Development of Photoaffinity Labelling Technologies for Small Molecule Drug Discovery – a Case Study Targeting Bromodomains

A doctorate thesis submitted to the department of Pure & Applied Chemistry

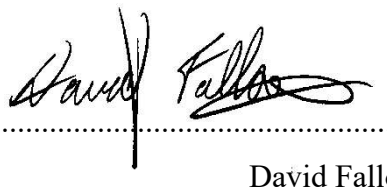
by

David Fallon

Under the supervision of Dr Jacob Bush and Prof. Nicholas Tomkinson.

This doctorate thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

13/08/2019



.....

David Fallon

Table of Contents

Table of Figures

Table of Tables

Abstract

The aim of this work was to investigate how the photoaffinity labelling (PAL) approach can positively impact and complement current methods of small molecule target-based drug discovery. For this broad application, an operationally simple and versatile synthetic protocol was required. An Ugi multicomponent reaction protocol for the expedient one-step synthesis of PAL probes was developed. The reaction couples an amine affinity function (compound of interest) with commonly used photoreactive groups and a variety of handle functionalities. Each component can be independently changed, providing access to a wide variety of PAL probes. For proof-of-concept studies, a series of pan-BET selective bromodomain PAL probes were obtained by parallel synthesis. Subsequent studies on the effect of different photoreactive groups, linker lengths, and irradiation wavelengths on photocrosslinking efficiency provided valuable insights into photoaffinity probe design. The observed trends in labelling were interpreted with additional consideration of the protein topology surrounding the active site, guided by crystallography and LC-MSMS studies. Optimal probes were progressed to MS-based proteomics to capture the BET family of proteins from live cells and reveal their potential on- and off-target profiles. From the optimisation studies above, PAL probes with high labelling efficiencies of recombinant BRD4 BD1 and BD2 protein were identified. The most promising probe was used to develop a novel biochemical PAL displacement assay, where the affinities of other competitor compounds to both BD1 and BD2 could be determined in the same experiment. This dual-domain PAL displacement assay represents the exciting and unexplored potential for PAL probes in biochemical assays.

Additionally, preliminary investigations into the impact of PAL on fragment-based biochemical screening were performed. A series of photoreactive BET-targeting fragments were prepared and their labelling efficiencies to recombinant BRD4 BD1 were found to be concentration-dependent and correlated well with the affinity of the fragment.

Overall, this work demonstrates new and exciting ways that the PAL approach can assist small-molecule target-based drug discovery, such as photoreactive fragment hit finding and novel biochemical PAL screening methods with recombinant protein. Additionally, the synthetic aspect of PAL-based chemoproteomics is made more accessible through the described one-step Ugi protocol and the information obtained regarding optimal probe design.

Acknowledgements

I would like to thank my industrial supervisor, Dr Jacob Bush, for his inspiration and guidance offered throughout my PhD. I thank him for his infectious passion for chemical biology. Thank you to my academic supervisor, Prof. Nicholas Tomkinson for his excellent and complementary supervision, and his enduring positive attitude towards my project.

I have thoroughly enjoyed my time on the GSK/Strathclyde collaborative PhD scheme, which is a direct result of the dedication shown by the organisers; Prof. Harry Kelly, Prof. William Kerr, Andrea Malley and Dr Laura Paterson in maintaining such an excellent programme.

Thank you to the past and present members of the Flexible Discovery Unit for providing an inspiring and welcoming workplace.

I am extremely grateful for a Postgraduate and Early Career Researcher Exchanges (PECRE) grant from the Scottish Funding Council (H17014), which allowed me to spend three months with Cellzome in Heidelberg. This exchange was extremely beneficial to both my project, and to my personal development. I would like to thank my hosts, Dr Chris Eberl and Stephanie Lehmann, who introduced me to chemoproteomics. I would like to thank the full Cellzome team for providing an inviting and dynamic workplace and contributing to a placement that I fondly enjoyed.

For their scientific insight and fruitful discussions, I would like to thank Emma Grant, Dr Vipul Patel, Dr Samuel Dalton, Dr Sebastian Campos, Dr Marcus Bantscheff, Dr Marcel Mülbaier, Johanna Vappiani, Sascha Knecht, Dr Chun-wa Chung, Ken Fantom, Alex Phillipou, Dr Daniel Thomas, Dr Mike Hann, Dr Roland Annan, Prof. John Murphy and Prof. Edward Tate.

I would like to commend the Agilent G6224 time-of-flight mass spectrometer, which on many late nights showed me the very fine line between love and anger.

Finally, thank you to all my friends and family outside the lab, particularly my partner Lynn, who provided immeasurable support over the course of my PhD journey.

Contributions

Section 2.3.7 X-ray crystallography was performed by Emily Lowndes, Jennifer Love and Chun-wa Chung.

Section 2.3.8 Protein digestion and LC-MSMS analysis was performed by Ken Fantom. Data analysis aided by Francesca Zappacosta.

Section 7.2.1.2 BRD4 BD1 and BRD4 BD2 recombinant protein were provided by Emma Jones.

Section 7.2.2 Potency assays (BRD4 BD1 and BD2 TR-FRET, MCP-1 whole blood) were performed by Alex Phillipou. Physicochemical measurements (ChromLogD, HSA binding, solubility and artificial membrane permeability) were performed by Iain Reid, Ferdausi Mazumder, Terrence Johnson and Robert Armstrong.

Section 7.4.2 HL-60 cell culturing was performed by Jürgen Stuhlfauth and Nilma Garcia-Altrieth.

Section 7.4.1.1 HL-60 lysate (method A) was prepared by Doris Uhlenbruch.

Section 7.4.1.2 Sample preparation (TMT labelling and desalting), and LC-MSMS instrument operation were performed by Kerstin Kammerer. Data analysis was aided by Chris Eberl.

Abbreviations

Abbreviation	Explanation
ABPP	Activity-based proteomic profiling
Ac	Acetyl
AfBPP	Affinity-based proteomic profiling
AGC	Automatic gain control
AMP	Artificial membrane permeability
ASU	Asymmetric unit
ATOM	Accelerating Therapeutics for Opportunities in Medicine
ATR	Attenuated total reflectance
BCA	Bicinchonic acid
BET	Bromodomain and extra-terminal domain
Boc	*Tert*-butyloxycarbonyl
BRD4 BD1	N-terminal domain (1st bromodomain) of bromodomain containing protein 4
BRD4 BD2	C-terminal domain (2nd bromodomain) of bromodomain containing protein 4
BSA	Bovine serum albumin
CAC	carnitine/acylcarnitine carrier (transporter protein)
CAD	Charged aerosol detection
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHEMBL	Chemistry database maintained by the European Molecular Biology Laboratory
CML	Chronic myelogenous leukemia
COSY	Correlated spectroscopy
CuAAC	Copper-catalysed alkyne-azide cycloaddition
CV	Column volume
DARTS	Drug affinity responsive target stability
DDA	Data-dependent aquisition
DIPEA	*N,N*-diisopropylethylamine
DME	1,2-Dimethoxyethane
DMF	*N,N*-dimethylformamide
DMSO	Dimethyl sulfoxide
DTT	1,4-Dithiothreitol
ELSD	Evaporative light scattering detection
ESI	Electrospray ionisation
Et	Ethyl
FAAH	Fatty acid amide hydrolase
FAC	Final active concentration

FBDD	Fragment-based drug design
FDR	False discovery rate
FP	Fluorescence polarisation
FRET	Fluorescence resonance energy transfer
FT	Fourier transform
GSH	Glutathione
HATU	1-(bis(dimethylamino)methylene)-1*H*-[1,2,3]triazolo[4,5-*b*]pyridine-1-ium 3-oxide hexafluorophosphate(V)
HCD	Higher energy collision dissociation
HDAC	Histone deacetylase
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HL-60	Human leukaemia cell line
HMBC	Heteronuclear multiple-bond correlation
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
HSA	Human serum albumin
HSQC	Heteronuclear single quantum coherence
IEDDA	Inverse electron demand Diels Alder
IGEPAL CA-630	octylphenoxypolyethoxyethanol
IMCR	Isonitrile-based multicomponent reaction
IMDM	Iscove's modified Dulbecco's media
KAc	Acetylated lysine
K_d	Dissociation constant
LCMS	Liquid chromatography-mass spectrometry
LDS	Lithium dodecyl sulfate
LiP-SRM	Limited proteolysis-selective reaction monitoring
LPS	Lipopolysaccaride
MCP-1	Monocyte chemoattractant protein 1
MDAP	Mass-directed AutoPrep
Me	Methyl
MES	2-(*N*-morpholino)ethanesulfonic acid
MS-based proteomics	Mass spectrometry-based proteomics
NCE	Normalised collision energy
nESI	Nano-electrospray ionisation
NMR	Nuclear magnetic resonance
OAc	Acetate
PAGE	Polyacrylamide gel electrophoresis
PAL	Photoaffinity labelling
PARP1	Poly(ADP-ribose)-polymerase 1

PBS	Phosphate-buffered saline
PEG	Polyethylene glycol
PIM5	Protease inhibitor mix 5
PRM	Parallel reaction monitoring
PSM	Peptide spectrum match
PTMs	Post-translational modifications
PTP	Protein tyrosine phosphatase
qIT	Quantitative reversible tethering
Qusm	Quantified unique spectra match
RBF	Round-bottomed flask
RMS	Root mean square
SBDD	Structure-based drug design
SCX	Strong cation exchange
SDS	Sodium dodecyl sulfate
SI	Supplementary information
SMILES	Simplified molecular-input line-entry system
SPG	Crystallisation buffer composed of succinic acid, sodium dihydrogen phosphate and glycine (2:2:7 molar ratio)
SPROX	Stability of proteins from rates of oxidation
TBTA	Tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TCO	*Trans*-cyclooctene
TEV	Tobacco etch virus (protease)
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TIC	Total ion chromatogram/count
TLC	Thin-layer chromatography
TMS (NMR context)	Trimethylsilane
TMT	Tandem mass tag
ToF	Time-of-flight
Tosyl or Ts	Toluenesulfonyl
TPP	Thermal proteomic profiling
TR-FRET	Time-resolved fluorescence energy transfer
UPLC	Ultra-performance liquid chromatography
UV	Ultraviolet

Chapter 1 Introduction

1.1 Current challenges in small molecule drug discovery

During the first decade of the 21st century the number of approvals for new small molecule therapeutics declined, despite increased R&D spending.[1,2] Numerous analyses from various pharmaceutical companies have suggested two key causes of attrition in clinical phases; a lack of efficacy in alleviating the disease, or toxic adverse events, most often attributed to an incomplete knowledge of which proteins are engaged by the potential new medicine.[3-5]

The aim of early-phase drug discovery is to provide candidate molecules that have sufficient evidence of target engagement and efficacy in the most relevant disease model possible, along with a demonstrated safety profile so that the compound can be administered to humans.[6] Advances within each area of pre-clinical discovery are required to collectively increase the productivity of this stage and improve on the quality of candidates entering clinical trials. For example, the ATOM (Accelerating Therapeutics for Opportunities in Medicine) project promises that the application of artificial intelligence (AI) and machine learning to areas such as multiparameter lead optimisation can reduce the timescale of pre-clinical discovery from an average of six years to one year.[7] The development of more informative assays at earlier stages of lead optimisation may also help to reduce candidate cycle times.[8,9] To minimise any delays in decision-making, these assays should provide unambiguous data to support the termination, continuation or clear change of direction for a program.

Multidisciplinary approaches such as chemical biology can aid pre-clinical drug discovery.[1,10-12] Of these, photoaffinity labelling (PAL) has become a powerful approach, particularly when coupled with modern mass spectrometry capabilities.[13-17] PAL allows for the capture of non-covalent target interactions formed by compounds of interest, and thus has positively impacted a range of hit identification platforms (Figure 1.1), including fragment-based and phenotypic screening and DNA-encoded library technologies.[18-20] The most extensive and powerful application of PAL has been in combination with MS-based chemoproteomics. This methodology can be used to demonstrate target engagement with intended on-target proteins, while also

identifying potential off-targets for pre-candidate molecules of interest in live cells.[21-23]

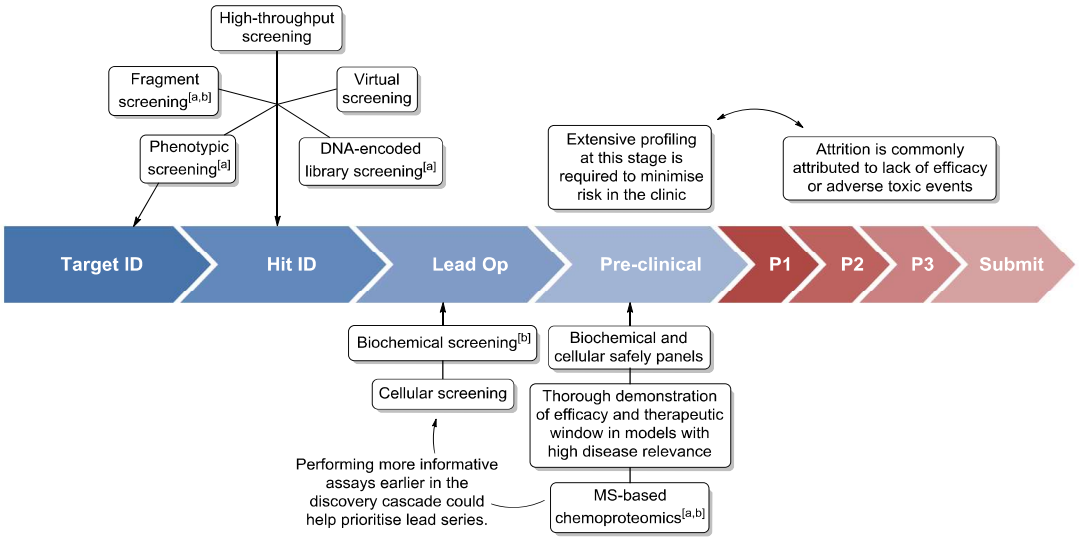


Figure 1.1 Overview of the current strategies adopted by pharmaceutical companies to develop new small molecule therapeutics. [a]Areas that have been enhanced by the application of the photoaffinity labelling (PAL) approach.[18-20] [b]Areas investigated in this work using PAL.

The work presented in this thesis focuses on making the photoaffinity labelling approach more accessible to small molecule drug discovery by introducing simplified synthetic strategy (Chapter 2), a novel photoaffinity-based recombinant assay platform (Chapter 3), simplified methods for photoaffinity-based proteomic profiling (Chapter 4) and investigating the application of PAL to fragment-based screening (Chapter 5).

1.2 MS-based chemoproteomics

MS-based chemoproteomics has become one of the most widely adopted approaches used to identify and profile protein targets engaged by small molecules of interest.[24] By providing information on the proteins engaged by a small molecule, chemoproteomics is an extremely powerful tool for drug discovery. Some of the main applications are summarised in Table 1.1.[25]

Table 1.1 Common applications of MS-based chemoproteomics in small molecule drug discovery.[25]

Application (in lysates or live cells)	Comment
Profiling of potential off-targets	Chemoproteomic probes designed from compounds of interest can be used to find off-targets and inform on potential safety issues.
Target engagement	Target engagement can be directly observed with the chemoproteomic probe or inferred by displacing the probe from the target protein with excess compound of interest.
Target identification	Phenotypic hits or natural products with unknown modes of action can be interrogated by finding the proteins they engage with.
Drug repurposing	With advancing technology and techniques, new targets for existing drugs are commonly found and can inform on new potential indications.
Biomarker discovery	Chemoproteomic profiling between normal and pathological states can identify changes in functional proteins and find new biomarkers for that disease.

The ultimate aim of chemoproteomics is to observe the engagement of any small molecule (chemo-) across the whole proteome (-proteomics).[26-28] Recent advances in LC-MSMS technologies continue to push towards this goal, allowing thousands of proteins to be identified and quantified within complex biological samples in a single experiment.[28,29] MS-based analyses of whole proteomes are possible, however to reduce complexity and significantly shorten analysis time, MS-based chemoproteomic workflows generally involve enrichment of the subset of the proteome that the small molecule engages with, prior to identification and relative quantification using LC-MSMS analysis.[17,24,30] Bottom-up/shotgun proteomics, where protein identification is performed by identifying digested peptides, is the most common method used in these approaches.[31] A general MS-based chemoproteomic workflow is shown in Figure 1.2 and can be divided into three key components; capture of the proteins engaged by a small molecule, enrichment, and detection.

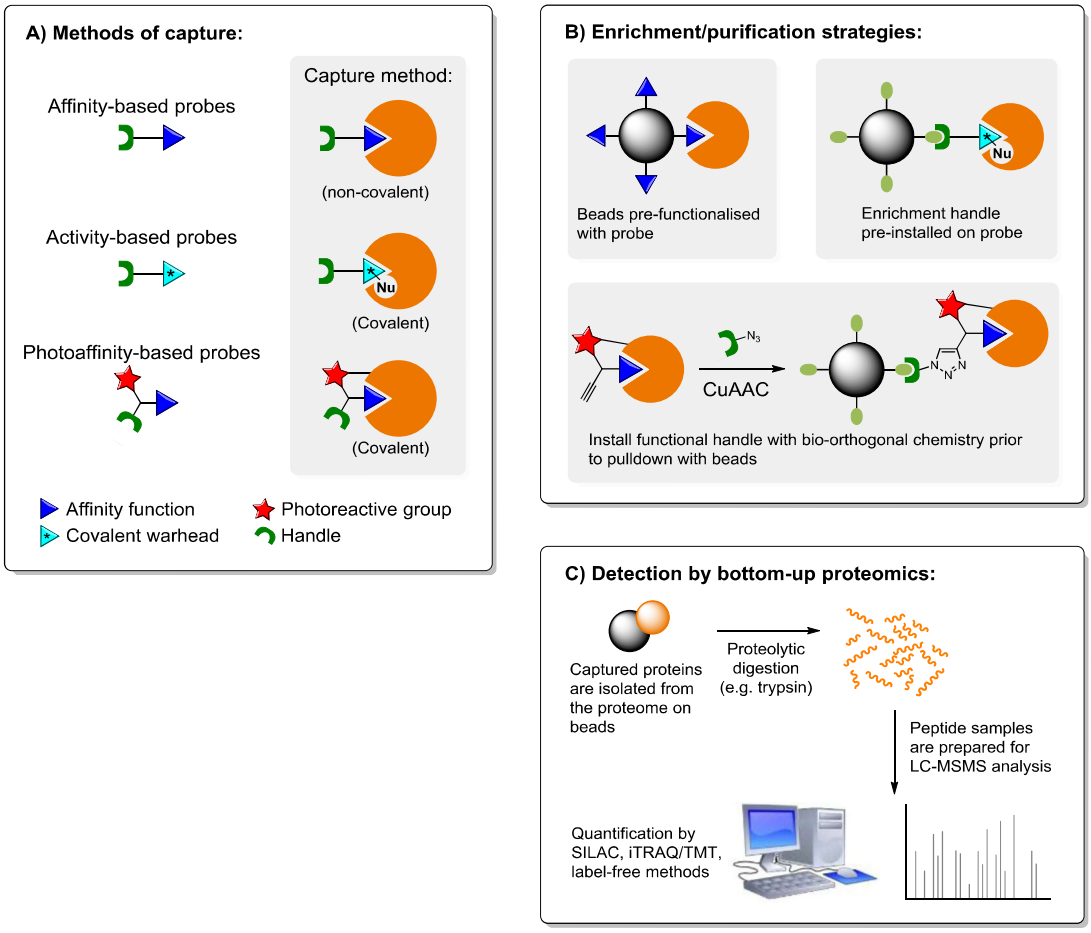


Figure 1.2 MS-based chemoproteomic workflows generally involve three experimental stages: A) engagement of the small molecule probe with target proteins. B) Enrichment of these engaged proteins. C) Detection of the enriched proteins by bottom-up proteomics.

The three most commonly used capture methods are performed using affinity-, activity- and photoaffinity-based probes. These three methods are discussed in detail in Section 1.2.1. Early enrichment strategies involved capturing the engaged proteins from cell lysates by immobilising the probe on solid-phase matrices or having a purification handle (*e.g.* biotin) pre-installed on the probe.[27,32,33] One limitation of these strategies is that they are only applicable for profiling cell lysates, and not live cells due to poor membrane permeability. Advances in bio-orthogonal click chemistry has allowed for a two-stage enrichment strategy, where live cells can be treated with the probe so that the most biologically relevant target proteins are captured. The cells are then lysed and click chemistry is used to install the purification tag.[21,34-40] For detection, various methods can be used such as in-gel fluorescence scanning or western blotting, however gel-free MS-based proteomic analysis is the most informative

detection method.[17,18,28,31,41] An introduction to MS-based proteomics is given in Section 1.3.

1.2.1 Capture methods for MS-based chemoproteomics

1.2.1.1 Affinity-based probes

Affinity-based purification offers the most direct approach to identifying proteins that bind to small molecule compounds of interest.[33] Affinity-matrix purification has been the most commonly used approach, where beads are functionalised with the small molecule of interest, before being incubated with cell lysate. The beads capture the target proteins *via* non-covalent engagement of the small molecule. The beads are gently washed to reduce non-specifically bound proteins. The remaining proteins are then eluted, digested and identified by LC-MSMS. Bantscheff and co-workers developed a competitive screening platform based on affinity-based purification known as Kinobeads™ (Figure 1.3).[32] The authors used a bead matrix functionalised with a variety of non-covalent pan-kinase inhibitors, which pulled down a wide range of endogenously expressed protein kinases from lysate, along with some additional ATP and purine-binding proteins. The experiment was repeated with cell lysate containing a compound of interest over a range of concentrations. By comparing the proteins identified by the beads alone *versus* the proteins that have reduced enrichment with increasing competitor compound, quantitative binding data for the competitor compound to the bead-captured portion of the kinome was obtained. The on-/off-targets of known kinase-targeting small molecule drugs such as imatinib, dasatinib and bosutinib were profiled. This demonstrated the power of the approach for identifying potential off-targets. Competitive affinity-based protein profiling (AfBPP) has since been extended to other protein families, for example, to histone deacetylases (HDACs).[42]

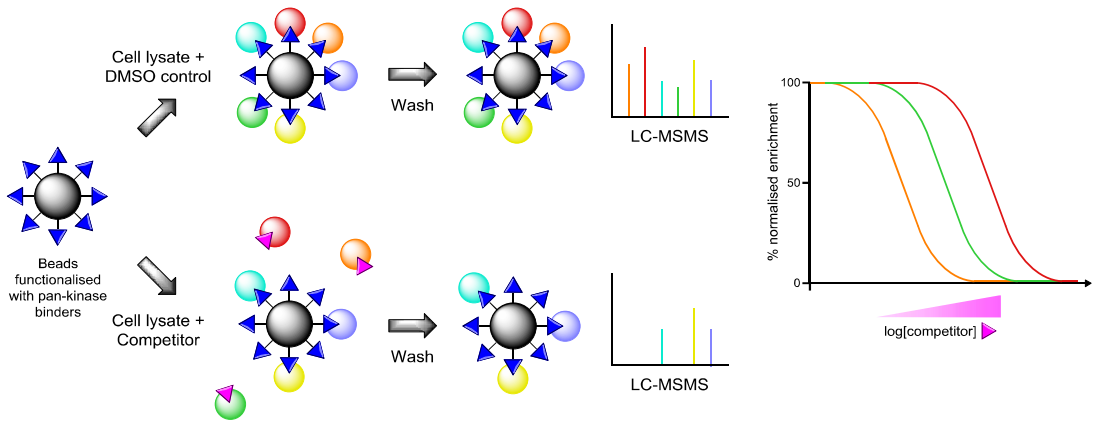


Figure 1.3 Kinobead™ assay platform developed by Bantscheff and co-workers.[32] Beads were functionalised with a mixture of seven potent pan-kinase inhibitors. In the absence of a competitor compound, the beads pulled down a large proportion of the endogenously expressed protein kinases from cell lysate. The experiment was performed in parallel containing a range of competitor ligand concentrations. The dose-dependent reduction in kinase enrichment produced a selectivity profile for the competitor compound.

