Efficient Ligand Discovery Using Sulfur(VI) Fluoride Reactive Fragments

Arron Aatkar[1,2], Aini Vuorinen[1,3], Oliver E. Longfield[1,2], Katharine Gilbert[1,2], Rachel Peltier-Heap[4], Craig D. Wagner,[4] Francesca Zappacosta[4], Katrin Rittinger[3], Chun-wa Chung[1], David House[1,3], Nicholas C. O. Tomkinson[2] & Jacob T. Bush[1,3]

Sulfur(VI) fluorides (SFs) have emerged as valuable electrophiles for the design of 'beyond cysteine' covalent inhibitors, and offer potential for expansion of the liganded proteome. Since SFs target a broad range of nucleophilic amino acids, they deliver an approach for the covalent modification of proteins without requirement for a proximal cysteine residue. Further to this, libraries of reactive fragments present an innovative approach for the discovery of ligands and tools for proteins of interest by leveraging a breadth of mass spectrometry analytical approaches. Herein, we report a screening approach that exploits the unique properties of SFs for this purpose. Libraries of SF-containing reactive fragments were synthesised, and a direct-to-biology workflow was taken to efficiently identify hit compounds for CAII and BCL6. The most promising hits were further characterised to establish the site(s) of covalent modification, modification kinetics, and target engagement in cells. Crystallography was used to gain a detailed molecular understanding of how these reactive fragments bind to their target. It is anticipated that this screening protocol can be used for the accelerated discovery of ‘beyond cysteine’ covalent inhibitors.

[1]GSK, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK. [2]Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow, G1 1XL, UK. [3]The Francis Crick Institute, London, NW1 1AT, UK. [4]GSK, South Collegeville Road, Collegeville, PA 19426, USA. Correspondence and requests for materials should be addressed to J.T.B (email: jacob.x.bush@gsk.com).

Introduction

The impact of cysteine-targeting covalent modifiers has spurred interest in the development of complementary 'beyond-cysteine' approaches to target additional amino acid residues and thus expand applicability across the proteome (Fig. 1a).[1–5] Sulfur(VI) fluorides (SFs), have emerged as useful electrophiles for this application, targeting multiple nucleophilic amino acid residues, including: lysine,[6] tyrosine,[7] and serine.[8] The prevalence of these residues in almost all protein pockets makes SFs promising functional groups for the development of covalent inhibitors for proteins, and expansion of the liganded proteome.[9–16] Recently, several SF-containing modulators have been reported, which enabled covalent modification of protein pockets without targeting a cysteine. Examples include 'XO44' for broad-spectrum kinase profiling,[17] various SF-containing ligands for targeting G protein-coupled receptors including the human adenosine A_3 receptor,[18] and 'EM12-SO_2F'/'EM12-FS' which modulate cereblon.[19] These were developed by structure-based, rational installation of the SF group on optimised non-covalent scaffolds. The development of complementary 'bottom-up' approaches will be useful for discovering ligands for targets that have low tractability to non-covalent ligands.[20]

Reactive fragment screening has emerged as a useful strategy for the discovery of chemical probes for protein targets of interest.[21] These approaches couple the utility of fragments in enabling the efficient coverage of chemical space, with a reactive functionality that traps weak protein-ligand interactions to improve binding, and enable robust detection by intact protein liquid chromatography-mass spectrometry (LC-MS) (Fig. 1b).[22,23] Covalent capture provides access to a suite of follow-up studies that further characterise the interaction, including determination of the site(s) of binding, measurement of kinetic parameters, and assessment of in-cell target engagement.[3,24,25] To date, the approach has been limited to cysteine-targeting covalent inhibitors for challenging targets, including HOIP, and $KRAS^{G12C}$, which led to the discovery of the FDA approved therapeutic AMG 510.[3,26] Chemistries that enable the screening of electrophilic libraries to target alternative nucleophilic residues would greatly expand the number of proteins that are amenable to reactive fragment screening technologies.

Here we report a SF reactive fragment screening approach for the identification of covalent ligands for proteins of interest. This approach enabled the rapid discovery of novel ligands for multiple protein pockets without reliance upon the presence of a cysteine residue. This strategy employed a high-throughput chemistry direct-to-biology (HTC-D2B) workflow, providing an expedient and accessible method for the rapid and iterative generation of SF reactive fragment libraries.[27–29]

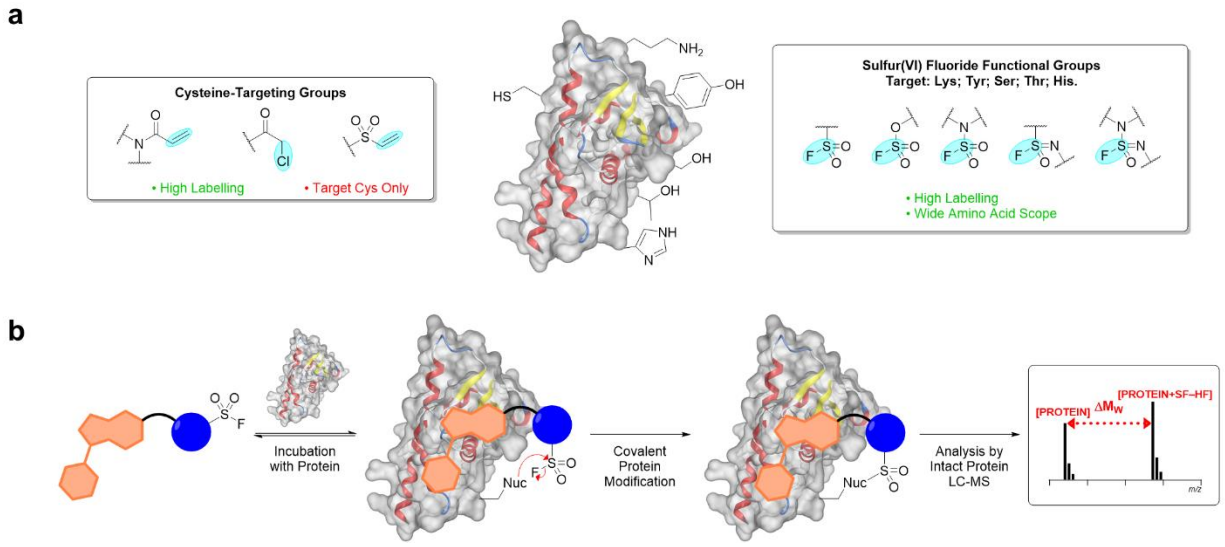


Fig. 1 | Approaches to covalent protein modification in chemical biology and drug discovery. (a) Summary of electrophiles commonly used for the covalent modification of proteins, including cysteine-targeting reactive groups, and SF functional groups. **(b)** Schematic representation of the mechanism associated with a SF-containing ligand targeting a nucleophilic amino acid residue, and subsequent read-out by intact protein LC-MS.

Results

The SF-based reactive fragment screening approach was developed in two stages: first, the development of an appropriate SF fragment library, and second, the application of this library in screens against three protein targets. A modular library was employed by linking a diverse set of amine-functionalised fragments to a SF-containing reactive moiety. Three SFs were selected that spanned a range of intrinsic reactivities, based on previous profiling of SF functionalities.[14] These included aromatic sulfonyl fluorides **1** (*meta*-substituted) and **2** (*para*-substituted with a methylene spacer), as well as sulfamoyl fluoride **3** (azetidine-linked) (Fig. 2a).

High-throughput chemistry synthesis of a SF reactive fragment library

A HTC protocol for the generation of the reactive fragment libraries was pursued to enable the rapid generation of SF fragments in 384-well plates. This allowed for screening in a D2B format as crude reaction mixtures, circumventing the requirement for purification and thus accelerating reactive fragment library screens.[29]

A succinimide-activated (OSu) amide coupling was employed using DMSO and *N*-ethylmorpholine (NEM) as the solvent and base, respectively.[29] The conditions were initially trialled on a panel of 12 diverse amine-functionalised fragments by addition of SF reactive moieties (**1a**, **2a**, and **3a**) in both dry DMSO, and DMSO:water (9:1) to assess robustness to hydrolysis under the reaction conditions. Reactions with OSu esters **2a** and **3a** afforded good conversion to the desired products and tolerance of 10% water, while *meta*-substituted OSu ester **1a** gave poorer conversions due to hydrolysis of the SF group to the sulfonic acid (*see Fig. S1 & S2*).[14]

Two 352-membered SF reactive fragment libraries were subsequently synthesised employing *para*-substituted reactive moiety **2**, and azetidine-linked reactive moiety **3**. A set of 352 amine-functionalised fragments were selected from the GSK compound collection by first filtering for fragment-like properties (aromatic ring count ≤2; HBDs/HBAs ≤4; heavy atoms ≤15; $150<M_W\leq250$), and then selecting for maximal chemical diversity by clustering on chemical fingerprints (*see Fig. S3*).[30] LC-MS analysis of six wells selected at random indicated good conversion to the desired products, and conversions were found to be highly reproducible across three library syntheses conducted on separate occasions (*see Fig. S4*).

