

Novel irreversible covalent BTK inhibitors discovered using DNA-encoded chemistry

John P Guilinger, Archana Archana, Martin Augustin, Andreas Bergmann, Paolo A Centrella, Matthew A Clark, John W Cuzzo, Maike Däther, Marie-Aude Guié, Sevan Habeshian, Reiner Kiefersauer, Stephan Krapp, Alfred Lammens, Lukas Lercher, Julie Liu, Yanbin Liu, Klaus Maskos, Michael Mrosek, Klaus Pflügler, Markus Siegert, Heather A Thomson, Xia Tian, Ying Zhang, Debora L Konz Makino, Anthony D Keefe

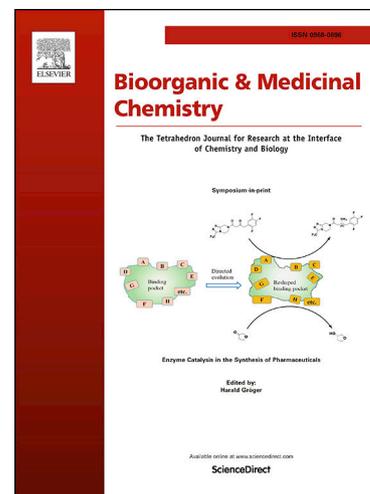
PII: S0968-0896(21)00231-5  
DOI: <https://doi.org/10.1016/j.bmc.2021.116223>  
Reference: BMC 116223

To appear in: *Bioorganic & Medicinal Chemistry*

Received Date: 28 February 2021  
Revised Date: 11 May 2021  
Accepted Date: 13 May 2021

Please cite this article as: J.P. Guilinger, A. Archana, M. Augustin, A. Bergmann, P.A. Centrella, M.A. Clark, J.W. Cuzzo, M. Däther, M-A. Guié, S. Habeshian, R. Kiefersauer, S. Krapp, A. Lammens, L. Lercher, J. Liu, Y. Liu, K. Maskos, M. Mrosek, K. Pflügler, M. Siegert, H.A. Thomson, X. Tian, Y. Zhang, D.L. Konz Makino, A.D. Keefe, Novel irreversible covalent BTK inhibitors discovered using DNA-encoded chemistry, *Bioorganic & Medicinal Chemistry* (2021), doi: <https://doi.org/10.1016/j.bmc.2021.116223>

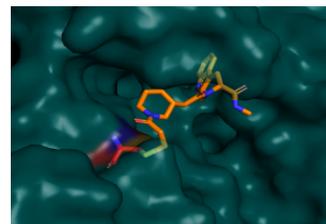
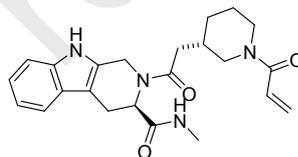
This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



**Novel irreversible covalent BTK inhibitors discovered using DNA-encoded chemistry**

Leave this area blank for abstract info.

JP Guilinger, A Archana, M Augustin, A Bergmann, PA Centrella, MA Clark, JW Cuzzo, M Däther, MA Guié, S Habeshian, R Kiefersauer, S Krapp, A Lammens, L Lercher, J Liu, Y Liu, K Maskos, M Mrosek, K Pflügler, M Siegert, HA Thomson, X Tian, Y Zhang, DL Konz Makino and AD Keefe  
*X-Chem Inc., 100 Beaver Street, Waltham, Massachusetts, 02453, USA*





## Novel irreversible covalent BTK inhibitors discovered using DNA-encoded chemistry

John P Guilinger\*<sup>1</sup>, Archana Archana<sup>2</sup>, Martin Augustin<sup>2</sup>, Andreas Bergmann<sup>2</sup>, Paolo A Centrella<sup>1</sup>, Matthew A Clark<sup>1</sup>, John W Cuzzo<sup>3</sup>, Maik Däther<sup>2</sup>, Marie-Aude Guié<sup>1</sup>, Sevan Habeshian<sup>4</sup>, Reiner Kiefersauer<sup>2</sup>, Stephan Krapp<sup>2</sup>, Alfred Lammens<sup>2</sup>, Lukas Lercher<sup>2</sup>, Julie Liu<sup>5</sup>, Yanbin Liu<sup>6</sup>, Klaus Maskos<sup>2</sup>, Michael Mrosek<sup>7</sup>, Klaus Pflügler<sup>2</sup>, Markus Siegert<sup>2</sup>, Heather A Thomson<sup>8</sup>, Xia Tian<sup>9</sup>, Ying Zhang<sup>1</sup>, Debora L Konz Makino\*<sup>2</sup> and Anthony D Keefe\*<sup>1</sup>

<sup>1</sup> X-Chem Inc., 100 Beaver Street, Waltham, Massachusetts 02453, USA

<sup>2</sup> Proteros biostructures GmbH, Bunsenstr. 7a, 82152 Planegg-Martinsried, Germany

<sup>3</sup> Formerly X-Chem Inc., now ZebiAI Inc., 100 Beaver Street, Waltham, Massachusetts 02453, USA

<sup>4</sup> Formerly X-Chem Inc., now Laboratory of Therapeutic Proteins and Peptides École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

<sup>5</sup> Formerly X-Chem Inc., now Accent Therapeutics Inc., 65 Hayden Avenue, Lexington, Massachusetts 02421, USA

<sup>6</sup> Formerly X-Chem Inc., now Cyteir Therapeutics, 128 Spring St, Lexington, MA 02421, USA

<sup>7</sup> Formerly Proteros biostructures GmbH, now Bruker AXS GmbH, Karlsruhe, Germany

<sup>8</sup> Formerly X-Chem Inc., now deceased.

<sup>9</sup> Formerly X-Chem Inc., now Arrakis Pharmaceuticals, 830 Winter Street, Waltham, Massachusetts 02451, USA

### ARTICLE INFO

#### Article history:

Received

Received in revised form

Accepted

Available online

#### Keywords

Bruton's Tyrosine Kinase

DNA-Encoded Chemical Libraries (DECL)

Covalent Irreversible Inhibitor

Drug Discovery

Tryptoline

### ABSTRACT

Libraries of DNA-Encoded small molecules created using combinatorial chemistry and synthetic oligonucleotides are being applied to drug discovery projects across the pharmaceutical industry. The majority of reported projects describe the discovery of reversible, i.e. non-covalent, target modulators. We synthesized multiple DNA-encoded chemical libraries terminated in electrophiles and then used them to discover covalent irreversible inhibitors and report the successful discovery of acrylamide- and epoxide-terminated Bruton's Tyrosine Kinase (BTK) inhibitors. We also demonstrate their selectivity, potency and covalent cysteine engagement using a range of techniques including X-ray crystallography, thermal transition shift assay, reporter displacement assay and intact protein complex mass spectrometry. The epoxide BTK inhibitors described here are the first ever reported to utilize this electrophile for this target.

### 1. Introduction

Libraries of DNA-Encoded small molecules created using combinatorial chemistry processes and synthetic oligonucleotides are being applied to drug discovery projects across the pharmaceutical industry. Recent reviews of this field include Favalli et al 2018 and Goodnow et al 2017. The overwhelming majority of reported projects describe the discovery of reversible, i.e. non-covalent, target modulators with only a very small number of reports relating to the discovery of covalent irreversible modulators. This latter category has thus far been limited to reported experiments with numerically small libraries - Zambaldo et al 2016 (10,000 encoded compounds), Chan et al 2017 (136 encoded compounds simultaneously selected against 236 encoded proteins) and Zimmermann et al 2017 (148,135 proximally displayed building block pairs). One other report is further limited to only reporting successful DNA-encoded positive control

enrichment experiments (Zhu et al 2019). There has recently been a resurgence of interest in the discovery of covalent drugs and for reviews of recent developments in the covalent inhibitor discovery field in general see for example Lagoutte et al 2017, Ghosh et al 2019 and Zhang et al 2019. We were motivated to determine if DNA-encoded chemistry could be used to discover covalent irreversible compounds from numerically large DNA-encoded chemical libraries, i.e. within the numerical range that are routinely reported for DNA-encoded chemical libraries used for the purposes of non-covalent drug discovery. To support the achievement of this goal we synthesized DNA-encoded chemical libraries terminated in electrophiles with numerical sizes in the tens to greater than one hundred million compounds and then developed methods to successfully use them in affinity-mediated selection processes adapted to discover irreversible target engagers. We chose the target Bruton's Tyrosine Kinase (BTK) for this project as there is clear precedent for the discovery and

utility for this target, e.g. ibrutinib (Burger et al 2015) for which the mode-of-action is known to be via the formation of a covalent bond with the active site cysteine. BTK is primarily expressed in B-cells and is activated as a result of B-cell receptor signaling pathways that in turn cause the recruitment of BTK to the cell membrane and thereby result in the phosphorylation of tyrosine 551 (Spaargaren et al 2015). Ibrutinib is a potent and selective covalent BTK inhibitor that binds to cysteine 481 in the active site of BTK and is an effective monotherapy for CLL (Burger et al 2015). BTK is also a target for inflammatory diseases, for example rheumatoid arthritis (Di Paolo et al 2011). We first developed experimental protocols that were able to enrich on-DNA ibrutinib (irreversible covalent inhibitor) but not on-DNA dasatinib (reversible inhibitor) using an affinity-mediated selection process with BTK. We then applied these optimized protocols to DNA-encoded chemical libraries terminated with acrylamide and then subsequently to a range of encoded electrophiles including other  $\alpha$ ,  $\beta$ -unsaturated carboxamides, epoxides and haloalkanes. Here we report the successful discovery of novel acrylamide-terminated BTK inhibitors and the demonstration of their selectivity, potency and cysteine engagement using a range of techniques including x-ray crystallography. Subsequent library builds and additional affinity-mediated selection experiments resulted in the discovery of more potent epoxide-terminated BTK inhibitors that are the first epoxide-mediated covalent irreversible inhibitors of BTK to be reported.