One disadvantage of affinity matrix protein profiling is that the technique can only be used with cell lysate and not live cells. Therefore, the observed binding profile for the competitor may differ from the true engagement profile that the compound exhibits in live cells. An alternative to bead functionalisation is to incorporate a "clickable" handle onto the affinity-based probe. This allows the probe to be used in live cells. Bio-orthogonal chemistry is used to install a purification handle after lysis, such as biotin.[35,39] Biotin–streptavidin bead enrichment is then used to pull down the non-covalently captured proteins for downstream identification by LC-MSMS. Rutkowska and co-workers demonstrated the affinity-based capture of poly(ADP-ribose)-polymerase 1 (PARP1) with a *trans*-cyclooctene (TCO) derivative of Olaparib.[34] HeLa cells were incubated with a fixed concentration of affinity-based probe, and a range of free Olaparib concentrations. After cell lysis, the affinity-captured proteins were pulled down with beads functionalised with tetrazine, which performs a bio-orthogonal inverse-electron demand Diels Alder (IEDDA) reaction with TCO. PARP1 enrichment was dose-dependent with free Olaparib concentration, demonstrating specific target engagement in live cells. One issue with this approach is that the non-covalent interactions must survive cell lysis and purification to be detected. Due to the reversible nature of the non-covalent engagement, the affinity-based probe may re-equilibrate with proteins that were not

previously accessible when the cell was intact (*e.g.* within organelles, membranes or protein complexes). Furthermore, gentle bead washing is required to minimise elution of any proteins specifically captured by weak interactions. However, if the bead washing is not extensive enough, it becomes difficult to distinguish specific enrichment *versus* non-specific enrichment.[43]

1.2.1.2 Activity-based probes

AfBPP can fail to capture weak binding target proteins, due to the non-covalent nature of the capture method. Activity-based probes are used to covalently capture target proteins by reacting with nucleophilic residues present in the active site.[44] Activity-based probes are bifunctional molecules consisting of an electrophilic affinity function and handle that allows for downstream identification or enrichment (*e.g.* fluorophore, biotin or a bio-orthogonal handle). The affinity function is usually a known inhibitor of the enzyme, or a mimic of the enzyme's natural substrate. This recognition element additionally contains a chemically reactive "warhead", which is usually an electrophile that can covalently label functional nucleophilic residues (cysteine, lysine, serine *etc.*) within the active site. For this approach to be successful, covalent labelling of these nucleophilic residues must be governed by binding of the affinity function to the active site, which provides a high local concentration of probe to label proximal nucleophilic residues. This calls for rigorous probe design and careful choice of electrophilic warhead. A wide array of electrophilic warheads have been reported, offering various levels of reactivity and residue specificity.[45-51] Commonly used covalent warheads include fluorophosphonates, acrylamides, α,β-unsaturated esters, epoxides, sulfonyl fluorides, and more recently, fluorosulfates.[48,52-57] Those that offer residue specific labelling have been recently reviewed by Shannon *et al*.[58]

Activity-based proteomic profiling (ABPP) is most suited to profiling enzyme families which share a conserved nucleophilic residue within their active site.[27] Cravatt and co-workers have developed a range of activity-based fluorophosphate probes capable of profiling serine hydrolases.[52,53] Serine hydrolases constitute one of the largest and widely distributed enzyme classes across all three kingdoms of life. In 2016, a fatty

acid amide hydrolase (FAAH) inhibitor BIA 10-2474 (**1.1**) entered phase I clinical trials and resulted in the death of one participant and severe neurological symptoms in four others.[59] Cravatt and co-workers profiled **1.1**, along with other more well-characterised FAAH inhibitors in human cells to identify any serine hydrolase off-targets.[60] The authors identified that **1.1** covalently modified the intended target FAAH, but also interacted with several lipolytic enzymes that are highly expressed in human brain tissue. Interestingly, these proteins were not engaged by the other FAAH inhibitors. This suggested that the neurotoxicity observed for **1.1** in the phase I clinical trial may have been due to disruption of neuronal lipid networks caused by the inhibition of these observed off-targets. This study illustrated the power of ABPP to identify off-targets of potential drug candidates and demonstrated the importance of performing these profiling experiments prior to investigations in the clinic.

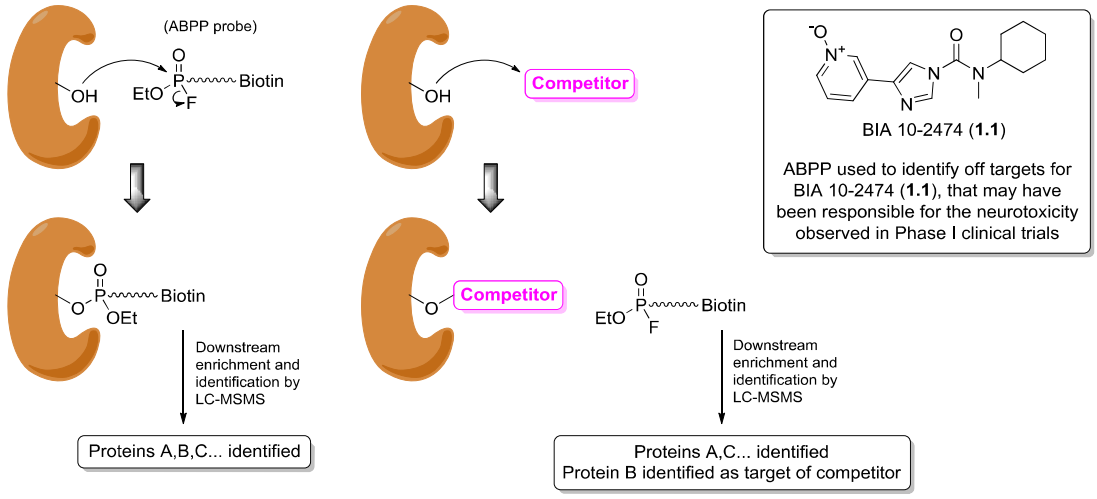


Figure 1.4 Profiling of serine hydrolases using a fluorophosphonate activity-based probe developed by Cravatt and co-workers.[52,60]

Taunton and co-workers recently reported a sulfonyl fluoride-based probe **1.2** for ABPP of kinases (Figure 1.5).[61] The authors demonstrated that pan-kinase probe **1.2** enriched 133 endogenously expressed kinases from live Jurkat cells by covalently labelling a conserved lysine within the ATP binding site. Dasatinib **1.3** was used as a proof-of-concept competitor compound to successfully demonstrate that **1.2** could be used for live-cell kinase profiling. This presents an advantage over other kinase profiling platforms such as Kinobeads™ for example, which is restricted to profiling cell lysate.[32]

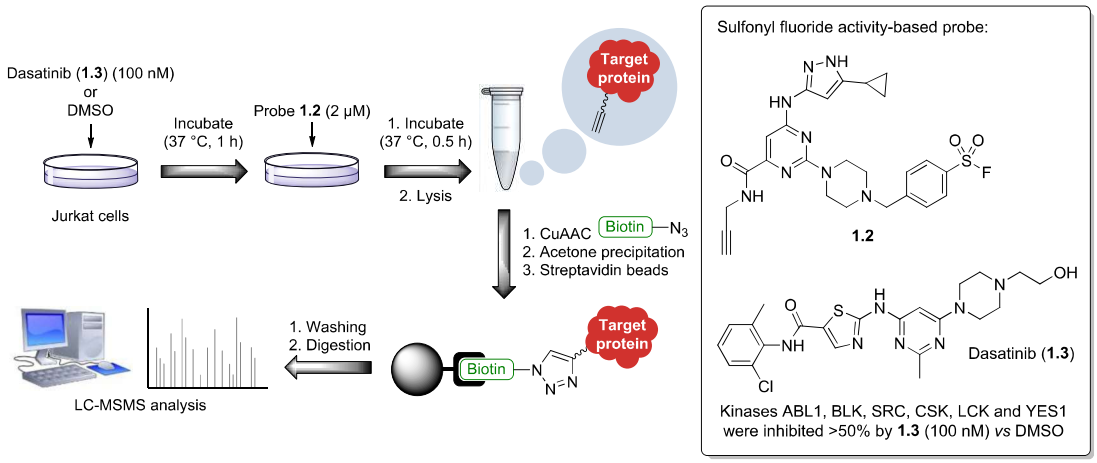


Figure 1.5 Sulfonyl fluoride activity-based probe **1.2** for profiling kinases in live cells developed by Taunton and co-workers.[61] Six kinases were inhibited (≥50%) by dasatinib **1.3** *versus* DMSO control in live Jurkat cells.

One caveat is that only proteins that contain a suitable nucleophilic residue within the binding site are amenable to ABPP. Any proteins that bind specifically to the activity-based probe but are not covalently captured due to having no appropriate nucleophilic residue in the binding pocket are not enriched and identified. Photoaffinity labelling (PAL) offers a more unbiased mechanism of covalent capture, whereby non-covalently engaged proteins are captured by reactive intermediates generated on the probe upon irradiation with UV light.

1.2.1.3 Photoaffinity probes

Photoaffinity probes are typically tri-functional molecules, consisting of an affinity function, a photoreactive group and a bio-orthogonal handle or reporter tag (Figure 1.6).[14] The PAL probe is introduced to a biological system (recombinant protein, lysates or live cells) and the sample is then irradiated with UV light. This irradiation activates the photoreactive group (most commonly an aryl azide, diazirine or benzophenone) to form an excited species (such as a carbene, nitrene or di-radical, respectively) capable of covalently labelling the target protein.[62,63] These photoreactive groups are discussed further in Section 1.2.2. The handle functionality is then used to detect covalently modified proteins using in-gel fluorescence scanning or MS-based proteomics.[15] These methods are discussed further in Section 1.2.4.

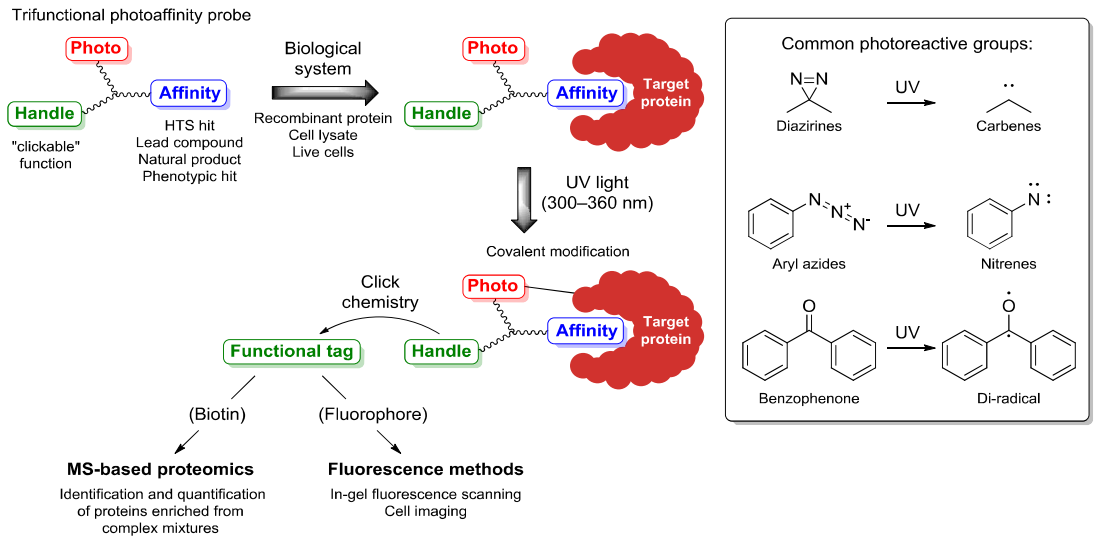


Figure 1.6 Overview of the PAL approach. The PAL probe is incubated with a biological system to allow for the non-covalent interactions formed between the PAL probe and the target proteins to reach equilibrium. These interactions are then captured by irradiating the samples with UV light, which excites the photoreactive group (most commonly a diazirine, aryl azide or benzophenone) to form a reactive species capable of covalently labelling proximal proteins before being quenched by water or buffer components. The handle functionality is then used to install a reporter tag *via* bio-orthogonal click chemistry. Installing biotin enables proteomic pulldown experiments, where the captured proteins are enriched with streptavidin beads and identified using bottom-up proteomics. Alternatively, a fluorophore can be installed, allowing for detection of the captured proteins by using a fluorescent readout.

Yao and co-workers recently synthesised a PAL probe derivative of dasatinib (**1.4**), a marketed dual-inhibitor of BCR-ABL and SRC tyrosine kinases used for the treatment of chronic myelogenous leukemia (CML) (Figure 1.7).[23,64] Photoaffinity-based pulldown experiments using **1.4** identified six previously unknown serine/tyrosine kinase targets from two different cell lines (K562 and HepG2). Dasatinib was found to be potent against these kinases in follow-up recombinant kinase screens. These off-targets had not been found using previous affinity-matrix purification efforts (*e.g.* Kinobeads™), and were not identified in the live-cell activity-based profiling of dasatinib (**1.3**) performed by Taunton and co-workers (discussed previously in Section 1.2.1.2).[32,61] This demonstrates that covalent capture by photoaffinity labelling has the potential to identify more biologically relevant off-targets compared to analogous affinity- or activity-based profiling efforts. Photoaffinity-based profiling has also been widely applied to identifying the protein targets of hit compounds from phenotypic screening efforts, and for bio-active natural products with unknown modes of action.[65-68]

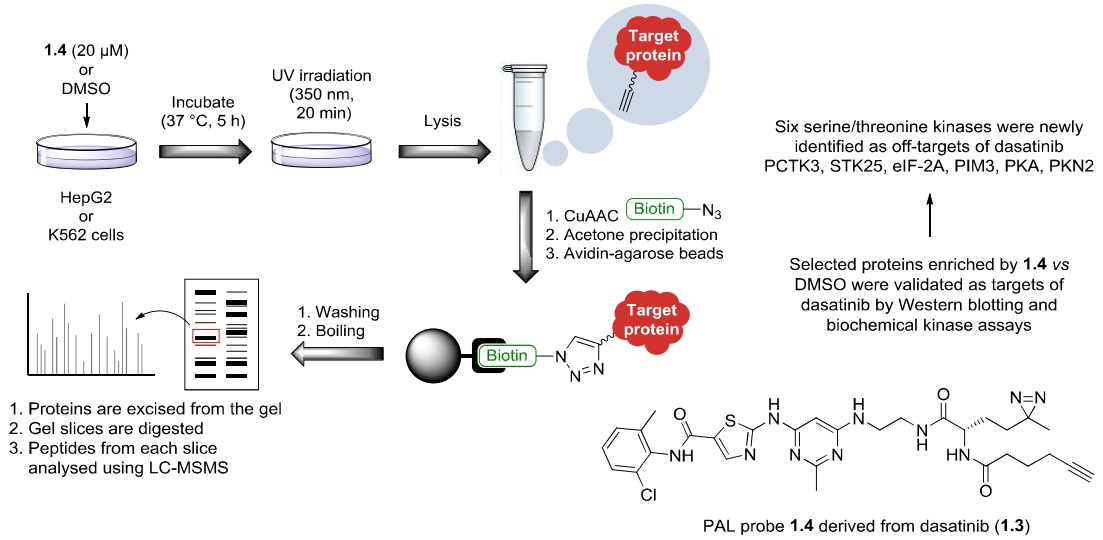


Figure 1.7 PAL probe **1.4** was derived from dasatinib (**1.3**) and used to identify possible protein targets of **1.3** in live HepG2 and K562 cells.[23] Six previously unknown kinase targets of dasatinib were identified and validated in subsequent biochemical assays.

One issue with using a PAL approach to covalently capture target proteins is that the photocrosslinking efficiency can be low. Furthermore, in comparison to affinity- or activity-based probes, relatively large changes to the parent affinity function must be made to incorporate the photoreactive group and bio-orthogonal handle, while retaining potency for the same target proteins engaged by the parent compound. This can be difficult, especially in the absence of SAR information.[8] A variety of photoreactive groups have been reported in the literature and are discussed in detail in Section 1.2.2. Several direct comparisons of various photoreactive groups have been reported, but are conflicting as to which photoreactive group provides the most efficient levels of labelling.[69-73] The amino acid insertion preferences for different photoreactive groups is dependent on the excited species formed under UV irradiation (*e.g.* carbenes, nitrenes, di-radicals). Therefore, it is widely advised to obtain a range of different PAL probes containing various photoreactive groups.[26,39,69,74-76] The synthesis of multiple tri-functional PAL probes can be tedious. Current approaches to synthesise generic PAL probes are reviewed in Section 1.4.

1.2.1.4 Comparison of capture methods for chemoproteomic profiling

The three capture methods previously discussed are compared in Table 1.2. An advantage of affinity-based purification is that the affinity-based probe is relatively simple and can retain the most structural similarity to the parent compound. However, due to the non-covalent capturing of target proteins, weakly bound proteins may be washed from the beads prior to detection. Furthermore, the proteomic profiles obtained from cell lysate may not fully reflect the profile of proteins that the compound engages in a live cell. Activity-based probes form a covalent bond with nucleophilic residues present in the active site of target proteins. This method is well-suited to profiling families of enzymes that contain conserved nucleophilic residues (kinases, serine hydrolases *etc.*). One disadvantage of this approach is that the proteins that are non-covalently engaged by the probe but not covalently captured are washed away during enrichment and not detected. Photoaffinity-based probes enable the covalent capture of non-covalent interactions. Furthermore, photoaffinity capture can be "switched-on" after the probe has had time to equilibrate within the biological sample. Photocrosslinking yields can be low due to quenching of the reactive intermediate by water or buffer components. One advantage shared by activity- and photoaffinity-based capture is that weakly bound proteins can be captured and detected. Recently, these two capture methods have been employed in fragment-based screening.[18,77-79] The application of PAL to fragment-based drug discovery (FBDD) is discussed in Chapter 5.

Table 1.2 Comparison of covalent capture methods used in chemoproteomics.

	Affinity-based	Activity-based	Photoaffinity-based
Applicable to live cells	Bead-based affinity matrix not applicable	Yes	Yes
Probe complexity	Bi-functional	Bi-functional	Tri-functional
Probe similarity to parent	+++	++	+
Capture mechanism	Non-covalent	Covalent	Covalent
"Switch-on" capture	No	No	Yes
Protein target range	Non-covalent interactions captured	Only covalent interactions captured	Non-covalent interactions captured
Capture efficiency	Variable	High	Variable
Weak protein–probe interactions captured	No	Yes	Yes

1.2.1.5 Capture-free MS-based chemoproteomics

Small molecule binding often induces a change to the biophysical properties of the target protein. Three MS-based profiling strategies that exploit these changes have recently been developed (Figure 1.8).[24,28,80] These are thermal proteome profiling (TPP), stability of proteins from rates of oxidation (SPROX) and drug affinity responsive target stability (DARTS).[81-83] These experiments involve no capture and enrichment of engaged proteins. Instead, the effect of a small molecule on the stability of the proteome (*versus* temperature, oxidation or proteolysis, respectively) is assessed. Thus, these experiments require no pulldown and enrichment step. However, as these experiments profile the whole proteome, sample fractionation and extensive mass spectrometry acquisition times are required (24–48 hours per experiment).[28]

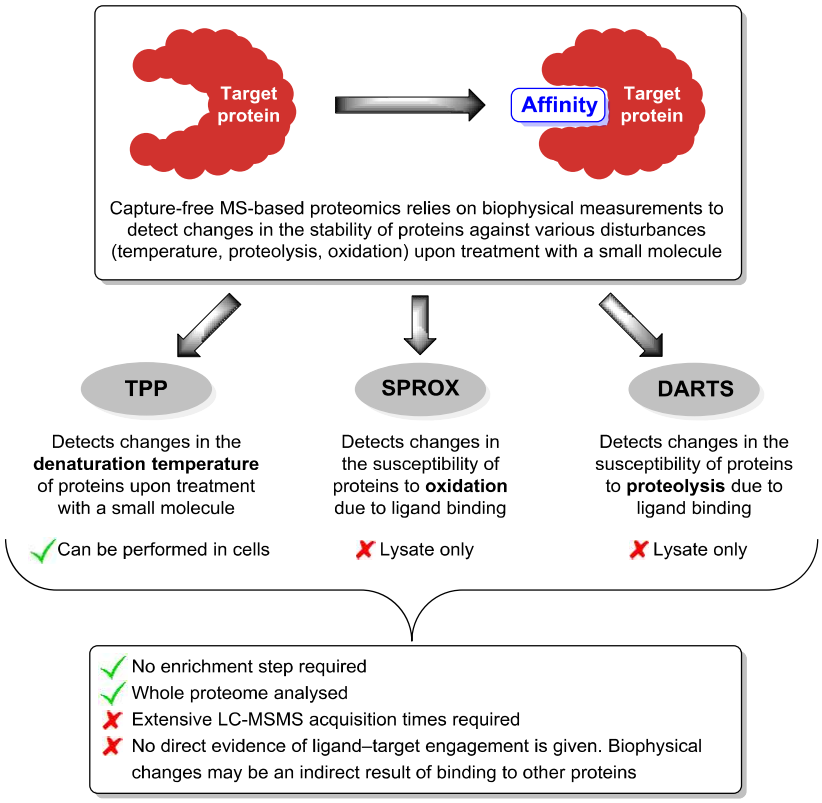


Figure 1.8 Overview of the capture-free methods used in chemoproteomics.[28]

TPP assesses the change in protein melting curves from live cells treated with and without a small molecule of interest. Proteins usually exhibit a higher denaturation temperature upon binding of a small molecule. In a TPP experiment, live cells are treated with and without a small molecule of interest and samples are aliquoted. Each aliquot is heated to a specific temperature (*e.g.* 10 aliquots between 37–63 °C). Depending on their stability, the proteins denature at specific temperatures. The cells are lysed, and the remaining soluble proteins at each temperature are quantified using multiplexed LC-MSMS analysis (discussed in Section 1.3.3). Shifts in melting temperatures are indicative of compound binding. The experiment can be performed with a range of concentrations to obtain affinity values for the compound of interest (known as 2D-TPP), at the expense of increased MS acquisition time.[84] One caveat to TPP is that not all proteins demonstrate a shift in melting point upon small molecule binding. Furthermore, because the method does not involve the direct capture of the target protein by the compound, follow up capture-based experiments are usually performed to validate target engagement.[28] Binding to a target protein that is a member

of a protein complex may induce temperature shifts in other interacting proteins, producing false positives. For example, the well-profiled compound dasatinib (**1.3**) (previously discussed as an example for affinity- and activity-based profiling) showed no melting shift for its known target BCR-ABL, but showed shifts for proteins involved in the downstream signalling pathway.[81]

In SPROX, aliquots of lysate are treated with and without a small molecule of interest. Each pair of aliquots is treated with an increasing concentration of denaturant, such as guanidinium hydrochloride.[85] The samples are then treated with hydrogen peroxide, which oxidises any solvent-exposed methionine residues. The samples are digested, and the extent of methionine oxidation is quantified using LC-MSMS. A decrease in methionine oxidation at a certain denaturant concentration between the treated and untreated samples is indicative of stabilisation of the target protein by compound binding. One drawback of this approach is the requirement for methionine (which has a low abundance of *ca*. 2%) to be present on the peptides being measured.[50,80]

DARTS operates on the principle that liganded protein targets are less susceptible to proteolysis than their apo forms, either through direct blocking of protease recognition sites by the ligand, or by indirectly affecting the rate of proteolysis by inducing a structural change in the target protein upon binding.[83,86] Cell lysate is treated with and without a small molecule of interest, and a small concentration of a non-specific protease is added for a limited time which digests exposed regions (such as loops and terminal ends). The amounts of intact protein between treated and untreated samples are then quantified by western blotting. Limited proteolysis-selective reaction monitoring (LiP-SRM) is an MS-based extension of DARTS, whereby the samples undergo a two-stage digestion process.[87] The samples are first digested with a non-specific protease as performed in DARTS, however the samples are then denatured and digested with trypsin. Differences in tryptic peptide identifications quantified by LC-MSMS indicate binding of the small molecule to those proteins. For DARTS or LiP-SRM, optimising the appropriate limited digestion conditions can be challenging.