SF screening applied to a range of purified proteins in 'direct-to-biology' approach

A panel of three proteins was selected for screening SF reactive fragment libraries: CAII, KRAS4B^{G12D}, and BCL6. These proteins were selected to sample broad structural diversity and biological function, while also being of therapeutic relevance. None of the proteins contain a catalytic nucleophilic amino acid residue, which allowed us to probe the utility of the SFs in targeting nucleophilic amino acid residues present in the vicinity of binding pockets.

Initially, the more reactive library, containing *para*-substituted aryl sulfonyl fluoride **2**, was screened against the three proteins (24 h incubation, 4 or 20 °C) and directly analysed by intact protein LC-MS (0.5 or 1 µM protein, 20 or 50 µM SF) (Fig. 2b). Across all screens, the majority of wells contained unmodified protein, indicating that the SF reactive moieties were not yielding non-specific covalent modifications. However, for some of the wells, the resultant mass spectra displayed additional peaks with mass shifts consistent with the covalent modification of the protein by the SF fragment, accompanied by the loss of HF as expected for the reaction between the nucleophilic amino acid residue with the SF group: [protein+SF–HF]. A range of modification yields were observed, and the hit threshold for each screen was defined as the mean percentage modification + 2 standard deviations (SDs). Hits were also prioritised based on the overall extent of modification, with some showing modification greater than 50% (e.g. **2b**–**e** with CAII), while others showed modification less than 50% (e.g. **2f**–**i** with BCL6) (Fig. 2c). All hits gave a single modification event on the protein, consistent with recognition-driven modification. Hits containing reactive moiety **2** from the screens against CAII and BCL6 were resynthesised and purified for use in further investigations. Disappointingly, no hits were observed for KRAS^{G12D}, consistent with the fact that this is considered to be a poorly tractable target.[25] The library containing the less reactive sulfamoyl fluoride **3** was screened against CAII but afforded no hits, suggesting an insufficient intrinsic reactivity of the electrophile.

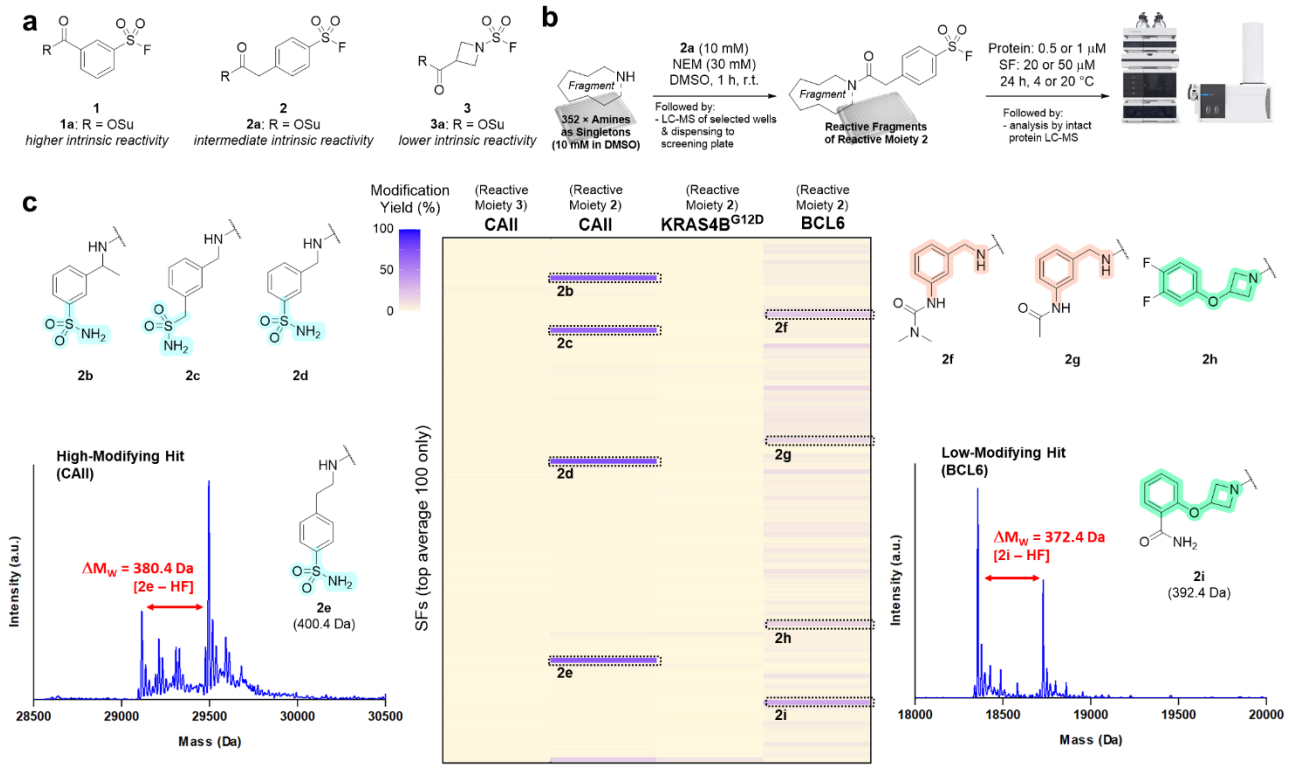


Fig. 2 | HTC-D2B protocol identifies hit compounds following screens against purified proteins. (a) Structures of SF reactive moieties **1**–**3** considered for the screening approach. **(b)** Overview of the high-throughput coupling of amine-functionalised fragments with the SF moiety, and subsequent screening of crude reaction mixtures against proteins of interest. **(c)** Heatmap summary of SF fragment screens against CAII, KRAS4B^{G12D}, and BCL6, with hit structures **2b**–**e** and **2f**–**i** shown alongside exemplar mass spectra.

Site(s) of binding identified for CAII hits through displacement and tandem MS studies

Carbonic anhydrase II (CAII) is a metalloenzyme responsible for the interconversion of carbon dioxide and bicarbonate, for which non-covalent inhibitors have been developed as treatments for glaucoma, oedema, and cancer.[31,32] CAII inhibitors are typically based on aromatic sulfonamide pharmacophores, which interact with the Zn^{2+} ion via the sulfonamide group.[33]

The HTC-D2B screen with CAII afforded four hits (**2b**–**e**) with modification yields of 66–75% (*see Fig. S5*). Three of these hits contained an aromatic sulfonamide, consistent with known inhibitors, while one of the four (**2c**) was an aliphatic sulfonamide, of which there are few prior reports. These four compounds were the only primary sulfonamides in the library, and the remainder of the library gave an average modification yield of <1%, indicating negligible non-specific modification, and high specificity of the hit fragment interactions.

The site of binding was investigated by displacement studies.[34] For this, we used the CAII inhibitor ethoxzolamide (**4**) (K_i = 8 nM), which is known to bind within the Zn^{2+} pocket.[31] The resynthesised and purified SFs (100 µM) were co-incubated with either ethoxzolamide (50 µM) or DMSO as a control, with CAII (0.5 µM). Inspection of the resultant mass spectra after a 24 h incubation revealed that the presence of ethoxzolamide had abolished covalent modification for all hits **2b**–**e**, indicating binding to the same site (Fig. 3a).

Tandem MS analyses were subsequently employed for identification of the amino acid residue(s) that were covalently modified. Samples of CAII modified by the resynthesised and purified SFs **2d** and **2e** were digested using trypsin and analysed by LC-MS/MS. This identified peptides $_{59}$ILNNGH*AFNVEFDDSQDKAVLK$_{80}$ and $_{59}$ILNNGH*AFNVEFDDSQDK$_{76}$ as the major site(s) of modification on His64 for both **2d** (366 Da) (*see Fig. S6*), and **2e** (380 Da) (Fig. 3b), respectively. Further to this, minor modifications were observed on the N-terminal peptide $_{1}$MSHHWGYGK$_{9}$ at His3 (**2d** and **2e**), and Tyr7 (**2d** only).

