Efficient large-scale exploration of fragment hit progression by exploiting binding-site purification of actives (B-SPA) through combining multi-step array synthesis and HT crystallography.

Harold Grosjean[#1,2], Anthony Aimon[#1,3], Storm Hassell-Hart[#6], Warren Thompson[#1,3], Lizbé Koekemoer[4,5], James Bennett[4,5], Anthony Bradley[5], Cameron Anderson[2], Conor Wild[1], William Bradshaw[4], Edward A. FitzGerald[8], Tobias Krojer[5], Oleg Fedorov[5], Philip C. Biggin[2], John Spencer*[6,7] and Frank von Delft*[1,3,4,5,9]

[#] Joint first authors

* Corresponding authors

Affiliation:

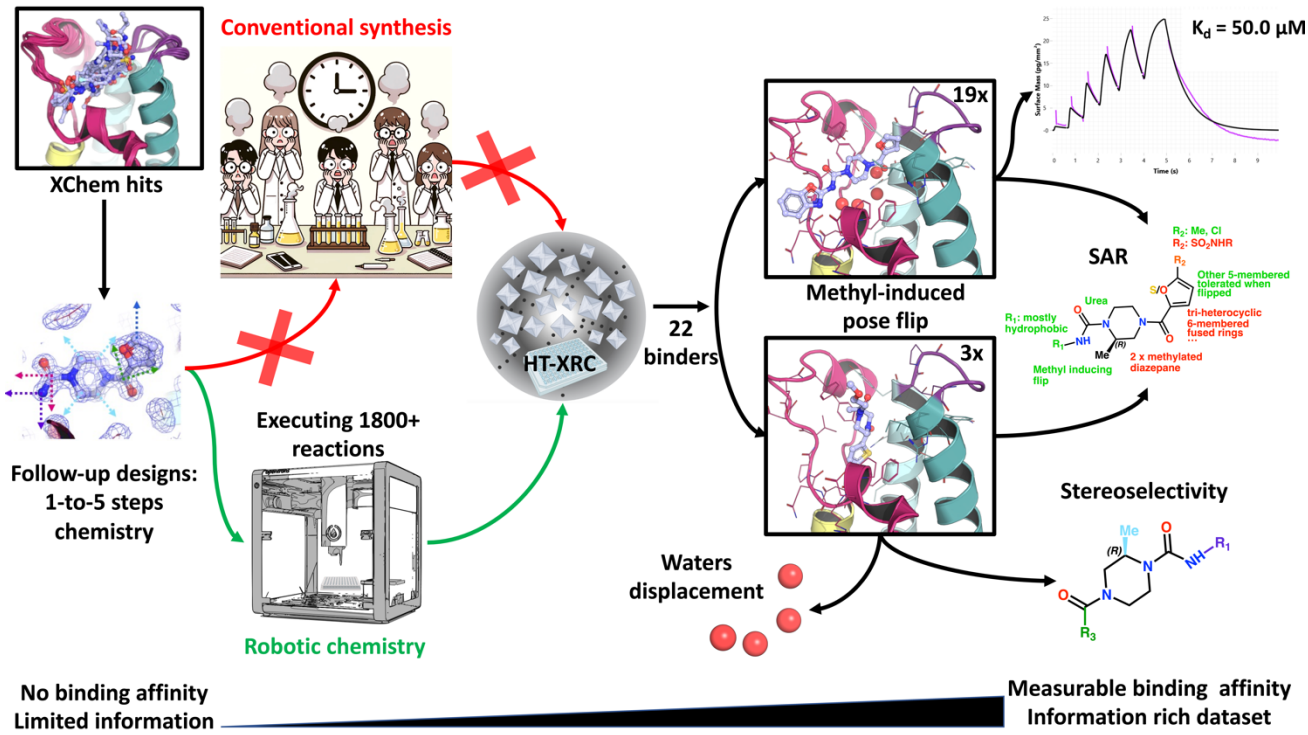
1. Diamond Light Source Ltd, Harwell Science and Innovation Campus, OX11 0QX, Didcot, UK
2. Structural Bioinformatics and Computational Biochemistry, Department of Biochemistry, University of Oxford, South Parks Road, OX1 3QU, Oxford, UK
3. Research Complex at Harwell, Harwell Science and Innovation Campus, OX11 0FA, Didcot, UK
4. Centre for Medicines Discovery, University of Oxford, Old Road Campus, Roosevelt Drive, OX3 7DQ, Headington, UK
5. Structural Genomics Consortium, University of Oxford, Old Road Campus, Roosevelt Drive, OX3 7DQ, Headington, UK
6. Department of Chemistry, School of Life Sciences, University of Sussex, Falmer, BN1 9QJ
7. Sussex Drug Discovery Centre (SDDC), School of Life Sciences, University of Sussex, Falmer, BN1
8. Creoptix AG, Zugerstrasse 76, 8820 Wädenswil, Switzerland
9. Department of Biochemistry, University of Johannesburg, Auckland Park 2006, South Africa

Author contribution: AA, SH-H, AB, TK: experimental design, robotics and synthesis of compounds. HG, TK, LK, SH-H: protein production and crystallization. HG, AA, CA, TK, CW, WJB; Xray analysis and modelling. LK, EAF, JB, OF: assays. WT, SH-H, AA: post-synthetic and quality control analysis. JS, FvD, PCB; project conceptualisation and oversight.

All authors contributed to data analysis, curation, drafting and proofreading the final manuscript.

Abstract

Fragment approaches are long-established in target-based ligand discovery. Nevertheless, their full transformative potential lies dormant, because progressing hits to potency remains difficult and underserved by methodology developments, which mostly focus on screening. The only credible progression paradigm is conventional design-make-test analyse (DMTA) medicinal chemistry, which is costly and thus necessitates picking winners early, thereby effectively discarding all the other hits. We here demonstrate the workability of an alternative strategy, namely immediate large-scale exploration of diverse hit-inspired compounds. The key insight is that it is effective to cheaply parallelize large numbers of non-uniform multi-step reactions, because even without compound purification, a high-quality readout of binding is available, namely crystallography of fragment screening. This has the sensitivity to detect even low-level binding of slightly active compounds, which the targeted binding site extracts directly from crude reaction mixtures (CRMs). In this proof-of-concept study, we expand a fragment hit from a crystal-based screen of the second bromodomain of human PHIP, using array synthesis on low-cost robotics to implement 6 independent multi-step reaction routes of up to 5 steps, attempting the synthesis of 1876 diverse expansions; designs were entirely driven by synthetic tractability. Expected product was present in 1108 CRMs, as detected by automated mass spectrometry; and 22 individual products were resolved in crystal structures of CRMs added to crystals. These provided an initial SAR map, revealed pose stability in 19 and instability in 3 products, and resolved stereochemical preference. Unexpectedly, in view of the naïve design approach, one resolved compound even showed on-scale biochemical potency (IC_{50}=34 µM) and biophysical affinity (K_d=50 µM) after resynthesis. This binding-site purification of actives (B-SPA) process is formulaic and engineerable, here yielding the output of >25 person-years in ~20 days, with solvent use reduced from >4,500L to <20L. Thus, this approach, coupled with algorithmically guided compound and reaction design and new formalisms for data analysis, lends itself to routine fragment progression.



Introduction

Fragment approaches have established themselves over the last few decades as a powerful approach for interrogating drug targets and discovering new biologically active compounds in a relatively short time.[1,2] By the end of 2021 there were 6 fragment-derived drugs in the clinic: in order of approval: Zelboraf (vemurafenib),[3] Venetoclax (ABT-199),[4] Erdafitinib,[5] Pexidartinib (PLX2297),[6] Sotorasib (AMG-510)[7] and Asciminib (ABL001).[8] Two of these moved through the drug discovery pipeline relatively quickly: Zelboraf progressed from an unselective fragment to a highly selective cancer drug in the span of 6 years; and Sotorasib took 8 years from a publication,[9] demonstrating the druggability of its target K-RAS^{G12C} to an FDA approved drug.

Fragment screening is usually performed using biophysical methods or X-Ray crystallography on rule of three (Ro3), diversity-orientated (DOS) or natural product-inspired libraries that sample a large portion of the accessible chemical space.[10–13] These tend also to be biased towards readily adaptable drug-like leads.[14,15] However, due to their size, fragments rarely have useful affinity, and the initial screen must be followed by a generally tricky design-make-test-analyse (DMTA) cycles[16], to progress one or more observed hits to biologically relevant potency. Design strategies are commonly classed as fragment linking, merging, or growing, and are ideally supported by advanced computational approaches such as molecular dynamics[17,18] or informed "SAR by catalogue"[19,20].

What makes the initial DMTA cycles particularly challenging is that one first needs to achieve compounds with measurable and preferably on-scale (<50 µM) activity or affinity before common medicinal chemistry design principles become applicable. For this reason, fragment-based efforts have come to rely on large libraries and careful assay cascades to ensure fragment hits are identified with activity or affinity that is reliably measurable, so that they can be progressed with confidence.[21] This draws on best-practice paradigms from High-Throughput Screening (HTS), where rigorous secondary and even tertiary assays are essential for confirming initial hits.[22]

Increasingly, however, fragment hits are being exploited that have no measurable affinity yet are readily identifiable and structurally compelling; these are common where protein crystallography is the primary screen and assay concentrations high, a now very accessible technique.[23,24] Examples include Resnick *et al.*[25] where such a fragment was merged with a covalent electrophilic fragment hit to achieve on-scale potency; and Boby *et al.*[26] report how the merging of two such fragments[27] led to an entire drug discovery effort. Algorithmic approaches have also become available: one approach is to ensure structural motifs observed in the fragments are directly recapitulated in the follow-up designs[28]; another uses the pharmacophoric information from multiple fragments to guide docking[29]; alternatively, interactions can be assessed and prioritised by Dynamic Undocking.[30]

Nevertheless, there invariably remains a long journey, both conceptually and experimentally, from merely on-scale potency to the nanomolar potency required of a lead compound or chemical probe, and the ability to cycle DMTA effectively remains the purview of only expert and well-funded organisations.[31] A key cost driver is the continued requirement for sufficiently purified compound, to minimise confounding assay results. Certainly, access to such compounds has been transformed over the last decade, thanks to a vastly expanded landscape of commercial vendors and providers, who have invested heavily in expertise, technology, building block collections and algorithms, thereby eliminating the need for up-front laboratory investment by individual discovery efforts, even for synthetically complex targets.[29] Additionally, progress has been made in recent years to developing low-cost HTS platforms[32] for the synthesis, purification and analysis of compounds.[33–36]

However, it appears that the scale of compound exploration must be at least an order of magnitude larger than can be supported by these evolutions on typical early discovery budgets. For instance, Gao *et al.*[37] have

used “on-the-fly" nanoscale library syntheses by acoustic dispensing to generate large arrays. Elsewhere, surface plasmon resonance (SPR) protocols have been used to identify hits from crude mixtures.[38] Such approaches not only increase the chances of directly identifying potent compound series, they also generate a critical quantity of data to support robust modelling of the structure-activity relationships (SAR).[39–41]

These approaches, however, retain key weaknesses: they can only explore focused regions of chemical space; and the required synthetic approach must be curtailed to accommodate the constraints of the assay read-outs.[42] They therefore cannot yet serve as a general approach for generating the data that optimally populates computational models to allow them to effectively identify high-potency compounds.