## 2. Results

### 2.1. Initial discovery of potent non-covalent BTK inhibitors

Our covalent discovery efforts were predated by the discovery of potent non-covalent inhibitors of BTK and we have already reported the discovery of **1**, a cell-active non-covalent inhibitor of BTK (Cuozzo et al 2017). This compound was discovered within library 1 (Table 1). Library 1 was synthesized by the acylation of a primary amine-terminated PEG linker displayed upon a headpiece oligonucleotide with 300 Fmoc amino acids followed by acylation with 157 formyl acids (carboxylates/aldehydes) followed by reductive amination with 2,341 amines (Table 1). Each chemical reaction was conducted in a segregated compartment and with concomitant ligation of an encoding oligonucleotide prior to pooling and splitting for each successive chemical step. Prior to library synthesis, each building block admitted into the library had been individually determined to have a yield of at least 50% in an on-DNA model reaction. Reaction yields were calculated by examination of the UV and TIC traces of the LCMS chromatograms. Double-substitutions and hydrolysis products were identified and excluded. Additionally any products in which the DNA was damaged by gain or loss of mass, such as depurination, phosphate hydrolysis, or covalent modification, were also excluded (Clark et al 2009). Each of the 110,261,100 unique compounds within this library is covalently attached to a uniquely identifying oligonucleotide tag combination that encodes its chemical history and can therefore be used to deduce its identity. Compound **1** (Table 2) when tested in a TR-FRET assay exhibited an  $IC_{50}$  value of 0.55 nM and it was also determined to be active in cell culture and whole human blood assays (Cuozzo et al 2017). During this study Compound **1** was determined to have an  $IC_{50}$  value of 3.8 nM using a Reporter Displacement Assay (Neumann et al 2011) with wild-type BTK, and a value of 3.0 nM for C481S BTK (Figure 1 and Table 3). Residence times were determined to be 9 and 22 minutes respectively (Table 3). Target engagement was confirmed by the observation of a large positive

and of +13.4°C for the C481S mutant (Figure 2 and Table 3). The similarity of the biophysical parameter values obtained using these two orthogonal methods on wild-type and C481S BTK strongly suggested that the presence of a reactive thiol group in the compound binding pocket does not affect the binding of Compound **1** to BTK. Intact protein Mass-Spectrometry analyses of Compound **1** pre-incubated with either wild-type or C481S BTK clearly indicate the non-covalent nature of this interaction (Figures 3 and 4, Supplementary Tables S2 and S3). Compound **1** engages BTK with a non-covalent interaction, and this was further confirmed by X-ray crystallography of Compound **1** in complex with BTK (Cuozzo et al 2017). Selectivity was assessed by screening against a panel of 135 different kinases at DiscoverX using the KINOMEScan™ profiling assay and this indicated that this compound inhibited multiple other kinases after a one-hour incubation at 10  $\mu$ M (Figure 5). This was perhaps unsurprising when one considers that this concentration is several hundred-fold above the observed  $IC_{50}$  for BTK. Nonetheless, this did provide a comparator for subsequent selectivity assessments with other compounds reported in this paper. Notably, Compound **1** includes a methyl amide stub that in the context of the encoded library formed the first few atoms of the linker between the encoded building block combination and the DNA tag. This linker stub was subsequently shown to be required for activity and the crystal structure shows this carbonyl hydrogen-bonded to the backbone carbonyl nitrogen of Phe413 in the glycine-rich loop of BTK (Figures 6 and 7). This interaction appears to enable the tryptoline to sit in a deep hydrophobic pocket while the quinoxaline simultaneously interacts with the hinge region. The selection output data revealed that while a range of structurally related hinge-binding elements were tolerated (Table 4) the central furan was the only building block co-enriched in this position with the corresponding thiophene being the one exception with weak but statistically significant co-enrichment also. The tryptoline was uniquely co-enriched, but this building block is a natural product with no closely related other building blocks included in this library (ECFP6 Tanimoto distance >0.8), including its enantiomer. Less potent non-tryptoline-containing active compounds were also discovered from this library in this selection experiment (Cuozzo et al 2017).

### 2.2. Subsequent discovery of structurally related non-covalent BTK inhibitors in a distinct DNA-encoded chemical library

Emboldened by our initial successes, BTK was chosen to be part of our further technology development efforts. The next DNA-encoded chemical library utilized for this project was a capped dipeptide library for which the first two synthesis steps were the same as the first synthesis step described for library 1. These were followed by an acylation step with 2,500 carboxylates to generate library 2 with a total of 225,000,000 unique compounds (Table 1). This library was also subjected to affinity-mediated selection with BTK, amplification and sequencing as described previously (Cuozzo et al 2017). Similarly to what had been observed with Library 1, a range of enriched features were observed in the processed selection output data. One cluster of enriched structurally related compounds was clearly closely related to compound **1** in that it was defined by the same tryptoline building block at the first chemical cycle. Co-enriched structurally related building blocks at the second chemical cycle included L-phenylalanine and substituted L-phenylalanines and other closely related amino acids (Table 4). The third and final cycle was defined by 5-carbamoyl-1H-pyrrole-3-carboxylic acid, a relatively unusual building block without other closely structurally related

The one exception to this statement being 5-(pyrrolidine-1-carbonyl)-1H-pyrrole-3-carboxylic acid which was not co-enriched and contains no terminal amide hydrogen atoms and would not therefore be able to provide these for hydrogen-bonding interactions to the hinge region. The synthesis of the most highly co-enriched building block combination yielded Compound 2 (Table 2) which was subsequently shown in the Reporter Displacement Assay to have an  $IC_{50}$  value of 23 nM and 56 nM on wild-type and C481S BTK, respectively (Table 3). Similarly to Compound 1, the removal of the methyl amide linker stub resulted in an inactive compound which is consistent with the hypothesis of a similar binding mode for both compounds.

### 2.3. Adaptation of DNA-encoded chemical library synthesis to support the discovery of irreversible covalent inhibitors

Next we undertook our first attempt at the synthesis of a library that could be used to discover covalent irreversible inhibitors. For this library we used a tripeptide scaffold with a terminal primary or secondary amine that was available for further subsequent functionalization. This library was synthesized in a similar fashion to Library 2, as described above, except that the first cycle of chemistry comprised acylation with 305 amino acids and the second and third with 293 amino acids for a total numerical diversity of 26,183,954 tripeptides. In order to transform this library into one capable of forming irreversible covalent bonds with nucleophiles such as thiols we chose to install acrylamide by acylation at the terminal amine with acrylic acid N-hydroxysuccinimide ester. Because acrylamide contains an  $\alpha,\beta$ -unsaturated carboxamide group it can function as an electrophile i.e. as a Michael acceptor. This library is listed as Library 3 in Table 1. Because only a single electrophile was installed there was no need to encode its identity using a building block identifying tag. The acrylamide electrophile is found in several designed covalent drugs including afatinib (Dungo et al 2013), ibrutinib (Burger et al 2015), rociletinib (Sequist et al 2015) and others (Jackson et al 2017) and in particular, acrylamide reacts with thiols such as cysteines (Jackson et al 2017).

### 2.4. Design of an affinity-mediated selection process for the discovery of irreversible target engagers within numerically large DNA-encoded chemical libraries

Most reported methodologies for the enrichment of rare target-engaging compounds existing within numerically large combinatorial libraries of building block combinations utilize multiple cycles of affinity-mediated selection to amplify the enrichment of binders with respect to non-binders and thereby render them observable and discoverable. We knew from our prior experience with the affinity-mediated selection of non-covalent libraries that a single cycle of selection was not sufficient to support discovery when working with library super-mixes containing upwards of tens of billions of compounds. If the washing efficiency of a single round of selection is 99.9% then the maximum enrichment possible is one thousand-fold. In reality this extent is significantly reduced by the dissociation of genuine binders during the wash phase, the likelihood that only a fraction of library members display fully elaborated compounds and the reality that only a fraction of the active molecules that are capable of engaging the target will do so. Covalent irreversible selection methods are reliant upon taking advantage of a single target-library member encounter and any methods introduced for enrichment amplification must operate within this very significant limitation. The strategy that we took advantage of was to use the irreversible

member to permit increased wash stringency by altering the wash conditions to make them denaturing. We tried a range of conditions while ensuring that the association between the target protein and the affinity matrix was maintained. Most protein affinity tags interact with immobilized capture reagents that are themselves proteins and as such these interactions are prone to becoming destabilized under denaturing conditions. Affinity tags that are reliant upon the maintenance of protein conformation include biotin/streptavidin, Strep-Tag or SBP-Tag/streptavidin, epitope/antibody e.g. FLAG-, myc- or C-tags, Protein A or G/antibody, glutathione/GST, amylose/MBP etc. One affinity tag that is not susceptible to denaturation-mediated disassociation is the polyhistidine tag captured via an immobilized transition metal ion such as  $Ni^{2+}$ . We chose to use hexahistidine-tagged BTK, immobilized  $Ni^{2+}$  matrices and on-DNA ibrutinib and dasatinib to discover conditions that would maximize enrichment of ibrutinib (irreversible) over dasatinib (reversible). The chemical structures of the on-DNA dasatinib and ibrutinib are shown in Supplementary Figure S1. Any protocol used for identifying irreversible covalent binders from a DNA-encoded library also needs to be able to support the recovery of sequence information from the encoding DNA as it emerges from the affinity-mediated selection, i.e. covalently bound to the target protein. Sufficiently denaturing washing conditions would of course also destabilize non-covalent target-library member interactions and should therefore provide a selection output that does not contain these, hence the additional utilization of on-DNA dasatinib. As a consequence of this methodological approach, separate selection experiments with distinct protocols would be needed to separately discover reversible and irreversible inhibitors and reversible and irreversible libraries therefore should not be combined into a single library super-mix.

In order to validate and optimize selection conditions to identify irreversible covalent inhibitors, the control compounds ibrutinib and dasatinib were conjugated to specific sequences of DNA with DNA tag architecture similar to that used for the DNA-encoded chemical libraries. Ibrutinib (Burger et al 2015) was conjugated to DNA as a positive control to detect irreversible covalent binding to BTK. Dasatinib (Lombardo et al 2004) was conjugated to DNA as a positive control to detect reversible non-covalent binding to BTK. A no-compound on-DNA negative control was also generated to detect non-specific interaction of the oligonucleotide tag architecture with BTK. We then used these on-DNA compounds to optimize affinity-mediated selection protocols for the discovery of covalent irreversible binders while discriminating against reversible binders. 50 nM of each on-DNA control was incubated as a mixture with no or 1  $\mu$ M BTK in protein storage buffer supplemented with 10 mM  $MgCl_2$  and no or 1 mg/mL sheared salmon sperm DNA as a general blocking agent for 1 hour. This mixture was then diluted 2-fold into a native buffer, a urea-based denaturing buffer or a guanidine-based denaturing buffer. Subsequently the BTK was captured on beaded agarose or magnetic bead IMAC matrices, washed in the same buffer used for dilution and then eluted using 400 mM imidazole in the same buffer used for dilution and washing. The eluate was then diluted and, along with a dilution series of the input of each on-DNA control, was subjected to qPCR analysis to determine the fraction of the input of each on-DNA control recovered in the elution for each of the different selection conditions. The first model selection (Figure 8a) used a guanidine-based denaturing buffer and a urea-based denaturing buffer and no sheared salmon sperm DNA as a general blocking agent during incubation of BTK with the on-DNA control mixture. This experiment was used to determine the optimal composition of the denaturing wash buffer. As expected,

on-I efficiency compared to the no-compound on-DNA control when BTK was present, and on-DNA ibrutinib was recovered with a similar efficiency compared to no compound on-DNA when no BTK was present indicating selective enrichment of the known irreversible covalent binder Ibrutinib (Figures 8a and 8b). Equivalent amounts of ibrutinib on-DNA were recovered in the elution using model selections with guanidine-based denaturing buffer vs. a urea-based denaturing buffer (Figure 8a). Similarly, equivalent amounts of no-compound on-DNA were recovered in the elution using model selections with guanidine-based denaturing buffer vs. urea-based denaturing buffer. Interestingly, dasatinib on-DNA amounts in the elution were substantially less using model selections with guanidine-based denaturing buffer when compared to the urea-based denaturing buffer, suggesting that the guanidine-based denaturing buffer was better able to denature BTK and/or otherwise wash away non-covalent binders such as dasatinib. The second model selection experiment (Figure 8b) used a native buffer, a guanidine-based denaturing buffer and also included sheared salmon sperm DNA and DNA-encoded chemical libraries during the incubation of BTK with the on-DNA control mixture to better mimic actual selection conditions. Equivalent amounts of on-DNA ibrutinib were recovered in the elution using model selections with magnetic bead IMAC vs. agarose IMAC (Figure 8b). However, both on-DNA dasatinib and no-compound on-DNA amounts were substantially less enriched in the elution using model selections with magnetic bead IMAC compared to the corresponding amounts with beaded agarose IMAC suggesting that magnetic bead IMAC enables better washing efficiencies of on-DNA dasatinib and no-compound on-DNA in the presence of BTK (Figure 8b). Taken together, these results indicate that on-DNA ibrutinib can be effectively enriched in model selections with optimized conditions yielding fold enrichments of >1000 for a single selection round when comparing the amount of on-DNA ibrutinib to the amount of no-compound on-DNA in model selections with BTK and magnetic bead IMAC under the optimized conditions (Figure 8b). We acknowledge that only approximately 3% of the input positive control molecules were recovered in this experiment (Figure 8b). However, the >1,000-fold relative enrichment observed for these when compared to the negative control will result in active molecules within library 4 being present at a 1-in-157 thousand frequency after selection, compared to a 1-in-157 million frequency prior to selection, assuming they behave similarly to this positive control. When 100 pmoles of library 4 is used for a selection experiment an average of 380,000 copies of each unique library member will be present in the input, and if 50% are passaged through to PCR we expect 5,000 copies to be present in the output. A 1-in-157 thousand frequency within the output translates to around six counts per million reads, and since we aim for 1-2 million reads per library per selection per time-point this signal strength is observable using our standard workflow. Numerically larger libraries will be usable with inputs of >100 pmoles, greater sequencing depth, or both.

