Data analysis was also conducted at low-c[51,52] so, in accordance with the crystallographic evidence, stoichiometry was set at n=1 set of sites, during fitting. Data was analysed using Microcal PEAQ Evaluation software (Malvern) using the one set of sites model to calculate the dissociation constant and the error of the fit. Experiments were carried out twice with two separate biological preparations of CDK2.

Chemical Synthesis

All compounds had purity ≥95% as determined by HPLC (UV detection) and ^{1}H NMR analysis.

Compounds **1**, **2**, **3**, **4**, **5**, **6**, **7**, **8**, **9**, **10**, **11**, **12**, **13**, **14**, **15**, **16**, **17**, **18**, **19**, **20**, **21**, **22**, **23**, **24**, **25**, **26**, **27**, **30**, **31**, **32** are commercially available.

4-Bromo-1-(2-hydroxyethyl)pyridin-2(1H)-one (28)

2-Bromoethanol (326 µL, 4.6 mmol) was added dropwise to a mixture of potassium carbonate (1.59 g, 11.5 mmol) and 4-bromo-2(1H)-pyridinone (0.20 g, 1.1 mmol) in anhydrous THF (10 mL) under nitrogen. The mixture was stirred at 80 ºC for 16 h, allowed to cool to room temperature, the precipitate filtered off and the filtrate evaporated in vacuo. The crude residue was purified by flash column chromatography (EtOAc:MeOH 9:1) to afford **28** (94 mg, 38%) as a white solid.

R_f 0.33 (100% EtOAc); mp. 134-135 ºC; IR: ν_{max} 3293, 3039, 2943, 1629 cm^{-1}; ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.23 (d, J = 7.2 Hz, 1H), 6.87 (d, J = 2.2 Hz, 1H), 6.38 (dd, J = 7.2, 2.2 Hz, 1H), 4.14-4.08 (m, 2H), 3.96 (q, J = 5.0 Hz, 2H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 162.2, 138.4, 136.0, 123.0, 110.5, 61.5, 52.9; HRMS-ESI (m/z): [M + H]$^+$ $C_7H_9BrNO_2^+$, 217.9811; found 217.9806.

4-Bromo-1-(2-methoxyethyl)pyridin-2(1H)-one (29)

2-Bromoethyl methyl ether (216 µL, 2.3 mmol) was added dropwise to a mixture of potassium carbonate (0.95 g, 6.9 mmol) and 4-bromo-2(1H)-pyridinone 6 (0.20 g, 1.1 mmol) in anhydrous MeCN (10 mL) under nitrogen. The mixture was stirred at 80 ºC for 16 h, allowed to cool to room temperature, the precipitate filtered off and the filtrate evaporated in vacuo. The crude residue was purified by flash column chromatography (EtOAc 100%) to afford **29** (119 mg, 45%) as a brown oil.

R_f 0.25 (1:1 Petrol Ether:EtOAc); IR ν_{max} 2891, 1625, 1111 cm^{-1}; ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.23 (d, J = 7.2 Hz, 1H), 6.84 (d, J = 2.2 Hz, 1H), 6.32 (dd, J = 7.2, 2.2 Hz, 1H), 4.12-4.12 (m, 2H), 3.64-3.68 (m, 2H), 3.30 (s, 3H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 161.3, 138.8, 135.7, 122.7, 109.7, 70.1, 59.0, 49.4; HRMS-ESI (m/z): [M + H]$^+$ $C_8H_{11}BrNO_2^+$, 231.9968; found 231.9966.

6-Iodo-9H-purin-2-amine (33)

Hydroiodic acid (5 mL, 47% aqueous) was added to 2-amino-6-chloropurine (0.42 g, 2.49 mmol) at 0 °C. The mixture was stirred at 0 °C for 8 hours, treated with 5 mL of deionised water at 5 °C for 1 hour. The precipitate was filtered and washed with water three

times. The filter cake was transferred to a 50 mL round bottom flask and of 6 M NaOH aqueous solution was added to the solid. The solution obtained was added slowly to a solution of acetic acid in water (2:25, 27 mL). The resulting suspension was stirred at room temperature for 2 h. The solid was collected by filtration, washed with water (2 × 5 mL), followed by petrol (5 mL). The solid was dried in vacuo to afford **33** (501 mg, 77%) as a yellow powder.

R_f 0.21 (DCM:MeOH 9:1); mp. 230-231 ºC; IR: ν_{max} 3488, 3293, 3170 cm^{-1}; ^1H NMR (500 MHz, $(CD_3)_2SO$) δ 12.81 (br s, 1H), 8.05 (s, 1H), 6.67 (s, 2H); ^{13}C NMR (125 MHz, $(CD_3)_2SO$) δ 160.1, 151.8, 141.2, 130.5, 122.4; HRMS-ESI (m/z): $[M + H]^+$ $C_5H_5IN_5^+$, 261.9584; found 261.9582.