These three capture-free profiling technologies offer powerful approaches to screen the binding of a small molecule across whole proteomes. However, only TPP is applicable to live cells. SPROX and DARTS can only be performed in cell lysates.

Furthermore, these methods currently require extensive MS acquisition times. As these methods do not directly capture the small molecule bound to the target protein, capture-based chemoproteomic follow-up studies are commonly performed to validate direct target engagement.[28]

Of the capture-based methods previously discussed, photoaffinity labelling offers the most biologically relevant capture method, where the non-covalent interactions formed by a PAL probe can be captured at a specific timepoint in live cells, with minimal bias for specific residues.

1.2.1.6 Requirements for effective photoaffinity labelling

By being able to capture weak transient interactions and not having to rely on the presence of nucleophilic residues, PAL offers the ability to capture the most comprehensive profile of non-covalent interactions within complex biological mixtures. However, achieving high levels of specific and efficient photocrosslinking can be challenging. Effective photoaffinity labelling (PAL) relies on several criteria:[26,62]

- The rate of the covalent reaction must be faster than the lifetime of the protein–probe complex.
- Any covalent photoadducts produced should be stable enough to survive downstream detection strategies.
- The excited species should react indiscriminately with any bond on the surface of the target protein to ensure all non-covalently interacting proteins are covalently captured.
- The PAL probe should be as structurally similar to the parent affinity function as possible to mimic the non-covalent binding interactions formed by the parent.
- The PAL probe must be chemically stable and only give covalent labelling upon UV excitation (>300 nm).

The first two criteria are mainly governed by the photoreactive group used. An ideal photoreactive group produces a reactive intermediate that can perform the covalent

labelling event faster than the off-rate of the PAL probe from the target protein. Otherwise, the reactive intermediate may dissociate from the active site of the target protein and non-specifically label other proteins. Similarly, any photoactivated PAL probe in solution must be quenched by buffer components before it can diffuse and cause non-specific labelling. By achieving covalent labelling at rates faster than the diffusion of the probe from the active site, PAL labelling allows for a snapshot to be taken of the non-covalent interactions formed by the PAL probe at the time of irradiation.

1.2.2 Photoreactive groups used in PAL studies

Photoaffinity labelling (PAL) was first introduced by Westheimer and co-workers in 1962, who reported that acetyldiazoacetate could inactivate chymotrypsin after irradiation with UV light.[88] Since then, many photoreactive groups have been reported, of which aryl azides, diazirines and benzophenones have been the most explored and employed.[26,62,75] The labelling mechanisms, impact on physicochemical properties and advantages/disadvantages of each group are discussed in Sections 1.2.2.1–1.2.2.3 and summarised in Table 1.3.

1.2.2.1 Aryl azides

Since being first introduced in 1969,[89] the photochemistry of aryl azides has been explored in detail (Figure 1.9).[90] Upon activation of aryl azide (**1.5**) with UV light (λ_{max} *ca*. 300 nm), nitrogen is released and a singlet nitrene **1.6** is formed which is capable of C–H insertion reactions to form **1.7**. The lifetime of this singlet nitrene is largely governed by the substituents present on the aromatic ring. The unsubstituted aryl nitrene can have a lifetime of around 0.1 ms in the singlet state, however with a *para* carbonyl substitution (aryl azides investigated in this work), this lifetime is limited to *ca.* 1 ns due to favoured rearrangement to dehydroazepine **1.8** *via* a benzazirine intermediate **1.9**.[91] Dehydroazepines have been shown to preferentially react with nitrogen nucleophiles to form **1.10**, (or form **1.11** *via* reaction through **1.9**) and can be quenched by HEPES buffer.[91] Alternatively, competitive intersystem

crossing to the lower energy triplet nitrene **1.12** (*ca.* 20 kcal mol^{-1}) can occur, and this transition is favoured at lower temperatures.[92,93] The triplet nitrene can persist in aqueous solution with a reasonable lifetime (*ca.* 1 ms) and can partake in more radical-like reactions, such as H-atom abstraction to form **1.13**.[94] These reactions may result in successful protein labelling (**1.14**) or may generate reactive intermediates (such as reactive oxygen species) with buffer components that can result in detrimental oxidation or cleavage of the protein.[95]

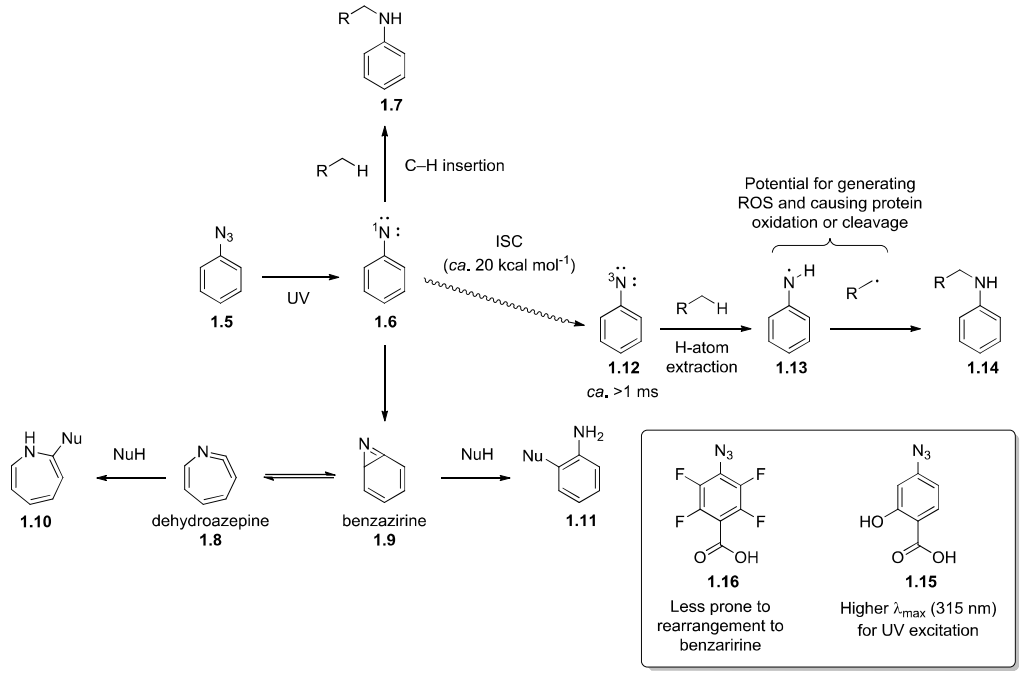


Figure 1.9 Overview of the photochemistry of aryl azides upon UV excitation.[26]

Species **1.6**, **1.12** and **1.8** have all been detected, and their respective ratios and lifetimes are highly influenced by substituents on the aromatic ring, aryl azide concentration, buffer components, temperature and irradiation wavelength.[90,91] Extensive studies have been conducted to find aryl azides that have improved photoreactive properties by installing various substituents on the aryl ring.[92,96] For example, *meta*-hydroxy substitution red-shifts the excitation wavelength (λ_{max} = 315 nm for **1.15**) which is less damaging to the biological system being irradiated.[91,97] Tetrafluoroaryl azide **1.16** has been shown to have a slower rate of rearrangement to the dehydroazepine, and therefore has a longer time to perform C–H insertion reactions from the singlet nitrene state. Additionally, when the dehydroazepine does form, it was

found to be more electrophilic (*versus* the non-fluorinated analogue) and may give higher levels of labelling through the nucleophilic addition pathway.[74,98] Thus, through this complex range of possible intermediates, aryl azide probes can behave as ideal photoreactive groups and label the target protein through fast and unbiased reactions (singlet nitrene C–H insertion) or label more regioselectively (amine labelling *via* the dehydroazepine intermediate). Due to these complex labelling pathways, the labelling efficiency is challenging to predict and should be found experimentally for each protein target.[74]

1.2.2.2 Diazirines

Since their introduction as photolabelling reagents by Smith and Knowles in 1973, aromatic diazirines have become one of the most widely used photoreactive groups.[99] One advantage offered by diazirines is that their excitation wavelength occurs between 350–380 nm, which is above the wavelengths that can damage biological systems (<300 nm).[26] However, intense irradiation (*ca.* 10 W cm^{-2}) is required to achieve fast photoactivation rates.[100] The photochemistry of aromatic diazirines is summarised in Figure 1.10.[26] Upon UV excitation, 3*H*-phenyldiazirine **1.17** produces a desired singlet carbene **1.18** (*ca*. 70%) and an undesired diazo intermediate **1.19** (*ca*. 30%) (Figure 1.10 A). The lifetime of this diazo intermediate usually exceeds the dissociation rate of the probe from the target protein and can result in non-specific labelling. Richards and co-workers found that the diazo intermediate **1.20** produced from aryl-CF_3 diazirine **1.21** was sufficiently stabilised by the electron-withdrawing trifluoromethyl moiety and showed no non-specific labelling under biochemical conditions (Figure 1.10 B).[101] Thus, **1.21** has become the most popular aromatic diazirine used in photolabelling studies.[101] From UV excitation of **1.21**, singlet carbene **1.22** is formed (*ca*. 65%) and is extremely reactive, with a lifetime of *ca*. 1 ns.[102] **1.22** can perform insertion reactions with N—H, O—H or C—H bonds to result in labelling, however the stability of these photo-adducts (**1.23**, **1.24** and **1.25** respectively) varies. **1.23** has been shown to release HF, and subsequent hydrolysis of the resulting enamine **1.26** under biochemical conditions gave the ketone **1.27**, with the loss of labelling.[103] Insertion into O—H or C—H bonds produces stable photo-adducts (**1.24** and **1.25**),

and a slight preference for O—H insertion has been observed.[103] In competition with these insertion events, the singlet carbene **1.22** can relax to a triplet ground state (**1.28**) *via* intersystem crossing.[104] This triplet carbene performs radical-like reactions, and can perform hydrogen abstraction to form **1.29**. The resulting radicals can recombine to give a labelling event (**1.25**), or a second abstraction can occur, resulting in no labelling and potential damage/oxidation of the protein through radical cascades.[26] Furthermore, the triplet carbene can be quenched by reaction with water or dissolved oxygen, both resulting in no labelling (**1.30**).[75]

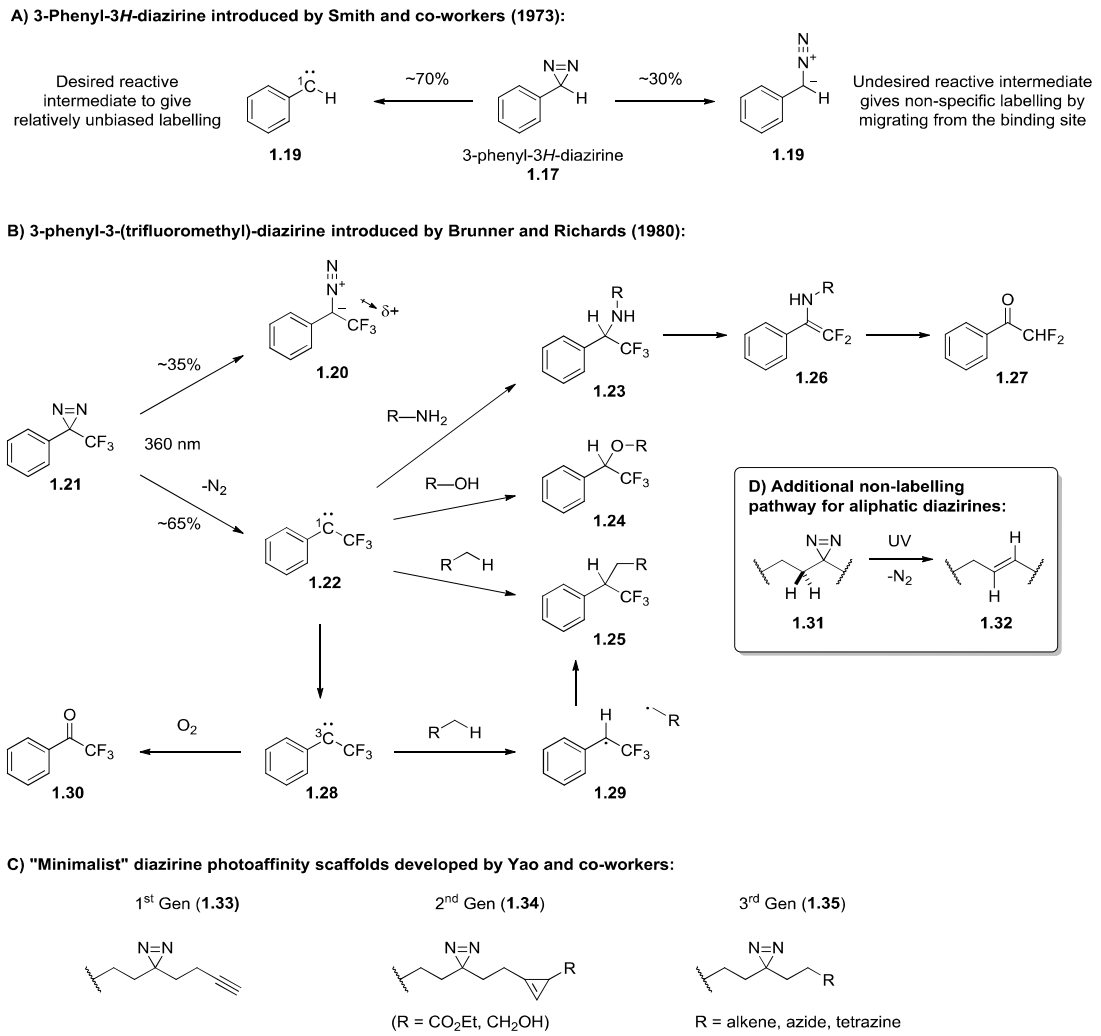


Figure 1.10 Overview of diazirine photochemistry used in PAL studies. A) 3-Phenyl-3*H*-diazirine **1.17** first introduced by Smith and co-workers. B) Photochemistry of the most popular aromatic diazirine **1.21**. (C, D) Despite less efficient labelling, there has been renewed interest in aliphatic diazirines due to their small size and minimal perturbation to the overall physicochemical properties of the PAL probe.

The possible reactions of aliphatic diazirines are similar, however, they generally produce higher levels of the undesired diazo intermediate over the preferred singlet carbene compared to aromatic diazirines.[62,105] Furthermore, aliphatic diazirines (**1.31**) can also undergo a 1,2-hydride shift from the intermediate carbene upon UV excitation to produce alkenes (**1.32**), resulting in no labelling.[106] Despite this, aliphatic diazirines have recently become one of the most popular photoreactive groups due to their small size and minimal impact on the overall physicochemical properties of the PAL probe.[107-109] The Yao group have developed various generations of "minimalist" diazirine photoreactive scaffolds (**1.33–1.35**), containing an aliphatic diazirine and click handle (Figure 1.10 D).[21,22,36,110] Each generation was produced to address new and advancing bio-orthogonal reactions.[36] Chosen for its small size, **1.33** has been recently incorporated into fragments by Cravatt *et al.* to give fully-functionalised fragments (FFFs), capable of labelling and reporting on target proteins in live cells.[18] The resurgent interest in aliphatic diazirines has produced more straight-forward syntheses and incorporation strategies.[107,111] The chemoselective preferences for aliphatic diazirines to insert into various amino acids has been studied at the primary level using peptides.[112] Labelling of all 20 naturally occurring amino acids was found, however there was a moderate chemoselectivity for Tyr, Glu, Asp and His residues. Higher levels of labelling (*ca.* 6-fold) could be achieved when the samples were flash-frozen (77 K).[112-114] An analogous study with intact proteins has not yet been reported, and would be expected to be highly dependent on the protein being studied.[112] Therefore, it has been suggested that the aliphatic diazirine should be incorporated in a variety of positions on the affinity function to sample for efficient levels of labelling.[107]

1.2.2.3 Benzophenone

Benzophenone has become a popular photoreactive group, and similar to the diazirines, can be excited using wavelengths above those that can cause damage to biological systems (λ_{max} *ca.* 350 nm). The photochemistry of benzophenone (**1.36**) relevant to PAL applications is summarised in Figure 1.11.

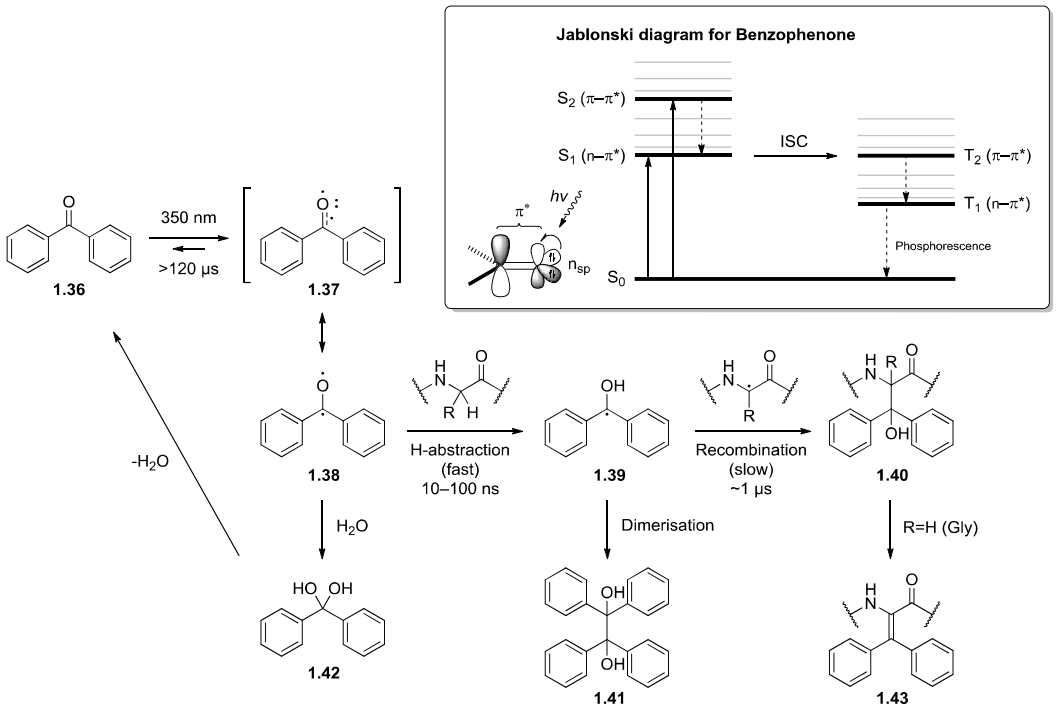


Figure 1.11 Photochemistry of benzophenone.[115]

An electron is excited from the non-bonding orbitals on oxygen to the antibonding π* orbital (n–π*) to form the lower energy singlet state S_1. An electron can also be excited from the bonding π orbital to the π* (π–π*) to form a higher energy S_2 state. The S_1 state is energetically close to the triplet T_2 (π–π*) state, and intersystem crossing readily occurs.[116] Relaxation to the lower energy T_1 (n–π*) state then occurs, and the majority of photochemical reactions originate from this state.[115] In this T_1 (n–π*) state described by **1.37**, an electron occupies a non-bonding (n_{sp2}) orbital on oxygen, and the other electron occupies the π* orbital. This imparts biradical character to the carbonyl bond (more commonly drawn as **1.38**), and the electron on oxygen can abstract hydrogen from a suitable donor to form the carbon-centred diphenyl radical **1.39**. In biological systems, this donor can be any moiety that forms a stabilised radical after abstraction (α-hydrogens of the peptide backbone, benzylic or tertiary centres, or hydrogens on amino acid residues with adjacent electron-donating heteroatoms).[26] For the resulting radicals to recombine and give covalent labelling (**1.40**), one of the radicals must intersystem cross to the opposite spin. This is forbidden by the spin selection rule and so is considered the rate-limiting step.[115] Alternatively, the diphenyl radical **1.39** can dimerise to form **1.41**, however this process has been observed to be

ca. 10-fold slower than the desired recombination reaction.[117] The diradical **1.38** can be scavenged by water to form diol **1.42**, which quickly dehydrates to re-form benzophenone **1.36**. Furthermore, if the diradical **1.38** fails to find a reactive partner, the T_1 (n–π*) state can relax back to the S_0 ground state (**1.36**). This recycling of the photoreactive group under irradiation is a key advantage of benzophenone *versus* the aryl azides or diazirines, which irreversibly expel nitrogen to form their reactive species.[63]

Benzophenone can label all amino acids, however numerous experiments have demonstrated a high chemoselectivity for methionine (up to 10-fold) at the γ- or ε-carbons (either side of the δ-sulfur).[76,115,118-120] This selective preference was profound, whereby benzophenone was found to label methionine up to 11 amino acids away from the probe's actual interaction site.[118] Therefore care must be taken when determining PAL probe binding sites through benzophenone labelling.[115] Moderate chemoselectivity for glycine has also been observed. After radical recombination with glycine, elimination was observed to give **1.43**.[26]

1.2.3 Comparison of photoreactive group properties

A comparison of the three most commonly used photoreactive groups (aryl azides, aryl and aliphatic diazirines, and benzophenone) is given in Table 1.3.[39,102,115,121]

Table 1.3 Comparison of the most commonly used photoreactive groups.