Here, we demonstrate an approach, binding-site purification of actives (B-SPA), that can indeed generate data not only at scale but also sampling chemical space suitably widely, and moreover at very low cost of synthesis. We have previously shown that crude reaction mixtures, from a single array of 2-component single-step reactions, yield eminently interpretable results of protein-ligand binding when using protein crystallography in combination with sensor-based affinity measurement (surface plasmon resonance, SPR).[43]

We now show that a more general synthetic approach remains effective, namely using array chemistry of diverse multi-step reactions that yield bespoke compounds in crude reaction mixtures. We exploit the fact that the binding site of the protein will extract the most potent species from complex mixtures of compounds, and that the effect can be observed by both biophysics[38] or crystallography[44], now significantly more sensitively in the latter through ongoing developments in signal extraction.[45]

We previously completed a crystallographic fragment screen (PDB deposition ID: G_1002162) against the second bromodomain of the Pleckstrin Homology Domain-Interacting Protein (PHIP). This multidomain protein is involved in various cellular processes including cellular growth and mobility[14] and has been implicated in aggressive cancers including BRAF-negative melanomas, breast and lung cancer[46,47]; nevertheless, the specific role of the second bromodomain (hereafter: PHIP(2)) is unknown.

In this study, we used array synthesis to rapidly survey the chemical space of the opportunities to grow fragment F709, a piperazine hit from this screen. Synthesis was implemented as a robotics prototype platform based on the low-cost OpenTrons OT-1™ liquid handler, while initial quality control (QC) of the crude reaction mixtures (CRM) used HPLC-mass spectrometric analysis, and all further bottlenecks (*viz.* extraction, solvent evaporation, purification, spectroscopic characterisation) were bypassed by the X-ray crystallographic analysis of the CRMs.[2,48] The workflow enabled the synthesis of a diverse >1000-member library of compounds in just 2 weeks, with analysis requiring a further 2 weeks, and yielding early structure activity relationships (SAR) derived from 3D structures, as well as compounds with on-scale biophysical affinity.

Results and discussion

A previously identified fragment offers good opportunities for robotic chemistry

Previous high-throughput crystallographic fragment screening efforts in our laboratory yielded multiple hits against PHIP(2). These experiments were used in the assembly of the DSiPoised fragment library[14], and the fragment hits used as the starting point for the SAMPL7 challenge and for this automated chemistry study.[49] We selected fragment F709 (**Fig. 1**), as it not only had unambiguous electron density at the binding site, but also revealed expansion vectors that could be diversely extended by the reaction repertoire of our robotically-enabled array chemistry, and that would sufficiently sample the chemical landscape to provide initial SAR. Overall, these vectors enabled fragment growth to sample 3 different regions of the binding site: i) across the ZA-channel, ii) within the central hydrophobic cavity and iii) opposite the water cavity (**Fig. 1**). F709 interacts with the protein notably via H-bonds with the backbone oxygen of proline 1340 and the side chain oxygen of

serine 1392. The 5-membered furan ring is also positioned perpendicularly to tyrosine 1395 thus promoting a pi-pi interaction and the piperazine ring occupies the central hydrophobic cavity (**Fig. 1**).

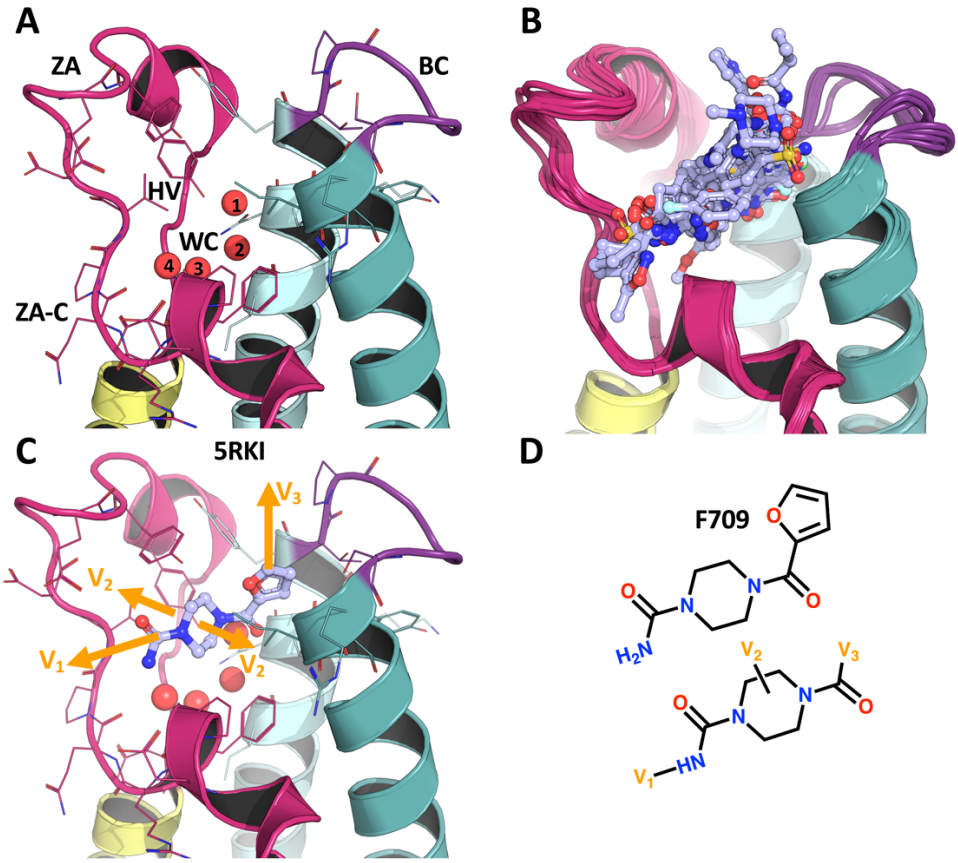


Figure 1: Crystallographic screening of PHIP(2) identified a fragment with good vectors for automated chemistry elaborations. *Panel A shows the structure of PHIP(2) acetylated-lysine binding site with the bromodomain waters displayed in red spheres. Helices Z, A, B and C are shown in pink, yellow, cyan and teal, respectively. ZA (in fuchsia) and BC (in purple) indicate the connecting loops between the corresponding α-helices. ZA-C, WC and HV indicate the ZA-channel, water cavity and hydrophobic void, respectively. Panel B shows all cocrystals identified by our previous fragment screening. Panel C shows the F709- bound cocrystal structure (PDB ID: 5RKI) with elaboration vectors in orange. Panel D shows the chemical structure of F709 and elaboration vectors in orange.*

Robotic chemistry successfully performs complex reactions with minimal resources

To determine feasibility, reactions were initially tested and developed using standard bench chemistry (**Suppl info. 2,3,5,6**) with a specific focus on solvent selection and solubility. Dimethylacetamide (DMA) proved to be a general solvent and was used in all the chemistry developed and executed on the OpenTrons OT-1. Thereafter, the reaction conditions were translated to the OpenTrons OT-1 as a liquid transfer-only process (**Suppl info. 7**). To test the validity of using the OpenTrons OT-1 for chemistry, three iterations (Iters 1.0, 1.1 and 2.0) of single and two-step chemistry for the formation of ureas and amides were explored (**Fig. 2**). 180 targets were attempted via the execution of 296 reactions, with a combined success rate of 83%, as defined by the identification of the expected molecular ion peak, by LCMS in the CRM (**Fig. 2**). Iter 1.1, exploring a "deletion" strategy, by effectively replacing the piperazine central core by a morpholine, and Iter 2.0, exploring an amide linkage vs the urea in Iter 1.0, yielded no structural hits. This demonstrates that we can rapidly scope structural features and "fast fail, fail cheap", quickly redefining our synthetic strategy.[50] Even negative results, obtained quickly, can have insightful impact in these labour-intensive processes (**Fig. 2**).

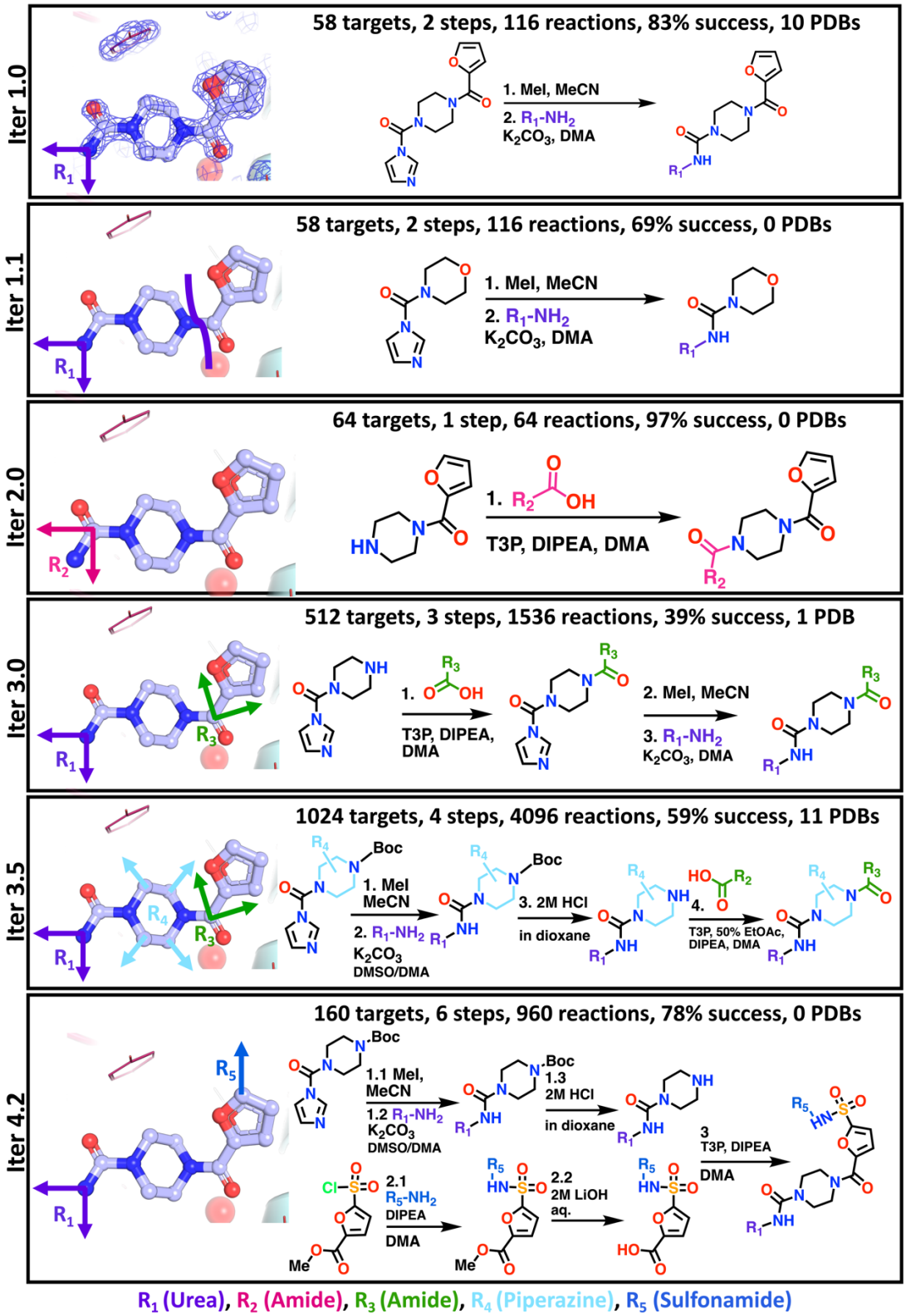


Figure 2: OpenTrons synthesis allows rapid and reliable chemistry for initial SAR scoping via fragment growth. *The different synthetic routes (iterations, Iters) and conditions are shown in each panel. The number of fragment growth targets, steps performed on the robot, total number of reactions performed for each step, success rates and number deposited PDB structures for those reactions are displayed at the top of each box. The success rates are defined by whether a product with expected mass can be detected by our quality control pipeline. The functional groups exploited at the various vectors are colour coded with labels bellow the figures and displayed with respect with the original fragment on the left.*

Building on the single-step iteration, three follow-up iterations were achieved with a total of 1696 multi-step targets, via a combined total of 6592 reactions, and a combined success rate of 57.51% (**Fig. 2**). Iter 4.2 was the most complicated synthesis attempted on the OpenTrons OT-1 with 160 attempts of two by two-step routes being combined towards the final product with an overall 78% success rate (**Fig. 2**). Iter 4.2 showed that it is possible to use relatively complex multi-step chemistry to yield significant fragment elaborations.