2-(3-(Pyrimidin-4-ylamino)phenyl)acetic acid (34)

4-Chloropyrimidine (100 mg, 0.87 mmol), 2-(3-aminophenyl) acetic acid (100 mg, 0.66 mmol) and DMF (4 mL) were stirred at 80 ºC for 3 h. The mixture was evaporated in vacuo and the crude product was purified by flash column chromatography (silica, DCM:MeOH:AcOH, 8:2:0.05) to afford **34** (15 mg, 10%) as a beige solid.

R_f 0.47 (DCM:MeOH:AcOH 9:1:0.05); mp. 186-190 ºC; IR: ν_{max} 3045, 2827, 1649, 1580, 1516, 1474 cm^{-1}; ^1H NMR (500 MHz, $(CD_3)_2SO$) δ 12.40 (br s, 1H), 9.89 (s, 1H), 8.65 (s, 1H), 8.28 (s, 1H), 7.63 (d, J = 7.8 Hz, 1H), 7.53 (s, 1H), 7.28 (t, J = 7.8 Hz, 1H), 6.95 (d, J = 7.8 Hz, 1H), 6.85 (d, J = 5.7 Hz, 1H), 3.56 (s, 2H); ^{13}C NMR (125 MHz, $(CD_3)_2SO$) δ 159.9, 158.0, 155.1, 139.6, 128.6, 123.5, 120.6, 118.2, 107.2, 41.0 (C_q absent $–CO_2H$); HRMS-ESI (m/z): $[M + H]^+$ $C_{12}H_{11}N_3O_2^+$, 230.0924; found 230.0925.

Methyl 2-(4-((2-methoxypyridin-4-yl)oxy)phenyl)acetate

4-Bromo-2-methoxypyridine (1.2 g, 6.4 mmol), 3,4,7,8-tetramethylphenantroline (139 mg, 0.59 moles), copper (I) iodide (54 mg, 0.28 mmol), caesium carbonate (2.9 g, 8.9 mmol) and methyl(4-hydroxyphenyl)acetate (1.5 g, 9.0 mmol) in toluene (12 mL) were stirred in a sealed tube under nitrogen atmosphere at 120 ºC for 3 days. The mixture was allowed to cool to room temperature, treated with water (12 mL) and extracted three times with EtOAc. The combined organic extracts were washed with brine, dried over sodium sulfate, filtered, and concentrated in vacuo. The crude residue was purified by flash column chromatography (silica; EtOAc: Petrol Ether 1:9 to 10:0) to afford the desired product (351 mg, 10%) as a pale yellow oil.

R_f 0.72 (EtOAc:Petrol Ether 9:1); ^1H NMR (500 MHz, $CDCl^3$) δ 8.03 (d, J = 5.9 Hz, 1H), 7.32 (d, J = 8.5 Hz, 2H), 7.04 (d, J = 8.5 Hz, 2H), 6.54 (dd, J = 5.9, 2.1 Hz, 1H), 6.18 (d, J = 2.1 Hz, 1H), 3.91 (s, 3H), 3.72 (s, 3H), 3.64 (s, 2H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 171.8, 167.0, 165.9, 153.3, 147.8, 131.0, 120.9, 107.4, 97.5, 53.8, 52.1, 40.5 (C_q absent).

2-(4-((2-Oxo-1,2-dihydropyridin-4-yl)oxy)phenyl)acetic acid (35)

Pyridine hydrochloride (146 mg, 1.2 mmol) and methyl 2-(4-((2-methoxypyridin-4-yl)oxy)phenyl)acetate (57 mg, 0.2 mmol) were combined and the mixture was stirred at 120 ºC until full conversion was observed by TLC. The mixture was allowed to cool to room temperature and purified by flash column chromatography (C_{18} silica, CH_3CN:H_2O 1:9 to 2:3) to give **35** (19 mg, 37%) as a beige solid.

R_f 0.61 (DCM:MeOH 9:1); mp. 260-261 ºC; IR: ν_{max} 3443, 1613, 1216, 1182 cm^{-1}; ^1H NMR (500 MHz, $(CD_3)_2SO$) δ 11.35 (s, 1H), 7.36 (d, J = 7.3 Hz, 1H), 7.34 (d, J = 8.4 Hz, 2H), 7.07 (d, J = 8.4 Hz, 2H), 5.97 (dd, J = 7.3, 2.5 Hz, 1H), 5.31 (d, J = 2.5 Hz, 1H), 3.53 (s, 2H); ^{13}C NMR (125 MHz, $(CD_3)_2SO$) δ 172.7, 167.7, 163.6, 151.6, 136.7, 132.8, 131.2 (×2), 120.7 (×2), 100.5, 98.6 (C_q absent); HRMS-ESI (m/z): $[M+H]^+$ $C_{13}H_{11}NO_4^+$, 246.0688; found 246.1010.