	Photoreactive group			
Property	**Aryl azides**	**Aryl diazirines**	**Alkyl diazirine**	**Benzophenone**
Irradiation wavelength	260–340 nm	350 nm	350–380 nm	350–360 nm
Mechanism of activation	Irreversible (-N_2)	Irreversible (-N_2)	Irreversible (-N_2)	Reversible
Desired reactive species	Singlet nitrene	Singlet carbene	Singlet carbene	Diradical triplet
Availability/ease of synthesis	++	+	++	+++
Stability	+	++	++	+++
Impact on physicochemical properties	++	++	+	+++
Amino acid preference	Amine residues (*via* dehydroazepine)	Cys, Trp, Tyr, His, and Phe residues	Tyr, Glu, Asp and His residues	High preference for Met

The development of new photoreactive functionalities for PAL is a continuing area of research. Moieties such as *α*-pyrones and pyrimidones, *N*-aryl tetrazoles and thienyl-substituted α-ketoamides have been recently shown to be suitable PAL reagents.[122-126] As these new groups are explored further, they may become complementary to the well-established groups discussed in Table 1.3.

1.2.4 Detection methods for identifying photocrosslinked target proteins

A variety of tags for detection or downstream manipulation of PAL captured proteins are available. Initially, radiolabels were incorporated into the PAL probe (*e.g.* ^{125}I, ^{3}H or ^{14}C).[88,127] The proteins were separated using polyacrylamide gel electrophoresis (PAGE) and protein bands that showed radioactivity corresponded to covalently modified target proteins. This method is highly sensitive, however requires specialised handling due to the radioactivity of the labels. A more straightforward method of gel-based detection involves incorporating a fluorescent dye into the PAL probe. The bands that fluoresce after gel electrophoresis report on the covalent labelling by the PAL probe. This method is known as in-gel fluorescence scanning.[18] To avoid the fluorescent dye impacting on the binding mode of the affinity function of the PAL

probe, the dye is more commonly incorporated using bio-orthogonal click chemistry after photocrosslinking has taken place.[17,21] Smaller click handles (such as alkynes) can be used, which offer minimal impact to the physicochemical properties, cell permeability and binding mode of the PAL probe.[36,128,129] Biotin can also be installed *via* click chemistry after irradiation.[39] Biotinylated proteins are then immobilised on beads coated with streptavidin. Biotin binds to streptavidin with one of the highest non-covalent affinities known ($K_d = 10^{-15}$ M).[130] This allows PAL captured proteins to be enriched from complex biological mixtures, such as cell lysate. The enriched proteins are then digested and identified using bottom-up proteomics. This technique is discussed in detail in Section 1.3.

Intact protein LCMS can be used to determine photocrosslinking yields to recombinant protein.[69] One advantage of this method is that it is handle-free, therefore simplified PAL probes containing only an affinity function and photoreactive group can be used. Furthermore, this method can identify the number of photocrosslinking events that have occurred and can inform on any other covalent modifications (*e.g.* oxidation) to the protein after the irradiation event.

1.2.4.1 Intact protein LCMS

Intact protein LCMS with electrospray ionisation (ESI) is a useful method used to identify and quantify the extent of covalent modifications to recombinant proteins.[131] An example workflow is shown in Figure 1.12. Volatile mobile phases are used to resolve the protein sample by reverse-phase chromatography under denaturing conditions. Volatile acidic modifiers (for positive mode MS) are used to increase chromatographic resolution and to assist in the positive ionisation of the analyte.[132] Gas phase ions are produced from the continuous eluent by ESI, where the eluent is nebulised into fine droplets from a capillary spray tip.[133] An electric field is created by applying a high potential difference between the spray tip and the source block (*ca.* 3–5 kV).[134] This imparts a positive charge to the surface of the emerging droplets. The source block and nitrogen drying gas are heated (*ca.* 150–400 °C). Under these conditions, solvent is rapidly evaporated from the droplets. The protein analyte picks up positive charges to form a distribution of singly, doubly or multiply charged ions

($[M+nH]^{n+}$).[134] For intact proteins, the number of charges that can be acquired is roughly equal to the number of basic sites (Arg, Lys, His and terminal NH_2).[132] As the droplets reduce in volume, the charge density on the surface of the droplet is increased to the point where charge repulsion overcomes the surface tension, and the droplet emits ions in an event known as a coulombic explosion.[133] The gaseous ions are then directed into the instrument for analysis. On average, one proton is picked up for every 1000 Da of intact protein molecular mass. Therefore, a distribution of charge states is observed, typically falling between *m/z* 800–3000. Time-of-flight (ToF) mass spectrometers are preferred for intact protein analysis, as these instruments can scan this full range with high accuracy faster (≤ 1 second) than other types of mass analysers (*e.g.* triple-quadrupole).[131,132]

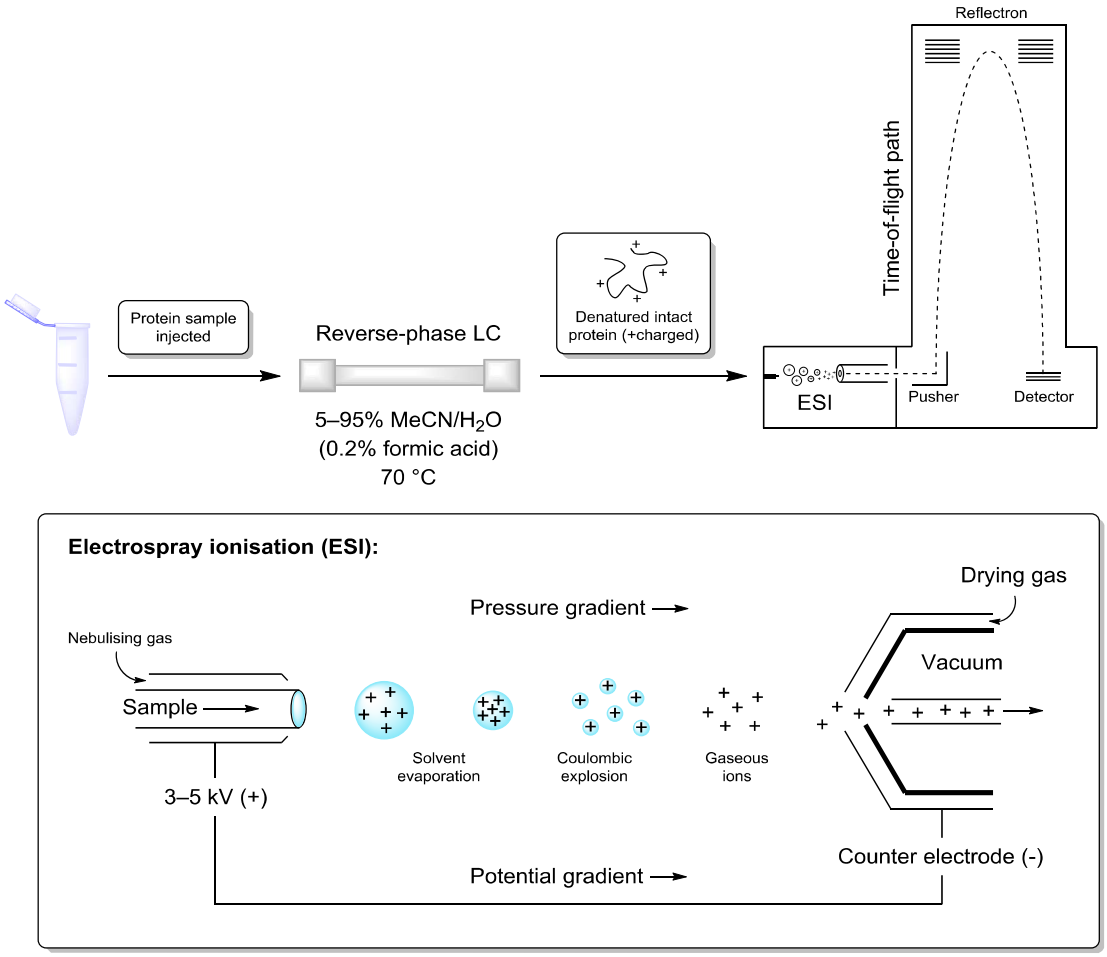


Figure 1.12 Schematic of intact protein LCMS with electrospray ionisation (ESI) and time-of-flight (ToF) analysis.[132,134]

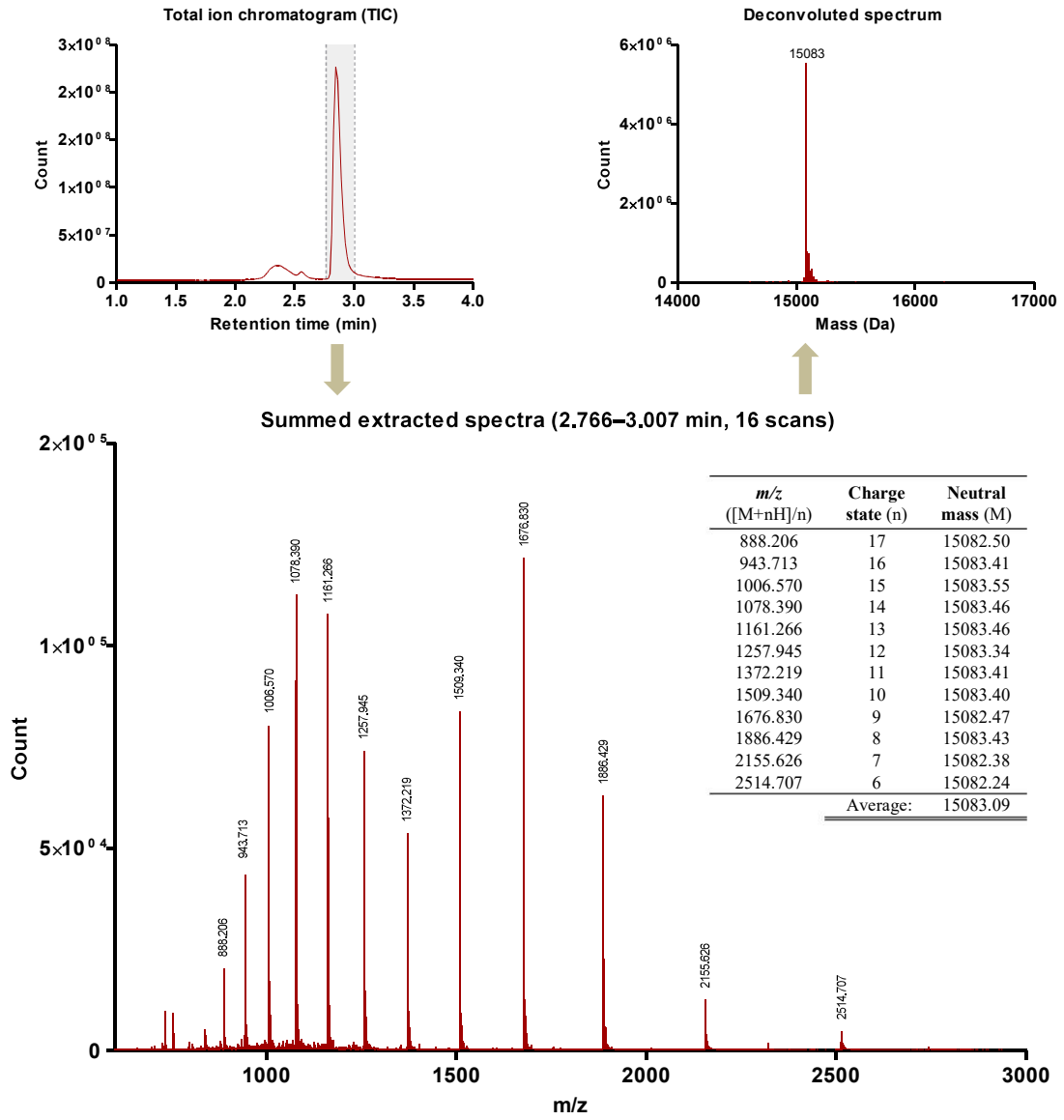
For ToF analysis, the ions are accelerated to a high kinetic energy using a pulse from an electric "pusher" field. The ions are directed into a field-free flight tube, focused and reflected by an electric field (reflectron) and then detected. Ions of the same charge are given the same kinetic energy by the potential energy pulse of the pusher field. The time for the ions to travel the fixed distance from the pusher pulse to the detector is directly proportional to the square root of the mass-to-charge ratio. The range of arrival times is recorded as one scan and converted to a range of *m/z* values following Equation 1.1.

Equation 1.1

$$\frac{m}{z} = \frac{(2U)t^2}{d^2}$$

where:
m/z = mass to charge ratio
U = potential energy applied (pusher field)
t = arrival time
d = flight path

An example of intact protein LCMS analysis is shown in Figure 1.13. The eluent from the LC is continuously scanned to produce a total ion chromatogram. The scans that contain the distribution of charge states for the protein analyte are selected and summed. A maximum entropy (MaxEnt) algorithm is then used to deconvolute the observed *m/z* values to give a deconvoluted mass spectrum reporting the neutral masses of the protein analytes present.[135-137]



m/z ([M+nH]/n)	Charge state (n)	Neutral mass (M)
888.206	17	15082.50
943.713	16	15083.41
1006.570	15	15083.55
1078.390	14	15083.46
1161.266	13	15083.46
1257.945	12	15083.34
1372.219	11	15083.41
1509.340	10	15083.40
1676.830	9	15082.47
1886.429	8	15083.43
2155.626	7	15082.38
2514.707	6	15082.24
	Average:	15083.09

Figure 1.13 Example data obtained using intact protein LCMS analysis. The scans within the TIC peak are summed to give the distribution of charge states for the protein analyte. This data is deconvoluted using a maximum entropy algorithm to produce a mass spectrum of neutral masses present in the sample.

Intact protein LCMS can usually produce deconvoluted spectra with an accuracy of *ca*. ±1 Da for every 1 kDa of molecular mass using current ToF instruments. Thus, this analytical method is highly suitable for determining covalent modifications to recombinant proteins, such as post-translational modifications, or chemical modifications by small molecule compounds.[79,131] In particular, intact protein LCMS is well-suited to determining the photocrosslinking yield of PAL probes to recombinant target proteins of interest.[69]

1.3 Bottom-up proteomics

Since the turn of the century, tandem mass spectrometry has become the gold standard analytical technique for chemoproteomics. This has been due to advancements in nanoscale liquid chromatography (nLC) and electrospray ionisation (nESI) sources, higher resolution mass spectrometers and improved bioinformatics for protein inference from peptide fragmentation spectra. For the identification of proteins from complex biological mixtures, bottom-up proteomics has become the dominant method.[31] Bottom-up proteomics (or shotgun proteomics) refers to the identification of intact proteins based on the indirect analysis of their peptides produced by proteolytic digestion.[138,139] Proteins are digested with a protease that cleaves at specific amino acid recognition sites. Trypsin is the most commonly used protease for a number of reasons.[132] Trypsin specifically cleaves the amide bond at the carboxyl end of lysine and arginine (but not if these amino acids are followed by proline). Based on the frequency of Lys and Arg, this produces peptides with average lengths suited to MS analysis. Furthermore, cleaving after Lys and Arg ensures that at least one protonatable amino acid is present on each peptide to aid ionisation. To aid complete digestion and refine the distribution of peptide lengths, endopeptidase LysC (which cleaves at lysine, even if followed by proline) is commonly used in combination with trypsin. The sample to be analysed can range in complexity. For example, digestion of *ca.* 200–2000 proteins enriched by a capture-based method may produce *ca*. 4000–40000 peptides. Whole proteome analyses conducted in capture-free proteomics may contain $>10^6$ peptides. For these large analyses, samples require pre-fractionation to reduce complexity, and each fraction (*ca*. 20–40) is run separately. Thus, for capture-free methods, extensive total acquisition times are required (24–48 h). For capture-based methods, a 1–3 h nLC run from a single sample usually provides enough chromatographic resolution for modern instruments to perform robust peptide identification with a data-dependent acquisition (DDA) method. As peptides elute from the nLC, they are ionised by nESI and detected in a full MS scan (MS1). This survey scan identifies the top N most abundant peptide precursor ions, which are then selected in turn and fragmented. The fragment ions are scanned (MS2) to provide a fragmentation spectrum for each precursor peptide ion.

The orbitrap is a relatively new mass analyser, first reported by Makarov in 2000.[140-142] Since then, orbitrap-based mass spectrometers have become the most powerful and routinely used instruments for bottom-up proteomics due to their high accuracy, resolution, sensitivity and dynamic range.[132] An orbitrap mass analyser is used in tandem with other mass analysers, such as a linear ion trap (LIT) in earlier Thermo LTQ instruments, or with a quadrupole in higher performing instruments such as the Thermo Q Exactive.[143-145] First unveiled in 2011, the Q Exactive series of instruments have become the workhorse for MS-based proteomics.[146] Taking the Q Exactive mass spectrometer as an example, a detailed review of how these hybrid instruments can analyse tens-of-thousands of peptides per sample is given in Section 1.3.1.

1.3.1 LC-MSMS analysis using a hybrid quadrupole-orbitrap mass spectrometer (Q Exactive)

A detailed schematic of a Q Exactive hybrid quadrupole-orbitrap mass spectrometer is shown in Figure 1.14.[145]

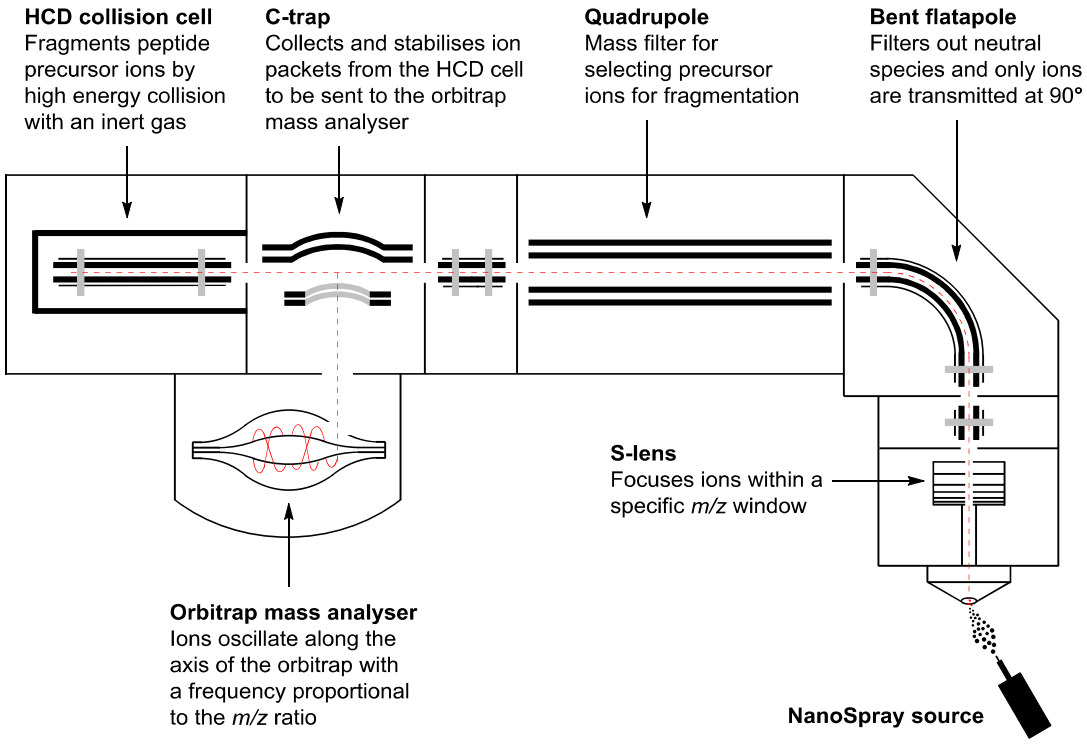


Figure 1.14 Schematic of a Q Exactive hybrid quadrupole-orbitrap mass spectrometer.[29]

The instrument first conducts a full MS survey scan as follows. As peptides ions are continuously produced from the nanospray source, they are focused into an ion beam by a series of stacked ion lenses (S-lens). The ions then enter a bent flatapole, where only charged ions are selected and neutral uncharged species are filtered out. The ions then enter the quadrupole. In full scan mode, the quadrupole performs no mass selection and all the ions are passed through to the C-trap. The ions are collected in static packets in the C-trap and stabilised, before being ejected into the orbitrap for analysis. Within the orbitrap, the ions are trapped by an electrostatic field in an orbit around a central spindle electrode. Ions oscillate along the spindle in the axial direction with a frequency dependent on the *m/z* ratio of the ion. The oscillating ions induce a current in the detector, and this current signal is Fourier transformed (FT) to produce *m/z* values for the ions present. Resolution is dependent on the length of time the ions are kept in the trap.[142] For a typical full MS survey scan, the ions are kept in the trap for 256 ms, which gives a resolution of 50,000 (at *m/z* 400), which is appropriate for proteomic applications.[145] The FT produces a survey scan, and within microseconds, the data-dependent algorithm selects the top N most abundant precursor ions. Most commonly, the top 10 most abundant ions are selected. The instrument then switches to MSMS mode. The selected precursor ion is selected by the quadrupole mass filter. The quadrupole consists of four parallel rods, to which a radiofrequency (RF) voltage and direct current (DC) voltage are applied. At a specific set of voltages, only certain *m/z* are transmitted through the quadrupole. Any other ions with a *m/z* different than the selected *m/z* spiral and collide with the rods. The selected ions are passed through to the higher energy collisional dissociation (HCD) cell, where the ions collide with an inert gas (nitrogen or argon) and fragment across amide bonds (Figure 1.15). If the charge is retained on the amino-terminal side of the fragmentation, a b-ion is produced. Inversely, if the charge is retained on the carboxy-terminal of the fragmentation, a y-ion is produced.[139]

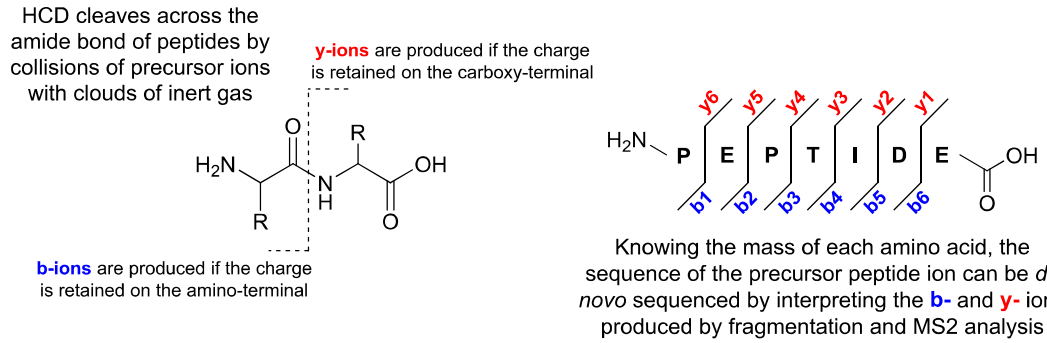


Figure 1.15 Fragmentation pattern induced by higher energy collisional dissociation (HCD).[139]

These fragmentation ions are then collected back into the C-trap, focused, and then ejected into the orbitrap for analysis. This MSMS fragment ion scan is performed within 64 ms (which gives a resolution of 12,500 at *m/z* 400). While this first MSMS scan is being acquired, the voltages of the quadrupole have already changed to select for the second most abundant precursor ion. Thus selection, fragmentation and collection in the C-trap occurs for the next selected precursor ion while the 64 ms scan of the fragment ions from the previous precursor ion is being performed. The instrument cycles through the top 10 precursor fragment ions in this fashion to complete one duty cycle, after which the instrument will perform another full MS survey scan and begin the cycle again. One complete duty cycle involving a full MS survey scan and ten MSMS fragment ion scans takes *ca*. 1 second to complete. As new peptide ions elute from the nanospray, the survey scan updates the list of top 10 precursor ions. If a precursor ion has already been fragmented in the previous duty cycle, the algorithm can ignore it and select the next most abundant ion for fragmentation. The data-dependent algorithm also includes exclusion criteria, such as only selecting for doubly charged (or greater) precursor ions for fragmentation to give more informative MSMS fragmentation spectra. This maximises the number of unique MSMS fragmentation spectra collected for in-depth identification of the peptides present in the sample. By performing fast duty cycles (*ca*. 1 sec for 10 MSMS spectra), intelligently selecting which peptide ions to fragment (DDA method) and being able to perform ion selection and scanning in parallel, the instrument can collect MSMS spectra fast enough to provide analytical coverage of the peptides continuously eluting from the nLC column. For a typical 90 min nLC method of a complex biological sample (*e.g.* peptides from a capture-based experiment) *ca*. 5000 full survey MS1

scans are collected, with up to 50000 accompanying MS2 scans. Bioinformatics packages are then used to infer protein identification from this large data set of collected MS2 fragmentation spectra.