Table 1: The automated chemistry workflow reduces time and solvent usage compared to human operations. *Comparison of the estimated time, columns/work-ups, and litres of solvent, between the automated and manual synthetic approaches to the target libraries. (Red = Manual synthesis, Green = Automated Synthesis).*

		Time		***Columns and workups***		***Solvent volume***	
Iteration	Number targets	Per target (days)	Total (days)	Per target	Total	Per target (mL)	Total (L)
1	58	3 / -	174 > 3	1	58 > 1	1000 > 5	58 > 0.29
1.1	58	3 / -	174 > 3	1	58 > 1	1000 > 5	58 > 0.29
2	64	1 / -	64 > 2	1	64 > 1	1000 > 5	64 > 0.32
3	512	5 / -	2560 > 4	2	1024 > 2	2500 > 10	1280 > 5.12
3.5	1024	5 / -	5120 > 4	2	2048 > 2	2500 > 10	2560 > 10.3
4.2	160	8 / -	1280 > 4	3	480 > 3	4500 > 15	720 > 2.4
Total targets	1876	**Total Time**	**25.68y > 20d**	**Total columns & workups**	**3732 >> 10**	**Total solvent volume (L)**	**4740 >> 18.70**
Saving		*Fraction Time*	***0.2%***	*Fraction*	***0.3%***	*Fraction solvent*	***0.4%***

The use of automated synthesis vastly outperforms a more conventional medicinal chemistry approach. The time and solvent requirements are dramatically reduced, and the method circumvents the requirements for purification/column chromatography (**Table 1**).

Based on the manual synthesis and purification of control compounds (**Suppl info. 6**), we estimate that the human time required to manually prepare an analogue varies between 1-8 working days, depending on the number of synthetic steps and purifications (**Table 1**). If the synthesis of the complete library was conducted in a linear fashion this would take about 25 years to complete. Although significant time reductions can be achieved by techniques such as parallel synthesis, bulk preparation of common intermediates, and increasing the number of chemists, we still anticipate that a project of this magnitude would take years to complete. In contrast, our platform was able to complete the work in a few weeks with a single trained chemist.

The use of automated synthesis also avoids the requirement for large-scale aqueous workups and purifications. Synthesis of the library under classical conditions would require up to 3732 chromatography stages, which when combined with the work-up stages results in a total solvent usage of 4740 L. In contrast, the automated protocol avoids the requirement for purification, with an estimated total solvent usage of 5-15 mL (**Table 1**) per analogue. This would require a total of 18.7 L of solvent, an over 250-fold reduction in costly and potentially polluting solvent usage.

A further advantage of the OpenTrons OT-1 is the small footprint. The unit can be easily contained in a single fumehood and automated protocols run for every stage of the synthesis (stock solution preparation, reaction set-up/running, work-up, final sample solution preparation, and QC sample preparation). In contrast, a classical synthetic approach would require substantial levels of equipment and space to optimise the library synthesis and almost continual operation by the chemist/chemists in the project team.

Overall, our workflow minimises the risk of potential human errors and reduces the labour associated with such work. The protocol significantly reduces the resources such as hardware and reagents and the consumption of expansive and polluting solvents required for workups and purifications.

QC analysis and bypassing purification bottleneck

Quality control (QC) of the 1876 attempted targets required analysing 3240 samples, with the post-workup step the most critical (**Table 2**). The analysis was initially performed manually, requiring 8 minutes per sample (preparation, LMCS run, analysis: 1 min, 5 min, 2 min respectively). Evaluation of those samples took approximately 27 working days. In practice, most of this time was concentrated on the analysis of the final, post-work-up, runs with approximately 17 days spent on the analysis (**Table 2**). Manual LCMS analysis is a bottleneck for executing high-throughput chemistry. To address this, *MSCheck*, an open source mass spectrum peak finding and peak scanning pip installable package, with potential to integrate into future automated chemistry software applications, was developed.[51] *MSCheck* utilises .mzML files, the open and generic XML format for mass spectra files converted from vendor files using ProteoWizard.[52,53] *MSCheck* searches for different parent ion matches eg. $M+H^+$, $M+Na^+$ and accepts a tolerance argument where masses matched can be within $M+H^+\pm$ tolerance. Peaks in the total ion count chromatogram are initially identified and the peak's full width at half maximum (FWHM) calculated using Scipy's signal peak analysis algorithms (*find_peaks* and *peak_widths* functions respectively). The mass spectrum patterns, as discrete data points in the total ion chromatogram, are analyzed around the peak above the FWHM height by searching for the sum of the parent mass of the target molecule and ion plus/minus the tolerance set (**Fig. 3**). *MSCheck* also generates a report in an .svg format that summaries the ion matches (**Suppl info. 9.1**).

Retrospective evaluation of iterations 1.0, 2.0, 3.0, 3.5 and 4.2 final showed that *MSCheck* matched with 82.9 % of the human-made analysis of synthesis outcomes (**Table 2**). *MSCheck*, including data preparation and script preparation, took five days to complete without the onerous task of constantly staring at a screen scanning for mass matches (**Table 2**). The workflow's efficiency can be improved by storing reaction details in a database, which would then allow for the quick identification and comparison of LCMS data with anticipated results. Currently, the bulk of the time is consumed in preparing the data files, a process that can be shortened to just a few minutes since the peak matching algorithm itself only takes a short time to execute. Currently, MSCheck identifies only prominent ion matches in the mass spectrum; enhancing it to also detect smaller peaks through a threshold or signal-to-noise ratio would increase analytical precision.

Table 2: MSCheck enables significant time gains for quality control with a good recall against human operations. *Summary of reaction LCMS samples run, and approximate analysis time required for the six iterations of the chemistry run. NC = Not completed, *Average of percentages*

Iteration	Number LCMS samples pre-workup	Number LCMS samples post-workup	Total LCMS samples per iteration	*MSCheck* match with human analysis (%)
1.0	58	58	116	81.0
1.1	58	58	116	NC
2.0	64	64	128	96.8
3.0	1024	1024	2048	78.1
3.5	0	512	512	74.4
4.2	0	320	320	84.4
Total	1204	2036	3240	82.9*
Estimated analysis time (days)	10	17	27	5

X-ray crystallography of crude reaction mixtures can be performed at high-throughput levels

High throughput X-ray crystallography requires that protein crystals must diffract at a good resolution and resist fracture during soaking. Previous efforts demonstrated that crude reaction mixtures can be prepared and soaked onto protein crystals and that reaction products can be resolved.[2] Here, crude reaction mixture preparation was also included in our robotic framework where the products were concentrated in the organic phase and solvent exchanged. Due to the intrinsic affinity of bromodomains for DMSO (dimethyl sulfoxide) and/ or DMA these solvents were replaced with ethylene glycol.

Reaction products were measured in 1077 out of 1876 syntheses by our quality control protocol, representing a 57.41% success rate across all iterations. We soaked the crude reaction mixtures onto protein crystals at the XChem fragment screening facility at Diamond Light Source (**Fig. 3**). This yielded 969 usable X-ray diffraction datasets for the crude reaction mixtures with successful syntheses, which translates into a crystal deterioration rate upon soaking of crude reaction mixtures of only 8.73% (**Fig. 3**). This demonstrates that our new automated synthesis workflow pairs well with the pre-existing downstream high-throughput crystallographic pipeline and can generate a high volume of usable data in a relatively small amount of time while maintaining crystal integrity. Furthermore, the quality control step supplements the workflow by allowing tractability of reaction outcomes facilitating the tracking of which electron density maps could be expected to have a bound reaction product.

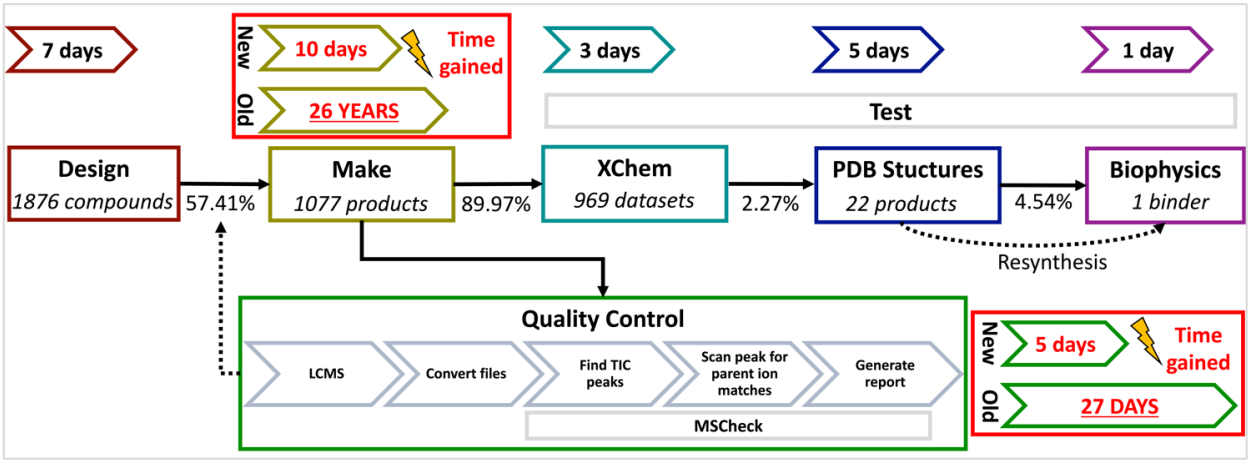


Figure 3: Automated chemistry and QC fit with the XChem workflow and reduces human labour. *Each step of the process is represented in a coloured box. The molecular and reaction designs were made with the Python package RDKit; preparation of CRMs on an OpenTrons OT-1; CRM QC via LC-MS and analysed with* MSCheck*; the crystals were prepared and resolved at the XChem, the product hits were identified from electron density maps with Coot and PanDDa; and the hits were confirmed via Creoptix and an alpha-screening assay. The quality control step is parallel to the main workflow because crystals were soaked with all crude reaction mixtures while the LC-MS outcome determined the number of successful reactions, as indicated by the dotted arrow. The new and old timings (in working days or weeks) are shown as arrow boxes with estimated values within them. New and old refer to the timings achieved by our protocol versus the estimated time it would take a human to process an equal number of compounds. The validation required ordering the pure compounds from Enamine. Success rates are shown above the arrow connecting the steps and are indicative of the number of successful outcomes for a given step over the number of successful outcomes of the previous step.*

Baker et al., (2020) processed 83 reaction mixtures in triplicate whereas our workflow performed a single crystallographic readout per compound. They demonstrated that performing triplicate experiments recovers more product-bound structures, thus it is likely that some binders were missed here.[2] The aim here was, however, to increase the number of synthesised follow-ups using cheap robotics prioritising volume over

quality whilst reducing times. Here a 22-fold gain in processing compounds or 7-fold gain in performing experiments compared to Baker et al., (2020).[2] Design of larger arrays are also possible,[37] but we purposely kept the size of the iterations at levels that are manageable for downstream XChem processing, which involves human interventions at certain steps such as crystal fishing.