6-chloro-N,9-bis(tetrahydro-2H-pyran-2-yl)-9H-purin-2-amine

3,4-Dihydropyran (2.7 mL, 29.5 mmol) was added to 6-chloro-2-aminopurine (1 g, 2.95 mmol) in anhydrous DMF (50 mL) followed by the addition of aqueous hydrochloric acid (50 µL, 5.7 M). The mixture was stirred at 60 ºC for 2 h. The mixture was evaporated in vacuo and the crude product was purified by flash column chromatography (silica, EtOAc:Petrol Ether 2:3) to afford the desired product (1.55 g, 78%) as a pale yellow oil that crystallised upon freezing.

R_f 0.44 (100% EtOAc); mp. 171-174 ºC; IR: ν_{max} 3269, 2940, 2855, 1716, 1614 cm^{-1}; ^1H NMR (500 MHz, $(CD_3)_2SO$) δ 8.42 (s, 1H), 8.14 (br s, 1H), 5.53 (d, J = 10.2 Hz, 1H), 5.20-5.12 (m, 1H), 4.02 (d, J = 11.1, 1H), 3.84 (d, J = 11.1 Hz, 1H), 3.46 (td, J = 2.8, 11.1 Hz, 1 H), 2.21-2.35 (m, 1H), 2.03-1.34 (m, 11H); ^{13}C NMR (125 MHz, $(CD_3)_2SO$) δ 158.1, 153.7, 150.0, 142.3, 124.7, 81.6, 80.4, 68.2, 66.0, 30.7, 30.1, 30.0, 25.4, 24.9, 22.9; HRMS-ESI (m/z): $[M + H]^+$ $C_{15}H_{21}ClN_4O_2^+$ 338.1378; found 338.1378.

Methyl 2-(3-((9-(tetrahydro-2H-pyran-2-yl)-2-((tetrahydro-2H-pyran-2-yl)amino)-9H-purin-6-yl)oxy)phenyl)acetate

DABCO (88 mg, 0.59 mmol) was added to a mixture of 6-chloro-N,9-bis(tetrahydro-2H-pyran-2-yl)-9H-purin-2-amine (100 mg, 0.30 mmol) and anhydrous DMSO (3 mL). The mixture was stirred at room temperature for 3 h. The mixture was added to a stirred suspension of DBU (49 µL, 0.33 mmol) and methyl 2-(3-hydroxyphenyl)acetate (98 mg, 0.40 mmol) in anhydrous DMF (2 mL). The mixture was stirred at 65 ºC for 72 h. The mixture was treated with aqueous saturated sodium bicarbonate (8 mL) and was extracted three times with EtOAc and the combined extracts washed three times with brine, dried over sodium sulfate, filtered, and concentrated in vacuo to give the crude product (145 mg) which was used as such in the next step without further purification.

2-(3-((2-Amino-9H-purin-6-yl)oxy)phenyl)acetic acid (36)

The crude methyl 2-(3-((9-(tetrahydro-2H-pyran-2-yl)-2-((tetrahydro-2H-pyran-2-yl)amino)-9H-purin-6-yl)oxy)phenyl)acetate (142 mg, 0.3 mmol) was dissolved in THF (5 mL) and 2M aqueous HCl (2.5 mL) and stirred at room temperature for 4 h. The mixture was evaporated in vacuo and the crude product was purified by reverse phase chromatography (C18 silica, water:CH_3CN:NH_4OH 70:30:0.1) to afford 36 (8 mg, 10%) as a white powder.

R_f 0.18 (DCM:MeOH 9:1); mp. 284-285 ºC; IR: ν_{max} 1233, 1265, 1376, 1568 cm^{-1}; ^1H NMR (500 MHz, CD_3OD) δ 7.97 (s, 1H), 7.38 (t, J = 8.1 Hz, 1H), 7.27-7.21 (m, 2H), 7.11 (dd, J = 8.1, 1.1 Hz, 1H), 3.60 (s, 2H); ^{13}C NMR (125 MHz, $(CD_3)_2SO$) δ 173.2, 160.22, 160.18, 153.0, 138.2, 129.6, 126.5, 122.8, 120.2, 41.9 (2 C_q and 1 C_t absent)*; HRMS-ESI (m/z): $[M + H]^+$ $C_{15}H_{14}N_3O_3^+$ 286.0935; found 286.0931.

*Weak signals at 139.8 ppm and 159.7 ppm were assigned by HMBC but are not sufficiently resolved.

2-(3-(2-Amino-9H-purin-6-yl)phenyl)acetic acid (37)

A mixture of 6-chloro-N,9-bis(tetrahydro-2H-pyran-2-yl)-9H-purin-2-amine (100 mg, 0.3 mmol), potassium carbonate (102 mg, 0.74 mmol), ethyl 2-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)acetate (86 mg, 0.3 mmol), tetrakis(triphenylphosphine)palladium(0) (17 mg, 0.015 mmol), DMF (1.5 mL) and deionised water (1.5 mL) was stirred at 140 ºC for 30 minutes under microwave irradiation under nitrogen. The mixture was treated with deionised water (5 mL), extracted three times with EtOAc and the combined extracts washed with brine, dried over sodium sulfate, filtered, and concentrated in vacuo. The crude residue was dissolved in THF (5 mL) and 2M aqueous HCl (2.5 mL) and stirred at room temperature for 4 h. The mixture was neutralised with 2M aqueous NaOH, extracted three times with EtOAc, and the combined extracts washed with brine, dried over sodium sulfate, filtered, and concentrated in vacuo to give a yellow oil that was triturated with anhydrous methanol (5 mL) to afford **37** (5 mg, 6%) as a white powder.