1.3.2 Protein inference from peptide fragmentation spectra

Bioinformatics packages such as SEQUEST, MaxQuant/Andromeda or Mascot are used to statistically match the experimental spectra to theoretical fragmentation spectra.[147-150] These theoretical fragmentation spectra are produced by *in silico* digestion of the known proteome contained in a database (*e.g.* UniProt). This provides a candidate list of possible peptides for each experimental spectrum obtained. These are ranked and filtered based on statistical cut-offs to produce a peptide spectrum match (PSM). To test the selection criteria, the experimental spectra are also matched to a decoy database, which is usually obtained by scrambling or reversing the protein sequences of the true database.[151] The number of matches between the experimental spectra and this decoy database provides a measure of false discovery rate (FDR) for the true database matches. If this rate is too high, the selection criteria for a PSM can be made more stringent to bring the FDR between 0.1–1%. One disadvantage of bottom-up proteomics is that digestion produces a large number of redundant peptides that are shared between proteins that have high sequence homology. Therefore, assignment can only be performed by matching unique peptides.[31] Two or more unique PSMs are generally required to identify a protein. Most biological applications of bottom-up proteomics require the measure of relative protein abundances between two or more different states (*e.g.* diseased *versus* healthy, or compound treatment *versus* DMSO control *etc.*). To meet this requirement, various methods for quantitative MS-based proteomics have been developed.

1.3.3 Relative quantification of proteins by LC-MSMS

A variety of methods have been developed to enable relative quantification of proteins in MS-based proteomic experiments.[152-156] The two most commonly used methods are

stable isotope labelling of amino acids in cell culture (SILAC) and isobaric mass tag labelling.[157,158]

In a SILAC approach, one population of cells are grown in a standard media. These cells are known as the "light" population. A second population of cells are grown in media containing the "heavy" amino acids $^{13}C_6$–Lys and $^{13}C^{15}N_4$–Arg. These isotopically labelled amino acids are incorporated into the cell population through metabolism. The cells go through several rounds of passage in the presence of the heavy media to maximise the incorporation of the heavy amino acids across the proteome. Each population of cells is treated with a condition (*e.g.* compound treatment *versus* DMSO control). The capture and enrichment steps are performed on both samples, and the protein concentration in both samples are normalised. The samples are then combined and digested before LC-MSMS analysis. Each peptide is relatively quantified between the two samples by comparing the intensity of the two isotopic paired peaks in the MS1 spectrum. To remove false positives, a "flip" experiment is usually performed whereby the treatment conditions are used on the opposite heavy or light cell populations. The intensities for the heavy and light isotopic peaks for each peptide should be directly inverse to the ratio of intensities observed in the first experiment. The duplication of cell culture activities is one disadvantage of using the SILAC approach. Furthermore, this approach only allows for relative quantification of proteins across two conditions (in duplex) per LC-MSMS analysis.

Isobaric mass tagging using tandem mass tag (TMT) reagents has become a prevalent strategy to provide multiplexed relative quantification of proteins during LC-MSMS analysis (Figure 1.16).[159] Each TMT reagent consists of an isotopically distinct mass reporter group, mass balance group, and reactive *N*-hydroxysuccinimide (NHS) ester. For a TMT 6-plex experiment, the peptide abundance for six different samples can be relatively quantified. Each sample is treated with a different TMT label after digestion. The reagent labels the free amino termini and lysine residues of the peptides by amide coupling with the NHS ester. The six samples are then combined and analysed in a single LC-MSMS experiment. Each TMT tag has an identical (isobaric) total molecular mass, and so the same peptides from each sample elute at the same retention time and are indistinguishable by molecular mass in the MS1 spectrum. When this peptide precursor ion is selected for fragmentation and MSMS analysis, the TMT

labels present also fragment across the internal amide bond present in the label. This produces an isotopically distinct “reporter” ion for each TMT label. The intensities of the reporter ions are compared, which gives a relative quantification of the peptide (and inferred protein) present in each of the six samples. By varying the distribution of carbon and nitrogen isotopes but maintaining the overall isobaric mass of the TMT label, TMT 10-plex, and most recently TMT 11-plex reagents can be used to give high multiplexing capability for MS-based proteomic experiments.[160,161]

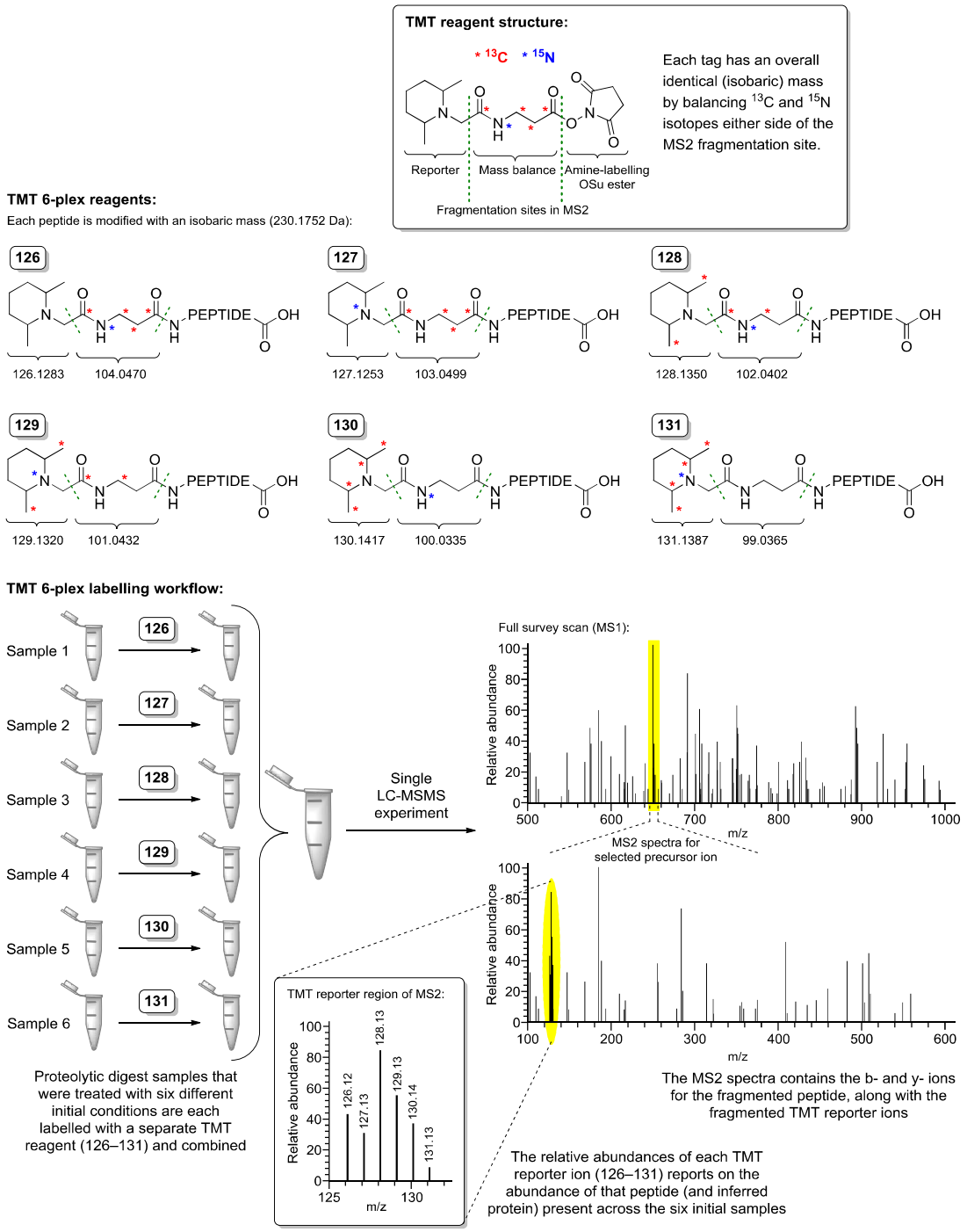


Figure 1.16 Schematic of TMT 6-plex labelling which allows for proteins to be relatively quantified across six different samples in one LC-MSMS experiment.[159]

1.3.3.1 Determining the site of photocrosslinking by LC-MSMS

Proteolytic digestion and LC-MSMS analysis can be used to identify the site of photocrosslinking, providing information on the binding site of the PAL probe to the target protein of interest.[26] This information is particularly useful if little is known about the interaction of a compound of interest and the target protein (*e.g.* HTS or fragment hits with novel binding modes), or if other methods to determine the binding site of compounds (*e.g.* X-ray crystallography) prove difficult. Typically, photocrosslinked target protein is purified by polyacrylamide gel electrophoresis (PAGE) and digested with a specific protease (most commonly trypsin and LysC). The peptides are analysed using LC-MSMS in data-dependent acquisition mode (Figure 1.17). Software packages such as Mascot can then be used to search for the site(s) of covalent modification. The amino acid sequence of the protein is entered, and a list of expected peptides is generated *in silico*. The mass of covalent modification by the PAL probe is entered as a possible covalent modification. The software then searches and identifies any MS1 precursor ions which have an *m/z* value corresponding to any of the expected peptides + the mass of the covalent modification. The b- and y-ions in the MS2 spectra for these precursor ions are then examined to identify which amino acid residue was modified by the PAL probe.

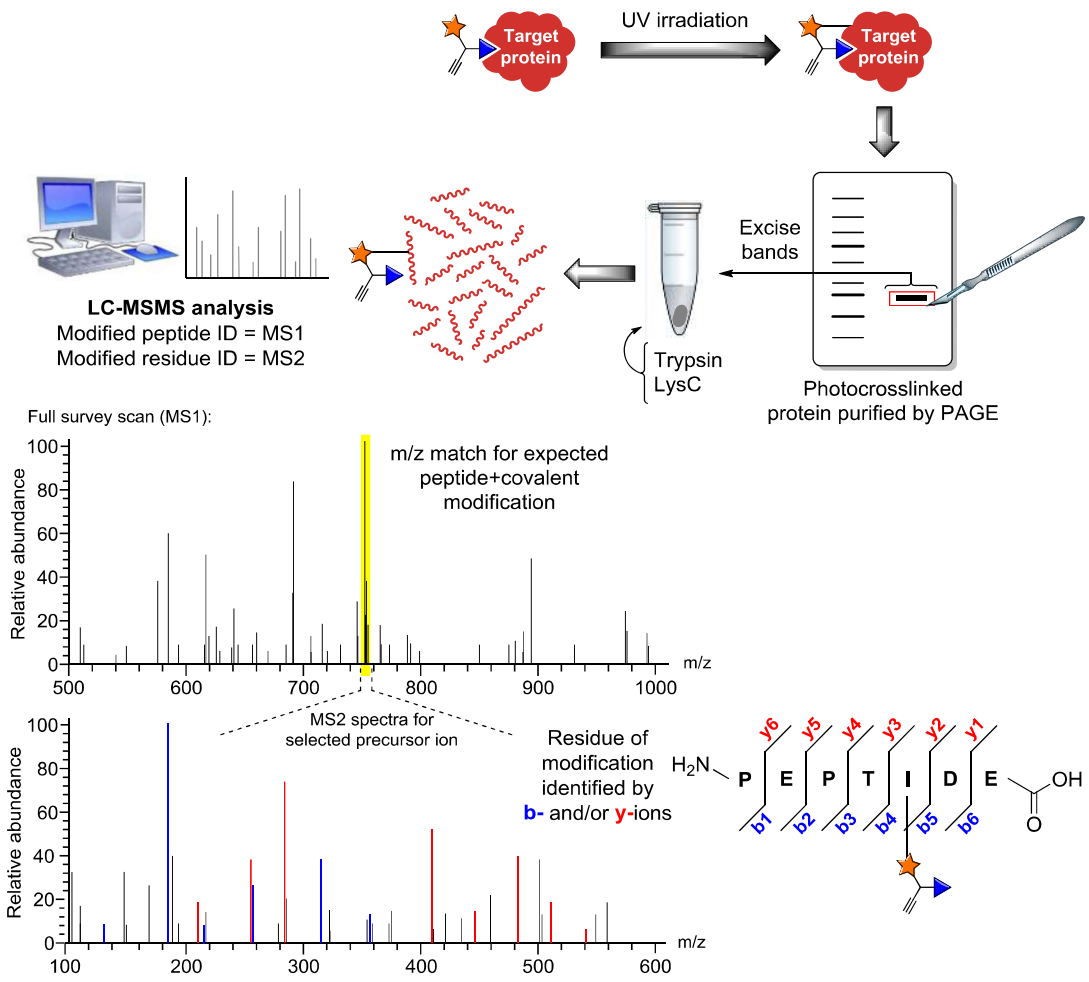


Figure 1.17 Workflow for identifying the site of photocrosslinking using tryptic digestion and LC-MSMS analysis.

1.3.4 Summary of PAL applications in small molecule drug discovery

PAL can be used with a variety of analytical techniques to investigate the affinity and specificity of various affinity functions of interest to target proteins (Figure 1.18). Intact protein LCMS can be used to determine the photocrosslinking yields of PAL probes to recombinant target proteins. Tryptic digestion and LC-MSMS analysis can be used to identify sites of photocrosslinking, which provides information on the binding site of the affinity function to the target protein. Bio-orthogonal click chemistry can be used to install purification handles such as biotin, enabling pulldown experiments with bottom-up proteomics to identify on- and -off target proteins from lysates or live cells. Alternatively, fluorophores can be installed, enabling

fluorescence-based detection methods such as in-gel fluorescence scanning or fluorescent cell imaging.

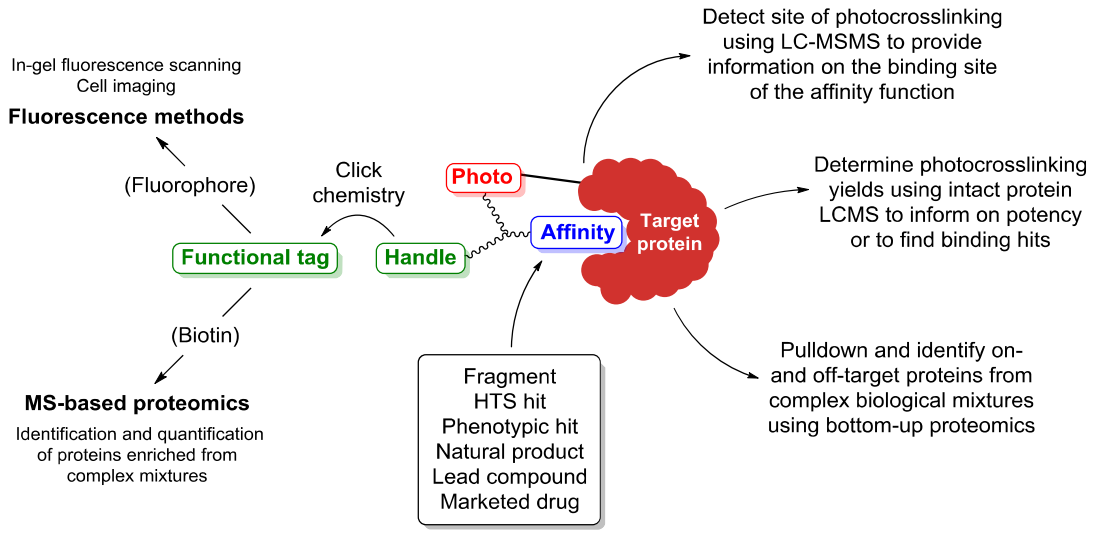


Figure 1.18 Overview of PAL applications and accompanying analytical techniques in small molecule drug discovery.

One limitation of the PAL approach is that photocrosslinking yields are often low, and it is not currently possible to predict high levels of photocrosslinking. The efficiency of labelling is dependent on which photoreactive group is used and how each photoreactive group engages with the topology of the protein around the active site. This relationship changes with each target protein and linker length from the affinity function. Furthermore, different photoreactive groups will have a range of impacts on the physicochemical properties of the PAL probe, which are important to consider if the PAL probe is to be used in live cells. Therefore, it is advised to synthesise a range of PAL probes and test for efficient levels of photocrosslinking and physicochemical properties.[26,39,69,74-76] Methods to synthesise PAL probes in one step with wide compatibility for various affinity functions, photoreactive groups and handles are lacking. The current approaches to PAL probe synthesis are reviewed in Section 1.4.

1.4 Synthetic strategies to access PAL probes

There are two general strategies used to construct PAL probes; 'embedding' or 'appending' the photoreactive functionality to the compound of interest.[14,15,107] Using

the embedding strategy, a photoreactive group is incorporated within the affinity function. (Figure 1.19) The group is installed in a way that minimises any changes to the binding mode. For example, installing a trifluoromethyl aryl diazirine (**1.21**) in place of a phenyl ring. This strategy maximises the potential contact between the photoreactive group and the surface of the binding site. If a bio-orthogonal handle is required, it is installed in a solvent-exposed vector that has minimal effect on the overall binding affinity of the PAL probe. Muranaka *et al.* adopted this strategy to design the PAL probe **1.44** from a known ligand **1.45** of the human adenosine A_{2A} receptor (Figure 1.19, right).[162] **1.44** was used to interrogate the site of binding by **1.45** to the A_{2A} receptor by using tryptic digestion and LC-MSMS analysis. One caveat of this approach is that it is specific to the affinity function of interest and requires a bespoke synthetic strategy for each incorporation. It can also be challenging to incorporate the photoreactive group without having a large influence on the binding mode of the affinity function.

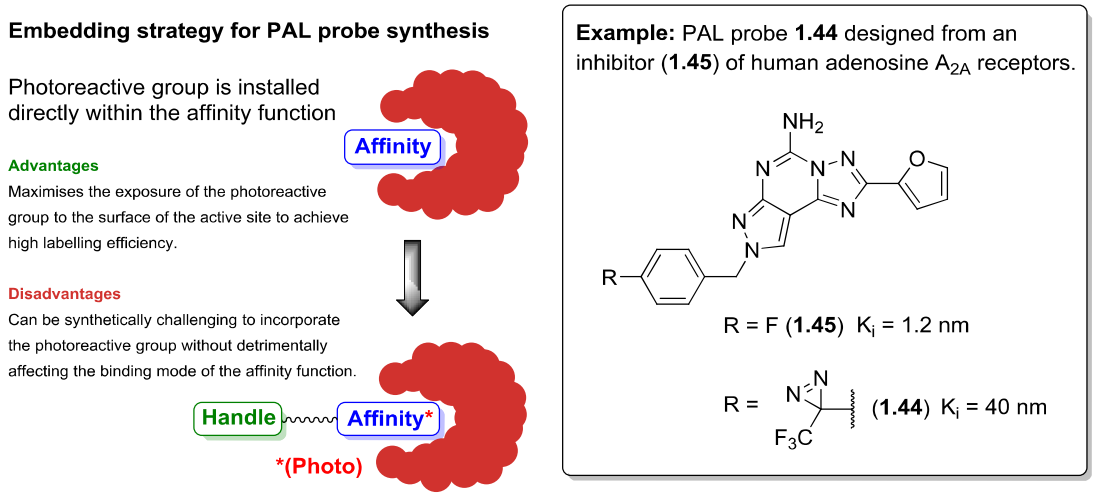


Figure 1.19 Embedding the photoreactive group within the affinity function by replacing a similar moiety is one strategy used to synthesise PAL probes.[107] This strategy maximises the contact between the photoreactive group and active site to achieve high labelling efficiencies, however it can be difficult to incorporate the photoreactive group without having a large impact on the binding affinity of the affinity function. The synthetic difficulty to incorporate the photoreactive group is also dependent on the affinity function being modified.

Alternatively, the photoreactive group and bio-orthogonal handle can be appended to a solvent-exposed vector of the affinity function together in the form of a photoaffinity scaffold (Figure 1.20). This strategy is less dependent on the affinity function of

interest, as the photoaffinity scaffold can be synthesised separately and then coupled to a solvent-exposed vector of the affinity function. One issue with this strategy is that by placing the photoreactive group outside of the binding pocket, there is an increased likelihood that the photoreactive group will be quenched by reaction with the solvent or buffer components. Despite this issue, this approach has become the most popular method to constructing PAL probes due to its general applicability to various affinity functions.

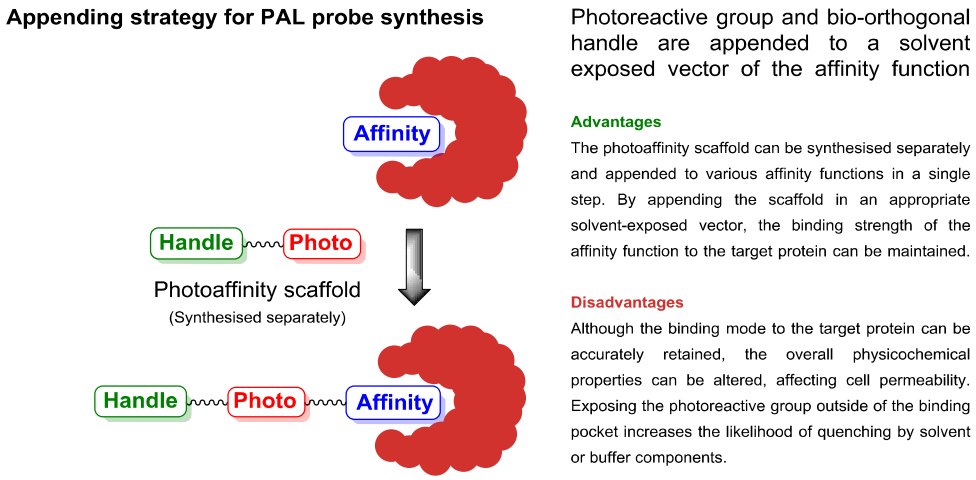


Figure 1.20 The 'appending' strategy for PAL probe synthesis.[107] The PAL scaffold containing the photoreactive group and bio-orthogonal handle is synthesised separately and then coupled to the affinity function of interest.