Our crystallographic experiments resolved a total of 29 unique compound-bound structures, which included 7 starting materials and 22 reaction products (**Fig. 4**). When looking at target-bound reaction products from successful syntheses and usable X-ray diffraction data after soaking, we achieved a hit rate of 2.27% (22/969) (**Fig. 3**). This is a lower hit rate than regular fragment screening with previous crystallographic XChem fragment screening against this crystal form yielded a 6.51% hit rate with similar soaking conditions.[49] The low hit rate is also likely due to inadequate compound designs as this exercise was mostly driven by the available chemistry around the starting fragment. We anticipate that better compound library designs will increase the hit rate of future experiments.

Structural analysis rationalises the binding landscape and resolves an unexpected pose

All starting materials bind at the same location between helices B and C (**Fig. 1**) and display a conserved binding mode where they interact with the protein via 2 pi-pi stackings and hydrogen bonding (**Fig. 1**). All products bound cocrystals maintain a 4-formylpiperazine-1-carboxamide scaffold, where the piperazine moiety occupies the central hydrophobic cavity. Poses for the products also retained the original furan or the analogous thiophene or pyrrole rings, with some hits having either a 2-chlorine or methyl furan.

The 19 unique products observed in crystals (**Fig. 4**) have the same binding mode defined by the starting fragment, referred as the lateral pose (**Fig. 1**). Iter 1.0, the exploration of the urea vector, was the only single-step iteration to yield ten follow-up crystal hits on the XChem platform (**Fig. 2**). Three of the binding events were from overlapping compounds from Iter 1.0. The importance of the furan ring for binding was highlighted when no structural hits were found for Iter 1.1's urea vector exploration where the furan ring was not included (**Fig. 2**). Iter 2.0 highlighted the importance of retaining the urea group of the initial fragments; no hits were found unless retained . Structurally, the urea group forms a hydrogen bond with the proline 1340 backbone oxygen (**Fig. 4**).

Unexpectedly, 3 products were resolved in an alternative orientation, termed the "diving" pose, relative to the original fragment F709 (**Fig. 1**) as shown in **Fig. 4**, demonstrating the effect of modifying the piperazine ring in Iter 3.5 (**Fig. 2**). These poses rotated by about 90° around the piperazine core (with respect with the original pose) and so that their 5-membered ring displaces waters 2 to 4 while the nearby amide displaces water 1 of the network (**Fig. 1**). There, the 5-membered rings seem to bind mostly via hydrophobic interactions with the amino acids comprising the water cavity. In addition, the cis conformation of the two amides was observed in the divers, while the lateral products displayed a trans conformation. A relatively large protein conformational change is paired with this novel diving binding pose. The ZA-loop adopts a generally more relaxed conformation, resulting in a more voluminous binding site that accommodates the flipped products, thus illustrating that some level of conformational motion is allowed by our crystal system (**Fig. 4**).

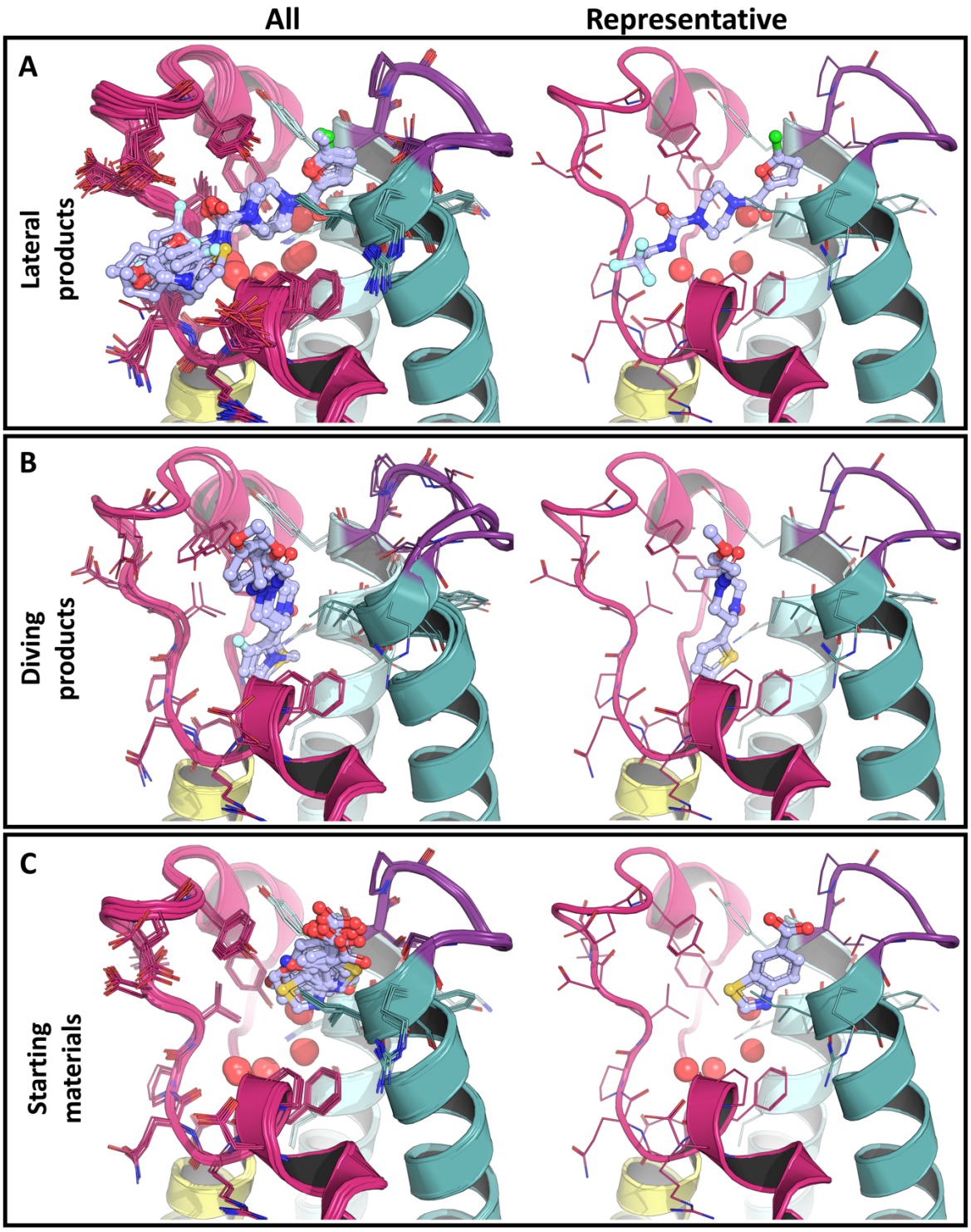


Figure 4: XChem screening of CRMs yielded bound product structures, with conserved and non-conserved binding poses, and starting materials. *The first and second panels show structures of the laterally (A) and diving (B) bound products, respectively. The third panel shows structures of the reaction starting material-bound (C) proteins. Each panel is subdivided into 2 columns. The first (All) aggregates all bound structures while the second (representative) show a single binder that has a binding mode representative of the others. The representative binders were arbitrarily selected to illustrate the corresponding binding mode.*

A striking feature of the diving products is that they all have methyl-substituted piperazine ring. When looking at their closest lateral neighbours based on chemical similarity distance, it seems that this alkyl substitution alone is required to change binding orientation (**Fig. 5**). The relatively low incidence of diving relative to lateral binders also indicates that a high number of experiments are needed to resolve this unexpected pose.