R_f 0.2 (DCM:MeOH 9:1); mp. 329-330 ºC; IR: ν_{max} 3380, 3163, 3041, 2113, 1911, 638 cm^{-1}; ^1H NMR (500 MHz, $(CD_3)_2SO$) δ 8.58 (d, J = 7.4 Hz, 1H), 8.50 (br s, 1H), 8.10 (s, 1H), 7.48 (t, J = 7.4 Hz, 1H), 7.40 (d, J = 7.4 Hz, 1H), 3.69 (s, 2H); ^{13}C NMR (125 MHz, (CD3)2SO) δ 174.8, 156.8, 148.6, 137.8, 135.5, 131.7, 130.6, 129.5, 41.5 (4 C_q absent)*; HRMS-ESI (m/z): $[M + NH_4]^+$ $C_{13}H_{12}N_5O_2 + NH_4$ 287.3025; found 287.1240.

*Weak signals at 123.5 ppm and 159.7 ppm were assigned by HMBC but are not sufficiently resolved.

3-Bromo-N-(4-chlorophenyl)benzamide

3-Bromobenzoic acid (400 mg, 2.00 mmol) was added to DMAP (242 mg, 1.83 mmol) in anhydrous DCM (20 mL), followed by the addition of EDC.HCl (460 mg, 2.40 mmol). After 15 min, a solution of 4-chloroaniline (280 mg, 2.20 mmol) in anhydrous DCM (5 mL) was added, and the mixture was stirred at rt. for 18 h. Water (10 mL) was added to quench the reaction, the organic layer separated and washed with 1.0M aqueous HCl (3 × 10 mL). The combined organic layers were dried over magnesium sulfate and the solvent concentrated in vacuo. Purification by flash chromatography (silica; 0-40% ethyl acetate in petrol ether) afforded the title compound as a white solid (270 mg, 87%).

R_f 0.78 (33% EtOAc:Petrol); mp. 136.0-138.7 ºC; UV λ_{max} (EtOH) 274 nm; IR ν_{max} 3361, 2112, 1652, 1595, 1517, 1397 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 7.34-7.31 (m, 2H), 7.36 (d, J = 8.0 Hz, 1H), 7.59-7.56 (m, 2H), 7.67 (ddd, J = 1.0 Hz, 2.0 Hz, 8.0 Hz, 1H), 7.76 (dd, J = 1.0 Hz, 8.0 Hz, 1H), 7.83 (br s, 1H), 7.98 (dd, J = 2.0 Hz, 1H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 164.3, 136.6, 136.1, 135.0, 130.4, 130.2, 130.0, 129.2, 125.6, 123.0, 121.6; HRMS-ESI (m/z): $[M + H]^+$ $C_{13}H_{10}ONClBr$ 309.9627; found 309.9629.

Ethyl (E)-3-(3-((4-chlorophenyl)carbamoyl)phenyl)acrylate

Palladium(II) acetate (6.8 mg, 0.026 mmol) and triphenylphosphine (7.9 mg, 0.026 mmol) were added to a solution of 3-bromo-N-(4-chlorophenyl)benzamide (80 mg, 0.259 mmol) in DMF (1.0 mL) at room temperature under nitrogen. Ethyl acrylate (0.04 mL, 0.259 mmol) and sodium acetate (27 mg, 0.325 mmol) were added, and the mixture was heated at 140 ºC for 18 h. After cooling to room temperature, water (20 mL/mmol) was added and the mixture passed through a pad of Celite. The organic layer was extracted with Et2O (3 × 10 mL), separated and washed with water (10 mL) and brine (2 × 10 mL) and dried over magnesium sulfate. The solvent was concentrated in vacuo to give the crude product. Purification by flash chromatography (silica; 0-100% DCM in petrol ether) afforded the title compound as a white solid (25 mg, 29%).

R_f 0.14 (100% DCM); mp. 148.7-155.0 ºC; UV λ_{max} (EtOH) 274 nm; IR ν_{max} 3312, 2981, 2928, 2109, 1703, 1650, 1595, 1535, 1490 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 8.35 (br s, 1H), 7.94 (s, 1H), 7.81 (d, J = 8.0 Hz, 1H), 7.64-7.59 (m, 4H), 7.44 (t, J = 8.0 Hz, 1H), 7.31-7.28 (m, 2H), 6.43 (d, J = 16.0 Hz, 1H), 4.23 (q, J = 7.0 Hz, 2H), 1.31 (t, J = 7.0 Hz, 3H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 166.7, 165.4, 143.2, 136.4, 135.4, 135.1,131.2, 129.8, 129.4, 129.1, 128.6, 126.6, 121.7, 60.8, 14.3; HRMS-ESI (m/z): $[M + H]^+$ $C_{18}H_{17}O_3NCl$ 330.0888; found 330.0891.