Yao *et al.* have developed a range of alkyl diazirine photoaffinity scaffolds (**1.46–1.48**, **1.34**, **1.35**) containing various bio-orthogonal handles (Figure 1.21).[21,22,36,110] These have been termed "minimalist" due to the minimal disturbance of the physicochemical properties to the affinity function upon introduction of the PAL scaffold. The authors have extensively demonstrated their application to photoaffinity-based proteomic profiling using kinase and bromodomain inhibitors as proof-of-concept studies.[21,22,110] The first-generation scaffold **1.46** containing an alkyl diazirine and alkyne has been recently adopted by the Cravatt group and coupled to fragment affinity functions (discussed further in Section 5.1.1.1).[18] The group has since developed second (**1.34**) and third (**1.35**) generation reagents, which allow for copper-free click ligations to be used.[36] One concern when using this approach is that it only incorporates the alkyl diazirine photoreactive group, and therefore is limited by low levels of singlet carbene generation and inefficient labelling.

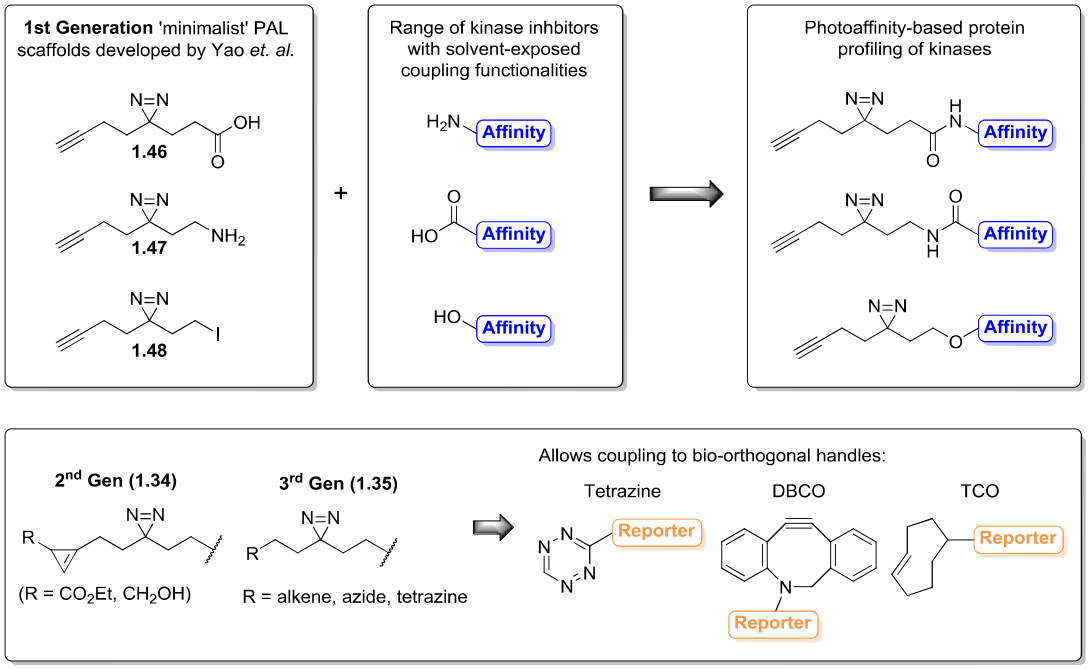


Figure 1.21 Minimalist alkyl diazirine PAL scaffolds developed by Yao *et al.*[21,22,110] The scaffold can be appended to the affinity function of interest *via* amide coupling or alkylation. Depending on the downstream requirements, a range of bio-orthogonal handles can be incorporated. One caveat to this approach is that it lacks versatility to incorporate other photoreactive groups that may offer higher levels of labelling efficiency than the aliphatic diazirine, which is known to produce relatively low levels of singlet nitrene.

Another approach to appending the photoreactive group and bio-orthogonal handle to the affinity function is to use multicomponent reactions.

1.4.1 Multicomponent reaction approaches to synthesise PAL probes

As PAL probes contain three functionalities, multicomponent reactions (MCRs) offer a convergent approach to their synthesis. To the best of our knowledge, only four methods of PAL probe synthesis by MCRs have been reported to date, and all are based on the Ugi four-component reaction. First reported by Ivar Karl Ugi in 1959, this reaction combines four components to furnish a bis-amide product **1.49** (Figure 1.22).[163] As the reaction incorporates four variable moieties (**1.50–1.53**) into a single compound, the Ugi reaction has been used extensively in combinatorial chemistry to generate large libraries of compounds.[164-167] The reaction proceeds by initial formation of an imine **1.54** from condensation of the amine **1.51** and aldehyde **1.52** components. This imine is protonated by the carboxylic acid component **1.50**. The activated

iminium is then attacked by the nucleophilic isonitrile component **1.53** to form a nitrilium ion **1.55**. The carboxylate then attacks this nitrilium ion to form intermediate **1.56**, which undergoes a Mumm rearrangement to form the desired bis-amide product **1.49**. The steps prior to the rearrangement are all reversible, and the reaction is driven by the irreversible rearrangement step.[168,169]

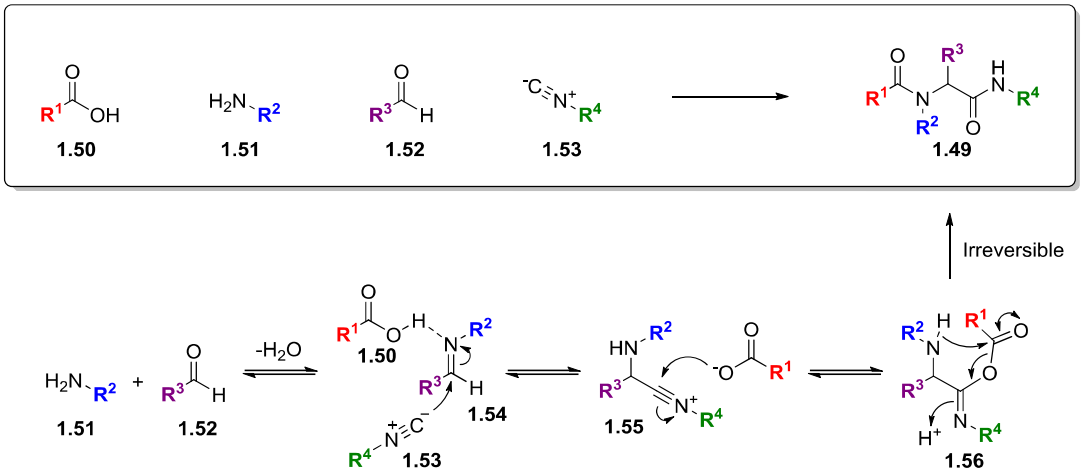


Figure 1.22 The Ugi four-component reaction and proposed mechanism.[168,169]

1.4.1.1 Ugi-reaction to synthesise PAL probes for protein tyrosine phosphatases

In 2012, Yao *et al*. reported the first use of a multicomponent reaction to synthesise PAL probes.[170] The authors used the Ugi four-component reaction to synthesise a panel of PAL probes that targeted protein tyrosine phosphatases (PTPs). A protected affinity function **1.57** was employed as the aldehyde component. Benzophenone was used as the photoreactive group and was incorporated *via* the isonitrile **1.58**. An alkyne bio-orthogonal handle **1.59** was incorporated *via* the carboxylic acid component. A set of 25 commercially available amines (**1.60**) were chosen to perform SAR studies for a secondary binding site within the PTP family of target proteins. The four components were combined at rt for 9 h in a solvent mixture of methanol and *N,N*-dimethylformamide (6:1 *v/v*). The resulting Ugi products containing a *tert*-butyl ester on the affinity function were deprotected with trifluoroacetic acid to produce the desired PAL probes in low to moderate yield (20–30%). The authors screened the 25 PAL probes for potency and levels of photolabelling against recombinant PTPB1. The probes that showed the highest levels of photolabelling were then irradiated in MCF-7

cell lysate and PTPB1 was identified as a protein target using both western blotting and MS-based proteomics.

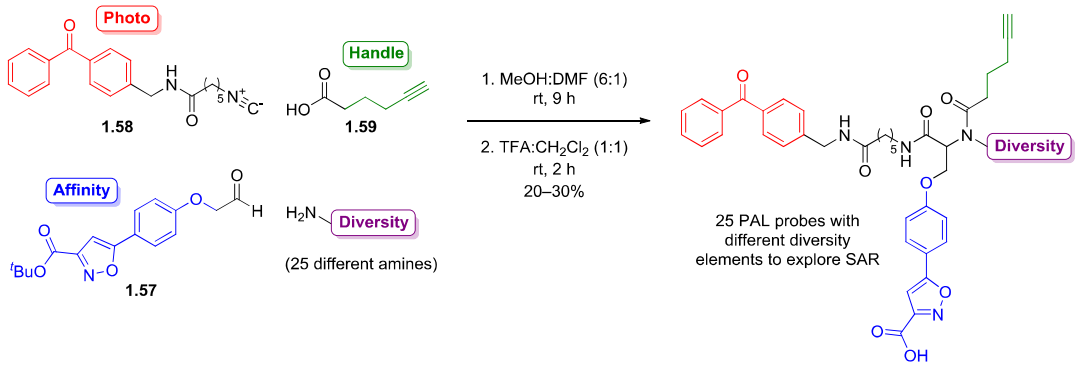


Figure 1.23 An Ugi four-component reaction protocol to synthesise PAL probes that targeted the protein tyrosine phosphatase family of proteins reported by Yao *et al.*[170] Benzophenone was chosen as the photoreactive group and was incorporated as the isonitrile component **1.58** in the Ugi reaction. The alkyne handle was incorporated as the carboxylic acid **1.59** and the pivalate ester-protected affinity function **1.57** was chosen as the aldehyde component. 25 amines were chosen for SAR studies to identify a second binding pocket within the PTP family of target proteins. The probes that showed the highest levels of labelling were chosen for photoaffinity-based proteomic profiling experiments in MCF-7 cell lysate and identified PTPB1 as an on-target, among other off-target proteins.

Several off-targets such as cathepsin D and prohibitin were also identified. One drawback of using this approach was the need to protect the carboxylic acid present in the affinity function to prevent it reacting in the place of the alkyne-carboxylic acid **1.59** during the Ugi reaction. The extent to which this PAL scaffold affected the physicochemical properties and cell permeability was unclear, as the proteomic experiments were only reported in cell lysates and not in live cells. Furthermore, only benzophenone was demonstrated as a suitable photoreactive group using this approach.

1.4.1.2 Ugi-azide reaction to synthesise PAL probe libraries

Cravatt and co-workers reported the use of the Ugi-azide multicomponent reaction to synthesise a library of 58 diverse PAL probes (Figure 1.24).[171] The reaction combined an array of isonitriles (**1.60**), aldehydes/ketones (**1.61**), and amines (**1.62**) with trimethylsilyl azide in equal amounts to produce functionalised tetrazole PAL probes in one synthetic step. The alkyl diazirine and alkyne were chosen as the photoreactive group and bio-orthogonal handle respectively, due to their minimal steric impact.

These two functionalities were spread across both the aldehyde/ketone and amine components to increase the diversity of PAL probe library. Initially, the proteins captured by the PAL probes were profiled using in-gel fluorescence scanning. The probes that showed selective labelling of proteins by in-gel fluorescence were progressed for MS-based proteomic experiments in live cells. The PAL probe library captured a diverse range of different proteins, the majority of which have currently no known chemical tools.

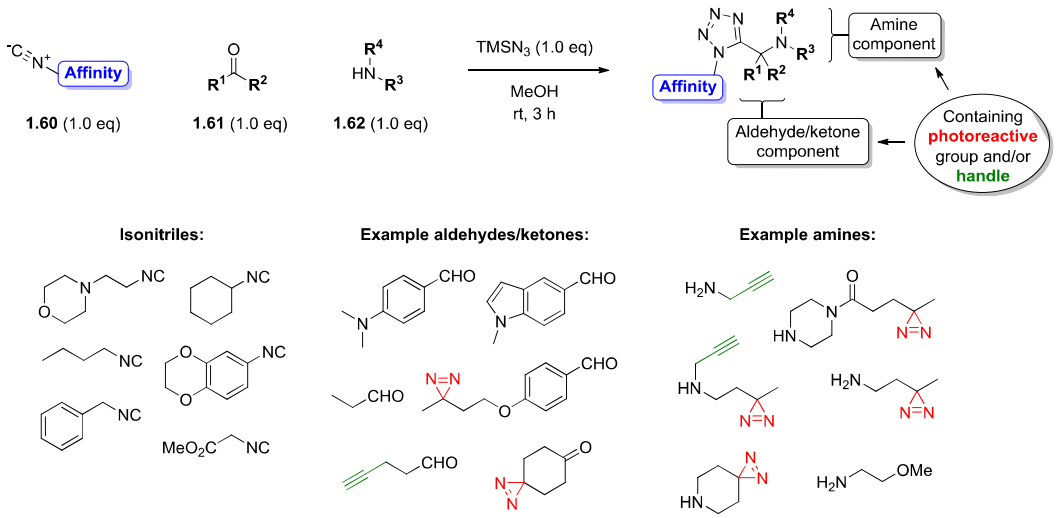


Figure 1.24 The Ugi-azide reaction applied by Cravatt and co-workers to synthesise a diverse range of PAL probes.[171]

One issue with this approach is that only the alkyl diazirine was used as the photoreactive group, which is known to give sub-optimal levels of labelling efficiency *versus* other common photoreactive groups. Furthermore, embedding the diazirine into each different aldehyde/ketone or amine component frequently required lengthy and bespoke syntheses.

1.4.1.3 Ugi reaction to synthesise PAL probe scaffolds

Bush *et al*. reported the use of the Ugi four-component reaction to synthesise PAL scaffolds (Figure 1.25).[69] The photoreactive group was incorporated as the carboxylic acid component, the handle as the amine component, and the isonitrile component contained functionality for the downstream attachment of the affinity function to the

scaffold. These three components were combined with paraformaldehyde and heated by microwave irradiation (100 °C, 20 min) in methanol. The protocol tolerated a variety of the most commonly used photoreactive groups including aryl azides (**1.16** and **1.63**) alkyl (**1.64**) and aryl (**1.65**) diazirines, and benzophenone (**1.66**) with good yields. An alkyne handle (**1.67**) for downstream CuAAC ligation could be used, or reporter tags such as biotin (**1.68**) or a rhodamine dye (**1.69**) could be installed directly *via* the amine functionality.

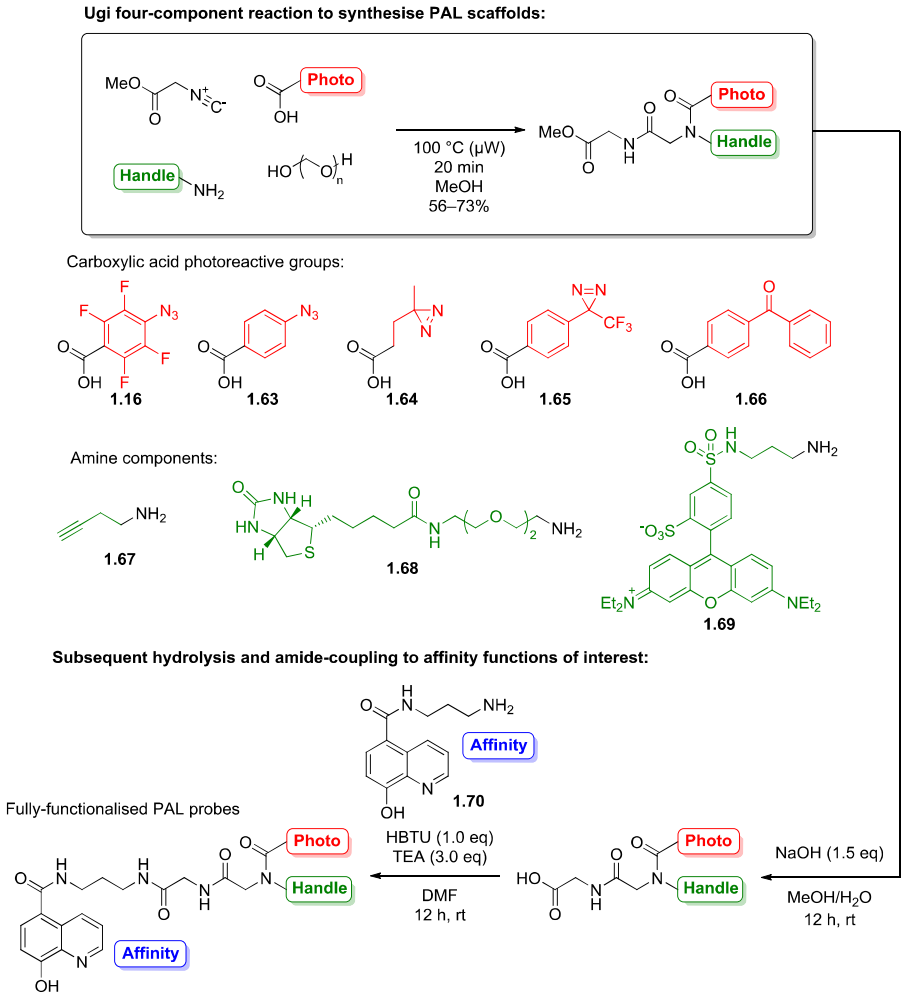


Figure 1.25 The Ugi four-component reaction protocol to synthesise PAL scaffolds reported by Bush *et al*.[69] The protocol was tolerant of various photoreactive groups, which are commercially available as carboxylic acids. A variety of handle functionalities were also tolerated. After synthesising the PAL scaffold, the methyl ester (incorporated as the isonitrile component) could be hydrolysed and the resulting carboxylic acid could be coupled to affinity functions of interest to give fully functionalised PAL probes in three synthetic steps.

The PAL scaffolds could then be linked to affinity functions of interest *via* the methyl ester functionality incorporated within the isonitrile component. The methyl ester was hydrolysed, and the resulting carboxylic acid PAL scaffolds containing the five different photoreactive groups (**1.16**, **1.63–1.66**) were coupled to a 2-oxoglutarate oxygenase inhibitor **1.70** as a proof-of-concept affinity function. The photocrosslinking efficiencies of the five photoreactive groups to recombinant 2-oxoglutarate oxygenase (PHD2) protein were compared using intact protein LCMS analysis. One caveat to this protocol is the need for three steps to synthesise the fully functionalised PAL probe. Furthermore, this arrangement of components introduces a relatively large atom spacing (the core Ugi scaffold added to the length of the isonitrile linker) between the affinity function and the photoreactive group. Depending on the length of the amine linker synthesised on the affinity function, the photoreactive group may be placed at an unnecessarily long distance from the active site, where it may be excessively exposed to solvent and buffer components. The effects of the PAL scaffold on the physicochemical properties and membrane permeability of the affinity function were not examined, however it can be assumed that the three additional amide bonds (two of which are secondary) introduced by this strategy would have a detrimental effect on cell permeability compared to the parent affinity function.[172]

1.4.1.4 Minimalist PAL scaffold for use in isonitrile-based multicomponent reactions

Recently, Lapinsky *et al.* reported the synthesis of an isonitrile variant (**1.71**) of the minimalist diazirine-alkyne PAL scaffold originally developed by the Yao group (Figure 1.26).[111] The authors developed this scaffold to obtain the 'best of both' strategies, whereby a wide range of diverse affinity functions can be constructed using an array of isonitrile-based multicomponent reactions (IMCRs), producing PAL probes that contain the minimalist diazirine-alkyne PAL scaffold. Five examples of different IMCRs employing **1.71** were reported. Only the synthetic advancement was communicated, and the performance of the probes in PAL labelling studies is yet to be reported. This synthetic development has expanded the toolbox of ways to incorporate the minimalist alkyl diazirine-alkyne PAL scaffold, however the PAL probes

synthesised using this approach are likely to share the same disadvantages described in Section 1.4. The strategy is limited to the alkyl diazirine photoreactive group, which can suffer from inefficient levels of labelling due to low singlet carbene formation.

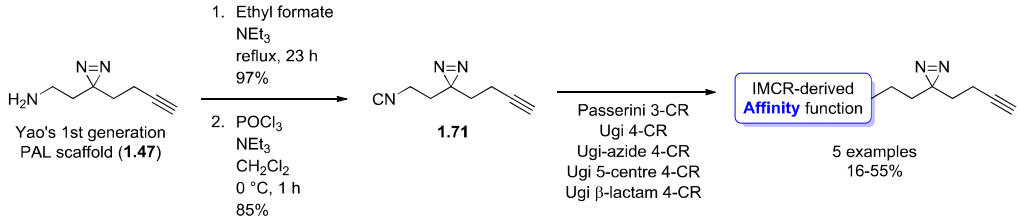


Figure 1.26 Isonitrile derivative **1.71** of Yao's minimalist diazirine-alkyne PAL scaffold reported by Lipinsky *et al.*[111] **1.71** can be used in a range of isonitrile-based multicomponent reactions to generate diverse affinity functions.

1.4.1.5 Summary of MCR approaches to the synthesis of PAL probes

The aforementioned methods to synthesise PAL probes each have limitations, including incompatibility for various photoreactive groups, low substrate scope, or multiple linear steps to access the final PAL probe. There remains a requirement for a more universal approach to PAL probe synthesis that ideally can be performed in a single step with a wide tolerance for various photoreactive groups, bio-orthogonal handles or reporter tags, and any affinity function of interest. This novel methodology would allow easier access to chemoproteomic experiments to find on- and off-targets of lead compounds or identify the protein targets of active compounds that have an unknown mode of action. Therefore, the development of a versatile, one-step synthesis of PAL probes would be highly applicable to the lead optimisation and pre-clinical stages of small molecule drug discovery.

1.5 Proteins of interest

To investigate and develop novel applications of photoaffinity probes to small molecule target-based drug discovery, the bromodomain and extra-terminal domain (BET) family of proteins were chosen as a representative class of target proteins for the proof-of-concept studies presented in this thesis.