### 2.5. Subsequent discovery of structurally related irreversible covalent BTK inhibitors in a DNA-encoded chemical library terminated with acrylamide

Having successfully demonstrated the enrichment of on-DNA ibrutinib using an optimized affinity-mediated selection protocol we then used this protocol in experiments with Library 3 (Table 1) and BTK. This protocol included incubation in native solution, introduction of guanidine as a denaturant, capture on IMAC magnetic beads, washing under denaturing conditions, liberation

and efficient amplification of the accompanying encoding oligonucleotides. Common perceptions of covalent inhibitors include the idea they are often not specific and so we chose to design indicators of specificity directly into our affinity-mediated selection campaign design. Firstly we chose a suite of seventeen diverse and independent protein off-targets, all of which contained cysteines and nine of which contained a cysteine in their active sites. Additional information about these seventeen off-targets is shown in Supplementary Table S1. These were subjected to affinity-mediated selection in parallel and building block combinations that enriched against BTK were only prioritized for hit re-synthesis if they failed to demonstrate enrichment against any of these seventeen off-targets. As a second independent approach to collecting compound promiscuity data we determined which building block combinations had ever been observed to enrich above a statistical significance threshold for any of the targets screened at X-Chem at any point prior to this screen. Again, synthesis prioritization was directed to compounds defined by building block combinations that had never before been seen to enrich against any target screened at X-Chem. When this experiment was conducted in 2015 the number of targets screened to-date was already several hundred. Additional parallel selections were designed to provide information on whether the enrichment signal increased over time, as would be expected for an irreversible interaction, and also whether the enrichment signal was reduced by pre-incubating the target with a saturating concentration of an orthosteric BTK inhibitor, as would be expected for a library compound binding to the active site. A representation of the number of compounds observed to enrich against BTK is shown in Figure 9. This figure shows that an abundance of compounds were observed to enrich in this affinity-mediated selection experiment. Individual clusters are colored according to the extent to which the building block combinations they represent have been observed to have enriched against targets previously screened at X-Chem with red representing the most promiscuous and green the least promiscuous. Our attention therefore was focused upon the clusters colored green. The most highly enriched cluster in this plot, circled in black, is defined by the same, i.e. previously-observed, tryptoline building block at the second cycle of chemistry, piperidine-3-acetic acid at the third cycle of chemistry displaying the acrylamide electrophile via acylation of its amine. Further selection output analysis revealed other structurally related building blocks at the third cycle of chemistry including other constrained saturated cyclic amino acids (Table 4). Two of these are represented by the green clusters circled in blue in Figure 9. All of these clusters of related compounds showed tolerance of a broad range of structurally unrelated building blocks at the first cycle of chemistry, which lead us to the conclusion that this region of the molecule may be dispensable. Synthesis of a combination of the most highly enriched building blocks at the second and third cycles of chemistry while truncating the first cycle of chemistry to a methyl amide resulted in Compound 3 (Table 2). The large number of parallel distinct selection conditions utilized in this experiment enabled us to make a number of predictions with regard to the selectivity characteristics of this compound prior to testing it in any assays. Figure 10 shows a profile of the relative enrichment of the building block combination defining Compound 3 across all of the parallel selections that were conducted. Aliquots of the BTK incubation solution were frozen at different time-points prior to their capture upon the affinity matrix and the profile indicates an increasing enrichment signal with increasing incubation time, as is consistent with the establishment of an irreversible interaction with the target, i.e. Compound 3 is predicted to exhibit a covalent mode of action. Parallel selections conducted with 18-hour incubations with and without the presence

of  $\epsilon$  inhibitor of BTK, show a starkly differential response with highly statistically significant enrichment observed for the latter and no enrichment observed for the former. This is consistent with Compound 3 binding BTK in its active site, i.e. that selections with each of the seventeen off-targets indicates that none of these resulted in a statistically significant signal (i.e. none exhibited a value of  $>10$  for the negative  $\log_{10}$  of the asymptotic significance value). This observation and the lack of enrichment observed for targets historically screened at X-Chem lead us to predict that compound 3 would be a highly selective inhibitor of BTK. Compound 3 was subsequently determined to have an apparent  $IC_{50}$  value of  $4.8 \mu\text{M}$  after 1 hour of incubation using the Reporter Displacement Assay with wild-type BTK, and was inactive against C481S BTK (Figure 1 and Table 3) as would be expected for a compound that elicits its activity by covalent modification of cysteine 481. Selectivity was then assessed by screening against a panel of 135 different kinases at DiscoverX using the KINOMEScan<sup>TM</sup> profiling assay and this indicated that Compound 3 inhibited few kinases after a one-hour incubation at  $10 \mu\text{M}$  (Figure 5). Covalent target engagement was confirmed by the high-resolution  $1.44 \text{ \AA}$  X-ray crystal structure of compound 3 in complex with wild-type BTK (Figures 6 and 7). And by a positive thermal transition shift of  $+6.5^\circ\text{C}$  for wild-type BTK and of  $<0.2^\circ\text{C}$  for the C481S mutant (Figure 2 and Table 3), as would be expected for a covalent modifier of cysteine 481. Target engagement was also confirmed by incubation of Compound 3 with BTK followed by intact-protein mass spectroscopy indicating that the protein had become an adduct with Compound 3 of the expected mass (Figure 3 and Table 3) and that no such adduct was observed after incubation with C481S BTK (Figure 4 and Table 3).

The overall structure of the complex of wild-type BTK (kinase domain) with Compound 3 (Figures 6 and 7) was almost identical to that of Compound 1 (Cuozzo et al 2017) with a  $C\alpha$  RMSD  $\approx 0.1$ . Compound 3 occupies the same binding site as was described for Compound 1. Similar observations have also been made for other inhibitors, for example for B43 (PDB: 3GEN, (Marcotte et al. 2010), in which the inhibitor resides in the ATP-binding pocket of BTK (Caldwell et al. 2019). Comparing the overall structure with previously reported structures of BTK (PDB 3PIZ and 3PJ2) (Kuglstatter et al, 2011), helix-C adopts the 'out' conformation in which it moves away from the active site and prevents the formation of the salt bridge between Lys430 and Glu445 that is important for activity (Supplementary Figures S2 and S3). Interestingly, we also observed that the binding of both Compound 1 and Compound 3 displaces the side chain of Lys430 further away from Glu445. The tryptoline moiety is common to both Compounds 1 and 3, and it occupies the hydrophobic pocket defined by Phe413, Ile432, Met437, Ile472 and Leu542 (Figures 6 and 7). The quinoxaline moiety of Compound 1 occupies the adenine-binding pocket and interacts with the hinge region by H-bonding between quinoxaline N-32 and the backbone amide nitrogen of Met477. In contrast to this, in the BTK structure with Compound 3 there are no hydrogen bonding interactions observed between the compound and the hinge region. Compound 3 also interacts with the glycine-rich loop and Asp539 in a DFG-in conformation via hydrogen bonding with the main-chain amide nitrogen of Gly411/Phe413 and the side-chain of Asp539, respectively. Further structural analysis of the BTK-Compound 3 complex confirms a covalent bond between this compound and cysteine 481 near to the ATP-binding pocket with continuous electron density between cysteine 481 and the Compound 3 warhead (Figure 7 and Supplementary Figure S2).

nds, additional structurally distinct series were also observed to have been enriched in the selection experiment. These include one exemplified by Compound 4. Similar to observations made for the set of compounds exemplified by Compound 3, the selection output data for Compound 4 also indicated that this compound exhibited time-dependent enrichment (data not shown), would be selective and would bind to BTK in a manner that inhibits its activity. This new cluster of compounds was defined by 3 and 4-cyano L-phenylalanine at the second cycle of chemistry and substituted L-proline derivatives at the third cycle of chemistry (Table 4). Similar to the way Compound 3 was represented in the selection output data, the first cycle of chemistry exhibited a high tolerance for a wide range of structurally unrelated building blocks and was ultimately determined not to be necessary for activity. We do not have crystallography data for the complex of compound 4 with BTK, but it is likely that the cyano phenylalanine side-chain occupies the same hydrophobic pocket as the tryptoline in compounds 1 and 3, and that the acrylamide reacts with the same active site cysteine. Compound 4 was determined to have an apparent  $IC_{50}$  value of  $36 \mu\text{M}$  after 1 hour of incubation using the Reporter Displacement Assay with wild-type BTK, and was observed to be inactive against C481S BTK (Table 3). Target engagement was confirmed by incubation of Compound 4 with BTK followed by intact-protein mass spectroscopy indicating that the protein had become an adduct with Compound 4 of the expected mass (Figure 3 and Table 3) and that no such adduct was observed after incubation with C481S BTK (Figure 4 and Table 3). Selectivity was assessed by screening against a panel of 135 different kinases at DiscoverX using the KINOMEScan<sup>TM</sup> profiling assay and this indicated that this compound is highly selective against over other kinases under these assay conditions (Figure 5).

## 2.6. Encoded Electrophile-Terminated Encoded Library Synthesis

The irreversible covalent inhibitors that we had discovered thus far all used acrylamide as a cysteine-reactive electrophile. This class of warhead ( $\alpha$ ,  $\beta$ -unsaturated carboxamide) has precedent for this target, not least with ibrutinib (Burger et al 2015). We next set out to determine if we could construct a DNA-encoded library terminated in a range of encoded electrophiles. There is increasing interest in alternatives to  $\alpha$ ,  $\beta$ -unsaturated carboxamides as electrophiles in covalent drugs (Gehring et al 2018, Ray and Murkin 2019) including epoxides, haloalkanes, sulfonyl fluorides, activated alkynes and several other classes. The strategy we adopted for the synthesis of library 4 (Table 1) utilized the same starting point that we used for library 3, a 26,183,954-member tripeptide library with a free terminal amine available for acylation. We chose six electrophilic building blocks for installation by acylation or carbamate formation at this terminal amine including five epoxides and one haloalkane. In order to encode the identity of each of these electrophiles we divided the nascent library into six aliquots and to each aliquot ligated a different fourth-chemistry-cycle tag prior to installing the same library-identifying tag, degenerate region and primer-binding sequence (Tailpiece) to each. Because we anticipated that specific electrophiles would have different generic reactivity, we assessed the relative reactivity of each electrophile by combining library aliquots containing each of the electrophiles at an equimolar ratio and then performed affinity-mediated selection using BTK and one of the off-target proteins. For each electrophile, the fractional counts in the processed and selection output were calculated and used to inform the final mixing ratios with electrophiles included

Based on these data the final library was generated using an equimolar ratio of each of the five epoxides with one tenth the amount for the haloalkane (i.e. a molar ratio of 1 haloalkane to 51 total library). Subsequent to the generation of this library we were able to show that highly similar selection output signatures could be generated by repeating the same selection experiments after storage of libraries at  $-80^{\circ}\text{C}$  for multiple years. This demonstrated sufficient stability of the electrophiles in this physical and chemical context to allow this and similar other libraries to be used on an ongoing basis.