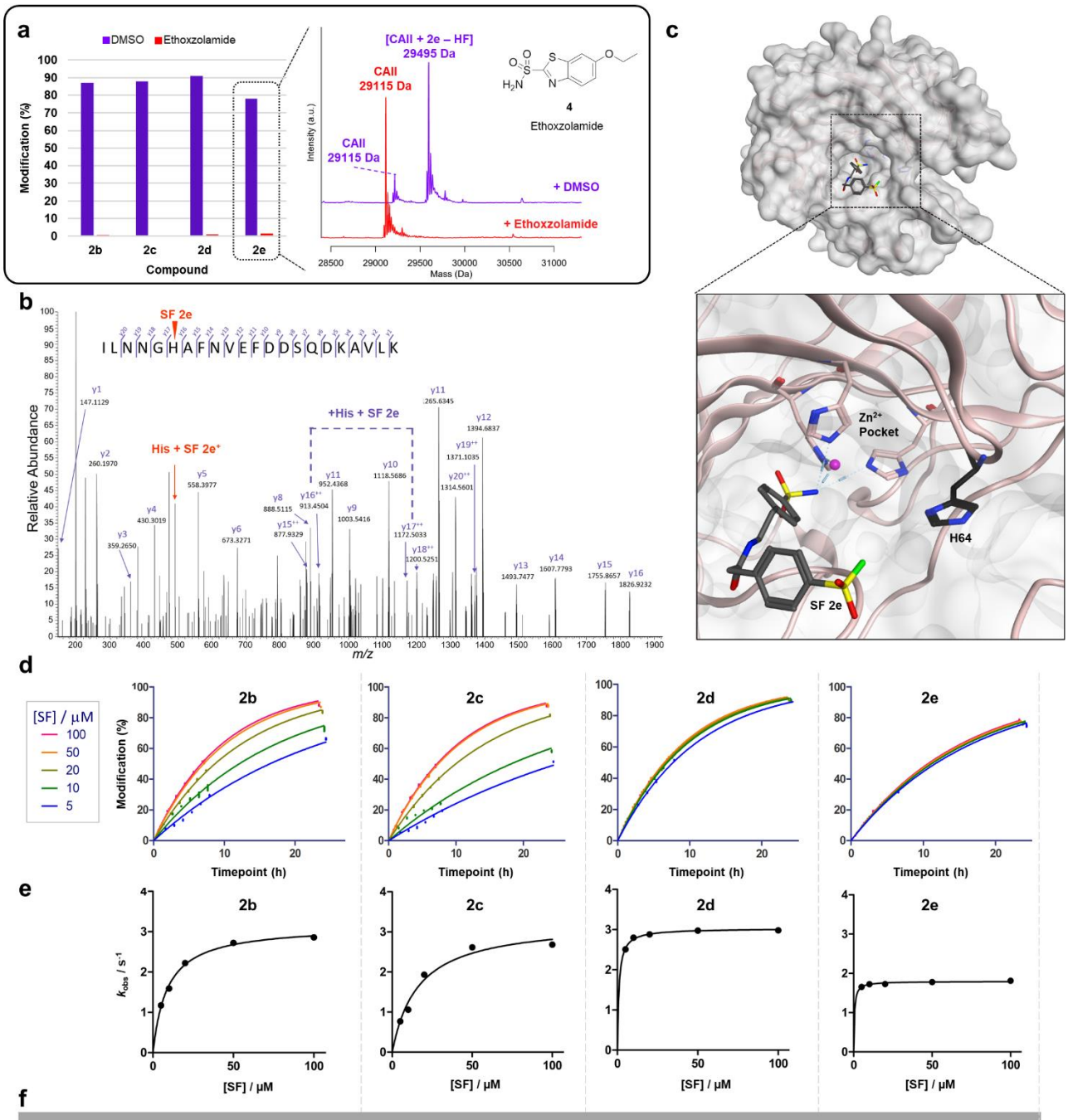
To rationalise these observations, virtual dockings were carried out on SFs **2d** and **2e**, based on the reported binding mode of ethoxzolamide (PDB: 3CAJ). For **2e**, the docking indicated that the sulfonamide was bound to the Zn^{2+} cofactor, and the His64 residue was proximal to the SF group, just 7 Å away, and appears well poised for reaction (Fig. 3c).[35] Similarly, for **2d**, the Zn^{2+}-sulfonamide interaction was observed, and the SF group was in the vicinity of His64, as well as residues His3 and Tyr7 which were also found to carry minor modifications (*see Fig. S7*).

Kinetic analyses reveal covalent modification efficiencies of CAII hits

The kinetics of binding were investigated by incubation of the resynthesised and purified SFs **2b**–**e** with CAII at a range of concentrations and analysis by intact protein LC-MS over a 24 h period (*see Fig. S8*). The concentration-response was analysed based on a two-step model of reversible ligand (L) binding and subsequent irreversible covalent modification of the protein (P) (P + L ⇌ P•L → PL) (*see Fig. S9*).[24] Time courses were fit to a single exponential function to give a k_{obs} for each concentration (Fig. 3d). These were then fitted using a Michaelis-Menten model to determine k_{inact} (corresponding to the rate of covalent bond formation) and K_I (corresponding to the recognition of the ligand for the protein pocket) (Fig. 3e & 3f). All fragments were found to have strong reversible affinity (K_I), and these were below the minimum concentration screened for **2d** and **2e** (<5 µM). This is consistent with the strong binding of aryl sulfonamides to the CAII Zn^{2+} binding pocket.[33] The SFs exhibited unexpectedly slow rates of covalent modification, with 50% modification achieved between 6.5–10.5 h for all 100 µM conditions. The slow rate of reaction may be attributed to a sub-optimal trajectory of the SF group toward the His64 residue in the reversible bound conformation, and/or low nucleophilicity of the residue.[14]

An advantage of the HTC-D2B approach taken with this screening strategy is the opportunity to rapidly explore iterative libraries of compounds which are structurally similar to the original hits identified. Such iteration-

based screens allow for faster design-make-test cycles in contrast to traditional methods in early-stage drug discovery, leading to the more efficient identification of potent and selective probes.[29] To explore whether any alternative sulfonamide-containing fragments would give high modification yields at a faster rate, a new library was designed. For this, 96 amine-functionalised fragment analogues of hits **2b**–**e** were selected, coupled to OSu ester **2a**, and the resultant library was incubated with CAII for 1 h prior to analysis by intact protein LC-MS. Inspection of the resultant mass spectra after this short incubation revealed many fragment hits that exhibited significantly higher modification yields (>70%, *see Fig. S10*) when compared with the original hits (<15% at 1 h, *see Fig. 3d*). This highlights the importance of electrophile trajectory in determining k_{inact}, and demonstrates the utility of the HTC-D2B approach as a means to rapidly optimise kinetic parameters.



f

SF	2b	2c	2d	2e
k_{inact} / 10^{-5} s^{-1}	3.2	3.3	3.0	1.8
K_I / µM	8.9	16	<5	<5
(k_{inact} / K_I) / (M^{-1} s^{-1})	3.6	2.0	>6.1	>3.6

Fig. 3 | Determination of the site(s) of modification, rationalisation with docking, and kinetic analyses for the measure of respective covalent modification efficiencies of CAII hits. (a) Modifications observed in the displacement studies and exemplar spectra observed following the displacement SF **2e** by ethoxzolamide. Final concentrations: 0.5 µM protein; 100 µM SF; 50 µM ethoxzolamide. **(b)** Exemplar MS/MS spectrum of peptide $_{59}$ILNNGH*AFNVEFDDSQDKAVLK$_{80}$ modified by SF **2e** confirming His64 as the site of covalent modification. **(c)** X-ray crystal structure of CAII (PDB: 3CAJ), and virtual docking showing SF **2e** in the CAII pocket, with sulfonamide bound to the active site Zn^{2+} cofactor and sulfonyl fluoride group proximal to the His64 residue. **(d)** Time courses (various concentrations plotted against time and fitted to a single exponential function to determine k_{obs}) showing concentration-dependent modification of SFs **2b**–**e** with CAII. **(e)** k_{obs} measurements plotted against the measured concentrations of SFs **2b**–**e** to determine k_{inact} and K_I. **(f)** Table displaying k_{inact}, K_I, and hence k_{inact}/K_I – a parameter to describe the overall modification efficiencies of SFs **2b**–**e**.

Assessment of target engagement in cells for CAII hits

Hits from the CAII screen were progressed to chemoproteomic profiling to determine cellular target engagement and to measure their off-target profiles. SFs **2d** and **2e** were functionalised with an alkyne handle at the linking amide group to give probes **2j** and **2k**. A structurally similar negative control was also designed by substitution of the sulfonamide for a methoxy group, **2l** (Fig. 4a). The three probes were first incubated with purified CAII, which confirmed that SFs **2j** and **2k** covalently modified CAII following addition of the alkyne group, and that the methoxy group-containing negative control **2l** did not. The proteins engaged by these three probes were subsequently studied in HEK293T cells.