(E)-3-(3-((4-Chlorophenyl)carbamoyl)phenyl)acrylic acid (38)

A mixture of ethyl (E)-3-(3-((4-chlorophenyl)carbamoyl)phenyl)acrylate (50 mg, 0.15 mmol), $LiOH.H_2O$, and THF:MeOH:H2O (4:2:1, 1 mL) was stirred at 0 ºC for 2 h. The mixture was treated with 2M aqueous HCl (2 mL) and extracted three times with EtOAc. The combined extracts were dried over sodium sulfate, filtered, and concentrated in vacuo to give a white solid that was purified by flash column chromatography (silica, MeOH: DCM 0:10 to 2:8) to afford **38** (14.4 mg, 31%) as a white powder.

R_f 0.10 (DCM:MeOH 9:1); m.p. 259-260 ºC; IR ν_{max} 3334, 2919, 2849, 2633, 2485, 2115, 1719, 1648, 1593, 1522 cm^{-1}; ^1H NMR (500 MHz, $(CD_3)_2SO$) δ 10.44 (1H, s), 8.23 (1H, s), 7.95 (1H, d, J = 7.8 Hz), 7.89 (1H, d, J = 7.8 Hz), 7.81 (2H, d, J = 8.8 Hz), 7.66 (1H, d, J = 16.0 Hz), 7.57 (1H, t, J = 7.8 Hz), 7.42 (2H, d, J = 8.8 Hz), 6.67 (1H, d, J = 16.0 Hz); 13C NMR (125 MHz, $(CD_3)_2SO$) δ 167.6, 165.2, 143.9, 138.0, 135.3, 134.6, 131.4, 129.4, 129.2 128.6, 127.5, 126.9, 122.1, 120.9; HRMS-ESI (m/z): $[M + H]^+$ $C_{16}H_{13}O_3N_{35}Cl$ 302.0579, found 302.0578.

Ancillary Information

The Supporting Information is available free of charge on the ACS Publications website. Supplementary figures for additional crystal structures of **2**, **14** and 2-amino-4-trifluoromethylpyridine; ITC traces of **1** and **8**; crystallographic data table and NMR spectra of synthesized compounds.

All crystallographic results were deposited to and are accessible from the Protein Data Bank (PDB) https://www.rcsb.org/. PDB accession codes for all structures derived from this study are as follows: **1** 6Q3C, **2** 6Q3B, **4** 6Q3F, **6** 6Q49, **7** 6Q48, **13** 6Q4B, **14** 6Q4A, **16** 6Q4C, **31** 6Q4D, **32** 6Q4F, **33** 6Q4E, **34** 6Q4J, **35** 6Q4I, **36** 6Q4H, **37** 6Q4G and **38** 6Q4K.

Corresponding Author

Michael J. Waring, email: mike.waring@ncl.ac.uk, Tel. +44 (0) 191 208 8591.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

We gratefully acknowledge financial support from Cancer Research UK (grant reference C2115/A21421), the Medical Research Council (grant reference MR/N009738/1), Astex Pharmaceuticals, the JGW Paterson Foundation (studentship award to DJW), Newcastle University (studentship award to JDL-F) and Jordan University of Science & Technology (studentship award to IA-K). We thank beamline staff at Diamond Light Source (Oxford) for help in refinement of CDK2 structures.

REFERENCES

REFERENCES

(1) Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **2002**, *1* (9), 727–730.

(2) Homans, S. W. Water, water everywhere--except where it matters? *Drug Discov. Today* **2007**, *12* (13–14), 534–539.

(3) Arrowsmith, C. H.; Audia, J. E.; Austin, C.; Baell, J.; Bennett, J.; Blagg, J.; Bountra, C.; Brennan, P. E.; Brown, P. J.; Bunnage, M. E.; Buser-Doepner, C.; Campbell, R. M.; Carter, A. J.; Cohen, P.; Copeland, R. A.; Cravatt, B.; Dahlin, J. L.; Dhanak, D.; Edwards, A. M.; Frederiksen, M.; Frye, S. V; Gray, N.; Grimshaw, C. E.; Hepworth, D.; Howe, T.; Huber, K. V. M.; Jin, J.; Knapp, S.; Kotz, J. D.; Kruger, R. G.; Lowe, D.; Mader, M. M.; Marsden, B.; Mueller-Fahrnow, A.; Müller, S.; O'Hagan, R. C.; Overington, J. P.; Owen, D. R.; Rosenberg, S. H.; Roth, B.; Roth, B.; Ross, R.; Schapira, M.; Schreiber, S. L.; Shoichet, B.; Sundström, M.; Superti-Furga, G.; Taunton, J.; Toledo-Sherman, L.; Walpole, C.; Walters, M. A.; Willson, T. M.; Workman, P.; Young, R. N.; Zuercher, W. J. The promise and peril of chemical probes. *Nat. Chem. Biol.* **2015**, *11* (8), 536–541.