### 2.7. Subsequent discovery of structurally-related irreversible covalent epoxide-containing BTK inhibitors in a DNA-encoded chemical library

We then used library 4 in an affinity-mediated selection experiment conducted with BTK in the same manner as that used for library 3, i.e. with a protocol optimized to enrich covalent irreversible binders only. The analysis of the output from this selection experiment again showed a broad range of enriched compounds. These included clusters of structurally closely related compounds. One such compound cluster contained compound 5 (Table 2) which, in common with compounds 1, 2 and 3, contained the tryptoline building block. Compound 5 was co-enriched with a set of closely structurally related compounds, all of which contained the tryptoline building block at the second cycle of chemistry. At the third cycle of chemistry these compounds contained a range of halogenated  $\beta$ -homophenylalanine derivatives acetylated with glycidic acid, a substituted epoxide (Table 4). Again, the selection output data suggested that the first cycle of chemistry was dispensable. The synthesis of the most highly enriched building block combinations at cycles two, three and four with the first cycle being truncated to a methyl amide provided Compound 5 (Table 2). Compound 5 exhibited an apparent  $\text{IC}_{50}$  value of  $0.10\ \mu\text{M}$  (1 hour incubation) when tested in the Reporter Displacement Assay against wild-type BTK (Table 3) and was inactive against C481S BTK (Table 3). The KINOMEScan™ profiling assay also indicated that Compound 5 is highly selective (Figure 5).

A separate set of co-enriched building blocks defined compounds with clear structural relatedness to Compound 4. For these compounds, the first cycle of chemistry again appeared to be dispensable; the second cycle of chemistry included the 4-cyano L- $\alpha$ -phenylalanine building block seen in compound 4, but also L- $\beta$ -phenylalanine derivatives (Table 4). In the third cycle of chemistry  $\alpha$ -amino acids with hydrophobic side-chains were co-enriched, which in turn were acetylated by glycidic acid or its methyl analog, with the latter being more highly enriched (Table 4). The synthesis of the most highly enriched building block combinations for each of the alternatives at chemistry cycle three with the first cycle being truncated to a methyl amide provided compounds 6 and 7.

Compounds 6 and 7 were determined to have apparent  $\text{IC}_{50}$  values of 26 and 40 nM respectively using the reporter displacement TR-FRET assay and one-hour incubations (Figure 1 and Table 3).  $K_{\text{inact}}$  and  $K_i$  values were also determined for both compounds with values of 0.0082 and 0.0094 per second and 1.3 and 1.9  $\mu\text{M}$  respectively (Table 3). Both were inactive in the Reporter Displacement Assay against C481S BTK (Figure 1 and Table 3). Target engagement was confirmed for both compounds by incubation with BTK followed by intact-protein mass spectroscopy indicating that the protein had become an adduct with each compound with the expected mass (Figure 3 and Table

compound (Figure 4 and Table 3). Target engagement was further confirmed by the observation of positive thermal shift values of  $+4.3^{\circ}\text{C}$  and  $+6.5^{\circ}\text{C}$  with wild-type BTK for Compounds 6 and 7 respectively (Figure 2 and Table 3).

## 3.1 Experimental Methods

### 3.1.1 Optimization of Affinity-Mediated Selection Conditions for Covalent Irreversible Binders

BTK (0 or 1  $\mu\text{M}$ ) (Signal Chem B10-10H-BULK) in 1x protein storage buffer Sodium Phosphate, pH 7 (50 mM), NaCl (150 mM), PMSF (0.1 mM), DTT (0.25 mM), glycerol (25%), supplemented with  $\text{MgCl}_2$  (10 mM) and supplemented with or without sheared salmon sperm DNA (ssDNA) (1mg/ml, Invitrogen) was incubated with a mixture of 50 nM of ibrutinib on-DNA, 50 nM dasatinib on-DNA and 50 nM no compound on-DNA control in a total volume of 50  $\mu\text{L}$  for 1 hour at room temperature. Then to each mixture was added 50  $\mu\text{L}$  of appropriate buffer which is either 1x non-denaturing buffer comprised of Sodium Phosphate pH 7.4 (50 mM), NaCl (200 mM), Imidazole (5 mM), ssDNA (1mg/ml, Invitrogen), TCEP (1 mM), Tween 20 (0.02%), 1x urea-based denaturing buffer comprised of 1x non-denaturing buffer with urea (6 M), or 1x guanidinium -based denaturing buffer comprised of 1x non-denaturing buffer with Guanidinium chloride (6 M). immobilized on agarose-based or magnetic resin IMAC, then washed 5 times with 200  $\mu\text{L}$  of appropriate buffer then incubation of the aspirated beads with 50  $\mu\text{L}$  0.8x appropriate buffer with an additional 400 mM imidazole, pH 7.5. For experiments with immobilization on agarose resin, pre-washed PhyTips containing 5  $\mu\text{L}$  of HisSelect resin (Phynexus) were used as previously described in Cuozzo et al 2017. For experiments with immobilization on magnetic resin, 50  $\mu\text{L}$  aliquot of Pierce HisPur™ Ni-NTA Magnetic Beads (ThermoFisher #88831) was aspirated and washed in 10 mM Tris (pH8.0) followed by washing in Denaturing Buffer followed by resuspension in 30  $\mu\text{L}$  of Denaturing Buffer was used. Elution, or a titration series of the input on-DNA control mix in 50  $\mu\text{L}$ , then subjected to qPCR analysis in 50  $\mu\text{L}$  qPCR volume using Power SYBR Green Master Mix (ThermoFisher 4368577) with 5  $\mu\text{L}$  of dilution of elution or standard curve of input as template and 0.5  $\mu\text{M}$  primer pair specific for each on-DNA control compound sequence.

### 3.1.2 Affinity-Mediated Selection for Irreversible Target Binders

A mixture of independently synthesized electrophile-terminated DNA-Encoded Chemical Libraries and 1  $\mu\text{M}$  of BTK (Signal Chem B10-10H-BULK) or 1  $\mu\text{M}$  off-target protein were combined in native solution and incubated with aliquots withdrawn and frozen at 2 and 18 hours. The Incubation Buffer was designed to model the cytoplasmic environment and was comprised of HEPES (20 mM), potassium acetate (134 mM), sodium acetate (8 mM), sodium chloride (4 mM), magnesium acetate (0.8 mM), sheared salmon sperm DNA (1mg/ml, Invitrogen), TCEP 0.5 mM and Tween 20 (0.02%) at pH 7.2 with 17  $\mu\text{M}$  total library concentration and 1  $\mu\text{M}$  BTK. Each 30  $\mu\text{L}$  aliquot, once withdrawn, was immediately combined with 30  $\mu\text{L}$  of a Denaturing Buffer comprised of guanidinium hydrochloride (6 M) di/mono sodium phosphate (pH 7.4, 50 mM), sodium chloride (200 mM), imidazole (2 mM), sheared salmon sperm DNA (1mg/ml, Invitrogen), TCEP 0.5 mM and Tween 20 (0.02%) and then frozen for storage. Once all samples had been generated and frozen each was heated to  $95^{\circ}\text{C}$  for two minutes and then cooled to room temperature and then incubated with pre-washed Ni-NTA magnetic beads. Each 50  $\mu\text{L}$  aliquot of Pierce HisPur™ Ni-NTA

washed in 10 mM Tris (pH8.0) followed by washing in Denaturing Buffer followed by resuspension in 30  $\mu$ l of Denaturing Buffer. Incubation samples were then captured upon the magnetic beads by incubation for 30 minutes with shaking at room temperature, non-captured mixture components were then removed with five 200  $\mu$ l washes with Denaturing Buffer. Elution of captured components was then achieved by incubation of the aspirated beads with the Elution Buffer (0.8x Denaturing Buffer with an additional 400 mM imidazole, pH 7.5). Additional magnetic aspirations were performed to ensure removal of all magnetic beads from the elution sample. Denaturants were removed from the elution sample by desalting column purification with Zeba Spin Desalting Columns, 40 kDa MWCO, 0.5mL (ThermoFisher #87766) according to the manufacturer's instructions. Additional selections were performed in parallel with no target, off-targets, BTK plus 0.1 mM staurosporine. Encoding oligonucleotides present in the processed eluates for each selection condition were then amplified using Platinum PCR Supermix™ (Invitrogen) with denaturation at 94°C, annealing at 55°C and extension at 72°C for 24 cycles using 5'- and 3'- primer oligonucleotides (each at 0.5  $\mu$ M) that each incorporate sequences complementary to the tailpiece or headpiece along with Illumina READ1 or READ2 sequences required to support clustering and subsequent single-read 100 base-pair sequencing on an Illumina HiSeq 2500. Sequencing was also performed for PCR-amplified samples of the naïve (unselected) library. Sequencing depth for irreversible selection outputs is targeted at 1-2 million single-end reads for each library/selection-condition/time-point combination. Sequence data were converted back into encoded chemical information computationally and demographic and statistical information were calculated for all individual building block combinations.

### 3.13 Synthesis of compounds 1-7

Compound 1 was synthesized according to (Cuozzo et al 2017), additional chemical synthesis information can be found in the Supplementary Information.

### 3.14 Intact Protein Complex Liquid Chromatography - Mass Spectrometry

For liquid chromatography mass spectrometry (LC-MS) analysis of ligands, proteins, and protein-ligand complexes, a Vanquish Flex LC-system (Thermo Fisher Scientific, Bremen, Germany) was used in combination with a maXis plus MS-system (Bruker, Bremen, Germany). The LC-system was equipped with a binary pump, a 25  $\mu$ l sample loop, a column oven and a variable wavelength detector (VWD, all Thermo Fisher Scientific). For analyte separation, a BIOshell™ A400 Protein C4 column (100 mm x 2.1 mm, 3.4  $\mu$ m, 400 Å, Merck, Darmstadt, Germany) as stationary protected by a BIOshell™ A400 Protein C4 guard cartridge (5 mm x 2.1 mm, 3.4  $\mu$ m, 400 Å, Merck) and the following gradient of solvent A (0.1% formic acid, FA, v/v) and solvent B (acetonitrile with 0.1% FA v/v) at a flowrate of 300  $\mu$ l/min was applied: t [min]/B [%]: 0/15; 2/15; 10/65; 10.5/98; 14.5/98; 15/15; 19/15. The solvent was actively preheated to 80 °C and the column oven was kept at 80 °C. Before entering the VWD, the solvent was cooled to 40 °C using a post-column cooler (Thermo Fisher Scientific). In the VWD, analytes were detected using two wavelengths of 214 nm and 280 nm and a data collection rate of 4 Hz. The MS-system was coupled on-line to the LC via the ESI interface. The maXis plus operated in the positive ion mode with activated focus mode for a mass spectrometric

end-plate offset was set to 500 V, the capillary voltage to 4500 V, the nebulizer gas to 2.0 bar, the dry gas to 10.0 l/min, and the dry temperature to 220°C. Furthermore, a Funnel 1 RF of 400 Vpp, a multipole RF of 400 Vpp, a quadrupole ion energy of 7.0 eV, a low mass of 322  $m/z$ , a collision energy of 10.0 eV, a collision RF of 1200.0 Vpp, a transfer time of 110.0  $\mu$ s, and a pre pulse storage of 10.0  $\mu$ s was applied. The first two minutes and the last seven minutes of each LC run were diverted to the waste. This time was used for external mass calibration using the ESI-L low concentration tuning mix (Agilent, Waldbronn, Germany) infused by a syringe pump (New Era Pump Systems, Inc., Farmingdale, NY, USA). In addition, internal mass calibration was performed using the compound at  $m/z$  922.009789 originating from the ESI-L low concentration tuning mix. The whole LC-MS system was controlled by Bruker Compass Hystar 5.1 with DCMS link plugin and otofControl 5.2 (both Bruker). For data evaluation and processing, DataAnalysis 5.1 and BioPharma Compass 2021 was used (both Bruker). Samples for LC-MS analysis were diluted to a protein concentration of 0.1 mg/ml in 5% acetonitrile (v/v) and 0.1% FA (v/v) prior to the injection of 5  $\mu$ l of this solution. Prior to analysis compounds were incubated with BTK for one hour.