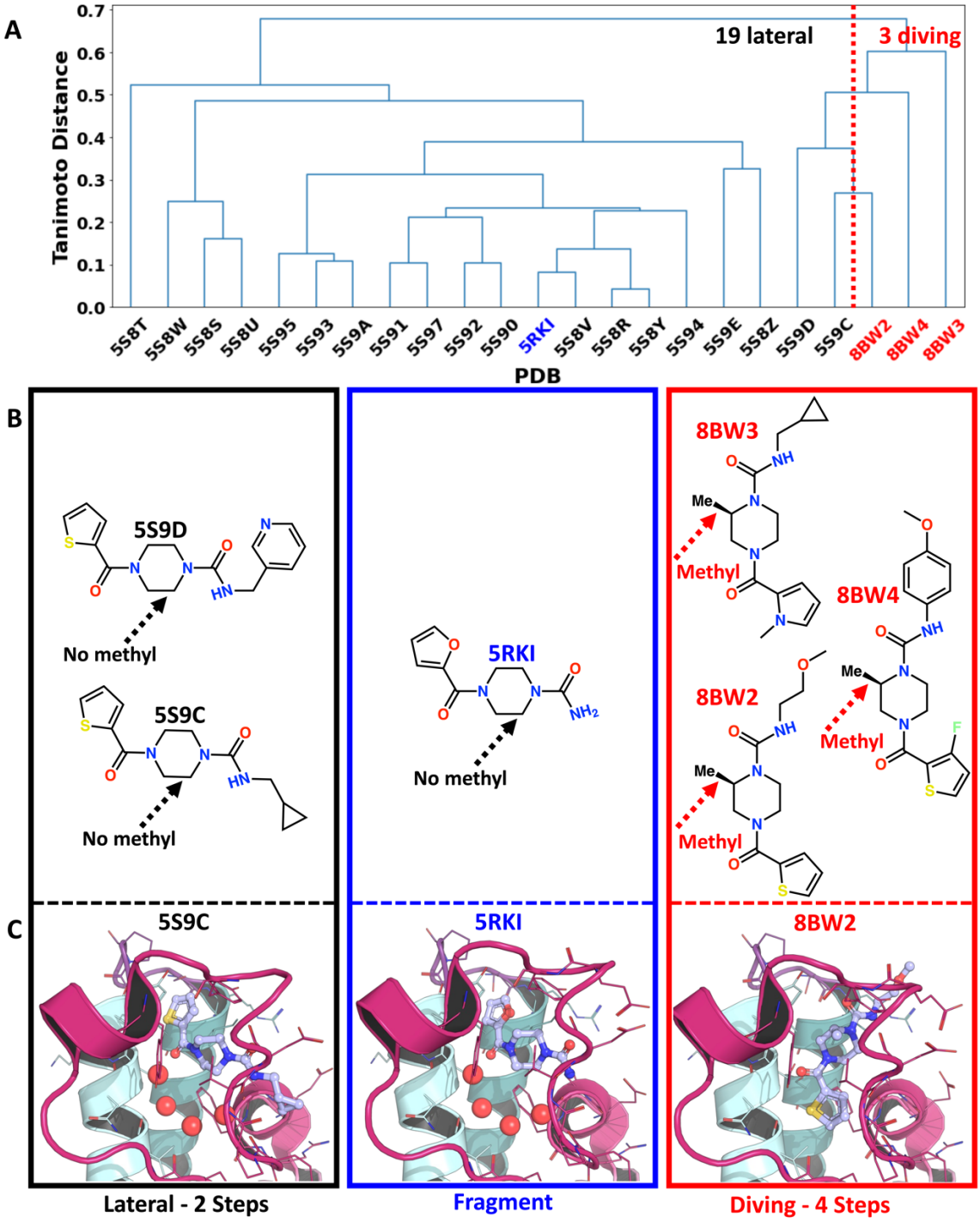


Figure 5: Unexpected diving binding pose appears to be triggered by the addition of a methyl to the core piperazine ring*. A dendrogram showing the chemical fingerprint (Tanimoto) similarity between bound reaction products is shown on top (A). The compounds are labelled by PDB accession IDs. Information related to starting fragment, lateral and diving compounds is highlighted in blue, black and red respectively. The 2D structures for the fragment (B), the divers' closest neighbour (laterals) and divers with a representative 3D structure (C) are shown in the first, second and third columns, respectively. The number of synthetic steps required to obtain those compounds is shown under the columns for lateral and diving bound reaction products.*

The synthesis of the 3 diving binder molecules made use of racemic building blocks (**Fig. 2**), resulting in CRMs containing a mixture of enantiomers and other by-products in unknown ratios. The quality control protocol, however, identified that products with the expected mass were present in the mixture. PanDDA event maps were used instead of traditional electron density maps to fit the compounds. The position of the methyl group around the piperazine ring was ambiguous and required further inspection (**Fig. 6**). Each compound has two stereoisomers with four possible configurations, two on each side of the piperazine.

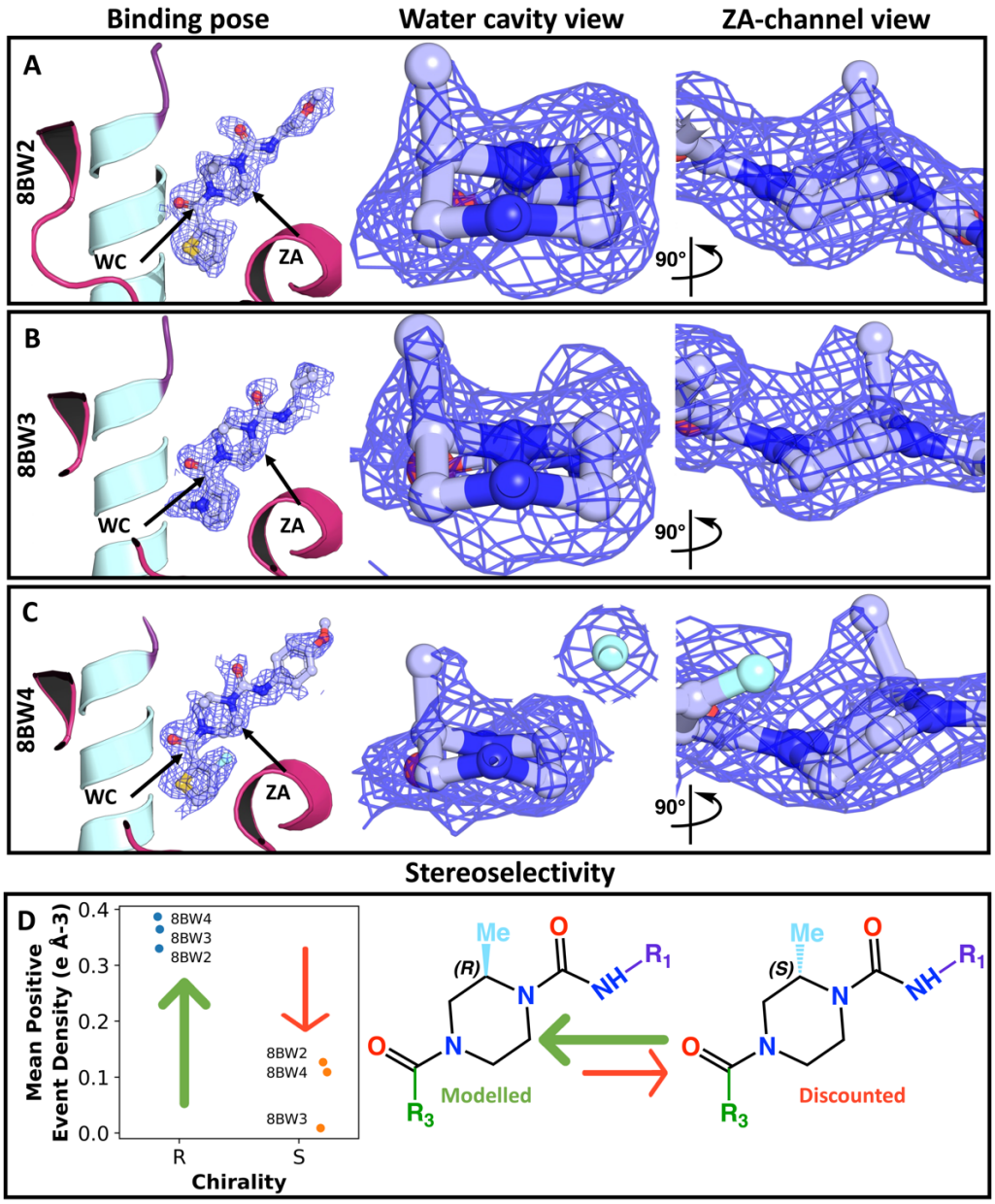


Figure 6: PanDDA event maps of racemate crude reaction mixtures indicate that the PHIP2 binding site is stereoselective for products containing an R-methylpiperazine moiety. *Pannels A to C show the PanDDA event maps for the divers. The first column, labelled as the binding pose, presents the overall event density for binding. The second column shows the density of the methylpiperazine moiety as observed from the water cavity (WC). The third column depicts the same moiety's density from the ZA-channel (ZA) perspective, with viewpoints marked by arrows in the binding pose column. Pannel D demonstrates the two potential chiralities, with the green arrow specifying the chirality modelled in the binding site. Pannel D also illustrate the mean positive density values computed along the carbon-carbon bond of piperazine-methyl. These values are aligned with the PanDDA event map for both R and S stereoisomers after fitting and refinement, ultimately concluding with the presentation of the selected final structure.*

These maps showed a consistent protrusion at the same location, suggesting the presence of an additional group (**Fig. 6**). Other possible locations do not have a similar protrusion and positioning the group there would result in a clash with the protein and/or the compound itself. At the identified position, the methyl group interact with the binding site through hydrophobic interactions (**Fig. 4**).

Relative to the average noise, as measured by the standard deviation of the difference between the event map and mean map in the immediate vicinity of the protein model, it is not possible confidently say any single difference between the R and S isomers is significant. A paired sample T-test however reveals that when taken together the three repeat observations show that there is stronger electron density support for the R isomer at the 1% level.

One driver for the pose change might be entropic gain via water displacement[54]. Indeed previous modelling efforts suggested the PHIP(2) water network to be relatively unstable compared to other bromodomains, with a positive Gibbs binding free energy[55]. Moreover, "magic methyls" are able to significantly alter binding affinity.[56] Overall, these results indicate that the protein binding site defined by this crystal system is stereoselective for the methyl group oriented in an "up" (R) fashion and located on the same side as the compound's cis-oriented amide group oxygens (**Fig. 6**). To fully resolve stereo-selectivity, pure enantiomer soaking is required. The aim here was to rapidly generate crystallographic information from crude reaction mixtures and thus, experiments with pure enantiomers were beyond the scope of this study.

Overall, the crystallographic results identify which vectors can be elaborated, with what type of chemical group and to what extent, thereby providing an estimate of the crystallographic SAR landscape around fragment F709 (**Fig. 7**). Several modifications, including replacing the piperazine a with diazepane moiety or dimethyl-piperazine, replacing the furan with a 6-membered ring or tri-heterocyclic ring, and adding a sulphonamide to the furan, in Iter 4.2 (**Fig. 2**), consistently resulted in non-binding events despite positive quality control confirming that their synthesis has been successful. We also recognize that there might be a high proportion of false negatives as soaking of crude reaction mixtures and high-throughput methodologies, in general, maximises data quantity over quality. However, our sampling of the chemical space was thorough enough to cause a change of binding pose paired with protein conformational changes that would not have been identified by classical or guided docking methods. It is also interesting to note that such chemical exploration would not have been possible without the use of multi-step chemistry thus highlighting the importance of those more laborious routes.

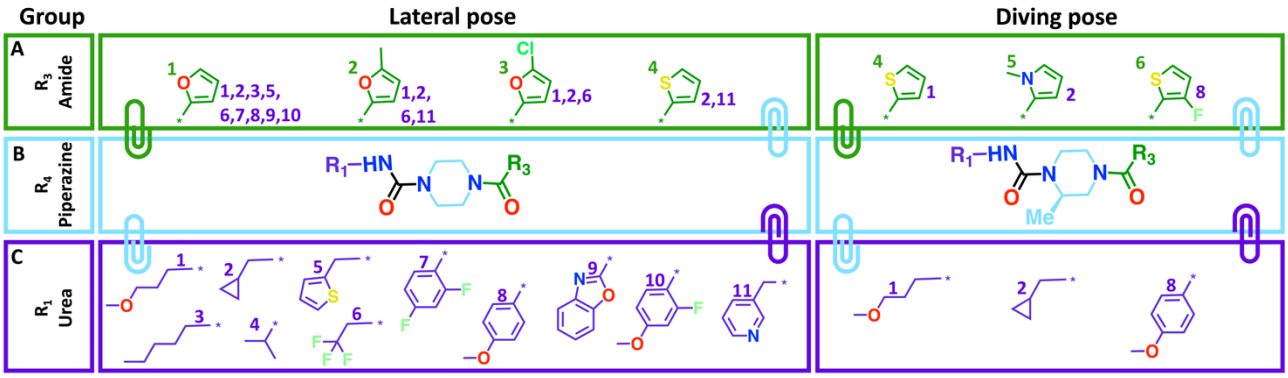


Figure 7: Combinatorial chemistry around vectors identifies R-groups combinations leading to crystallographic binding and change in binding pose. *The different groups causing crystal binding for each pose are shown in the panels. The numbers next to the group for the amide (A) and urea expansions (C) highlight the observed combination, around the central piperazine scaffold (B), between those groups.*

Crystallographic hit exhibits on-scale biochemically and biophysically binding

The starting template fragment, F709 did not yield any detectable signal in solution assays and elaborating fragments aims at increasing the compound's affinity for the target. Hence, once identified, crystallographic hits must be confirmed via orthogonal methods. For this purpose, we synthesised in-house, or ordered the pure and independently synthesised compounds via Enamine Ltd. These were then evaluated with Amplified Luminescence Proximity Homogeneous Assay (AlphaScreen™)[57] testing for biochemical activity. A time-resolved grating-coupled interferometry-based biosensor assay, using a pulsed injection scheme (waveRAPID®)[58], was used to estimate affinity and interaction kinetic rate constants (**Fig. 8**).