(4) Waring, M. J. Lipophilicity in drug discovery. *Expert Opin. Drug Discov.* **2010**, *5* (3), 235–248.

(5) Young, R. J.; Green, D. V. S.; Luscombe, C. N.; Hill, A. P. Getting physical in drug discovery II: the impact of chromatographic hydrophobicity measurements and aromaticity. *Drug Discov. Today* **2011**, *16* (17–18), 822–830.

(6) Kawasaki, Y.; Freire, E. Finding a better path to drug selectivity. *Drug Discov. Today* **2011**, *16* (21–22), 985–990.

(7) Sankararaman, S.; Sha, F.; Kirsch, J. F.; Jordan, M. I.; Sjölander, K. Active site prediction using evolutionary and structural information. *Bioinformatics* **2010**, *26* (5), 617–624.

(8) Lu, S.; Huang, W.; Zhang, J. Recent computational advances in the identification of allosteric sites in proteins. *Drug Discov. Today* **2014**, *19* (10), 1595–1600.

(9) Rathi, P. C.; Ludlow, R. F.; Hall, R. J.; Murray, C. W.; Mortenson, P. N.; Verdonk, M. L. Predicting "hot" and "warm" spots for fragment binding. *J. Med. Chem.* **2017**, *60* (9), 4036–4046.

(10) Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **2009**, *49* (2), 377–389.

(11) Ward, R. A. Using protein-ligand docking to assess the chemical tractability of inhibiting a protein target. *J. Mol. Model.* **2010**, *16* (12), 1833–1843.

(12) Gee, C. T.; Arntson, K. E.; Urick, A. K.; Mishra, N. K.; Hawk, L. M. L.; Wisniewski, A. J.; Pomerantz, W. C. K. Protein-observed 19F-NMR for fragment screening, affinity quantification and druggability assessment. *Nat. Protoc.* **2016**, *11* (8), 1414–1427.

(13) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **2005**, *48* (7), 2518–2525.

(14) Davies, T. G.; Tickle, I. J. Fragment screening using X-ray crystallography. *Top. Curr. Chem.* **2012**, *317*, 33–59.

(15) Radeva, N.; Krimmer, S. G.; Stieler, M.; Fu, K.; Wang, X.; Ehrmann, F. R.; Metz, A.; Huschmann, F. U.; Weiss, M. S.; Mueller, U.; Schiebel, J.; Heine, A.; Klebe, G. Experimental active-site mapping by fragments: hot spots remote from the catalytic center of endothiapepsin. *J. Med. Chem.* **2016**, *59* (16), 7561–7575.

(16) Betzi, S.; Alam, R.; Martin, M.; Lubbers, D. J.; Han, H.; Jakkaraj, S. R.; Georg, G. I.; Schönbrunn, E. Discovery of a potential allosteric ligand binding site in CDK2. *ACS Chem. Biol.* **2011**, *6* (5), 492–501.

(17) Hubbard, R. E.; Murray, J. B. Experiences in fragment-based lead discovery. *Methods Enzymol.* **2011**, *493*, 509–531.

(18) Hann, M. M.; Leach, A. R.; Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (3), 856–864.

(19) Leach, A. R.; Hann, M. M. Molecular complexity and fragment-based drug discovery: ten years on. *Curr. Opin. Chem. Biol.* **2011**, *15* (4), 489–496.

(20) Fink, T.; Reymond, J.-L. Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. *J. Chem. Inf. Model.* **2007**, *47* (2), 342–353.

(21) Blum, L. C.; Reymond, J.-L. 970 Million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.* **2009**, *131* (25), 8732–8733.

(22) Keserű, G. M.; Erlanson, D. A.; Ferenczy, G. G.; Hann, M. M.; Murray, C. W.; Pickett, S. D. 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.* **2016**, *59* (18), 8189–8206.

(23) Hunter, C. A. Quantifying intermolecular interactions: guidelines for the molecular recognition toolbox. *Angew. Chemie Int. Ed.* **2004**, *43* (40), 5310–5324.

(24) Tiefenbrunn, T.; Forli, S.; Happer, M.; Gonzalez, A.; Tsai, Y.; Soltis, M.; Elder, J. H.; Olson, A. J.; Stout, C. D. Crystallographic fragment-based drug discovery: use of a brominated fragment library targeting HIV protease. *Chem. Biol. Drug Des.* **2014**, *83* (2), 141–148.

(25) Wilcken, R.; Liu, X.; Zimmermann, M. O.; Rutherford, T. J.; Fersht, A. R.; Joerger, A. C.; Boeckler, F. M. Halogen-enriched fragment libraries as leads for drug rescue of mutant p53. *J. Am. Chem. Soc.* **2012**, *134* (15), 6810–6818.