### 3.15 Reporter Displacement Assay

Binding kinetics determinations were performed with the Proteros reporter displacement assay as previously described [Neumann et al. 2011, Schneider et al. 2013]. In the assay, the time- and competitor concentration-dependent displacement of a Cy5 labelled ATP-competitive kinase inhibitor probe is measured. Probe and enzyme are pre-incubated, before addition of unlabeled competitor. After brief mixing, data are recorded for 1 hour. Data were normalized to high (enzyme with probe, no competitor) and low (enzyme without probe, with competitor) controls for every time-point. For covalent irreversible compounds, the fluorescence decay over time was fit to a mono-exponential decay, excluding the earliest time-point. Obtained  $k_{obs}$  values were replotted against concentration and fit using  $k_{obs} = k_{inact} * [I] / (K_I + [I])$ .

### 3.16 Thermal Shift Assay

To determine the protein melting temperatures, 5  $\mu$ M BTK (389-659) wild-type or C481S mutant was incubated with 50  $\mu$ M of the respective compound for 120 min in 120  $\mu$ l buffer (20 mM HEPES pH 8.0, 0.5 mM TCEP, 150 mM NaCl and 0.01% Tween20, 0.5% DMSO) at room temperature. Subsequently, samples were loaded into nanoDSF capillaries in technical triplicates (PR-C006 NanoTemper Technologies) and loaded into the NanoTemper Prometheus NT.48. Fluorescence at 330 and 350 nm was monitored during a temperature ramp from 20.0 – 95.0°C at 2.5°C/min. The 350 nm / 330 nm ratio was analyzed and melting points were determined using the PR.ThermControl (NanoTemper Technologies, v.2.1.2) software.

### 3.17 Kinase Panel

Compounds were tested for selectivity against a panel of 135 different kinases at DiscoverX using the KINOMEScan™ profiling assay. This is a reporter displacement assay with detection of DNA-conjugate reporter molecules conducted by qPCR. Incubations with test molecules were conducted at 10  $\mu$ M for one hour at room temperature in PBS.

N-terminally GST-tagged BTK (aa389-659) was expressed and purified as previously described (Cuozzo et al 2017). For crystallization purposes the GST tag was cleaved off and the protein was subjected to crystallization using a grid screen. This grid screen was designed by systematically varying the PEG concentration and pH around previously published Polyethylene glycol based conditions (PEG 5kDa MME, ammonium sulfate and MES pH 6.0-7.0). Crystals of BTK were obtained in a few days. In order to obtain the co-crystal structure with Compound 3, these crystals were soaked into the final concentration of 10 mM compound 3 (dissolved in 100% DMSO) in the crystallization buffer. Soaked crystals were cryo-protected and flash-cooled in liquid nitrogen. The X-ray crystal diffraction dataset was collected at the SLS (Swiss Light Source, Villigen, Switzerland). Afterwards, these diffraction data were indexed and integrated using XDS (Kabsch et al 2010) and scaling was performed using XSCALE (Kabsch et al 2010). The structure was solved by a molecular replacement (MR) method using a previously published BTK structure as the MR model (Marcotte et al 2010). The structure was refined by iterative model building using COOT (Emsley & Cowtan, 2004) and refinement using the CCP4 software package. Data collection and refinement statistics are described in Table 5.

#### 4.1 Discussion

We have demonstrated the ability to synthesize numerically large electrophile-terminated oligonucleotide-encoded chemical libraries and then to apply affinity-mediated selection protocols to them to enrich irreversible covalent target-engagers to a sufficient extent to permit their identification within the selection output. For example, the most enriched building block combinations containing compound 3 encoded at the second and third cycles of chemistry with additional defined building blocks at the first cycle of chemistry were observed with count of seven in the selection output. Because there were 551,923 qualifying reads derived from this selection these individual fully elaborated compounds each are 0.00127% of the selection output. Because the numerical size of library 3 is 26,183,954 the DNA tag combination encoding Compound 3 was 0.00000382% within the selection input and so a single cycle of affinity-mediated selection effected an enrichment of (at least) 330-fold which with sufficient sequencing depth enabled the identification of active compounds. The corresponding number for compound 4 discovered within Library 3 was 420-fold (9 counts). For the discoveries of Compounds 5, 6 and 7 in Library 4 it might be expected that because of the larger numerical size of this library (157,103,724 encoded compounds) observed enrichment extents would be smaller. The observed fold-enrichment values for these compounds were 53, 440 and 520-fold respectively. The amount of library used for these experiments was 100 pmoles with 50% of the output used as input for PCR amplification.

The advances that made these discoveries possible were achieved by undertaking a number of protocol optimization experiments in which on-DNA ibrutinib (irreversible) and dasatinib (reversible) were used as positive and negative controls respectively and enrichment was quantified under a range of different conditions. Most notably increased wash stringency using buffer with denaturant increased the enrichments of the irreversible covalent binder over reversible binders compared to “standard” wash stringency using buffer with no denaturant (Figure 8). Ultimately, during selections, incubation time was used as a variable to help identify irreversible covalent binding events since non-covalent

time-independent enrichment signal at least at the incubation times used in our covalent selections (2-15 hours). Covalent inhibitors generally exhibit increasing enrichment at longer time points, see for example Figure 10. The success of our approach is exemplified by the characteristics of the compounds that we discovered. Modes of action were confirmed using a series of biophysical and crystallographic methods including assessments of potency and selectivity which also proved their ability to engage the target with a covalent bond. Notably, none of the covalent irreversible compounds discovered in this study required the first cycle of chemistry to be present to exhibit activity and in all cases it was evident from the large number of structurally unrelated building blocks co-enriched in this register that this was the case. This is in great contrast to our experience of working with the X-Chem DNA-Encoded Chemistry platform in general. Reviewing all of the four hundred and fifty-six non-covalent compounds that we have licensed to our external partners to-date (between December 2012 and December 2020) revealed that greater than 94% of these contained building blocks co-enriched at all of the cycles of encoded chemistry. For licensed covalent irreversible compounds the corresponding number is 92%. It is unclear to us why this project was so unusual in that it provided two different chemical series, both of which amenable to truncation when compared to their original library format. This may be because of one or more of the following reasons. The electrophile-terminated libraries utilized for this project contain four cycles of chemistry whereas those we apply to drug discovery projects generally contain only two or three cycles of chemistry. The chemistries used in the first and second cycles for the electrophile-terminated libraries 3 and 4 were identical meaning that individual molecules for which the installation of the building block at the first cycle did not occur would still be able to accept building blocks at all downstream chemistry cycles. All building blocks were validated for each of the chemistry steps within each library synthesis scheme with only those achieving yields of >50% being admitted to the library. Accordingly, a significant fraction of the resultant library is fully elaborated molecules, but it is also inevitable that truncates will also exist within the encoded mixture. It is also possible that this is a target-specific observation. Since these proof-of-concept experiments were conducted in 2015 (Clark et al 2016) we have offered affinity-mediated irreversible selection as a means to discover covalent inhibitors of therapeutic targets to our partners and observed that 92% of all licensed compounds thus far were comprised of all encoded cycles of chemistry.

#### 5.1 Conclusion

We have developed and now report a method to adapt DNA-encoded chemistry to the discovery of covalent irreversible inhibitors to targets of therapeutic interest. And we have demonstrated the utility of this method to discover multiple novel series of covalent irreversible inhibitors of Bruton's Tyrosine Kinase (BTK). The design of the selection campaign was adapted to inform upon the mode-of-action and selectivity of the enriched compounds and the compounds thus discovered were characterized in sufficient detail to demonstrate the predictive nature of these data. The compounds were fully characterized by Reporter Displacement Assay, Thermal Stability Assay, and intact protein mass-spectrometry of reversible and irreversible compounds with wild-type and C481S (unreactive) BTK proteins, and also using high-resolution X-ray crystal structures. These different series employ multiple covalent warheads to support bond formation including  $\alpha$ ,  $\beta$ -unsaturated carboxamides and epoxides. As far as we are aware, these epoxides are the first ever-

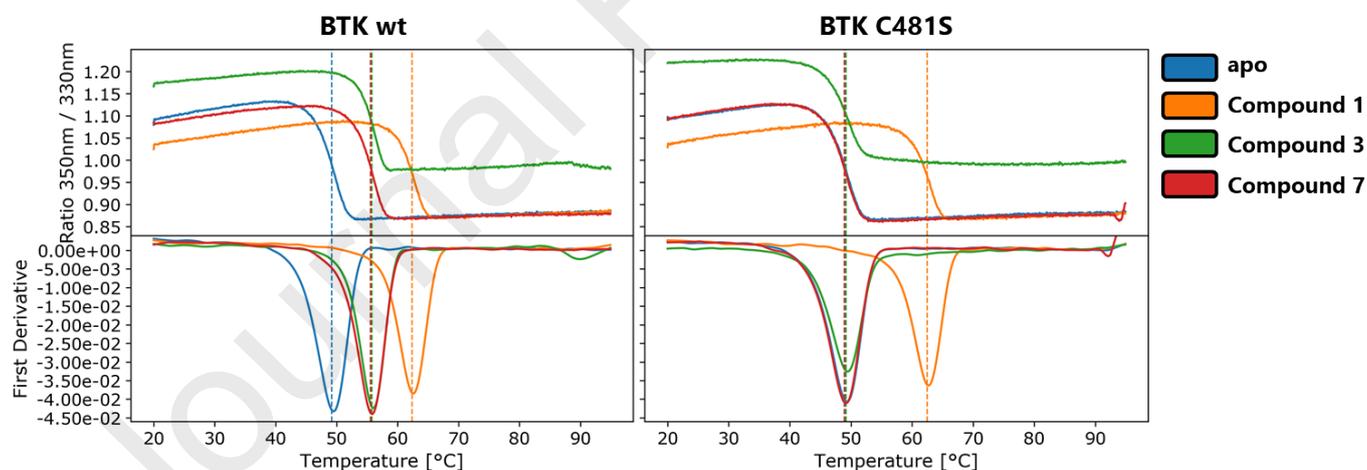
been operating this platform on behalf of our partners since 2015 and have licensed multiple series of covalent irreversible compounds to a range of targets, including those more challenging than BTK. This is the only screening method currently in existence that combines irreversible target engagement assessment with the opportunity to simultaneously screen hundreds of millions

for projects for which a covalent mode of action and novel chemical equity are both considered desirable.