1.5.1 Bromodomain and extra-terminal domain (BET) proteins

Epigenetics refers to the inheritable variations in phenotype that are a result of changes within chromosomes, without alterations in the DNA sequence.[173] The epigenetic code regulates the precise temporal control of gene expression. This allows for one genome to produce a vast array of differentiated cell types, and ultimately contributing to the development of complex multicellular organisms.[174-176] One mechanism of epigenetic regulation is the selection of when and where the genome can be accessed by transcriptional machinery by controlling the remodelling of chromatin (Figure 1.27).[177] This is primarily performed by controlling the superstructure of nucleosomes. Nucleosomes are the basic repeating unit of chromatin, composed of two superhelical turns of DNA wrapped around eight core histone proteins.[178] The histone proteins consist of a central globular domain, and their N-termini (known as histone tails) protrude from the nucleosome and are subject to a range of post-translational modifications (PTMs), such as acetylation, phosphorylation, methylation, ubiquitination and sumoylation.[177] This high degree of PTMs (known as the histone code) affects the quaternary structures of the nucleosomes and is responsible for the recruitment of chromatin remodelling enzymes. These enzymes have been grouped according to their function of "writing", "reading" or "erasing" the histone code.[179]

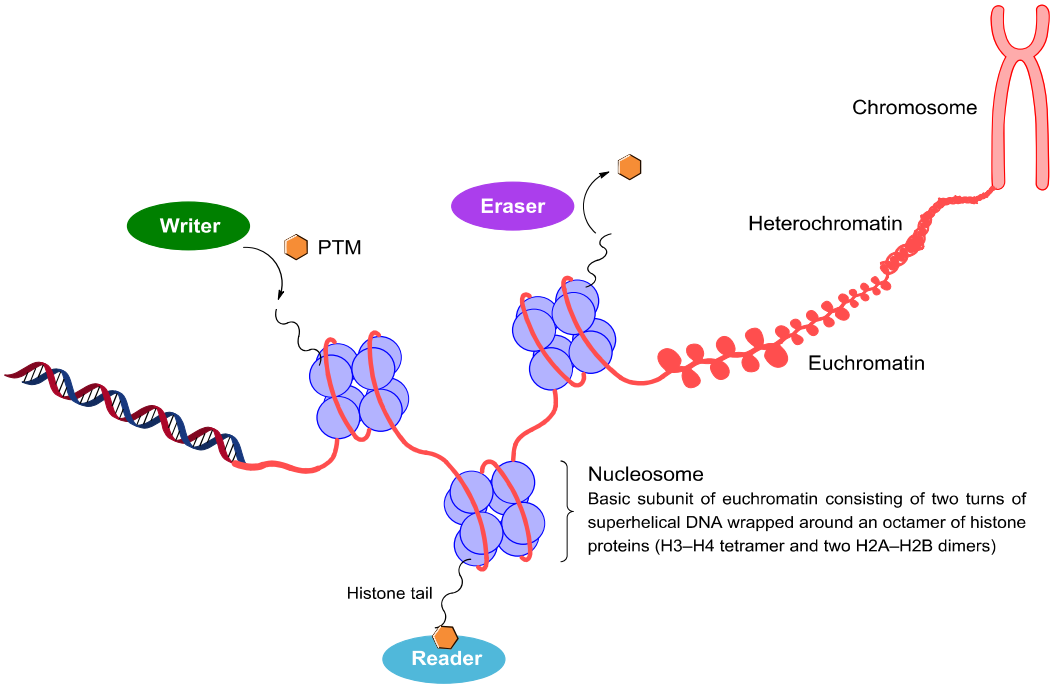


Figure 1.27 Overview of the histone code. Nucleosomes consist of an octamer of histone proteins and form the basic unit of euchromatin, which is the open transcriptional on-state of DNA.[178] The N-terminal domains of the histone proteins protrude from the nucleosome (histone tails) and are subject to a wide range of post translational modifications (PTMs). These are installed, recognised and removed by classes of writer, reader and eraser proteins, respectively. Specific recognition of PTMs (or combinations thereof) by reader domains leads to the recruitment of transcriptional machinery and gene expression.[179]

Bromodomain containing proteins (BRDs) are a class of epigenetic reader protein that specifically recognise acetylated lysine (KAc) marks on histone tails.[180] Upon recognition, BRDs act as a scaffold for the recruitment and assembly of complexes that alter chromatin accessibility to transcription factors and RNA polymerases.[179] There are 61 known bromodomains, each consisting of *ca*. 110 amino acids and are present across 46 proteins grouped into 8 structural classes (Figure 1.28).[181]

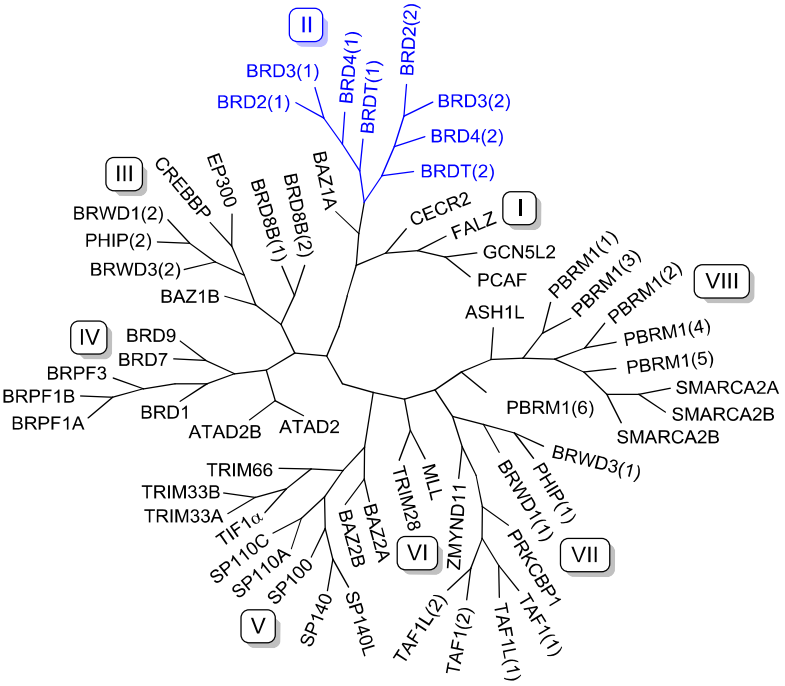


Figure 1.28 Eight families of bromodomain-containing proteins. The BET family (II, blue) were used as proof-of-concept target proteins in this work.[180]

The bromodomain and extra-terminal domain (BET) family of proteins consist of four proteins; BRD2, BRD3, BRD4 and BRDT. These proteins each contain two tandem bromodomains (BD1 and BD2) which recognise KAc marks on histones H3 and H4.[182,183] BRD2, -3 and -4 are ubiquitously expressed, whereas BRDT is only expressed in the testes.[184]

Disfunction of the BET family of proteins has been shown to play critical roles in a wide variety of autoimmune diseases, such as psoriasis, arthritis, and inflammatory bone resorption.[185-188] Small molecule BET inhibitors have shown efficacy against a range of cancers, including myeloma, glioblastoma, acute myeloid leukaemia (AML), and breast cancer.[189-192] The development of selective small molecule inhibitors for BET proteins has been an intense area of focus for many pharmaceutical companies over the last ten years, which has resulted in several small molecule BET inhibitors entering clinical trials in oncology.[183]

This work investigates various applications of PAL to small molecule target-based drug discovery. Representative of a class of target proteins that have been actively pursued in drug discovery programmes, the BET family were chosen as proof-of-concept target proteins. It is envisaged that the developments made in this work will be transferable to other target proteins.

1.6 Aims

The specific aims of this work were to:

Chapter 2:

- Simplify the process of PAL probe design and synthesis so that it is amenable to parallel chemistry.
- Determine the structure activity relationship between PAL probes containing various photoreactive groups and percentage labelling to recombinant protein.
- Rationalise any observed SAR by using X-ray crystallography and LC-MSMS analysis.

Chapter 3:

- Investigate the applicability of the PAL approach to biochemical screening with recombinant protein.

Chapter 4:

- Develop a photoaffinity-based proteomic profiling workflow enabled by CuAAC ligation.
- Demonstrate that PAL probes synthesised using the simplifed protocols developed in Chapter 2 can be applied to live-cell MS-based proteomics to capture and identify on- and off-target proteins.

Chapter 5:

- Investigate the dose-response relationship between photoaffinity fragments and percentage labelling to recombinant protein.
- Investigate the promise of this approach as new method for fragment-based lead discovery.

Chapter 2 Developments in the design and synthesis of PAL probes

2.1 Introduction

Photoaffinity probes have become powerful tools in the field of chemical biology, especially when coupled to modern mass spectrometry methods. One of the most useful applications of PAL probes in the context of small molecule drug discovery is in photoaffinity-based proteomic profiling, which involves the identification of on- and off- protein targets of a compound of interest (introduced in Section 1.2.1.3). The photoaffinity approach enables capture of transient non-covalent interactions, which are often missed by affinity-based probes. Additionally, photoreactive capture shows minimal regioselectivity for amino acid residues, which is a drawback of the more targeted electrophilic warheads used in activity-based probes. PAL probes can also be used to identify and characterise binding sites of labelled proteins using proteolytic digestion and LC-MSMS analysis.[14] A hindrance to the application of these two PAL applications is the requirement to synthesise the tri-functional PAL probe from the parent affinity function, while maintaining the key binding interactions and physicochemical properties of the parent compound. A variety of synthetic strategies (reviewed in Section 1.4) can be employed, including “embedding” or “appending” the photoreactive group and handle functionality to the affinity function. These strategies can be synthetically challenging, particularly in the case of natural products, or if little is known about the structure-activity relationship between the compound of interest and the target protein(s).[66] There remains a need for generic methodologies to enable one-step synthesis of PAL probes, which are compatible with a variety of photoreactive groups, handles, and affinity functions. Furthermore, studies to assess the impact of linker length on the photocrosslinking efficiency of various photoreactive groups have not been reported. This chapter describes the development of a highly versatile one-step synthesis of PAL probes, which is compatible with a range of affinity functions, photoreactive groups and handles. This protocol allowed for the expedient synthesis of a range of PAL probes which were used to investigate the effect of linker length on the photocrosslinking efficiency of various photoreactive

groups. The physicochemical properties and cellular potencies of the PAL probes were also assessed in comparison to the parent affinity function. By providing rapid synthetic access to chemical tools for photoaffinity-based proteomic profiling, it is envisaged that the protocol reported in this work will be highly applicable to the lead optimisation and pre-clinical stages of small molecule drug discovery to provide crucial information regarding the on- and off-target profiles of compounds of interest.

2.2 Development of a one-pot Ugi protocol for the synthesis of highly variable PAL probes

Of the approaches to PAL probe synthesis reviewed in Section 1.4.1, the Ugi four-component reaction protocol reported by Bush *et al.* demonstrates the widest synthetic tolerance for various photoreactive groups.[69] However, this procedure requires three synthetic steps to furnish a fully functionalised PAL probe. This arrangement of the four reaction components also introduces redundant atom spacing (the core Ugi scaffold and isonitrile linker) between the affinity function and photoreactive group (Figure 2.1). The components of the reaction could be rearranged so that the amine affinity function is employed directly in the Ugi reaction as the amine component and the handle could be installed as the isonitrile component. This would allow for a one-step protocol to synthesise PAL probes. This arrangement of components would also place the photoreactive group directly on the amine of the affinity function, allowing for the photoreactive group to be brought closer to the surface of the target protein if desired. To test this proposed Ugi protocol, recently reported BET inhibitors were chosen as affinity functions for proof-of-concept studies.

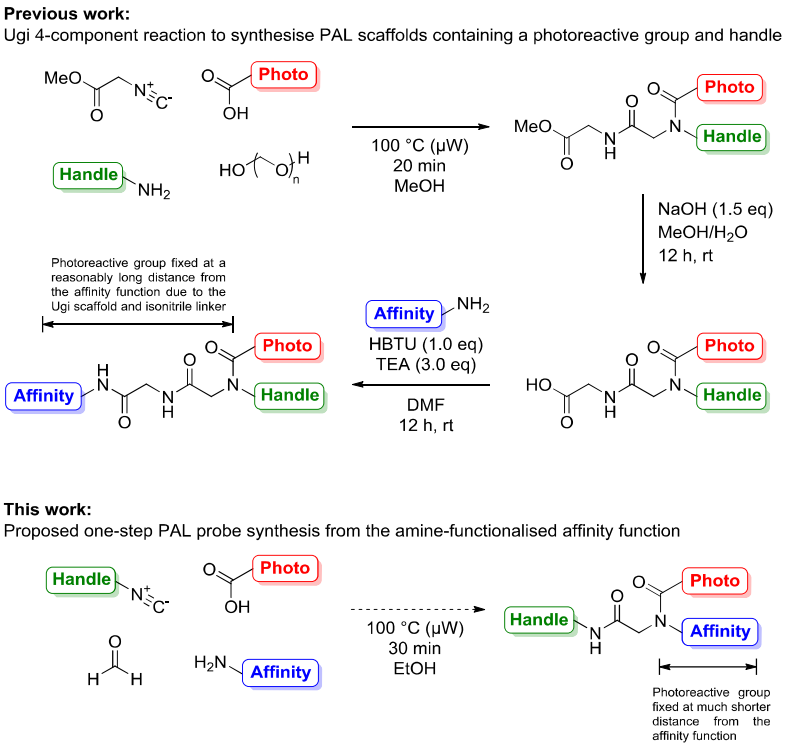


Figure 2.1 Previously reported synthesis of PAL scaffolds containing an ester, a photoreactive group and handle using the Ugi four-component reaction (top).[69] The proposed Ugi protocol investigated in this work uses the same amine affinity function, however the handle functionality is incorporated as the isonitrile component to allow the synthesis of fully functionalised PAL probes in one step.

2.2.1 Selecting an appropriate affinity function

To explore the potential of the proposed Ugi protocol, a recently reported BRD4 inhibitor **2.1** was selected as the affinity function (Figure 2.2, left).[193] Although **2.1** was shown to have only moderate affinity for BRD4 BD1 (AlphaScreen pIC_{50} = 5.0), the binding mode of this compound was known from a co-crystal structure of **2.1** in the acetyl lysine binding site of BRD4 BD1 (Figure 2.2). The dimethyl isoxazole warhead forms key hydrogen bonding interactions with N140, and with the conserved network of four water molecules found within the active site of BRD4 BD1.[194] A solvent-exposed vector from the morpholine of **2.1** was observed, and it was envisaged that PAL scaffolds could be incorporated from this vector with minimal impact on the binding mode or affinity of **2.1** to BRD4 BD1.

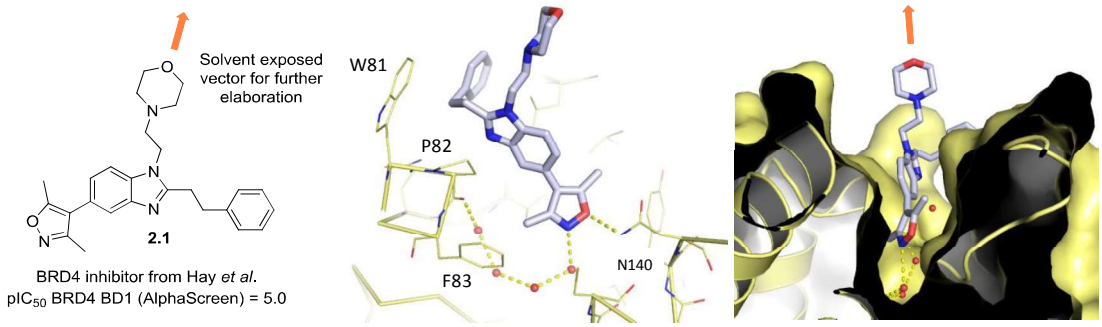
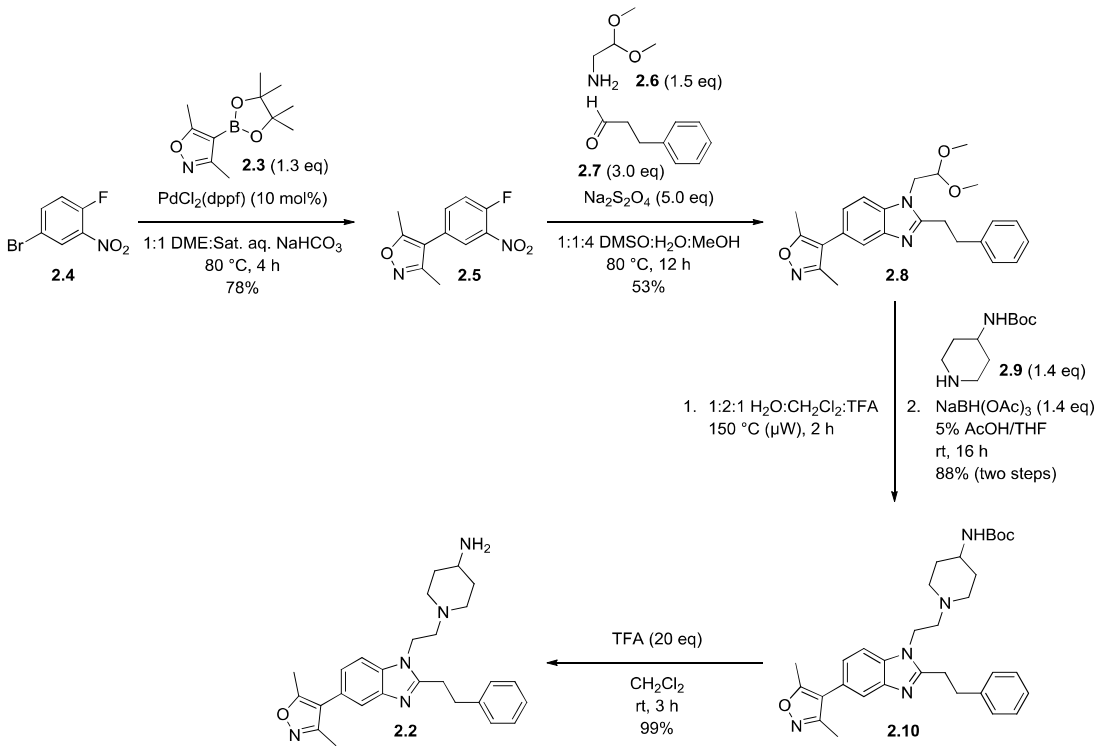


Figure 2.2 BRD4 inhibitor (**2.1**) reported by Hay and co-workers (left).[193] X-ray crystal structure of **2.1** co-crystallised with BRD4 BD1 (PDB:4NR8) (middle). The dimethyl isoxazole warhead forms a hydrogen bond to the conserved network of four water molecules commonly found within the BRD4 BD1 binding pocket.[194] A hydrogen bond can also be observed between the dimethyl isoxazole and N140. No key interactions are formed with the WPF shelf (W81, P92, F83). PAL scaffolds could be coupled to the solvent exposed vector from the morpholine of **2.1** (right).

A primary amine derivative (**2.2**) of **2.1** suitable for use in the proposed Ugi protocol was synthesised (Scheme 2.1). The dimethyl isoxazole warhead **2.3** was coupled to starting material **2.4** *via* a Suzuki reaction to afford **2.5**. A reductive ring-closing S_NAr reaction with amine **2.6**, aldehyde **2.7** and **2.5** afforded the intermediate **2.8**. Deprotection of the dimethoxy acetal **2.8** under acidic conditions to yield the aldehyde allowed for a subsequent reductive amination with amine **2.9** to give **2.10**. Boc deprotection under acidic conditions afforded **2.2**, containing a primary amine in the desired solvent-exposed vector. Moderate to good yields were achieved, with an overall yield of 36% over four steps.

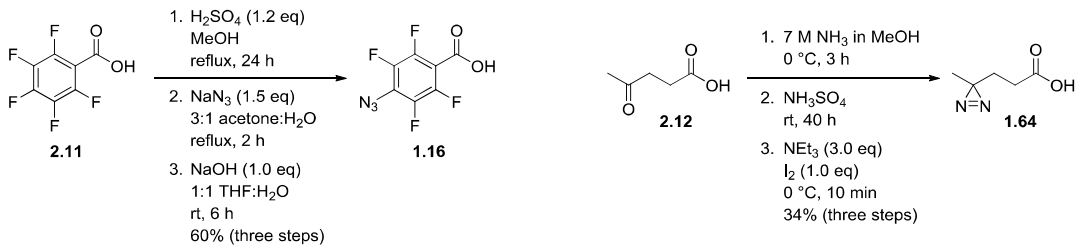


Scheme 2.1 Synthesis of **2.2**, an analogue of **2.1** with a primary amine directed in a solvent exposed vector.

Following the synthesis of an appropriate amine affinity function **2.2**, attention was turned to the synthesis of a selection of handles and photoreactive groups.

2.2.1.1 Synthesis of photoreactive groups

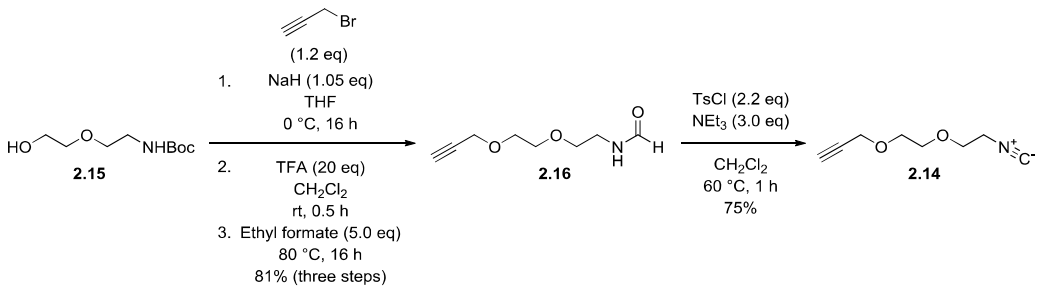
To investigate the tolerance of the proposed Ugi protocol for various photoreactive groups, five of the most commonly used groups were chosen. Aryl azide **1.63**, trifluoromethyl aryl diazirine **1.65** and benzophenone **1.66** are commercially available as carboxylic acids. The remaining two acids (tetrafluoroaryl azide **1.16** and alkyl diazirine **1.64**) were synthesised as shown in Scheme 2.2. Synthesis of **1.16** was achieved by esterification of pentafluorobenzoic acid **2.11**, followed by an S_NAr reaction with sodium azide.[96] Subsequent hydrolysis produced **1.16** in good overall yield (60%). Alkyl diazirine **1.64** was synthesised from levulinic acid **2.12** in three steps. Ketone **2.12** was condensed with 7 M ammonia in methanol, and the resulting imine was trapped with hydroxylamine-O-sulfonic acid to afford the diaziridine, which was oxidised using iodine to give **1.64** in moderate overall yield (34%).



Scheme 2.2 Synthesis of tetrafluoroaryl azide **1.16** and alkyl diazirine **1.64**.

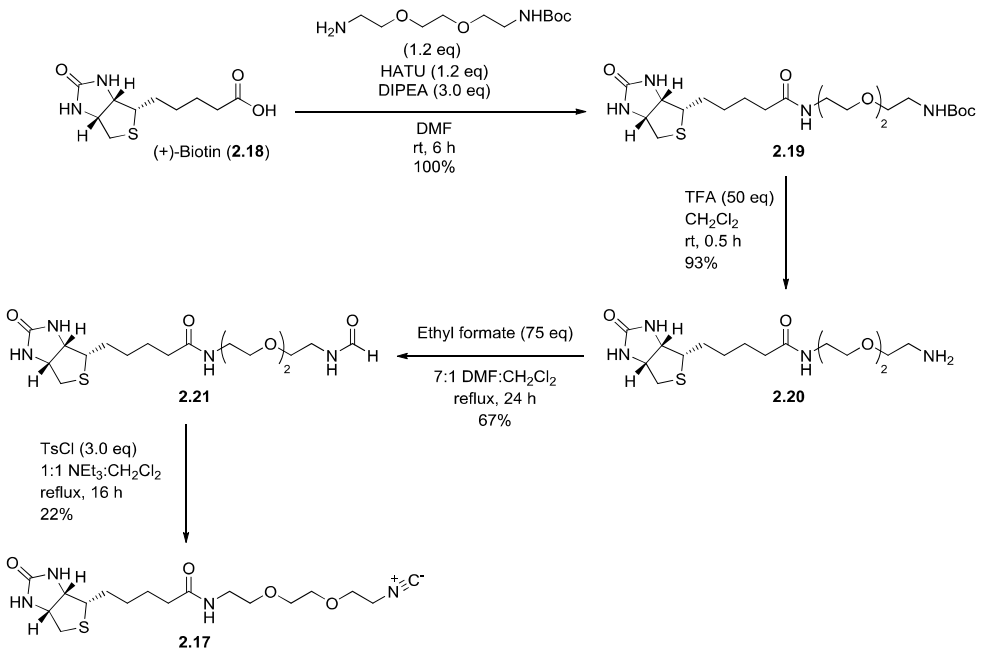
2.2.1.2 Synthesis of handle components

Three handles were chosen to investigate the tolerance of the proposed Ugi protocol for various handle functionalities; an ethyl ester to enable further chemical modification, an alkyne to enable bio-orthogonal click chemistry, and biotin to enable avidin bead-based enrichment. An isonitrile form of the handle was required to enable application in the proposed Ugi protocol. The simple ethyl ester isonitrile **2.13** was commercially available, and the remaining two handles were synthesised. Alkyne-isonitrile **2.14** was synthesised from **2.15** in two steps *via* **2.16**, in high overall yield (61%) (Scheme 2.3).