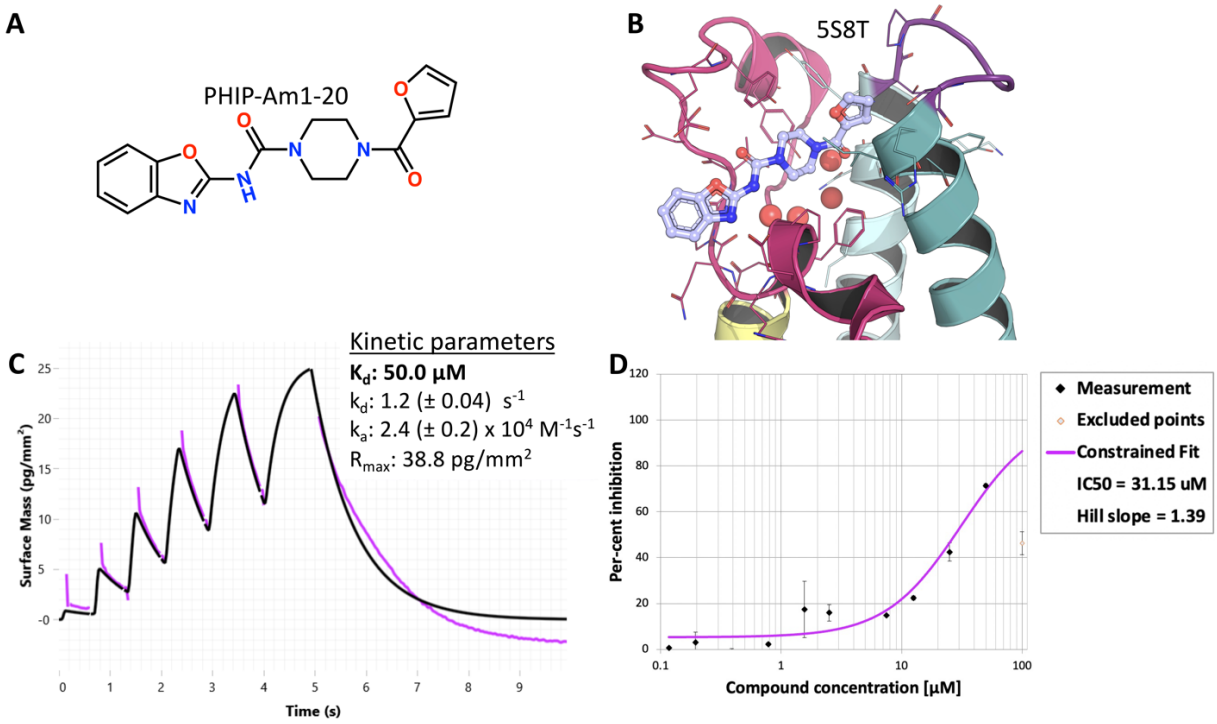


Figure 8: Structural and assay binding of reaction product PHIP-AM1-20. *Panels A and B show the chemical structure and crystallographic binding pose of PHIP-AM1-20, respectively. The PDB code of the bound complex is written above B. Panel C shows kinetic parameters determined from the interaction kinetic curves by global fitting using a 1:1 interaction kinetic model modelled with the WAVEcontrol software. Panel D shows the dose response from the AlphaScreening assay. For both C and D, raw data in black and fitted binding curves in purple.*

Compound PHIP-AM1-20, product of Iter 1 (**Fig. 2**), had a measurable effect in both assays with a K_d and a IC50 of 50.03 µM and 31.15 µM corresponding to ligand efficacies of 2.00 and 1.25 µM/heavy atom for the kinetic analysis and alpha-screen, respectively (**Fig. 8**). The binding event appears to be a reversible 1-step 1:1 interaction. This represents a 2- to 3-fold increase in binding affinity when compared to the best binder obtained by Cox et al., (2016) that were elaborations around the first DSI-poised fragment binders identified against PHIP(2).[14] Overall, at scale potency was achieved starting from an undetectable fragment.

Conclusion and future directions

This conceptual fragment to drug-like molecule growth exercise (**Fig. 1**) has demonstrated that it is possible to save time, solvent usage, and many synthetic, extraction, work-up and purification bottlenecks (**Table 1**) by combining crude array synthesis and X-ray crystallographic structural determination of the resulting molecules (**Fig. 3**). A crude fragment screen of the DSiP library at Diamond vs PHIP(2), led to a piperazinyl hit F709 (**Fig. 1**). Consequently, we have developed several thousand multi-step, robotic-driven reactions with minimum automated work-up using F709 as the starting fragment (**Fig. 2**). To reduce crude compound analysis time (**Table 2**), a semi-automated LCMS analysis tool *MSCheck* has been developed to supersede the previous cumbersome manually evaluated LCMS chromatogram methods (**Fig. 3**).[51] Crude reaction mixtures were submitted to XChem for X-ray analysis at high-throughput levels and without compromising crystal integrity thus, bypassing purification steps (**Fig. 3**). This resulted in the identification of crystal binders (**Fig. 4**) which enabled us to map the crystallographic SAR landscape around PHIP(2) defined by our enumeration. We also resolved a previously unidentified binding pose that displaces all 4 bromodomain water (**Fig. 5**) and found one of the crystallographic binders to be active in a biochemical assay (**Fig. 8**).

This binding-site purification of actives (B-SPA) technique, which has shown promise in generate Structure-Activity Relationship (SAR) information economically from large scale crystallographic readouts of fragment elaborations in crude reaction mixtures and identifying reaction products with on-scale activity. Our findings also suggest that B-SPA could be a valuable tool in further streamlining drug discovery workflows by passing costly and polluting purification steps.

Our method has proven to be effective at generating large amounts of crystallographic data but lacks an equally high-throughput and automated means of validating those hits via assays. Here, we still relied upon the purchase of pure compounds (**Fig. 3**), or resynthesis in-house using conventional chemistry, but efforts have been made in measuring K_d from crude reaction mixtures, which may combine efficiently with the novel pulsed injection schemes implemented for the time-resolved kinetic analysis using grating-coupled interferometry. Automated methods will also be needed to systematically analyse and rationalise the data resulting from these increasingly large high-throughput crystallographic screenings.

This case study serves as an exemplar on how to generate structural data quickly with minimal bottlenecks, caveated by the fact that improved binding interactions are not necessarily commensurate with increased activity nor are they a substitute for biological activity derived from an assay. Future efforts and workflows will benefit from regular STOP/GO decisions informed by holistic structural and biological data input to modify the direction of the research and maximise the chances of success.

Acknowledgment

The authors also thank the Diamond Light Source for beam time (proposal mx19301) and the staff of beamlines I03 and I04-1 for assistance with crystal testing and data collection.

Funding

EPSRC (EP/P026990/1) is thanked for funding (JS, SHH, FvD). The Diamond Light Source is also thanked for funding (COL0108 - HG). The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada, Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD grant no. 115766], Janssen, Merck KgaA Darmstadt Germany, MSD, Novartis Pharma AG, Ontario Ministry of Economic Development and Innovation, Pfizer, FAPDF, CAPES, CNPq, São Paulo Research Foundation-FAPESP, Takeda, and Wellcome [106169/ZZ14/Z]. Wellcome Institutional Strategic Support Fund (ISSF) is thanked for funding this work at Sussex (JS, SHH).

Data access and availabilities

Supporting data uploaded to Zendo (10.5281/zenodo.7586212) and includes: a summary of X-ray and LCMS results for the reactions executed on the OpenTrons, output reports and summaries from MSCheck (semi-automated LCMS analyzer tool) and the Python scripts used to execute single and multistep chemistry on the OpenTrons.

References

(1) de Esch, I. J. P.; Erlanson, D. A.; Jahnke, W.; Johnson, C. N.; Walsh, L. Fragment-to-Lead Medicinal Chemistry Publications in 2020. *J. Med. Chem.* **2022**, *65* (1), 84–99. https://doi.org/10.1021/acs.jmedchem.1c01803.

(2) Baker, L. M.; Aimon, A.; Murray, J. B.; Surgenor, A. E.; Matassova, N.; Roughley, S. D.; Collins, P. M.; Krojer, T.; von Delft, F.; Hubbard, R. E. Rapid Optimisation of Fragments and Hits to Lead Compounds from Screening of Crude Reaction Mixtures. *Commun Chem* **2020**, *3* (1), 1–11. https://doi.org/10.1038/s42004-020-00367-0.

(3) Kim, A.; Cohen, M. S. The Discovery of Vemurafenib for the Treatment of BRAF-Mutated Metastatic Melanoma. *Expert Opinion on Drug Discovery* **2016**, *11* (9), 907–916. https://doi.org/10.1080/17460441.2016.1201057.

(4) Souers, A. J.; Leverson, J. D.; Boghaert, E. R.; Ackler, S. L.; Catron, N. D.; Chen, J.; Dayton, B. D.; Ding, H.; Enschede, S. H.; Fairbrother, W. J.; Huang, D. C. S.; Hymowitz, S. G.; Jin, S.; Khaw, S. L.; Kovar, P. J.; Lam, L. T.; Lee, J.; Maecker, H. L.; Marsh, K. C.; Mason, K. D.; Mitten, M. J.; Nimmer, P. M.; Oleksijew, A.; Park, C. H.; Park, C.-M.; Phillips, D. C.; Roberts, A. W.; Sampath, D.; Seymour, J. F.; Smith, M. L.; Sullivan, G. M.; Tahir, S. K.; Tse, C.; Wendt, M. D.; Xiao, Y.; Xue, J. C.; Zhang, H.; Humerickhouse, R. A.; Rosenberg, S. H.; Elmore, S. W. ABT-199, a Potent and Selective BCL-2 Inhibitor, Achieves Antitumor Activity While Sparing Platelets. *Nat Med* **2013**, *19* (2), 202–208. https://doi.org/10.1038/nm.3048.

(5) Murray, C. W.; Newell, D. R.; Angibaud, P. A Successful Collaboration between Academia, Biotech and Pharma Led to Discovery of Erdafitinib, a Selective FGFR Inhibitor Recently Approved by the FDA. *Med. Chem. Commun.* **2019**, *10* (9), 1509–1511. https://doi.org/10.1039/C9MD90044F.

(6) Lamb, Y. N. Pexidartinib: First Approval. *Drugs* **2019**, *79* (16), 1805–1812. https://doi.org/10.1007/s40265-019-01210-0.

(7) Skoulidis, F.; Li, B. T.; Dy, G. K.; Price, T. J.; Falchook, G. S.; Wolf, J.; Italiano, A.; Schuler, M.; Borghaei, H.; Barlesi, F.; Kato, T.; Curioni-Fontecedro, A.; Sacher, A.; Spira, A.; Ramalingam, S. S.; Takahashi, T.; Besse, B.; Anderson, A.; Ang, A.; Tran, Q.; Mather, O.; Henary, H.; Ngarmchamnanrith, G.; Friberg, G.;

Velcheti, V.; Govindan, R. Sotorasib for Lung Cancers with *KRAS* p.G12C Mutation. *N Engl J Med* **2021**, *384* (25), 2371–2381. https://doi.org/10.1056/NEJMoa2103695.