Journal Pre-proofs

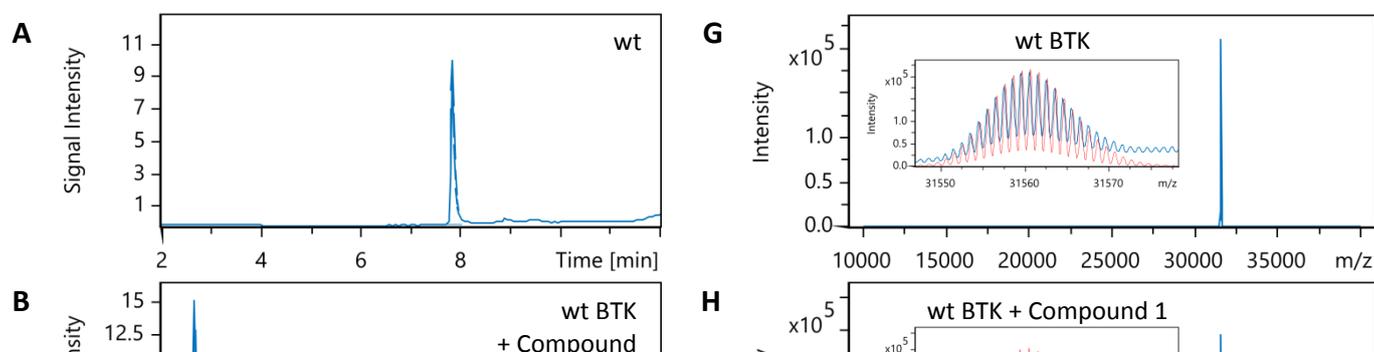
**Figure 1** Reporter Displacement Assay data for Compounds 1, 3 and 7 against wild-type and C481S BTK indicating the activity of Compound 1 against both and of Compounds 3 and 7 against wild-type BTK only. This observation is consistent with the hypothesis that the mode of action of Compounds 3 and 7 is through irreversible covalent engagement of Cysteine 481, whereas for Compound 1 the interaction is non-covalent and reversible in nature and not dependent upon engagement of Cysteine 481. Time-dependent normalized signals at different compound concentrations (in  $\mu\text{M}$ ) are shown on the right (see legend). Experiments were performed in technical triplicates.

**Figure 2**



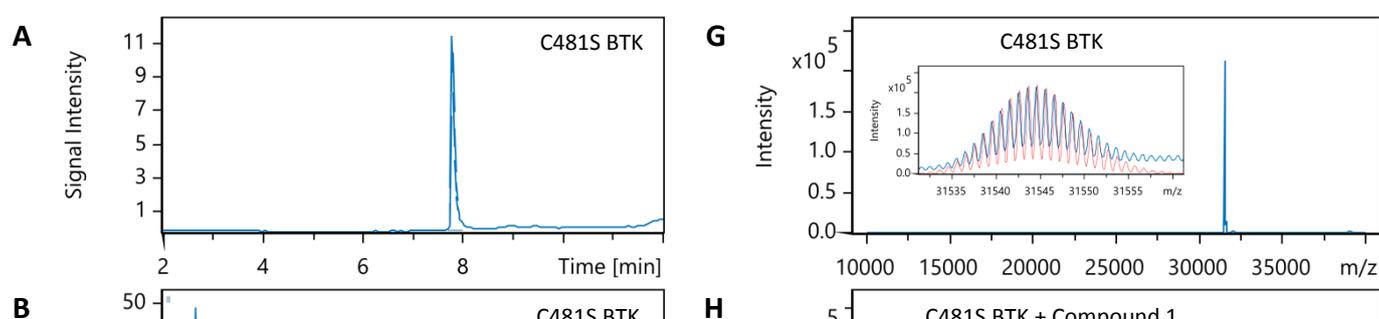
**Figure 2** Thermal Shift Assay data derived from the intrinsic tryptophan fluorescence of BTK measured using nanoDSF with a NanoTemper Prometheus NT.48 instrument. Large positive thermal transition shifts are observed for Compounds 1 (orange), 3 (green) and 7 (red) against wild-type BTK (apo: blue) whereas only for Compound 1 for C481S BTK. This observation is consistent with the hypothesis that the mode of action of Compounds 3 and 7 is through irreversible covalent engagement of Cysteine 481, whereas for Compound 1 the interaction is non-covalent and irreversible in nature and not dependent upon engagement of Cysteine 481. Raw 350 / 330 nm ratio data are shown in the top panel. The first derivative of the ratio is shown in the bottom panel. Melting temperatures, as calculated from the maximum in the first derivative, are shown as dashed lines in the respective colors. Data are averages from three technical replicates

**Figure 3**



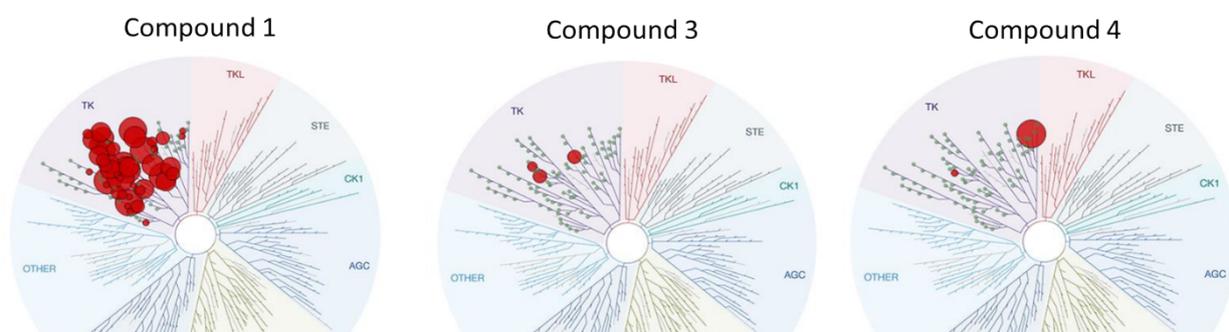
**Figure 3** A-F: UV chromatograms of wild-type BTK after incubation with compound: A: BTK only; B: BTK and Compound 1; C: BTK and Compound 3; D: BTK and Compound 4; E: BTK and Compound 6; F: BTK and Compound 7. BTK eluted with a retention time of 7.9 min - 8.0 min. Chromatographic peaks highlighted with a hash correspond to the elution of the respective unreacted compounds. G-L: deconvoluted mass spectra of wild-type BTK; G: BTK only; H: BTK and Compound 1 indicating no covalent engagement, only the mass of the unreacted BTK was detected; I: BTK and Compound 3 indicating a covalent adduct; J: BTK and Compound 4 indicating a covalent adduct; K: BTK and Compound 6 indicating a covalent adduct; L: BTK and Compound 7 indicating a covalent adduct. Masses marked with an asterisk indicate unreacted BTK; in the inserts a zoom into the corresponding region of interest is shown; the blue line indicates the measured spectrum and the red line indicates the simulated spectrum, respectively

**Figure 4**



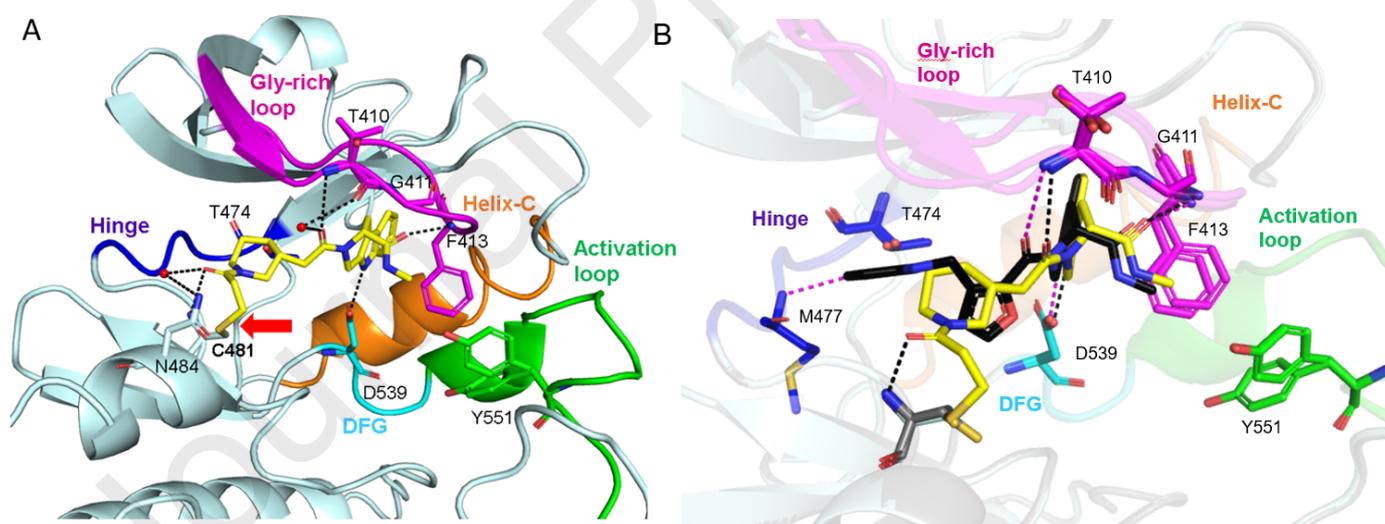
**Figure 4** A-F: UV chromatograms of C481S BTK after incubation with compound: A: BTK only; B: BTK and Compound 1; C: BTK and Compound 3; D: BTK and Compound 4; E: BTK and Compound 6; F: BTK and Compound 7. BTK eluted with a retention time of 7.8 min – 7.9 min. Chromatographic peaks highlighted with a hash correspond to the elution of the respective unreacted compounds. G-L: deconvoluted mass spectra of C481S BTK; G: BTK only; H: BTK and Compound 1; I: BTK and Compound 3; J: BTK and Compound 4; K: BTK and Compound 6; L: BTK and Compound 7. None of these compounds are able to form a covalent adduct with BTK481S. In the inserts a zoom into the corresponding region of interest is shown; the blue line indicates the measured spectrum and the red line indicates the simulated spectrum, respectively.

**Figure 5**



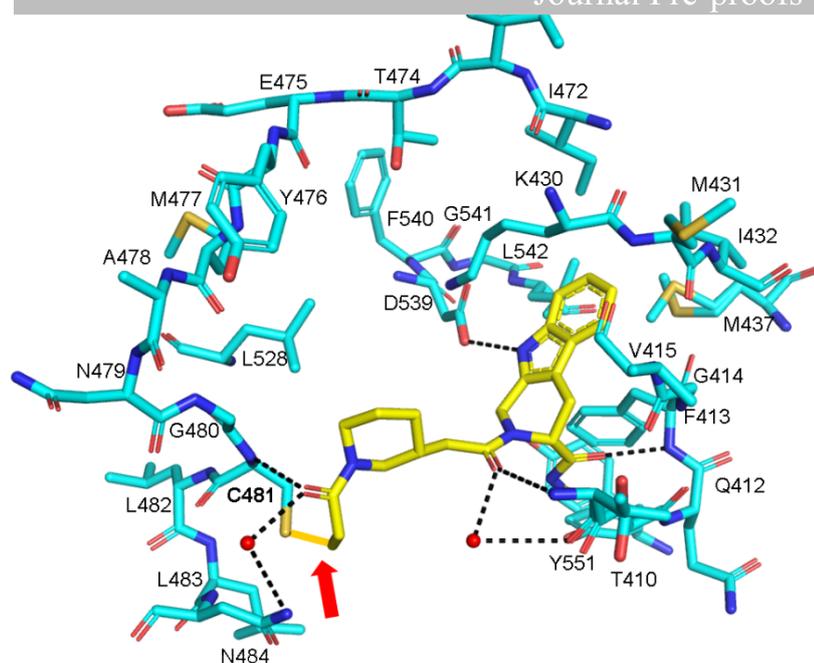
**Figure 5.** Compounds were screened for selectivity against a panel of 135 different kinases at DiscoverX using the KINOMEScan™ profiling assay. This is a reporter displacement assay with detection of displaced DNA-conjugate reporter molecules conducted by qPCR. Incubations with test molecules were conducted at 10  $\mu$ M for one hour at room temperature in PBS. Scale indicates extent of remaining activity under assay conditions with 0% corresponding to full inhibition.