HEK293T cells were treated for 1 h with probes **2j**, **2k**, or **2l** (10 μM), or DMSO vehicle. Competition-based experiments were also conducted where cells were pre-treated with parent hits **2d** (40 μM), or **2e** (40 μM) for 1 h, before addition of alkyne-containing probes **2j** (10 μM) or **2k** (10 μM), respectively. Following incubation, treated cells were lysed, conjugated with biotin-azide by Cu-click, and biotinylated proteins were enriched using NeutrAvidin beads. Enriched proteins were digested with LysC and trypsin, prior to analysis by LC-MS/MS (Fig. 4b).

All three probes (**2j**–**l**) were found to enrich multiple proteins by comparison to the DMSO control (523, 501, and 447 respectively, $\log_2$ ratio ≥0.58, p-value ≤0.05, #unique peptides ≥2), highlighting the promiscuity of the reactive fragments in this environment. A good correlation was observed between all three probes and the proteins enriched, indicating that most of the enrichment was driven by the SF moiety, rather than the fragment portion of the probes (Fig. 4c). The most significantly enriched proteins included: FABP5, APMAP, and CRABP2. Previous reports identified that FABP5 and CRABP2 were targeted by an arylfluorosulfate probe by reaction at a tyrosine residue in each of these proteins.[36] While some enrichment of CAII was observed for **2j** and **2k**, this was poorly resolved amongst the many other enriched proteins. To further investigate CAII engagement of the sulfonamide-containing active probes (**2j** and **2k**) in live cells, we compared the enriched proteins to those enriched by the methoxy-containing negative control (**2l**). This highlighted CAII as one of the few differentially enriched proteins, suggesting that the aryl sulfonamide fragment hits were driving cellular engagement of CAII (Fig. 4d & 4e).

Interestingly, competition experiments with **2d** and **2e** showed relatively few significantly competed proteins, suggesting that many of the interactions involved sub-stoichiometric binding. This was true of CAII, where neither probes **2j** or **2k** showed competition with parents **2d** or **2e**, respectively. This is consistent with the slow rate of covalent modification of CAII by **2d** and **2e** in the biochemical kinetic analyses, which suggested that the protein would only be partially modified after a 1 h incubation. Together, these results indicate that while the aryl sulfonamide fragment hits did engage CAII in cells, it was with low stoichiometry and poor selectivity over many additional off-targets.

Further investigation of the off-targets revealed that some proteins were competed in the presence of parent, indicating high levels of engagement; these included: FABP5, PEBP1, ABHD6, and LYPLA1. (Fig. 4f & 4g).[37] ABHD6 and LYPLA1 catalyse the hydrolysis of esters and thioesters, respectively, and have nucleophilic catalytic residues which are likely to react with the SF moiety.[38,39] Similarly, PEBP1 is a phospholipid binder, and is the prototype of a novel family of serine protease inhibitors.[40] More specifically, each of these proteins bind molecules containing hydrophobic alkyl chains, consistent with previous reports involving SF-containing probes targeting FABP5 in cells.[36] The high levels of occupancy at these targets suggests they may be amenable to the development of potent SF-based covalent inhibitors.

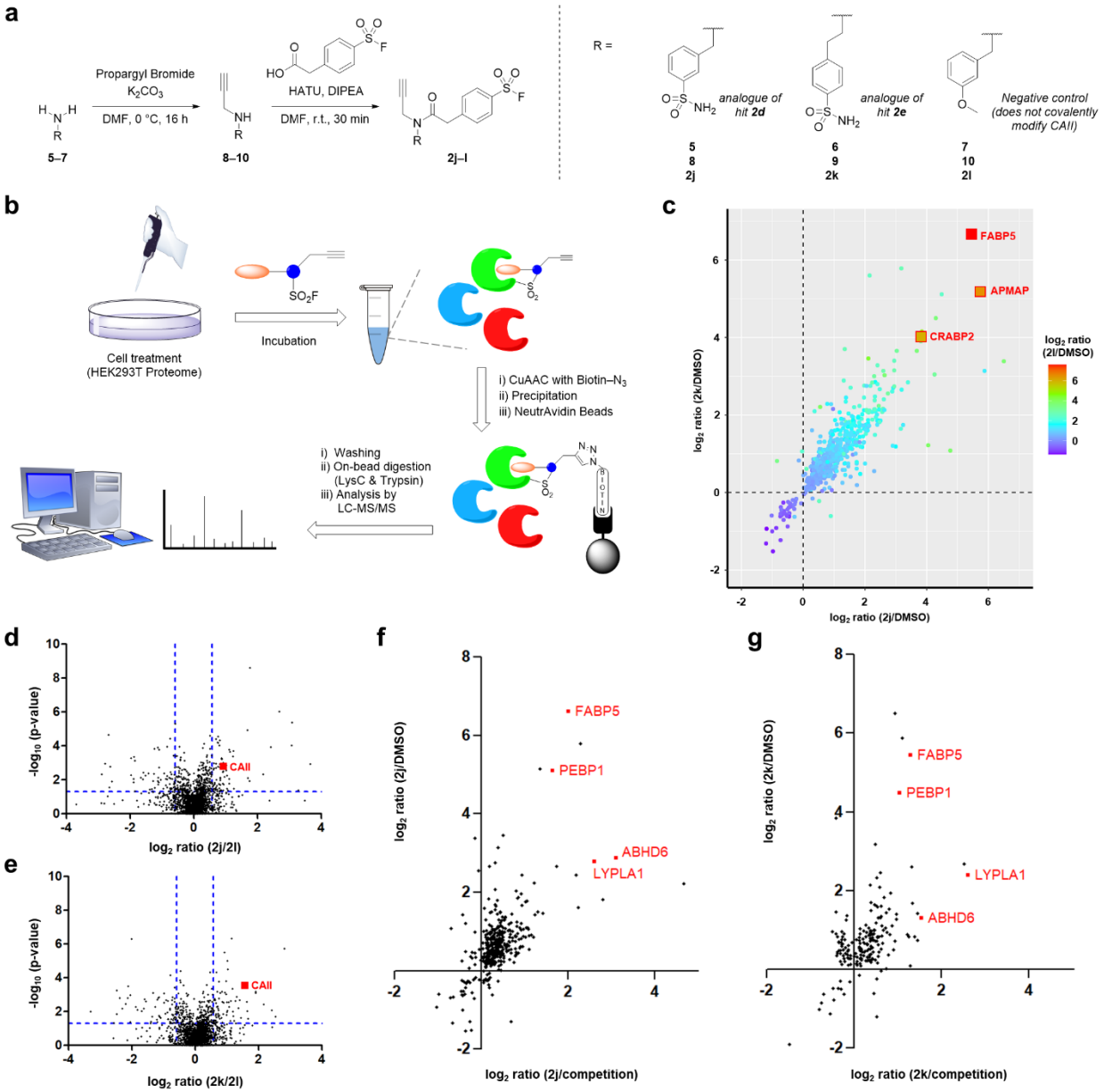


Fig. 4 | Assessing target engagement in cells with chemoproteomics. (a) Scheme depicting the synthesis of probes **2j** and **2k** (alkyne-functionalised analogues of CAII hits **2e** and **2d**, respectively), and negative control **2l**. **(b)** Schematic representation of MS-based proteomic workflow used to assess target engagement in cells. **(c)** Plot of **2j** and **2k** log_2 ratios coloured by **2l** log_2 ratio, highlighting commonly enriched proteins. **(d)** Volcano plot highlighting CAII enrichment by SF **2j**, plotted as a log_2 ratio compared to negative control **2l**. Blue dashed lines correspond to thresholds: log_2 ratio ≥0.58; p-value ≤0.05. **(e)** Volcano plot highlighting CAII enrichment by SF **2k**, plotted as a log_2 ratio compared to negative control **2l**. Blue dashed lines correspond to thresholds: log_2 ratio ≥0.58; p-value ≤0.05. **(f)** Log_2 representation of fold-difference between probe **2j**/competition and **2j**/DMSO, highlighting commonly enriched proteins. **(g)** Log_2 representation of fold-difference between probe **2k**/competition and **2k**/DMSO, highlighting commonly enriched proteins.