(8) *Novartis receives FDA Breakthrough Therapy designations for investigational STAMP inhibitor asciminib (ABL001) in chronic myeloid leukemia*. Novartis. https://www.novartis.com/news/media-releases/novartis-receives-fda-breakthrough-therapy-designations-investigational-stamp-inhibitor-asciminib-abl001-chronic-myeloid-leukemia (accessed 2022-07-01).

(9) Ostrem, J. M.; Peters, U.; Sos, M. L.; Wells, J. A.; Shokat, K. M. K-Ras(G12C) Inhibitors Allosterically Control GTP Affinity and Effector Interactions. *Nature* **2013**, *503* (7477), 548–551. https://doi.org/10.1038/nature12796.

(10) Balamurugan, R.; Dekker, F. J.; Waldmann, H. Design of Compound Libraries Based on Natural Product Scaffolds and Protein Structure Similarity Clustering (PSSC). *Mol. BioSyst.* **2005**, *1* (1), 36–45. https://doi.org/10.1039/B503623B.

(11) Laraia, L.; Waldmann, H. Natural Product Inspired Compound Collections: Evolutionary Principle, Chemical Synthesis, Phenotypic Screening, and Target Identification. *Drug Discovery Today: Technologies* **2017**, *23*, 75–82. https://doi.org/10.1016/j.ddtec.2017.03.003.

(12) Pahl, A.; Waldmann, H.; Kumar, K. Exploring Natural Product Fragments for Drug and Probe Discovery. *CHIMIA* **2017**, *71* (10). https://doi.org/10.2533/chimia.2017.653.

(13) Karageorgis, G.; Foley, D. J.; Laraia, L.; Waldmann, H. Principle and Design of Pseudo-Natural Products. *Nat. Chem.* **2020**, *12* (3), 227–235. https://doi.org/10.1038/s41557-019-0411-x.

(14) Cox, O. B.; Krojer, T.; Collins, P.; Monteiro, O.; Talon, R.; Bradley, A.; Fedorov, O.; Amin, J.; Marsden, B. D.; Spencer, J.; Delft, F. von; Brennan, P. E. A Poised Fragment Library Enables Rapid Synthetic Expansion Yielding the First Reported Inhibitors of PHIP(2), an Atypical Bromodomain. *Chem. Sci.* **2016**, *7* (3), 2322–2330. https://doi.org/10.1039/C5SC03115J.

(15) Kidd, S. L.; Fowler, E.; Reinhardt, T.; Compton, T.; Mateu, N.; Newman, H.; Bellini, D.; Talon, R.; McLoughlin, J.; Krojer, T.; Aimon, A.; Bradley, A.; Fairhead, M.; Brear, P.; Díaz-Sáez, L.; McAuley, K.; Sore, H. F.; Madin, A.; O'Donovan, D. H.; Huber, K. V. M.; Hyvönen, M.; Delft, F. von; Dowson, C. G.; Spring, D. R. Demonstration of the Utility of DOS-Derived Fragment Libraries for Rapid Hit Derivatisation in a Multidirectional Fashion. *Chem. Sci.* **2020**, *11* (39), 10792–10801. https://doi.org/10.1039/D0SC01232G.

(16) Davis, A.; Ward, S. E. *The Handbook of Medicinal Chemistry: Principles and Practice*; Royal Society of Chemistry, 2014.

(17) Beberg, A. L.; Ensign, D. L.; Jayachandran, G.; Khaliq, S.; Pande, V. S. Folding@home: Lessons from Eight Years of Volunteer Distributed Computing. In *2009 IEEE International Symposium on Parallel & Distributed Processing*; IEEE: Rome, Italy, 2009; pp 1–8. https://doi.org/10.1109/IPDPS.2009.5160922.

(18) Bissaro, M.; Sturlese, M.; Moro, S. The Rise of Molecular Simulations in Fragment-Based Drug Design (FBDD): An Overview. *Drug Discovery Today* **2020**, *25* (9), 1693–1701. https://doi.org/10.1016/j.drudis.2020.06.023.

(19) Wang, Y.; Haight, I.; Gupta, R.; Vasudevan, A. What Is in Our Kit? An Analysis of Building Blocks Used in Medicinal Chemistry Parallel Libraries. *J. Med. Chem.* **2021**, *64* (23), 17115–17122. https://doi.org/10.1021/acs.jmedchem.1c01139.

(20) Klingler, F.-M.; Gastreich, M.; Grygorenko, O.; Savych, O.; Borysko, P.; Griniukova, A.; Gubina, K.; Lemmen, C.; Moroz, Y. SAR by Space: Enriching Hit Sets from the Chemical Space. *Molecules* **2019**, *24* (17), 3096. https://doi.org/10.3390/molecules24173096.

(21) Kirsch, P.; Hartman, A. M.; Hirsch, A. K. H.; Empting, M. Concepts and Core Principles of Fragment-Based Drug Design. *Molecules* **2019**, *24* (23), 4309. https://doi.org/10.3390/molecules24234309.

(22) Sundberg, T. B.; Liang, Y.; Wu, H.; Choi, H. G.; Kim, N. D.; Sim, T.; Johannessen, L.; Petrone, A.; Khor, B.; Graham, D. B.; Latorre, I. J.; Phillips, A. J.; Schreiber, S. L.; Perez, J.; Shamji, A. F.; Gray, N. S.; Xavier, R. J. Development of Chemical Probes for Investigation of Salt-Inducible Kinase Function *in Vivo*. *ACS Chem. Biol.* **2016**, *11* (8), 2105–2111. https://doi.org/10.1021/acschembio.6b00217.

(23) Collins, P. M.; Ng, J. T.; Talon, R.; Nekrosiute, K.; Krojer, T.; Douangamath, A.; Brandao-Neto, J.; Wright, N.; Pearce, N. M.; von Delft, F. Gentle, Fast and Effective Crystal Soaking by Acoustic Dispensing. *Acta Crystallogr D Struct Biol* **2017**, *73* (3), 246–255. https://doi.org/10.1107/S205979831700331X.
(24) Douangamath, A.; Powell, A.; Fearon, D.; Collins, P. M.; Talon, R.; Krojer, T.; Skyner, R.; Brandao-Neto, J.; Dunnett, L.; Dias, A.; Aimon, A.; Pearce, N. M.; Wild, C.; Gorrie-Stone, T.; von Delft, F. Achieving Efficient Fragment Screening at XChem Facility at Diamond Light Source. *JoVE* **2021**, No. 171, 62414. https://doi.org/10.3791/62414.
(25) Resnick, E.; Bradley, A.; Gan, J.; Douangamath, A.; Krojer, T.; Sethi, R.; Geurink, P. P.; Aimon, A.; Amitai, G.; Bellini, D.; Bennett, J.; Fairhead, M.; Fedorov, O.; Gabizon, R.; Gan, J.; Guo, J.; Plotnikov, A.; Reznik, N.; Ruda, G. F.; Díaz-Sáez, L.; Straub, V. M.; Szommer, T.; Velupillai, S.; Zaidman, D.; Zhang, Y.; Coker, A. R.; Dowson, C. G.; Barr, H. M.; Wang, C.; Huber, K. V. M.; Brennan, P. E.; Ovaa, H.; Von Delft, F.; London, N. Rapid Covalent-Probe Discovery by Electrophile-Fragment Screening. *J. Am. Chem. Soc.* **2019**, *141* (22), 8951–8968. https://doi.org/10.1021/jacs.9b02822.
(26) Boby, M. L.; Fearon, D.; Ferla, M.; Filep, M.; Koekemoer, L.; Robinson, M. C.; The COVID Moonshot Consortium; Chodera, J. D.; Lee, A. A.; London, N.; Von Delft, A.; Von Delft, F. *Open Science Discovery of Potent Non-Covalent SARS-CoV-2 Main Protease Inhibitors*; preprint; Biochemistry, 2020. https://doi.org/10.1101/2020.10.29.339317.
(27) Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.; Owen, C. D.; Resnick, E.; Strain-Damerell, C.; Aimon, A.; Ábrányi-Balogh, P.; Brandão-Neto, J.; Carbery, A.; Davison, G.; Dias, A.; Downes, T. D.; Dunnett, L.; Fairhead, M.; Firth, J. D.; Jones, S. P.; Keeley, A.; Keserü, G. M.; Klein, H. F.; Martin, M. P.; Noble, M. E. M.; O'Brien, P.; Powell, A.; Reddi, R. N.; Skyner, R.; Snee, M.; Waring, M. J.; Wild, C.; London, N.; Von Delft, F.; Walsh, M. A. Crystallographic and Electrophilic Fragment Screening of the SARS-CoV-2 Main Protease. *Nat Commun* **2020**, *11* (1), 5047. https://doi.org/10.1038/s41467-020-18709-w.
(28) Wills, S.; Sanchez-Garcia, R.; Dudgeon, T.; Roughley, S. D.; Merritt, A.; Hubbard, R. E.; Davidson, J.; Von Delft, F.; Deane, C. M. Fragment Merging Using a Graph Database Samples Different Catalogue Space than Similarity Search. *J. Chem. Inf. Model.* **2023**, *63* (11), 3423–3437. https://doi.org/10.1021/acs.jcim.3c00276.
(29) Müller, J.; Klein, R.; Tarkhanova, O.; Gryniukova, A.; Borysko, P.; Merkl, S.; Ruf, M.; Neumann, A.; Gastreich, M.; Moroz, Y. S.; Klebe, G.; Glinca, S. Magnet for the Needle in Haystack: "Crystal Structure First" Fragment Hits Unlock Active Chemical Matter Using Targeted Exploration of Vast Chemical Spaces. *J. Med. Chem.* **2022**, *65* (23), 15663–15678. https://doi.org/10.1021/acs.jmedchem.2c00813.
(30) Piticchio, S. G.; Martínez-Cartró, M.; Scaffidi, S.; Rachman, M.; Rodriguez-Arevalo, S.; Sanchez-Arfelis, A.; Escolano, C.; Picaud, S.; Krojer, T.; Filippakopoulos, P.; Von Delft, F.; Galdeano, C.; Barril, X. Discovery of Novel BRD4 Ligand Scaffolds by Automated Navigation of the Fragment Chemical Space. *J. Med. Chem.* **2021**, *64* (24), 17887–17900. https://doi.org/10.1021/acs.jmedchem.1c01108.
(31) Wesolowski, S. S.; Brown, D. G. The Strategies and Politics of Successful Design, Make, Test, and Analyze (DMTA) Cycles in Lead Generation. In *Methods and Principles in Medicinal Chemistry*; Holenz, J., Ed.; Wiley, 2016; pp 487–512. https://doi.org/10.1002/9783527677047.ch17.
(32) Johnson, C. N.; Erlanson, D. A.; Jahnke, W.; Mortenson, P. N.; Rees, D. C. Fragment-to-Lead Medicinal Chemistry Publications in 2016. *J. Med. Chem.* **2018**, *61* (5), 1774–1784. https://doi.org/10.1021/acs.jmedchem.7b01298.
(33) Chow, S.; Liver, S.; Nelson, A. Streamlining Bioactive Molecular Discovery through Integration and Automation. *Nat Rev Chem* **2018**, *2* (8), 174–183. https://doi.org/10.1038/s41570-018-0025-7.
(34) Bentley, M. R.; Ilyichova, O. V.; Wang, G.; Williams, M. L.; Sharma, G.; Alwan, W. S.; Whitehouse, R. L.; Mohanty, B.; Scammells, P. J.; Heras, B.; Martin, J. L.; Totsika, M.; Capuano, B.; Doak, B. C.; Scanlon, M. J. Rapid Elaboration of Fragments into Leads by X-Ray Crystallographic Screening of Parallel Chemical Libraries (REFiL $_X$). *J. Med. Chem.* **2020**, *63* (13), 6863–6875. https://doi.org/10.1021/acs.jmedchem.0c00111.
(35) Thomas, R. P.; Heap, R. E.; Zappacosta, F.; Grant, E. K.; Pogány, P.; Besley, S.; Fallon, D. J.; Hann, M. M.; House, D.; Tomkinson, N. C. O.; Bush, J. T. A Direct-to-Biology High-Throughput Chemistry Approach to