**Figure 6**



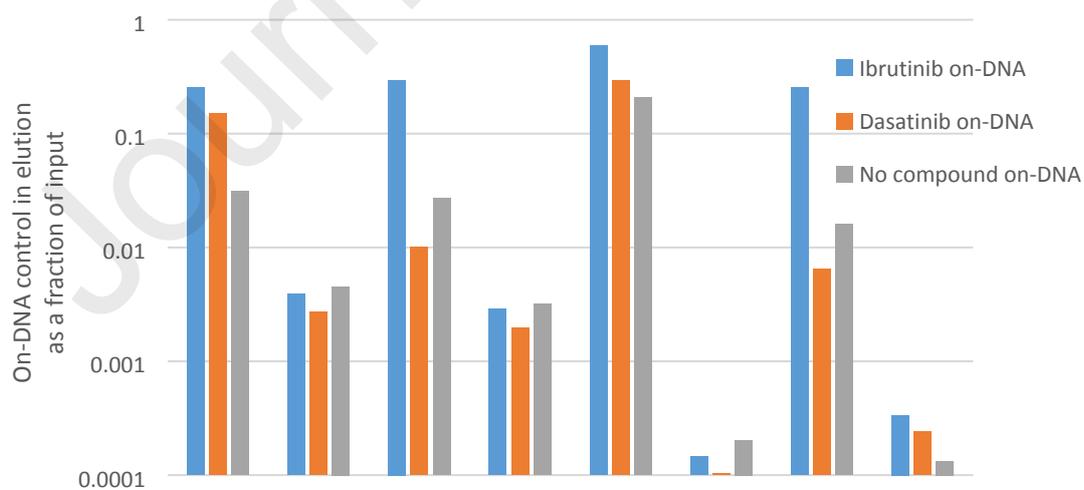
**Figure 6.** The 1.44 Å resolution BTK structure in complex with covalent irreversible inhibitor Compound 3. (A) Close-up view of the binding of Compound 3. Red arrow indicates the covalent bond between Compound 3 and Cys481. (B) Close-up view of overlay of active site of BTK bound to Compound 3 (yellow) and Compound 1 (black). Protein structure is shown as cartoon, inhibitor and selected residue side chains are depicted using stick model. Water molecules that bridge the interaction between the protein and the inhibitor are shown as red spheres. Magenta and black dashed lines represent hydrogen bonds between BTK kinase domain with Compounds 1 and 3, respectively.

**Figure 7**



**Figure 7** Overall binding pocket of Compound 3 with BTK kinase domain. Protein residues and compound 3 are shown in cyan and yellow sticks respectively. Water molecules are represented as red spheres and hydrogen bonds as dashed line. The red arrow indicates the covalent bond between Compound 3 and Cys481.

**Figure 8a**



**Figure 8b**

Figure 8c

vs. input by qPCR analysis

**Figure 8.** Quantitative PCR-derived enrichment data for on-DNA ibrutinib and dasatinib under model selection conditions. 8a shows relative enrichment for different denaturants and capture matrices in the absence of sheared salmon sperm DNA and a DNA-Encoded Chemical Library super-mix, while 8b shows the same in the presence of sheared salmon sperm DNA and a DNA-Encoded Chemical Library super-mix. Under these conditions, guanidine and IMAC magnetic beads provided superior discrimination between ibrutinib (covalent, irreversible) and dasatinib (non-covalent, reversible). 6c indicates the experimental scheme.

Figure 9

**Figure 9.** Affinity-mediated selection output of Library 3 (Table 1) against 1  $\mu$ M BTK incubated for 18 hours. The x- and y-axes represent building blocks in the second and third cycles of chemistry respectively. The vertical axis is a relative enrichment metric for each combination of building blocks B and C (negative  $\log_{10}$  of the asymptotic significance value) with a value of  $>10$  considered to be statistically significant. Instances (compounds) are randomly displaced by 0.5% to indicate the extent to which building block variation at the first cycle of chemistry is tolerated. The extent to which enriched building block combinations have been seen to enrich against prior targets is represented by colour with green representing the least promiscuous (building block combinations never previously seen to have been enriched against any target, i.e. negative  $\log_{10}$  of the asymptotic significance value has never previously exceeded 10) and red the most promiscuous. Clusters colored in green had never been seen to enrich before in other screens against other targets. Compound 3 was found within the enriched cluster circled in dark blue. Closely related compounds to compound 3 with variation at building block C (see Table 4) are circled in light blue. Compound 4 was found within the enriched cluster circled in red.

**Figure 10**

**Figure 10.** Profile of the enrichment of the building block combination defining Compound 3 across all affinity-mediated selections conducted within the selection campaign. The vertical axis is a relative enrichment metric for each combination of building blocks B and C (negative  $\log_{10}$  of the asymptotic significance value) with a value of  $>10$  considered to be statistically significant. Each off-target contained at least one cysteine, additional off-target information is shown in Supplementary Table S1.

**Table 1.** DNA-Encoded chemical libraries used in this study

Library	First Chemical Step	Second Chemical Step	Third Chemical Step	Electrophile	Unique Compounds in Library
1	Acylation with 300 Amino Acids	Acylation with 157 Formyl Acids	Reductive Amination with 2,341 Amines	None	110,261,100
2	Acylation with 300 Amino Acids	Acylation with 300 Amino Acids	Acylation with 2,500 Carboxylates	None	225,000,000
3	Acylation with 305 Amino Acids	Acylation with 293 Amino Acids	Acylation with 293 Amino Acids	Acrylamide	26,183,954
4	Acylation with 305 Amino Acids	Acylation with 293 Amino Acids	Acylation with 293 Amino Acids	Six, including range of epoxides	157,103,670

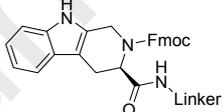
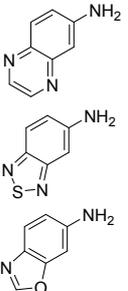
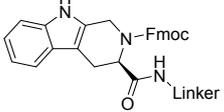
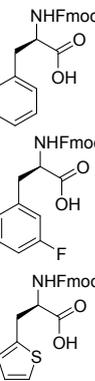
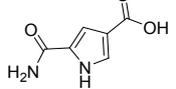
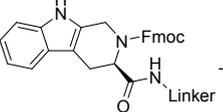
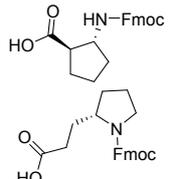
**Table 2.** Compounds reported in this study



Compound	BTK Wild-Type							BTK C481S			
	IC <sub>50</sub> at 1h / $\mu$ M	Residence Time / min	$k_{inact}$ / per second	$K_i$ / $\mu$ M	$k_{inact} / K_i$ (per second per M)	$\Delta T_m$ / $^{\circ}$ C	MS/%	IC <sub>50</sub> at 1h / $\mu$ M	Residence Time / min	$\Delta T_m$ / $^{\circ}$ C	MS/%
1	0.00378 $\pm$ 0.00036	9 $\pm$ 2	NA	NA	NA	13.2 $\pm$ 0.1	0	0.00304 $\pm$ 0.00036	22 $\pm$ 3	13.4 $\pm$ 0.1	0
2	0.0227 $\pm$ 0.0023	34 $\pm$ 4	NA	NA	NA		NA	0.0562 $\pm$ 0.0036	14 $\pm$ 1		NA
3	4.76 $\pm$ 0.54	NA	ND	ND	59.2 $\pm$ 1.0	6.55 $\pm$ 0.15	97.9	>5.92E+01	NA	0.10 $\pm$ 0.06	0
4	36.4 $\pm$ 4.6	NA	NA	NA	NA		94.4	>5.92E+01	NA		0
5	0.232 $\pm$ 0.034	NA	1.54E-02 $\pm$ 0.09E-02	9.27 $\pm$ 1.97	1660 $\pm$ 365		NA	>5.92E+01	NA		NA
6	0.0581 $\pm$ 0.0048	NA	8.22E-03 $\pm$ 1.98E-04	1.25 $\pm$ 0.11	6580 $\pm$ 617	4.51 $\pm$ 0.06	92.3	>5.92E+01	NA	0.23 $\pm$ 0.33	0
7	0.0842 $\pm$ 0.0062	NA	9.41E-03 $\pm$ 3.78E-04	1.93 $\pm$ 0.32	4880 $\pm$ 826	6.36 $\pm$ 0.10	95.2	>5.92E+01	NA	-0.15 $\pm$ 0.08	0

**Table 3.** Summary of data determined for Compounds 1-7 using the Reporter Displacement Assay (RDA), Thermal Shift Assay (nanoDSF) and intact protein LC-MS. IC<sub>50</sub> values, residence times,  $k_{inact}$ ,  $K_i$  and  $k_{inact} / K_i$  were determined in the Proteros RDA.  $\Delta T_m$  values were obtained by subtracting the  $T_m$  of the apo BTK (wild-type or C481S) from the  $T_m$  in presence of 50  $\mu$ M compound as determined by nanoDSF. The amounts of protein bound to compound (MS/%) were calculated by comparing the peak heights in the deconvoluted MS spectrum of unreacted and reacted protein.

**Table 4.** Tolerated building block variability as exhibited by co-enriched structurally related compounds in the selection output

Library	Compound	First Chemical Cycle Building Blocks (A)	Second Chemical Cycle Building Blocks (B)	Third Chemical Cycle Building Blocks (C)	Terminal Electrophile
1	1				None
2	2				None
3	3	Variable, dispensable			

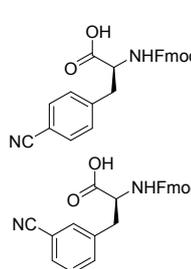
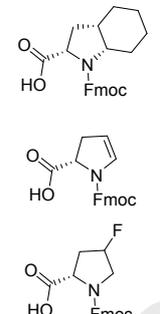
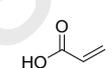
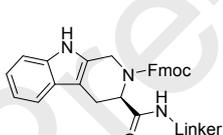
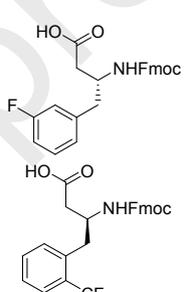
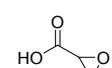
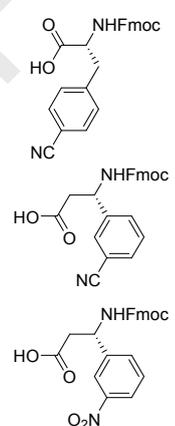
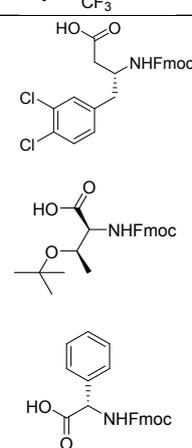
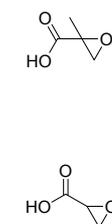
Journal Pre-proofs					
3	4	Variable, dispensable			
4	5	Variable, dispensable			
4	6 & 7	Variable, dispensable			

Table 5

X-1	
Wavelength [Å]	1.0000
Detector	PILATUS 2M
Space group	P 21 21 2
Cell: a; b; c; [Å]	71.62; 103.23; 37.86
$\alpha$ ; $\beta$ ; $\gamma$ ; [°]	90.0; 90.0; 90.0
Resolution [Å]	1.44 (1.69-1.44)
Unique reflections	50687 (19233)
Multiplicity	3.2 (3.1)
Completeness [%]	98.0 (98.9)
Rsym [%]	5.6 (44.6)
Rmeas [%]	6.7 (53.7)
$\langle I/\sigma(I) \rangle$	13.54 (2.78)

**Refinement**

Resolution [Å]	58.84-1.44
Number of reflections (working /test)	45594 / 5090
Rcryst [%]	16.1
Rfree [%]	20.4
Total number of atoms:	
Protein	2390
Water	333
Ligand	30
Deviation from ideal geometry:	
Bond lengths [Å]	0.009
Bond angles [°]	1.45
Ramachandran plot	
Most favoured regions [%]	95.1
allowed regions [%]	4.9
Disallowed regions [%]	0.0

$I/\sigma(I)$  corresponds to the average of the intensity divided by its average standard deviation.