Follow-up studies on BCL6: site(s) of binding studies and hit expansion

The transcription factor B-cell lymphoma 6 (BCL6) has been identified as a driver of oncogenesis in lymphoid malignancies. BCL6 is implicated in several protein-protein interactions with corepressors, and disruption of these interactions are currently being investigated as a strategy for cancer treatment.[41] A range of small molecules that target BCL6 have been reported, including reversible inhibitors such as compound **11** (GSK137), and macrocyclic compound **12**.[41,42] Compounds with alternative mechanisms have also been reported, including degraders such as compound **13** (BI-3802), and more recently, rationally designed covalent inhibitor **14** (TMX-2164) that targets Tyr57 (Fig. 5a & 5b).[43,44]

The top hits from the SF screen with BCL6 represented two distinct chemotypes: *meta*-substituted benzamides (**2f** and **2g**) and azetidine aryl ethers (**2h** and **2i**). These were resynthesised and purified for follow-up studies. Tandem MS was used to identify the amino acid residue(s) responsible for covalent modification, and interestingly, the two different chemotypes were found to target two different residues (Fig. 5c & 5d). The results showed that *meta*-substituted benzamides **2f** (373 Da) and **2g** (344 Da) modified Tyr57 on the peptide $_{47}$TVLMACSGLFY*SIFTDQLKR$_{66}$, and azetidine aryl ethers **2h** (365 Da) and **2i** (372 Da) modified His115 on the peptide $_{98}$EGNIMAVMATAMYLQMEH*VVDTCR$_{121}$ (*see Fig. S11*). Tyr57 has previously been targeted by a BCL6 inhibitor (**14**, TMX-2164), however, the targeting of His115 is a novel modification; these residues are located on opposite sides of the binding site of previously reported BCL6 inhibitors (**11**–**14**) (*see Fig. 5b*).

An iterative screen was subsequently carried out with the intention of expanding the pool of hit compounds and identifying BCL6 hits with higher covalent modification yields. For this, a new 352-membered library of SF-based reactive fragments was designed based on a similarity search of available amine-functionalised fragments using the top four original hits (**2f**–**i**). The library was generated by HTC, incubated with BCL6 in a D2B fashion, and subsequently analysed by intact protein LC-MS. The extent of covalent modifications observed for this screen were markedly higher, with many further hits discovered (**2m**–**t**). This included five reactive fragments which gave covalent modification yields greater than the maximum modification yields observed in the original screen, and three hits with over 50% modification (Fig. 5e & 5f). The *meta*-substituted benzamides gave the higher modification yields, and several different groups in the *meta*-position were well-tolerated. Other hits showed that different *N*-containing saturated heterocycles with alternative heteroatom links to aryl groups were also tolerated by the binding pocket.

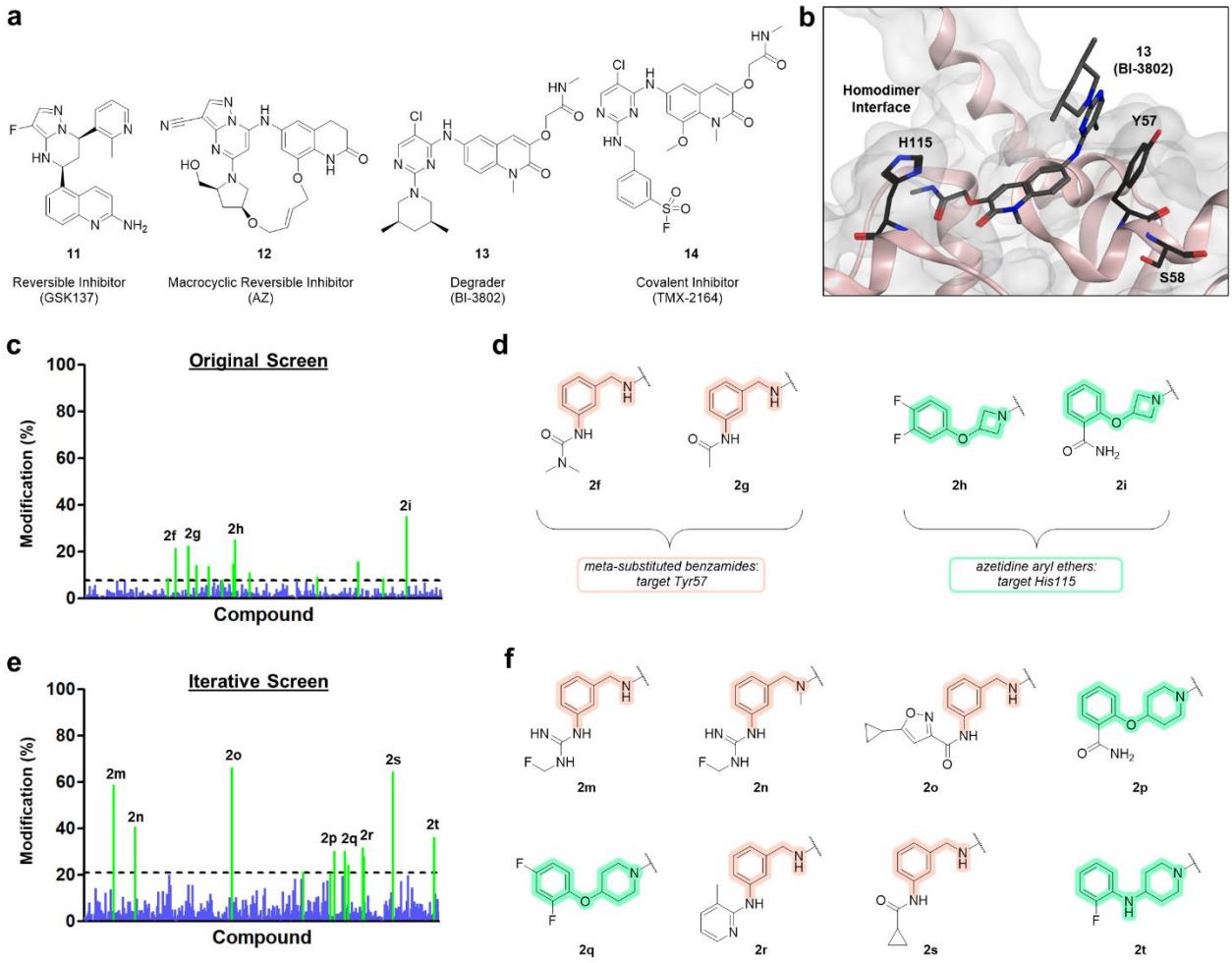


Fig. 5 | BCL6 hits validated and expanded upon with iterative screen. (a) Structures of reported BCL6 binders **11**–**14**. **(b)** Crystal structure of compound **13** (BI-3802) in complex with BCL6 BTB/POZ domain, highlighting proximal nucleophilic amino acid residues (PDB: 5MW2). **(c)** Summary of original screen against BCL6 with reactive moiety **2**. Dashed line shows hit threshold at 10%; hits coloured green; non-hits coloured blue. **(d)** Structures of hits **2f**–**i** involving two chemotypes: *meta*-substituted benzamide fragments (highlighted peach), and azetidinyl/piperidinyl fragments (highlighted turquoise). **(e)** Summary of iterative screen against BCL6 with library of hit fragment analogues. Dashed line shows hit threshold at 21%; hits coloured green; non-hits coloured blue. **(f)** Structures of hits **2m**–**t** with two consistent chemotypes.

Structural and biophysical investigations into BCL6 hits

With additional hits identified from the iterative screen, we sought to obtain further structural and biophysical information to better understand how these reactive fragments bind to BCL6. A total of eight SF hits were resynthesised and purified: the four from the original screen (**2f**–**i**), and four from the iterative screen (**2p**, **2q**, **2s**, **2t**). To confirm that these eight hits targeted the same binding site as previously reported binders, GSK137 (biochemical pIC50 = 8) was used as a known inhibitor for a displacement experiment.[42] The SFs (100 µM) were incubated for 24 h with BCL6 in the presence of GSK137 (100 µM) or DMSO as a control, and the samples were analysed by intact protein LC-MS. The resultant mass spectra showed that covalent modification was attenuated by 78–96% for all eight SFs, indicating that all hits were competing for the same site (Fig. 6a).[42]

We subsequently analysed protein-fragment interactions using differential scanning fluorimetry (DSF) to explore the impact of these binding events on protein stability.[45] The azetidinyl/piperidinyl fragments exhibited negligible thermal shifts (*see Fig. S12b & S12c*), however, the *meta*-substituted benzamide complexes were highly thermally stabilised (ΔT_m ~7 °C) relative to unmodified BCL6, which implied the presence of intermolecular interactions from ligand binding (Fig. 6b). This level of stabilisation was comparable to that observed for the potent, non-covalent inhibitor GSK137 (ΔT_m ~12 °C, *see Fig. S12d*).[42] Kinetic analyses of the

four hits from the original screen (**2f**–**i**) revealed that for the *meta*-substituted benzamides, the reversible affinities of these fragments were relatively weak, with K_I = 56 and 80 µM for **2f** and **2g**, respectively (*see Fig. S13*). Therefore, it was interesting to observe a thermal shift similar to an inhibitor with biochemical pIC50 = 8; this illustrates the impact of covalent modification on protein-ligand binding.