(26) Blaney, J.; Nienaber, V.; Burley, S. K. Fragment-based Lead Discovery and Optimization Using X-ray Crystallography, Computational Chemistry, and High-throughput Organic Synthesis. In Fragment-based Approaches in Drug Discovery (eds. R. Mannhold, H. Kubinyi, G. Folkers, W. Jahnke and D. A. Erlanson) 2006; pp. 215–248, Wiley-VCH Verlag, Weinheim, Germany.

(27) Bauman, J. D.; Harrison, J. J. E. K.; Arnold, E. Rapid experimental SAD phasing and hot-spot identification with halogenated fragments. *IUCrJ.* **2016**, *3*, 51–60.

(28) Sharma, P. S.; Sharma, R.; Tyagi, R. Inhibitors of cyclin dependent kinases: useful targets for cancer treatment. *Curr. Cancer Drug Targets* **2008**, *8* (1), 53–75.

(29) Ludlow, R. F.; Verdonk, M. L.; Saini, H. K.; Tickle, I. J.; Jhoti, H. Detection of secondary binding sites in proteins using fragment screening. *Proc. Natl. Acad. Sci.* **2015**, *112* (52), 15910–15915.

(30) Pearce, N. M.; Krojer, T.; Bradley, A. R.; Collins, P.; Nowak, R. P.; Talon, R.; Marsden, B. D.; Kelm, S.; Shi, J.; Deane, C. M.; von Delft, F. A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.* **2017**, *8*, 15123.

(31) Antolín, A. A.; Mestres, J. Distant polypharmacology among MLP chemical probes. *ACS Chem. Biol.* **2015**, *10* (2), 395–400.

(32) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **2004**, *9* (10), 430–431.

(33) Arris, C. E.; Boyle, F. T.; Calvert, A. H.; Curtin, N. J.; Endicott, J. A.; Garman, E. F.; Gibson, A. E.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Lawrie, A. M.; Newell, D. R.; Noble, M. E.; Sausville, E. A.; Schultz, R.; Yu, W. Identification of novel purine and pyrimidine cyclin-dependent kinase inhibitors with distinct molecular interactions and tumor cell growth inhibition profiles. *J. Med. Chem.* **2000**, *43* (15), 2797–2804.

(34) Wyatt, P. G.; Woodhead, A. J.; Berdini, V.; Boulstridge, J. A.; Carr, M. G.; Cross, D. M.; Davis, D. J.; Devine, L. A.; Early, T. R.; Feltell, R. E.; Lewis, E. J.; McMenamin, R. L.; Navarro, E. F.; O'Brien, M. A.; O'Reilly, M.; Reule, M.; Saxty, G.; Seavers, L. C. A.; Smith, D.-M.; Squires, M. S.; Trewartha, G.; Walker, M. T.; Woolford, A. J.-A. Identification of N -(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1 H -pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. *J. Med. Chem.* **2008**, *51* (16), 4986–4999.

(35) Xing, L.; Klug-Mcleod, J.; Rai, B.; Lunney, E. A. Kinase hinge binding scaffolds and their hydrogen bond patterns. *Bioorg. Med. Chem.* **2015**, *23* (19), 6520–6527.

(36) Garman, E. F. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (4), 339–351.

(37) Blanksby, S. J.; Ellison, G. B. Bond dissociation energies of organic molecules. *Acc. Chem. Res.* **2003**, *36* (4), 255–263.

(38) Ramagopal, U. A.; Dauter, Z.; Thirumuruhan, R.; Fedorov, E.; Almo, S. C. Radiation-induced site-specific damage of mercury derivatives: phasing and implications. *Acta Crystallogr. D. Biol. Crystallogr.* **2005**, *61* (Pt 9), 1289–1298.

(39) Martin, M. P.; Alam, R.; Betzi, S.; Ingles, D. J.; Zhu, J.-Y.; Schönbrunn, E. A novel approach to the discovery of small-molecule ligands of CDK2. *ChemBioChem* **2012**, *13* (14), 2128–2136.

(40) Brown, N. R.; Noble, M. E. M.; Lawrie, A. M.; Morris, M. C.; Tunnah, P.; Divita, G.; Johnson, L. N.; Endicott, J. A. Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity. *J. Biol. Chem.* **1999**, *274* (13), 8746–8756.

(41) Betzi, S.; Alam, R.; Martin, M.; Lubbers, D. J.; Han, H.; Jakkaraj, S. R.; Georg, G. I.; Schönbrunn, E. Discovery of a potential allosteric ligand binding site in CDK2. *ACS Chem. Biol.* **2011**, *6* (5), 492–501.