$R_{cryst}/R_{free} = \sum |F_{obs} - F_{calc}| / \sum |F_{obs}|$ , where  $F_{obs}$  and  $F_{calc}$  are the observed and calculated structure factors, respectively.  $R_{free}$  is the same as  $R_{work}$ , calculated for the 10% of the data that were randomly omitted from refinement

**Table 5** Data collection and refinement statistics for structure determination of the complex of BTK kinase domain with Compound 3 using X-ray diffraction and crystallography.

**References**

- Burger, J.A., Tedeschi, A., Barr, P.M., Robak, T., Owen, C., Ghia, P., Bairey, O., Hillmen, P., Bartlett, N.L., Li, J. and Simpson, D., 2015. Ibrutinib as initial therapy for patients with chronic lymphocytic leukemia. *New England Journal of Medicine*, 373(25), pp.2425-2437.
- Caldwell, R.D., Qiu, H., Askew, B.C., Bender, A.T., Brugger, N., Camps, M., Dhanabal, M., Dutt, V., Eichhorn, T., Gardberg, A.S. and Goutopoulos, A., 2019. Discovery of evobrutinib: an oral, potent, and highly selective, covalent Bruton's tyrosine kinase (BTK) inhibitor for the treatment of immunological diseases.
- Chan, A.I., McGregor, L.M., Jain, T. and Liu, D.R., 2017. Discovery of a covalent kinase inhibitor from a DNA-encoded small-molecule library  $\times$  protein library selection. *Journal of the American Chemical Society*, 139(30), pp.10192-10195.
- Clark, M.A., Acharya, R.A., Arico-Muendel, C.C., Belyanskaya, S.L., Benjamin, D.R., Carlson, N.R., Centrella, P.A., Chiu, C.H., Creaser, S.P., Cuozzo, J.W. and Davie, C.P., 2009. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nature chemical biology*, 5(9), pp.647-654
- Clark, M., Keefe, A.D., Guilinger, J., Zhang, Y. and Tian, X., X-CHEM Inc, 2020. Covalent BTK inhibitors and uses thereof. U.S. Patent Application 16/097,093.
- Cuozzo, J.W., Centrella, P.A., Gikunju, D., Habeshian, S., Hupp, C.D., Keefe, A.D., Sigel, E.A., Soutter, H.H., Thomson, H.A., Zhang, Y. and Clark, M.A., 2017. Discovery of a Potent BTK Inhibitor with a Novel Binding Mode by Using Parallel Selections with a DNA-Encoded Chemical Library. *ChemBioChem*, 18(9), pp.864-871.
- Di Paolo, J.A., Huang, T., Balazs, M., Barbosa, J., Barck, K.H., Bravo, B.J., Carano, R.A., Darrow, J., Davies, D.R., DeForge, L.E. and Diehl, L., 2011. Specific Btk inhibition suppresses B cell- and myeloid cell-mediated arthritis. *Nature chemical biology*, 7(1), p.41.
- Dungo, R.T. and Keating, G.M., 2013. Afatinib: first global approval. *Drugs*, 73(13), pp.1503-1515.
- Emsley, P. and Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta crystallographica section D: biological crystallography*, 60(12), pp.2126-2132.
- Favalli, N., Bassi, G., Scheuermann, J. and Neri, D., 2018. DNA-encoded chemical libraries—achievements and remaining challenges. *FEBS letters*, 592(12), pp.2168-2180.
- Gehring, M. and Laufer, S.A., 2018. Emerging and re-emerging warheads for targeted covalent inhibitors: applications in medicinal chemistry and chemical biology. *Journal of medicinal chemistry*, 62(12), pp.5673-5724.
- Goodnow, R.A., Dumelin, C.E. and Keefe, A.D., 2017. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nature Reviews Drug Discovery*, 16(2), pp.131-147.
- Greig, S.L., 2016. Osimertinib: first global approval. *Drugs*, 76(2), pp.263-273.
- Ghosh, A.K., Samanta, I., Mondal, A. and Liu, W.R., 2019. Covalent inhibition in drug discovery. *ChemMedChem*, 14(9), p.889.
- Jackson, P.A., Widen, J.C., Harki, D.A. and Brummond, K.M., 2017. Covalent modifiers: A chemical perspective on the reactivity of  $\alpha$ ,  $\beta$ -unsaturated carbonyls with thiols via hetero-Michael addition reactions. *Journal of medicinal chemistry*, 60(3), pp.839-885.
- Kabsch, W., 2010. Xds. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), pp.125-132.
- Kuglstatter, A., Wong, A., Tsing, S., Lee, S.W., Lou, Y., Villaseñor, A.G., Bradshaw, J.M., Shaw, D., Barnett, J.W. and Browner, M.F., 2011. Insights into the conformational flexibility of Bruton's tyrosine kinase from multiple ligand complex structures. *Protein Science*, 20(2), pp.428-436.
- Lagoutte, R., Patouret, R. and Winssinger, N., 2017. Covalent inhibitors: an opportunity for rational target selectivity. *Current Opinion in Chemical Biology*, 39, pp.54-63.

- Benina, K., Castaneda, S., Comchus, L.A., Das, J., Dowcyko, A.M. and Fairchild, C., 2004. Discovery of N-(2-chloro-6-methylphenyl)-2-(6-(4-(2-hydroxyethyl)-piperazin-1-yl)-2-methylpyrimidin-4-ylamino) thiazole-5-carboxamide (BMS-354825), a dual Src/Abl kinase inhibitor with potent antitumor activity in preclinical assays. *Journal of medicinal chemistry*, 47(27), pp.6658-6661.
20. Marcotte, D.J., Liu, Y.T., Arduini, R.M., Hession, C.A., Miatkowski, K., Wildes, C.P., Cullen, P.F., Hong, V., Hopkins, B.T., Mertsching, E. and Jenkins, T.J., 2010. Structures of human Bruton's tyrosine kinase in active and inactive conformations suggest a mechanism of activation for TEC family kinases. *Protein Science*, 19(3), pp.429-439.
  21. Neumann, L., von König, K. and Ullmann, D., 2011. HTS reporter displacement assay for fragment screening and fragment evolution toward leads with optimized binding kinetics, binding selectivity, and thermodynamic signature. *Methods in enzymology*, 493, pp.299-320.
  22. Project, C.C., 1994. The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography*, 50(Pt 5), pp.760-763.
  23. Ray, S. and Murkin, A.S., 2019. New electrophiles and strategies for mechanism-based and targeted covalent inhibitor design. *Biochemistry*, 58(52), pp.5234-5244.
  24. Schneider, E.V., Böttcher, J., Huber, R., Maskos, K. and Neumann, L., 2013. Structure-kinetic relationship study of CDK8/CycC specific compounds. *Proceedings of the National Academy of Sciences*, 110(20), pp.8081-8086.
  25. Sequist, L.V., Soria, J.C., Goldman, J.W., Wakelee, H.A., Gadgeel, S.M., Varga, A., Papadimitrakopoulou, V., Solomon, B.J., Oxnard, G.R., Dziadziuszko, R. and Aisner, D.L., 2015. Rociletinib in EGFR-mutated non-small-cell lung cancer. *New England Journal of Medicine*, 372(18), pp.1700-1709.
  26. Spaargaren, M., De Rooij, M.F.M., Kater, A.P. and Eldering, E., 2015. BTK inhibitors in chronic lymphocytic leukemia: a glimpse to the future. *Oncogene*, 34(19), pp.2426-2436.
  27. Zambaldo, C., Daguer, J.P., Saarbach, J., Barluenga, S. and Winssinger, N., 2016. Screening for covalent inhibitors using DNA-display of small molecule libraries functionalized with cysteine reactive moieties. *MedChemComm*, 7(7), pp.1340-1351.
  28. Zhang, T., Hatcher, J.M., Teng, M., Gray, N.S. and Kostic, M., 2019. Recent advances in selective and irreversible covalent ligand development and validation. *Cell chemical biology*, 26(11), pp.1486-1500.
  29. Zhu, Z., Grady, L.C., Ding, Y., Lind, K.E., Davie, C.P., Phelps, C.B. and Evindar, G., 2019. Development of a Selection Method for Discovering Irreversible (Covalent) Binders from a DNA-Encoded Library. *SLAS DISCOVERY: Advancing Life Sciences R&D*, 24(2), pp.169-174.
  30. Zimmermann, G., Rieder, U., Bajic, D., Vanetti, S., Chaikuad, A., Knapp, S., Scheuermann, J., Mattarella, M. and Neri, D., 2017. A Specific and Covalent JNK-1 Ligand Selected from an Encoded Self-Assembling Chemical Library. *Chemistry—A European Journal*, 23(34), pp.8152-8155.
- JPG designed the covalent selection protocols, JPG performed the non-covalent selections, MAG processed the selection output sequencing data, ADK and HAT analyzed the non-covalent selection output data, ADK and JPG analyzed the covalent selection output data, YL and YZ coordinated the outsourced off-DNA compound synthesis, YL performed compound synthesis, JWC coordinated the sourcing of BTK, JWC and JL established the initial TR-FRET (data not shown), JL ran the initial TR-FRET (data not shown), AA interpreted the crystallography results, MA established purification protocols for BTK, AB provided scientific and technical support for RDA and nanoDSF assays, MD performed the RDA assay, RK optimized crystallization conditions, SK coordinated the crystallography, AL solved the Compound 3 complex crystal structure, LL designed and coordinated the RDA and nanoDSF experiments, KM purified the wild-type and C481S BTK proteins, MM solved the Compound 1 complex crystal structure, KP performed nanoDSF and RDA assays, MS performed and analyzed all intact protein LC-MS studies, DKM designed overall compound assay strategy and coordinated all Proteros activities.

### Statement on Conflicts of Interest

All authors were employees of X-Chem Inc. or Proteros Biostructures GmbH at the time they conducted work on this project and declare no conflicts of interest.

### Supplementary Material

Supplementary materials exist in a separate document.

### Acknowledgments

X-Chem Inc. and Proteros Biostructures GmbH are acknowledged for their support of this work. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

### Author Contributions

MAC, PAC and YZ designed the library chemical schemes, PAC synthesized libraries 1 and 2, SH synthesized the non-electrophilic core of libraries 3 and 4, XT installed the electrophiles, ADK and HAT designed the non-covalent selection