Crystal structures of SFs **2f** and **2s** within BCL6 were generated to further investigate the binding mode of the fragments; this used a previously published protocol.[42] Co-crystal structures of **2f** and **2s** with BCL6 (solved to 1.6 Å and 1.8 Å resolution, respectively) revealed Tyr57 covalently conjugated to the fragments via a sulfonate ester, consistent with tandem MS studies for SF **2f** (Fig. 6c & 6d). The residue Arg27 was observed to be in close proximity to the sulfonyl groups of **2f** and **2s** which may have catalysed the covalent modification of Tyr57 via hydrogen bonding interactions. While the *meta*-substitutions of the benzamides showed opposite trajectories, a common binding surface of the benzamide aryl ring was observed, suggesting that a tri-substituted aryl group might be tolerated here (Fig. 6e). The twisted conformation of **2s** could also represent a pre-modified state of the protein-SF complex, which may have become disrupted for **2f** upon covalent modification with Tyr57 to afford the extended conformation observed. The structures of GSK137 and **2s** were overlaid, revealing that the two binding sites had significant overlap in the final crystallographic state (Fig. 6f). This suggested that **2s** may have similar inhibitory action to GSK137 on the BCL6 BTB/POZ domain.[42]

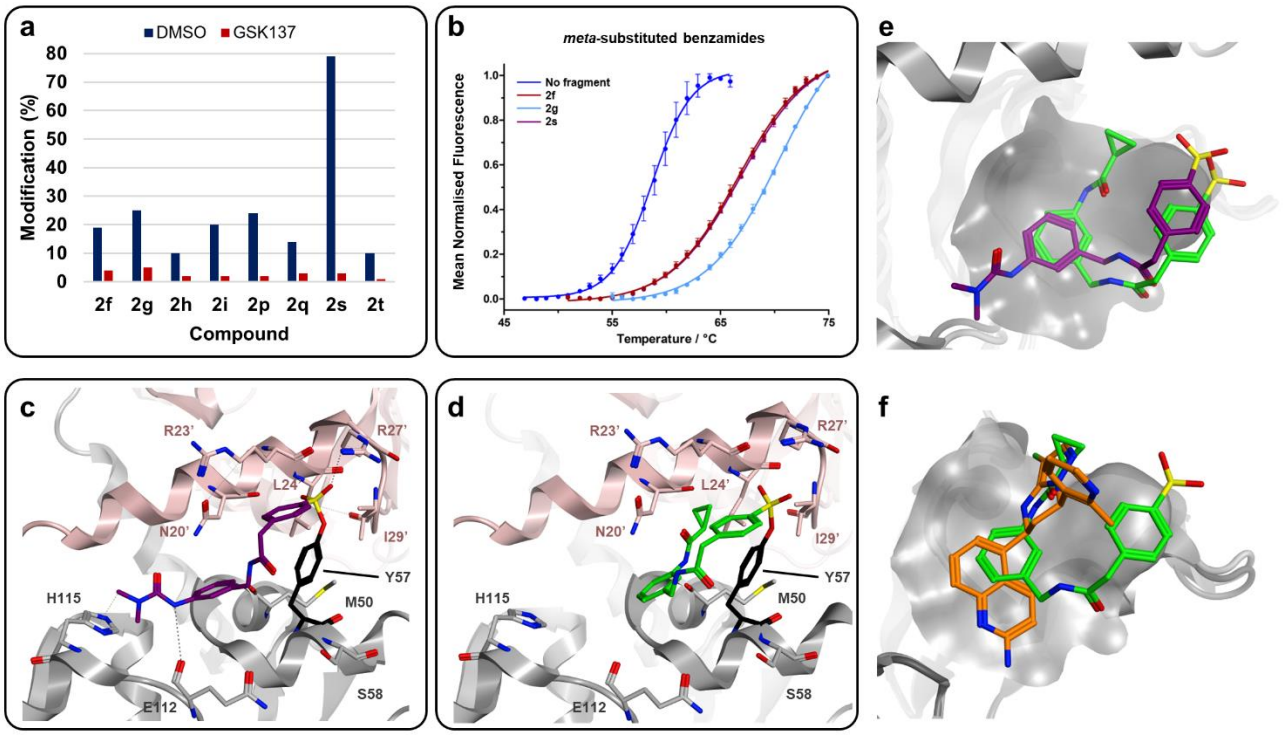


Fig. 6 | Investigations of BCL6 hits provide further structural insights into binding modes of hit SFs. (a) Incubation of GSK137 with BCL6 abolishes covalent modification by *meta*-substituted benzamide and azetidinyl/piperidinyl SF fragments. **(b)** Covalent modification by the *meta*-substituted benzamide hits increases the stability of BCL6 in the conventional DSF assay. Each point represents an average of three replicates. **(c)** Co-crystal structure of SF **2f** with BCL6 BTB/POZ dimer. One monomer is shown in grey; one monomer is shown in pink. Compound **2f** is shown in stick representation with carbon atoms coloured purple. **(d)** Co-crystal structure of SF **2s** with BCL6 BTB/POZ dimer. One monomer is shown in grey; one monomer is shown in pink. Compound **2s** is shown in stick representation with carbon atoms coloured green. **(e)** Overlaid crystal structures of **2f** and **2s** showing the aryl group of respective benzamides sharing a common binding surface. Tyr57 residue omitted for clarity. Compounds **2f** and **2s** are shown in stick representation, with carbon atoms coloured purple and green, respectively. **(f)** Overlaid crystal structures of GSK137 (PDB: 7BDE) and **2s** which have overlapping binding sites in the final crystallographic state. Tyr57 omitted for clarity. Compounds GSK137 and **2s** are shown in stick representation, with carbon atoms coloured orange and green, respectively.

Discussion

Covalent inhibitors are of high interest in chemical biology and drug discovery for expansion of the liganded proteome and for liganding challenging therapeutic targets.[46] The majority of efforts to develop covalent inhibitors have focussed on the targeting of cysteine residues in the vicinity of protein pockets. However, to tackle the increasing number of genetically validated targets that drug discovery teams are presented with, approaches to covalent ligand discovery without reliance on cysteine are highly sought after. SFs offer the opportunity to target a broad repertoire of nucleophilic amino acids, and hence, are effective reactive moieties to consider for 'beyond cysteine' covalent inhibitors. To date, SF-containing chemical probes have been rationally designed based on structure-guided approaches; this typically entails the installation of a SF group onto a potent reversible scaffold.[47] Here, we have demonstrated that sulfonyl fluoride reactive fragments offer an expedient approach to identify novel 'beyond-cysteine' covalent ligands.

Screening of a modest library of 352 fragments afforded hits for two of the three proteins assessed, which were found to interact with functionally relevant pockets. An advantage of screening with reactive fragments over classical reversible fragments, is the facile identification of the site(s) of covalent modification using tandem MS. Hits identified here were found to modify tyrosine and histidine residues. The observation of histidine modification under the denaturing mass spectrometry workflows was perhaps surprising, since previous reports have suggested that histidine adducts of SFs can be unstable. The multiple examples of histidine modification provide evidence that this is not always the case and that SFs are suitable electrophiles for targeting histidine residues.[10,19,48,49]

Kinetic analyses of the CAII hits provided a measure of covalent modification efficiencies and parameters k_{inact} and K_I. Interestingly, the CAII original hits exhibited good non-covalent affinity (K_I) but slow rates of covalent modification (k_{inact}). This suggested either low nucleophilicity of the histidine residue or sub-optimal geometry for covalent modification. A HTC-D2B screen of analogues close to the original hits identified fragments that gave fast rates of modification (>70% covalent modification in <1 h), highlighting the role of electrophile position in the design of covalent inhibitors, and the utility of the HTC-D2B approach as a means to optimise both k_{inact} as well as K_I.

The chemoproteomic analyses of the CAII hits revealed that they engaged CAII in cells, however, at low stoichiometry, likely due to the low k_{inact} rates for these hits. The two CAII hits and the negative control were also found to interact with many additional proteins in cells (>100), many of which were common to all three SF compounds. This highlighted that specificity is likely to be a key challenge to overcome when developing SF-based inhibitors. Previous work in our group has demonstrated that the reactivity of the SF electrophile is highly tuneable.[14] Therefore, it is anticipated that selective target engagement in cells would be achievable through reduction of the intrinsic reactivity of the SF, while optimising k_{inact} and K_I for the target. The HTC-D2B approach described here provides a useful strategy for performing this optimisation.