(42) Potterton, L.; Agirre, J.; Ballard, C.; Cowtan, K.; Dodson, E.; Evans, P. R.; Jenkins, H. T.; Keegan, R.; Krissinel, E.; Stevenson, K.; Lebedev, A.; McNicholas, S. J.; Nicholls, R. A.; Noble, M.; Pannu, N. S.; Roth, C.; Sheldrick, G.; Skubak, P.; Turkenburg, J.; Uski, V.; von Delft, F.; Waterman, D.; Wilson, K.; Winn, M.; Wojdyr, M. CCP 4 i 2: the new graphical user interface to the CCP 4 program suite. *Acta Crystallogr. Sect. D Struct. Biol.* **2018**, *74* (2), 68–84.

(43) Waterman, D. G.; Winter, G.; Gildea, R. J.; Parkhurst, J. M.; Brewster, A. S.; Sauter, N. K.; Evans, G. Diffraction-geometry refinement in the DIALS framework. *Acta Crystallogr. Sect. D Struct. Biol.* **2016**, *72* (4), 558–575.

(44) Winter, G. Xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **2010**, *43* (1), 186–190.

(45) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40* (Pt 4), 658–674.

(46) De Bondt, H. L.; Rosenblatt, J.; Jancarik, J.; Jones, H. D.; Morgan, D. O.; Kim, S. H. Crystal structure of cyclin-dependent kinase 2. *Nature* **1993**, *363* (6430), 595–602.

(47) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.* **2011**, *67* (Pt 4), 355–367.

(48) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66* (Pt 4), 486–501.

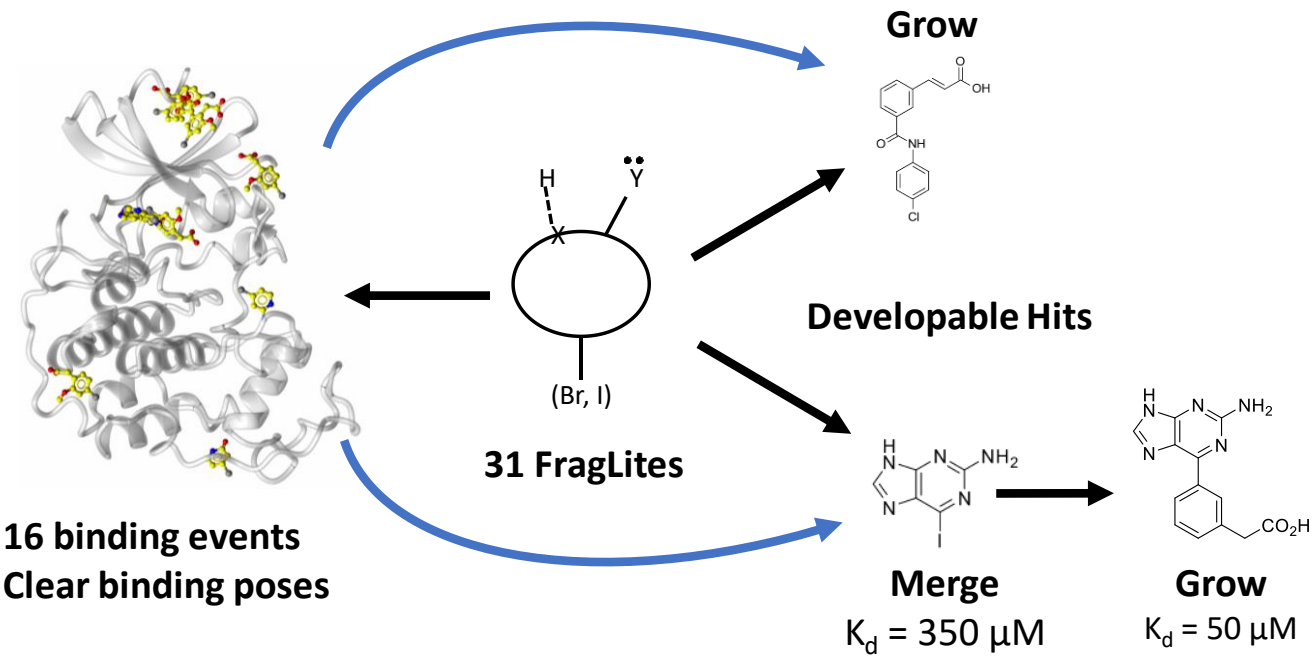
(49) Long, F.; Nicholls, R. A.; Emsley, P.; Graæulis, S.; Merkys, A.; Vaitkus, A.; Murshudov, G. N. AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr. Sect. D, Struct. Biol.* **2017**, *73* (Pt 2), 112–122.

(50) McNicholas, S.; Potterton, E.; Wilson, K. S.; Noble, M. E. M. AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr. D. Biol. Crystallogr.* **2011**, *67* (Pt 4), 386–394.

(51) Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L. N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **1989**, *179* (1), 131–137.

(52) Turnbull, W. B.; Daranas, A. H. On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc.* **2003**, *125* (48), 14859–14866.

TOC graphic


Grow
H
Y
X
(Br, I)
Developable Hits
31 FragLites
16 binding events
Clear binding poses
Merge
K_d = 350 µM
Grow
K_d = 50 µM