Reactive Fragment Screening. *Chem. Sci.* **2021**, *12* (36), 12098–12106. https://doi.org/10.1039/D1SC03551G.
(36) Abdiaj, I.; Cañellas, S.; Dieguez, A.; Linares, M. L.; Pijper, B.; Fontana, A.; Rodriguez, R.; Trabanco, A.; Palao, E.; Alcázar, J. End-to-End Automated Synthesis of C(Sp^3)-Enriched Drug-like Molecules *via* Negishi Coupling and Novel, Automated Liquid–Liquid Extraction. *J. Med. Chem.* **2023**, *66* (1), 716–732. https://doi.org/10.1021/acs.jmedchem.2c01646.
(37) Gao, L.; Shaabani, S.; Reyes Romero, A.; Xu, R.; Ahmadianmoghaddam, M.; Dömling, A. 'Chemistry at the Speed of Sound': Automated 1536-Well Nanoscale Synthesis of 16 Scaffolds in Parallel. *Green Chem.* **2023**, 10.1039.D2GC04312B. https://doi.org/10.1039/D2GC04312B.
(38) Murray, J. B.; Roughley, S. D.; Matassova, N.; Brough, P. A. Off-Rate Screening (ORS) By Surface Plasmon Resonance. An Efficient Method to Kinetically Sample Hit to Lead Chemical Space from Unpurified Reaction Products. *J. Med. Chem.* **2014**, *57* (7), 2845–2850. https://doi.org/10.1021/jm401848a.
(39) McCloskey, K.; Sigel, E. A.; Kearnes, S.; Xue, L.; Tian, X.; Moccia, D.; Gikunju, D.; Bazzaz, S.; Chan, B.; Clark, M. A.; Cuozzo, J. W.; Guié, M.-A.; Guilinger, J. P.; Huguet, C.; Hupp, C. D.; Keefe, A. D.; Mulhern, C. J.; Zhang, Y.; Riley, P. Machine Learning on DNA-Encoded Libraries: A New Paradigm for Hit Finding. *J. Med. Chem.* **2020**, *63* (16), 8857–8866. https://doi.org/10.1021/acs.jmedchem.0c00452.
(40) Scantlebury, J.; Vost, L.; Carbery, A.; Hadfield, T. E.; Turnbull, O. M.; Brown, N.; Chenthamarakshan, V.; Das, P.; Grosjean, H.; Von Delft, F.; Deane, C. M. A Small Step Toward Generalizability: Training a Machine Learning Scoring Function for Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2023**, *63* (10), 2960–2974. https://doi.org/10.1021/acs.jcim.3c00322.
(41) Zeller, M. J.; Favorov, O.; Li, K.; Nuthanakanti, A.; Hussein, D.; Michaud, A.; Lafontaine, D. A.; Busan, S.; Serganov, A.; Aubé, J.; Weeks, K. M. SHAPE-Enabled Fragment-Based Ligand Discovery for RNA. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119* (20), e2122660119. https://doi.org/10.1073/pnas.2122660119.
(42) St. Denis, J. D.; Hall, R. J.; Murray, C. W.; Heightman, T. D.; Rees, D. C. Fragment-Based Drug Discovery: Opportunities for Organic Synthesis. *RSC Med. Chem.* **2021**, *12* (3), 321–329. https://doi.org/10.1039/D0MD00375A.
(43) Baker, L. M.; Aimon, A.; Murray, J. B.; Surgenor, A. E.; Matassova, N.; Roughley, S. D.; Collins, P. M.; Krojer, T.; von Delft, F.; Hubbard, R. E. Rapid Optimisation of Fragments and Hits to Lead Compounds from Screening of Crude Reaction Mixtures. *Commun Chem* **2020**, *3* (1), 1–11. https://doi.org/10.1038/s42004-020-00367-0.
(44) Aguda, A. H.; Lavallee, V.; Cheng, P.; Bott, T. M.; Meimetis, L. G.; Law, S.; Nguyen, N. T.; Williams, D. E.; Kaleta, J.; Villanueva, I.; Davies, J.; Andersen, R. J.; Brayer, G. D.; Brömme, D. Affinity Crystallography: A New Approach to Extracting High-Affinity Enzyme Inhibitors from Natural Extracts. *J. Nat. Prod.* **2016**, *79* (8), 1962–1970. https://doi.org/10.1021/acs.jnatprod.6b00215.
(45) Pearce, N. M.; Krojer, T.; Bradley, A. R.; Collins, P.; Nowak, R. P.; Talon, R.; Marsden, B. D.; Kelm, S.; Shi, J.; Deane, C. M.; von Delft, F. A Multi-Crystal Method for Extracting Obscured Crystallographic States from Conventionally Uninterpretable Electron Density. *Nat Commun* **2017**, *8* (1), 15123. https://doi.org/10.1038/ncomms15123.
(46) De Semir, D.; Nosrati, M.; Bezrookove, V.; Dar, A. A.; Federman, S.; Bienvenu, G.; Venna, S.; Rangel, J.; Climent, J.; Meyer Tamgüney, T. M.; Thummala, S.; Tong, S.; Leong, S. P. L.; Haqq, C.; Billings, P.; Miller, J. R.; Sagebiel, R. W.; Debs, R.; Kashani-Sabet, M. Pleckstrin Homology Domain-Interacting Protein (PHIP) as a Marker and Mediator of Melanoma Metastasis. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (18), 7067–7072. https://doi.org/10.1073/pnas.1119949109.
(47) de Semir, D.; Bezrookove, V.; Nosrati, M.; Dar, A. A.; Wu, C.; Shen, J.; Rieken, C.; Venkatasubramanian, M.; Miller, J. R.; Desprez, P.-Y.; McAllister, S.; Soroceanu, L.; Debs, R. J.; Salomonis, N.; Schadendorf, D.; Cleaver, J. E.; Kashani-Sabet, M. PHIP as a Therapeutic Target for Driver-Negative Subtypes of Melanoma, Breast, and Lung Cancer. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115* (25), E5766–E5775. https://doi.org/10.1073/pnas.1804779115.
(48) Khan Tareque, R.; Hassell-Hart, S.; Krojer, T.; Bradley, A.; Velupillai, S.; Talon, R.; Fairhead, M.; Day, I. J.; Bala, K.; Felix, R.; Kemmitt, P. D.; Brennan, P.; von Delft, F.; Díaz Sáez, L.; Huber, K.; Spencer, J.

Deliberately Losing Control of C−H Activation Processes in the Design of Small-Molecule-Fragment Arrays Targeting Peroxisomal Metabolism. *ChemMedChem* **2020**, *15* (24), 2513–2520. https://doi.org/10.1002/cmdc.202000543.

(49) Grosjean, H.; Işık, M.; Aimon, A.; Mobley, D.; Chodera, J.; von Delft, F.; Biggin, P. C. SAMPL7 Protein-Ligand Challenge: A Community-Wide Evaluation of Computational Methods against Fragment Screening and Pose-Prediction. *J Comput Aided Mol Des* **2022**, *36* (4), 291–311. https://doi.org/10.1007/s10822-022-00452-7.

(50) DiPiro, J. T.; Chisholm-Burns, M. A. Fail Fast. *AJPE* **2013**, *77* (8), 159. https://doi.org/10.5688/ajpe778159.

(51) Thompson, W. Waztom/Mscheck: Pip Install Version, 2022. https://doi.org/10.5281/ZENODO.7018551.

(52) Pedrioli, P. G. A.; Eng, J. K.; Hubley, R.; Vogelzang, M.; Deutsch, E. W.; Raught, B.; Pratt, B.; Nilsson, E.; Angeletti, R. H.; Apweiler, R.; Cheung, K.; Costello, C. E.; Hermjakob, H.; Huang, S.; Julian, R. K.; Kapp, E.; McComb, M. E.; Oliver, S. G.; Omenn, G.; Paton, N. W.; Simpson, R.; Smith, R.; Taylor, C. F.; Zhu, W.; Aebersold, R. A Common Open Representation of Mass Spectrometry Data and Its Application to Proteomics Research. *Nat Biotechnol* **2004**, *22* (11), 1459–1466. https://doi.org/10.1038/nbt1031.

(53) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nat Biotechnol* **2012**, *30* (10), 918–920. https://doi.org/10.1038/nbt.2377.

(54) Schauperl, M.; Czodrowski, P.; Fuchs, J. E.; Huber, R. G.; Waldner, B. J.; Podewitz, M.; Kramer, C.; Liedl, K. R. Binding Pose Flip Explained via Enthalpic and Entropic Contributions. *J. Chem. Inf. Model.* **2017**, *57* (2), 345–354. https://doi.org/10.1021/acs.jcim.6b00483.

(55) Aldeghi, M.; Ross, G. A.; Bodkin, M. J.; Essex, J. W.; Knapp, S.; Biggin, P. C. Large-Scale Analysis of Water Stability in Bromodomain Binding Pockets with Grand Canonical Monte Carlo. *Commun. Chem.* **2018**, *1* (1), 19. https://doi.org/10.1038/s42004-018-0019-x.

(56) Barreiro, E. J.; Kümmerle, A. E.; Fraga, C. A. M. The Methylation Effect in Medicinal Chemistry. *Chem. Rev.* **2011**, *111* (9), 5215–5246. https://doi.org/10.1021/cr200060g.

(57) Philpott, M.; Yang, J.; Tumber, T.; Fedorov, O.; Uttarkar, S.; Filippakopoulos, P.; Picaud, S.; Keates, T.; Felletar, I.; Ciulli, A.; Knapp, S.; Heightman, T. D. Bromodomain-Peptide Displacement Assays for Interactome Mapping and Inhibitor Discovery. *Mol Biosyst* **2011**, *7* (10), 2899–2908. https://doi.org/10.1039/c1mb05099k.

(58) Kartal, Ö.; Andres, F.; Lai, M. P.; Nehme, R.; Cottier, K. waveRAPID—A Robust Assay for High-Throughput Kinetic Screens with the Creoptix WAVEsystem. *SLAS Discovery* **2021**, *26* (8), 995–1003. https://doi.org/10.1177/24725552211013827.