The SF fragment screen also identified novel hits for BCL6, a therapeutic target under investigation for conditions such as blood, breast, and lung cancers.[42] Interestingly, the screen afforded two hit series that were found to covalently modify residues on opposite sides of the same pocket, one to a tyrosine and one to a histidine, highlighting the versatility of the approach in targeting nucleophilic amino acid residues within a binding pocket. In this case, the pool of hit compounds was rapidly expanded upon by an iterative screen, and the hits represented the two novel chemotypes for BCL6. The attenuation of covalent modification when incubated in the presence of GSK137, a published BCL6 inhibitor, implied that the fragments target a known small-molecule binding surface.[42] Co-crystal structures of BCL6 with hit SFs were obtained, which confirmed that the binding conformations overlapped with that of GSK137, positing the fragment as a covalent probe to aid the discovery of novel inhibitors of the BCL6 BTB/POZ domain.

Collectively, these results demonstrate that SF reactive fragment screening offers an efficient approach for the discovery of 'beyond-cysteine' covalent ligands. The HTC-D2B approach enables rapid iterative design-make-test cycles to drive toward more potent and selective chemical tools. This HTC-D2B workflow facilitates the exploration of alternative SF groups as the reactive moiety, which can be exploited to reduce the intrinsic reactivity, and thus promiscuity, of the fragments, while simultaneously optimising k_{inact} and K_I for the target of interest. This will support the development of covalent chemical probes to expand the liganded proteome, and ultimately, further our understanding of disease biology.

References

1 H. Mukherjee and N. P. Grimster, *Curr. Opin. Chem. Biol.*, 2018, **44**, 30–38.

2 E. Resnick, A. Bradley, J. Gan, A. Douangamath, T. Krojer, R. Sethi, P. P. Geurink, A. Aimon, G. Amitai, D. Bellini, J. Bennett, M. Fairhead, O. Fedorov, R. Gabizon, J. Gan, J. Guo, A. Plotnikov, N. Reznik, G. F. Ruda, L. Díaz-Sáez, V. M. Straub, T. Szommer, S. Velupillai, D. Zaidman, Y. Zhang, A. R. Coker, C. G. Dowson, H. M. Barr, C. Wang, K. V. M. Huber, P. E. Brennan, H. Ovaa, F. Von Delft and N. London, *J. Am. Chem. Soc.*, 2019, **141**, 8951–8968.

3 H. Johansson, Y. C. I. Tsai, K. Fantom, C. W. Chung, S. Kümper, L. Martino, D. A. Thomas, H. C. Eberl, M. Muelbaier, D. House and K. Rittinger, *J. Am. Chem. Soc.*, 2019, **141**, 2703–2712.

4 A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C. D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Ábrányi-Balogh, J. Brandão-Neto, A. Carbery, G. Davison, A. Dias, T. D. Downes, L. Dunnett, M. Fairhead, J. D. Firth, S. P. Jones, A. Keeley, G. M. Keserü, H. F. Klein, M. P. Martin, M. E. M. Noble, P. O'Brien, A. Powell, R. N. Reddi, R. Skyner, M. Snee, M. J. Waring, C. Wild, N. London, F. von Delft and M. A. Walsh, *Nat. Commun.*, 2020, **11**, 1–11.

5 L. H. Jones, *Annu. Rep. Med. Chem.*, 2021, **56**, 95–134.

6 J. Pettinger, K. Jones and M. D. Cheeseman, *Angew. Chem. Int. Ed.*, 2017, **56**, 15200–15209.

7 E. C. Hett, H. Xu, K. F. Geoghegan, A. Gopalsamy, R. E. Kyne, C. A. Menard, A. Narayanan, M. D. Parikh, S. Liu, L. Roberts, R. P. Robinson, M. A. Tones and L. H. Jones, *ACS Chem. Biol.*, 2015, **10**, 1094–1098.

8 D. A. Shannon, C. Gu, C. J. Mclaughlin, M. Kaiser, R. A. L. van der Hoorn and E. Weerapana, *ChemBioChem*, 2012, **13**, 2327–2330.

9 J. Dong, L. Krasnova, M. G. Finn and K. Barry Sharpless, *Angew. Chem. Int. Ed.*, 2014, **53**, 9430–9448.

10 A. Narayanan and L. H. Jones, *Chem. Sci.*, 2015, **6**, 2650–2659.

11 D. E. Mortenson, G. J. Brighty, L. Plate, G. Bare, W. Chen, S. Li, H. Wang, B. F. Cravatt, S. Forli, E. T. Powers, K. B. Sharpless, I. A. Wilson and J. W. Kelly, *J. Am. Chem. Soc.*, 2018, **140**, 200–210.

12 G. J. Brighty, R. C. Botham, S. Li, L. Nelson, D. E. Mortenson, G. Li, C. Morisseau, H. Wang, B. D. Hammock, K. B. Sharpless and J. W. Kelly, *Nat. Chem.*, 2020, **12**, 906–913.

13 H. Mukherjee, J. Debreczeni, J. Breed, S. Tentarelli, B. Aquila, J. E. Dowling, A. Whitty and N. P. Grimster, *Org. Biomol. Chem.*, 2017, **15**, 9685–9695.

14 K. Gilbert, A. Vuorinen, A. Aatkar, P. Pogány, J. Pettinger, E. K. Grant, J. M. Kirkpatrick, K. Rittinger, D. House, G. A. Burley and J. T. Bush, *ChemRxiv*, 2022, DOI:10.26434/CHEMRXIV-2022-J8B8D-V2.

15 T. I. Oprea, C. G. Bologa, S. Brunak, A. Campbell, G. N. Gan, A. Gaulton, S. M. Gomez, R. Guha, A. Hersey, J. Holmes, A. Jadhav, L. J. Jensen, G. L. Johnson, A. Karlson, A. R. Leach, A. Ma'ayan, A. Malovannaya, S. Mani, S. L. Mathias, M. T. McManus, T. F. Meehan, C. Von Mering, D. Muthas, D. T. Nguyen, J. P. Overington, G. Papadatos, J. Qin, C. Reich, B. L. Roth, S. C. Schürer, A. Simeonov, L. A. Sklar, N. Southall, S. Tomita, I. Tudose, O. Ursu, D. Vidović, A. Waller, D. Westergaard, J. J. Yang and G. Zahoránszky-Köhalmi, *Nat. Rev. Drug Discov.*, 2018, **17**, 317–332.

16 C. G. Parker, A. Galmozzi, Y. Wang, B. E. Correia, K. Sasaki, C. M. Joslyn, A. S. Kim, C. L. Cavallaro, R. M. Lawrence, S. R. Johnson, I. Narvaiza, E. Saez and B. F. Cravatt, *Cell*, 2017, **168**, 527–541.

17 Q. Zhao, X. Ouyang, X. Wan, K. S. Gajiwala, J. C. Kath, L. H. Jones, A. L. Burlingame and J.

Taunton, *J. Am. Chem. Soc.*, 2017, **139**, 680–685.

18 X. Yang, J. P. D. Van Veldhoven, J. Offringa, B. J. Kuiper, E. B. Lenselink, L. H. Heitman, D. Van Der Es and A. P. Ijzerman, *J. Med. Chem.*, 2019, **62**, 3539–3552.

19 J. T. Cruite, G. P. Dann, J. Che, K. A. Donovan, S. Ferrao, S. B. Ficarro, E. S. Fischer, N. S. Gray, F. Huerta, N. R. Kong, H. Liu, J. A. Marto, R. J. Metivier, R. P. Nowak, B. L. Zerfas and L. H. Jones, *RSC Chem. Biol.*, 2022, **3**, 1105–1110.

20 A. J. Carter, O. Kraemer, M. Zwick, A. Mueller-Fahrnow, C. H. Arrowsmith and A. M. Edwards, *Drug Discov. Today*, 2019, **24**, 2111–2115.

21 W. Lu, M. Kostic, T. Zhang, J. Che, M. P. Patricelli, L. H. Jones, E. T. Chouchani and N. S. Gray, *RSC Chem. Biol.*, 2021, **2**, 354–367.

22 G. M. Keserü and G. M. Makara, *Nat. Rev. Drug Discov.*, 2009, **8**, 203–212.

23 F. Giordanetto, C. Jin, L. Willmore, M. Feher and D. E. Shaw, *J. Med. Chem.*, 2019, **62**, 3381–3394.

24 J. Pettinger, M. Carter, K. Jones and M. D. Cheeseman, *J. Med. Chem.*, 2019, **62**, 11383–11398.

25 E. K. Grant, D. J. Fallon, M. M. Hann, K. G. M. Fantom, C. Quinn, F. Zappacosta, R. S. Annan, C. wa Chung, P. Bamborough, D. P. Dixon, P. Stacey, D. House, V. K. Patel, N. C. O. Tomkinson and J. T. Bush, *Angew. Chem. Int. Ed.*, 2020, **59**, 21096–21105.

26 B. A. Lanman, J. R. Allen, J. G. Allen, A. K. Amegadzie, K. S. Ashton, S. K. Booker, J. J. Chen, N. Chen, M. J. Frohn, G. Goodman, D. J. Kopecky, L. Liu, P. Lopez, J. D. Low, V. Ma, A. E. Minatti, T. T. Nguyen, N. Nishimura, A. J. Pickrell, A. B. Reed, Y. Shin, A. C. Siegmund, N. A. Tamayo, C. M. Tegley, M. C. Walton, H. L. Wang, R. P. Wurz, M. Xue, K. C. Yang, P. Achanta, M. D. Bartberger, J. Canon, L. S. Hollis, J. D. McCarter, C. Mohr, K. Rex, A. Y. Saiki, T. San Miguel, L. P. Volak, K. H. Wang, D. A. Whittington, S. G. Zech, J. R. Lipford and V. J. Cee, *J. Med. Chem.*, 2020, **63**, 52–65.

27 Y. Shin, J. W. Jeong, R. P. Wurz, P. Achanta, T. Arvedson, M. D. Bartberger, I. D. G. Campuzano, R. Fucini, S. K. Hansen, J. Ingersoll, J. S. Iwig, J. R. Lipford, V. Ma, D. J. Kopecky, J. McCarter, T. San Miguel, C. Mohr, S. Sabet, A. Y. Saiki, A. Sawayama, S. Sethofer, C. M. Tegley, L. P. Volak, K. Yang, B. A. Lanman, D. A. Erlanson and V. J. Cee, *ACS Med. Chem. Lett.*, 2019, **10**, 1302–1308.

28 A. I. Green, F. Hobor, C. P. Tinworth, S. Warriner, A. J. Wilson and A. Nelson, *Chem. Eur. J.*, 2020, **26**, 10682–10689.

29 R. P. Thomas, R. E. Heap, F. Zappacosta, E. K. Grant, P. Pogány, S. Besley, D. J. Fallon, M. M. Hann, D. House, N. C. O. Tomkinson and J. T. Bush, *Chem. Sci.*, 2021, **12**, 12098–12106.

30 D. Rogers and M. Hahn, *J. Chem. Inf. Model.*, 2010, **50**, 742–754.

31 C. T. Supuran, *Nat. Rev. Drug Discov.*, 2008, **7**, 168–181.

32 C. L. Lomelino, C. T. Supuran and R. McKenna, *Int. J. Mol. Sci.*, 2016, **17**, 1150.

33 T. Gokcen, I. Gulcin, T. Ozturk and A. C. Goren, *J. Enzyme Inhib. Med. Chem.*, 2016, **31**, 180–188.

34 E. K. Grant, D. J. Fallon, H. C. Eberl, K. G. M. Fantom, F. Zappacosta, C. Messenger, N. C. O. Tomkinson and J. T. Bush, *Angew. Chem. Int. Ed.*, 2019, **58**, 17322–17327.

35 A. Di Fiore, C. Pedone, J. Antel, H. Waldeck, A. Witte, M. Wurl, A. Scozzafava, C. T. Supuran and G. De Simone, *Bioorganic Med. Chem. Lett.*, 2008, **18**, 2669–2674.

36 W. Chen, J. Dong, L. Plate, D. E. Mortenson, G. J. Brighty, S. Li, Y. Liu, A. Galmozzi, P. S. Lee, J. J. Hulce, B. F. Cravatt, E. Saez, E. T. Powers, I. A. Wilson, K. B. Sharpless and J. W. Kelly, *J. Am. Chem. Soc.*, 2016, **138**, 7353–7364.

37 D. J. Fallon, S. Lehmann, C. wa Chung, A. Phillipou, C. Eberl, K. G. M. Fantom, F. Zappacosta, V. K.

Patel, M. Bantscheff, C. J. Schofield, N. C. O. Tomkinson and J. T. Bush, *Eur. J. Chem.*, 2021, **27**, 17880–17888.

38 F. Li, X. Fei, J. Xu and C. Ji, *Mol. Biol. Rep.*, 2009, **36**, 691–696.

39 A. Wang, H. C. Yang, P. Friedman, C. A. Johnson and E. A. Dennis, *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids*, 1999, **1437**, 157–169.

40 U. Hengst, H. Albrecht, D. Hess and D. Monard, *J. Biol. Chem.*, 2001, **276**, 535–540.

41 N. Kerres, S. Steurer, S. Schlager, G. Bader, H. Berger, M. Caligiuri, C. Dank, J. R. Engen, P. Ettmayer, B. Fischerauer, G. Flotzinger, D. Gerlach, T. Gerstberger, T. Gmaschitz, P. Greb, B. Han, E. Heyes, R. E. Iacob, D. Kessler, H. Kölle, L. Lamarre, D. R. Lancia, S. Lucas, M. Mayer, K. Mayr, N. Mischerikow, K. Mück, C. Peinsipp, O. Petermann, U. Reiser, D. Rudolph, K. Rumpel, C. Salomon, D. Scharn, R. Schnitzer, A. Schrenk, N. Schweifer, D. Thompson, E. Traxler, R. Varecka, T. Voss, A. Weiss-Puxbaum, S. Winkler, X. Zheng, A. Zoephel, N. Kraut, D. McConnell, M. Pearson and M. Koegl, *Cell Rep.*, 2017, **20**, 2860–2875.

42 A. C. Pearce, M. J. Bamford, R. Barber, A. Bridges, M. A. Convery, C. Demetriou, S. Evans, T. Gobbetti, D. J. Hirst, D. S. Holmes, J. P. Hutchinson, S. Jayne, L. Lezina, M. T. McCabe, C. Messenger, J. Morley, M. C. Musso, P. Scott-Stevens, A. S. Manso, J. Schofield, T. Slocombe, D. Somers, A. L. Walker, A. Wyce, X. P. Zhang and S. D. Wagner, *J. Biol. Chem.*, 2021, **297**, 100928.

43 W. McCoull, R. D. Abrams, E. Anderson, K. Blades, P. Barton, M. Box, J. Burgess, K. Byth, Q. Cao, C. Chuaqui, R. J. Carbajo, T. Cheung, E. Code, A. D. Ferguson, S. Fillery, N. O. Fuller, E. Gangl, N. Gao, M. Grist, D. Hargreaves, M. R. Howard, J. Hu, P. D. Kemmitt, J. E. Nelson, N. O'Connell, D. B. Prince, P. Raubo, P. B. Rawlins, G. R. Robb, J. Shi, M. J. Waring, D. Whittaker, M. Wylot and X. Zhu, *J. Med. Chem.*, 2017, **60**, 4386–4402.

44 M. Teng, M. Teng, S. B. Ficarro, S. B. Ficarro, S. B. Ficarro, H. Yoon, H. Yoon, J. Che, J. Che, J. Zhou, E. S. Fischer, E. S. Fischer, J. A. Marto, J. A. Marto, J. A. Marto, T. Zhang, T. Zhang, N. S. Gray and N. S. Gray, *ACS Med. Chem. Lett.*, 2020, **11**, 1269–1273.

45 M. Vedadi, F. H. Niesen, A. Allali-Hassani, O. Y. Fedorov, P. J. Finerty, G. A. Wasney, R. Yeung, C. Arrowsmith, L. J. Ball, H. Berglund, R. Hui, B. D. Marsden, P. Nordlund, M. Sundstrom, J. Weigelt and A. M. Edwards, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 15835–15840.

46 L. Boike, N. J. Henning and D. K. Nomura, *Nat. Rev. Drug Discov.*, 2022, DOI:10.1038/s41573-022-00542-z.

47 L. H. Jones and J. W. Kelly, *RSC Med. Chem.*, 2020, **11**, 10–17.

48 L. Gambini, C. Baggio, P. Udompholkul, J. Jossart, A. F. Salem, J. J. P. Perry and M. Pellecchia, *J. Med. Chem.*, 2019, **62**, 5616–5627.

49 J. Che and L. H. Jones, *RSC Med. Chem.*, 2022, **13**, 1121–1126.