Scheme 2.3 Synthesis of alkyne-isonitrile **2.14**.

Biotin-isonitrile **2.17** was synthesised from biotin (**2.18**) in four steps. High conversions and yields were achieved for the first three steps [**2.19** (100%), **2.20** (93%) and **2.21** (67%) respectively], however the final step was low-yielding as a result of complications during purification due to the polarity of biotin (22%). This gave a moderate overall yield of 14% for **2.17** over the four steps.



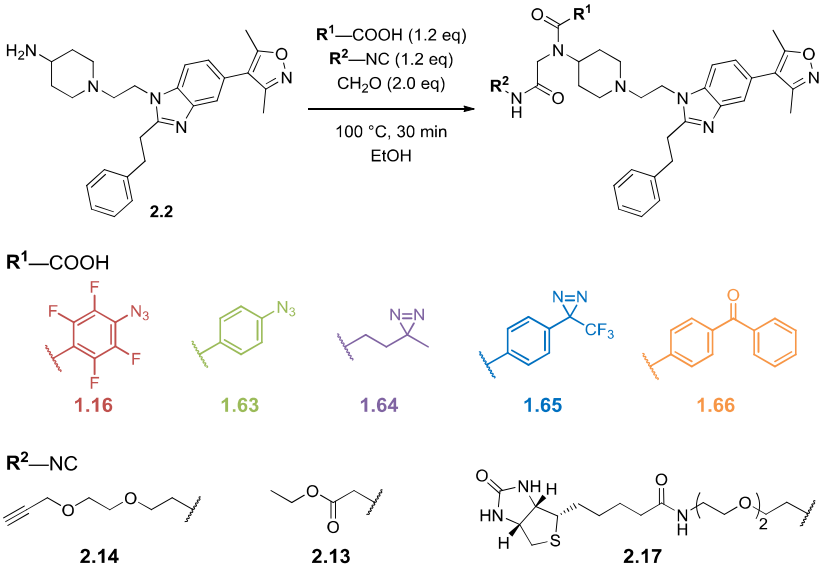
Scheme 2.4 Synthesis of biotin-isonitrile **2.17**.

Once the required affinity function, photoreactive groups and handle moieties were obtained, the synthetic feasibility of the proposed Ugi protocol was investigated.

2.2.2 One-step Ugi protocol to synthesise low-affinity BET PAL probes

Initially, the compatibility of the Ugi protocol with various photoreactive groups (**1.16**, **1.63–1.66**) was investigated. Amine affinity function **2.2** (1.0 eq), isonitrile-alkyne handle **2.14** (1.2 eq), paraformaldehyde (2.0 eq) and one of the five carboxylic acid photoreactive groups (**1.16**, **1.63–1.66**) (1.2 eq) were combined in ethanol and heated using microwave irradiation (100 °C, 30 min) (Table 2.1).[69]

Table 2.1 Exploring the tolerance of various photoreactive groups and handle moieties in the Ugi protocol and assessing the impact of the Ugi scaffold on affinity to BRD4 BD1 and BD2. TR-FRET assays were performed by Alex Phillipou.



R^1	R^2	Compound (%)[a]		pIC_{50} BD1[b,d]	pIC_{50} BD2 [c,d]
–	–	**2.2**[e]	–	5.7	5.0
1.16	**2.14**	**2.25**	(0)	–	–
1.63	**2.14**	**2.26**	(0)	–	–
1.64	**2.14**	**2.22**	(54)	5.9	5.4
1.65	**2.14**	**2.23**	(49)	6.0	5.9
1.66	**2.14**	**2.24**	(64)	6.1	<4.3[f]
1.16	**2.13**	**2.27**	(67)	5.9	<4.3[f]
1.63	**2.13**	**2.28**	(66)	6.3	5.7
1.64	**2.17**	**2.29**	(28)	5.8	5.3

[a] Isolated yields. [b] TR-FRET assay with recombinant BRD4 BD1. [c] TR-FRET assay with recombinant BRD4 BD2. [d] Mean of at least two replicates. [e] Parent compound included for comparison. [f] Result was below the lower limit of the assay.

Diazirines (**1.64** and **1.65**) and benzophenone (**1.66**) afforded the desired products (**2.22**, **2.23** and **2.24**) in moderate to good yields (54%, 49% and 64% respectively). Good conversions were also achieved with aryl azides (**1.16** and **1.63**), however, the products (**2.25** and **2.26**) were found to be unstable upon purification, potentially due to Huisgen cycloaddition of the azide and alkyne groups upon concentration (explored further in Section 2.2.5). Introduction of the ethyl ester isonitrile **2.13** as a representative non-alkyne handle, afforded the PAL probes **2.27** and **2.28** in good yields (67% and 66% respectively).

The Ugi protocol was also tested with a more complex handle. Reaction of isonitrile-biotin **2.17** with affinity function **2.2** and alkyl diazirine **1.64** gave the biotinylated PAL probe **2.25** in moderate yield (28%). This Ugi protocol demonstrated

tolerance for five commonly used photoreactive groups, and various isonitrile handles. Thus, the Ugi reaction provides a highly efficient synthetic method to access a wide diversity of PAL probes in one step.

The effect of the added Ugi scaffold and various photoreactive groups on target affinity relative to the parent amine affinity function **2.2** was investigated. The affinity of each PAL probe to recombinant BRD4 BD1 and BD2 was measured in a TR-FRET assay (Table 2.1, TR-FRET assays were performed by Alex Phillipou). All probes were found to have similar or improved potency over the parent amine **2.2** for BD1 (pIC_{50} = +0.1 to +0.6). The effect on BD2 was more varied (pIC_{50} = -0.8 to +0.8), resulting in some BD1 biased probes (*e.g.* **2.24**, >63 fold selective).

Having demonstrated that the PAL probes synthesised using the one-step Ugi protocol had maintained or improved affinity for BRD4 BD1 in a biochemical assay, the photocrosslinking yields of PAL probes **2.27**, **2.28**, **2.22**, **2.23** and **2.24** to BRD4 BD1 were assessed.

2.2.3 Photocrosslinking timecourses for PAL probes (2.27, 2.28, 2.22, 2.23 and 2.24) to BRD4 BD1

To compare the photocrosslinking efficiencies of the five different photoreactive groups, probes **2.27**, **2.28**, **2.22**, **2.23** and **2.24** (20 μM) were incubated with recombinant BRD4 BD1 (3 μM) and irradiated (302 nm) over a timecourse. The samples were then analysed using intact protein LCMS (introduced in Section 1.2.4.1). An example workflow for the intact protein LCMS analysis is shown for the aryl azide probe **2.28** in Figure 2.3. In the total ion chromatogram (TIC), the MS scans within the peak for protein were summed (regions shaded in grey) to give a spectrum containing peaks corresponding to the distribution of charge states ($[M+nH]^{n+}$) for the protein species. A maximum entropy deconvolution algorithm was then used to find the mass of the neutral intact protein species present. Prior to UV irradiation, the mass of 15083 Da for unmodified BRD4 BD1 was found (expected mass from amino acid sequence = 15083.43 Da). After irradiation (302 nm, 10 min) in the presence of probe **2.28**, the mass of the unmodified protein was observed, along with covalently modified protein (15787 Da). The mass of this modified protein species corresponded to the mass of

BRD4 BD1 + **2.28** – N_2 (15083.43 + 731.35 – 28.01 = 15786.77 Da). The percentage labelling was found to increase with irradiation time (Figure 2.4, left). The percentage labelling of all five probes (**2.27**, **2.28**, **2.22**, **2.23** and **2.24**) to BRD4 BD1 at each timepoint was calculated using Equation 2.1, and plotted against irradiation time (Figure 2.4, right).

Equation 2.1

$$\%\ labelling = \frac{Peak\ area\ for\ single\ labelled\ protein}{(Peak\ area\ for\ unlabelled\ protein + single\ labelled\ protein)} \times 100$$

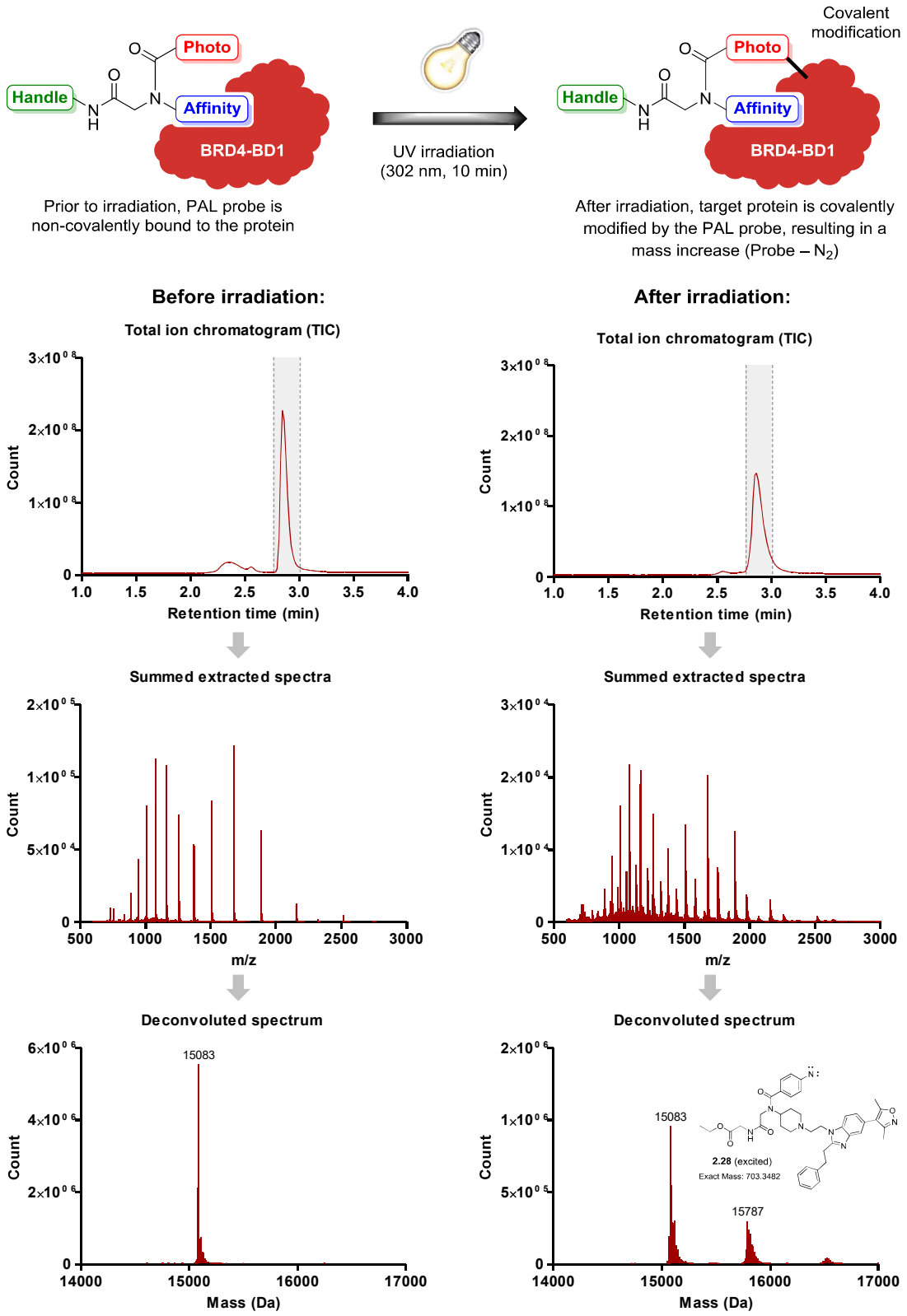


Figure 2.3 Example workflow (with probe **2.28**) for intact protein LCMS analysis. The scans within the protein peak in the TIC (area shaded in grey) are summed to give a mass spectrum containing peaks corresponding to a distribution of charge states for the protein species present. These spectra are then deconvoluted using a maximum entropy algorithm to give a mass and intensity for the unlabelled and modified protein species, providing a measure of percentage labelling calculated using Equation 2.1.

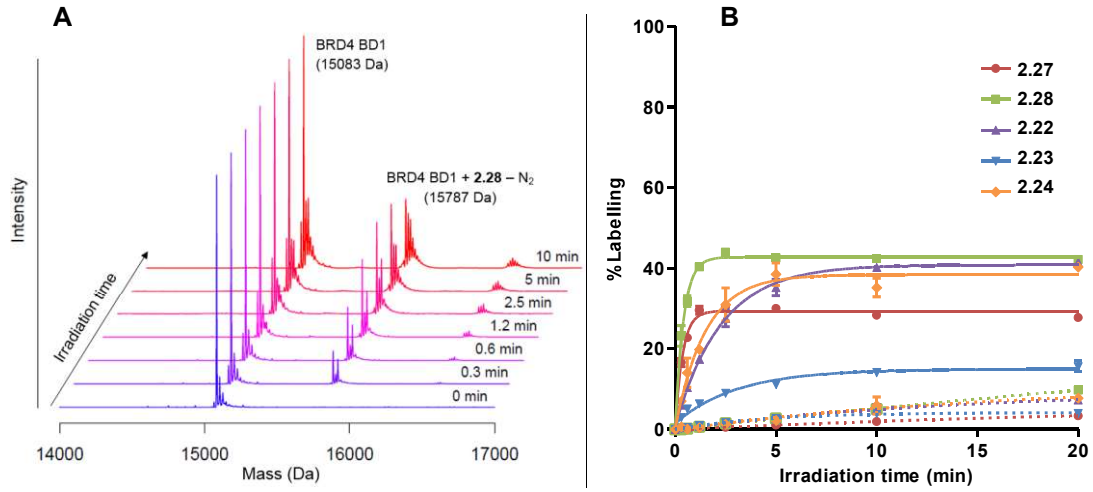


Figure 2.4 A) Example of intact protein LCMS analysis of BRD4 BD1 upon irradiation with probe **2.28** at various timepoints. B) Photocrosslinking timecourses showing percentage labelling for probes (**2.27**, **2.28**, **2.22**, **2.23** and **2.24**) at 302 nm (solid line) and 365 nm (dashed line).

All five photoreactive groups afforded appreciable photocrosslinking yields (>15%) following irradiation at 302 nm, with aryl azide (**2.28**), alkyl diazirine (**2.22**) and benzophenone (**2.24**) reaching *ca.* 40% labelling (Figure 2.4). The rate of activation was fastest for the aryl azide probes (**2.27** and **2.28**) (*ca.* 2 min to complete activation), while the diazirine (**2.22** and **2.23**) and benzophenone (**2.24**) probes required longer irradiation times (5–10 min) to reach maximum percentage labelling. The timecourse was also performed with 365 nm irradiation (Figure 2.4, dashed lines). The rates and yields of photocrosslinking at this wavelength were considerably slower, with <10% photolabelling after 20 min for all five probes. This was surprising, as 365 nm is a widely used irradiation wavelength in PAL studies, and diazirine and benzophenone photoreactive groups are reported to have a λ_{max} *ca.* 350 nm.[21,41,105,115] This prompted an investigation into the irradiance (energy delivered per unit area per unit time) from both 302 nm and 365 nm lamps.

2.2.4 Irradiance test for 302 nm and 365 nm lamps

To investigate if the low levels of labelling observed after 365 nm *vs* 302 nm irradiation was due to a large discrepancy in the power output between the lamps, the irradiance of both lamps was determined. For both lamps, the time taken to deliver 100 mJ cm^{-2} was measured repeatedly (Figure 2.5). For the first few measurements, the time taken to deliver 100 mJ cm^{-2} decreased for both lamps. This was likely due to warming of

the bulbs during the initial few measurements. After the bulbs became warm (*ca*. 2 min), the time to deliver 100 mJ cm^{-2} became constant. The energy delivered per cm per second was then calculated. The irradiance of the 365 nm lamp was found to be approximately twice that of the 302 nm lamp (7.55 *vs* 3.78 mW cm^{-2} respectively). Therefore, the lower photocrosslinking yields achieved with 365 nm *versus* 302 nm could not have been due to a lower irradiance from the 365 nm lamp.

Recently, Wilson and co-workers reported full photoactivation of diazirines with 365 nm irradiation in under 10 seconds using a bespoke UV-LED lamp.[100] This lamp had a reported irradiance of 15 W cm^{-2}, which is *ca*. two-thousand times more powerful than the 365 nm lamp used in this work. In light of this, to achieve full photoactivation using irradiation at 350–360 nm (λ_{max} of diazirines and benzophenones), lamps with much higher irradiance outputs may be required.

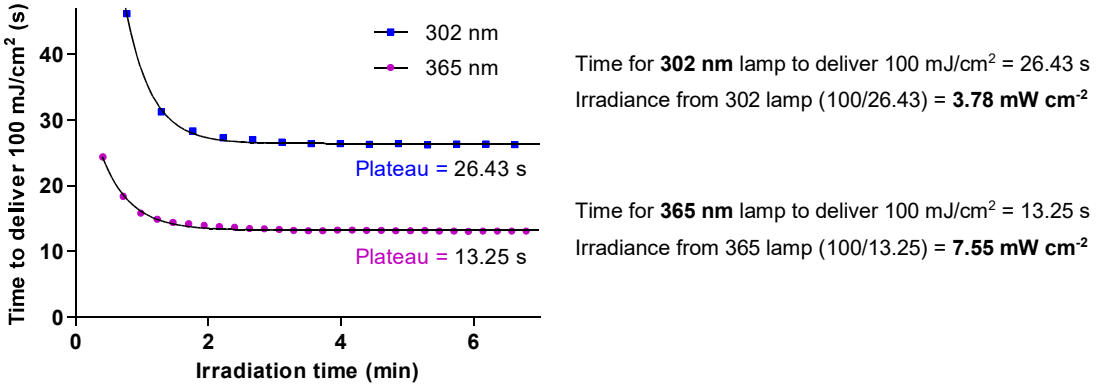


Figure 2.5 Comparison of the power per cm^2 (irradiance) of two UV lamps (302 and 365 nm) used in this work. The amount of time taken for the lamp to deliver 100 mJ of energy per cm^2 was measured repeatedly to construct a timecourse. The power per cm^2 of the 302 nm lamp was found to be approximately half of the output from the 365 nm lamp. The data also demonstrated that the bulbs should be warmed for at least two minutes prior to irradiating samples.

2.2.5 Investigations into the observed alkyne-azide incompatibility

The substrate scope studies for the Ugi-four component reaction protocol (Section 2.2.2) identified that the aryl azide photoreactive groups **1.16** and **1.63** were incompatible with the alkyne-isonitrile handle **2.14**. When **2.14** was used in combination with aryl azides **1.16** and **1.63**, the desired products (**2.25** and **2.26**) were identified in good conversion using LCMS analysis of the crude reaction mixture. However, when the samples were concentrated after reverse phase purification,

degradation was observed. To investigate this degradation, the isolated products were monitored using LCMS (Figure 2.6). For this study, the compounds were isolated as solids, and a sample was freshly prepared for LCMS analysis at each indicated timepoint.

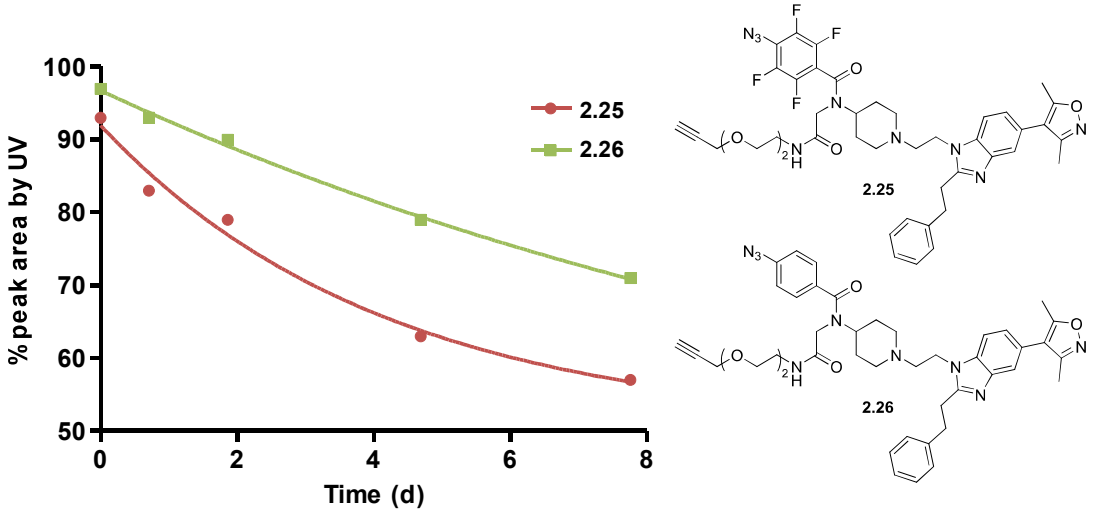


Figure 2.6 Percentage peak area (UV chromatogram) for the products **2.25** and **2.26** after isolation determined using LCMS analysis. Samples were isolated as solids and stored at rt. A fresh LCMS sample was prepared from a stock of solid sample for analysis at each timepoint.

To ensure this observed degradation was independent of the affinity function used, a simplified *para*-methoxy benzylamine analogue **2.30** was also prepared (Figure **2.7**). The crude reaction mixture was divided into two purification streams (normal phase flash chromatography and reverse phase preparative HPLC) to ensure the observed degradation of the isolated product was not caused by the purification method. The isolated product from both purification methods produced an identical degradation profile (Figure 2.7). Example LCMS analyses are shown in Figure 2.8. After isolation of **2.30** (0 min timepoint in Figure 2.7), LCMS analysis showed a peak consistent with the desired product (95% by UV, $[M+H]^+$ 538.3 Da). At the 16-day timepoint, the desired product had been reduced to 21% peak area by UV, with new broad peaks appearing at later retention times with masses indicative of dimerization ($[M+H]^+$ 1075.6 Da). Alkyne-azide probes have been used previously in the literature, and have often yielded poor photolabelling of protein targets, which may have been due to degradation of the PAL probe as observed here.[41,195] Due to the observed instability of the alkyne-azide PAL probes **2.25**, **2.26** and **2.30** after isolation, these probes were not used in any further PAL studies presented in this thesis.

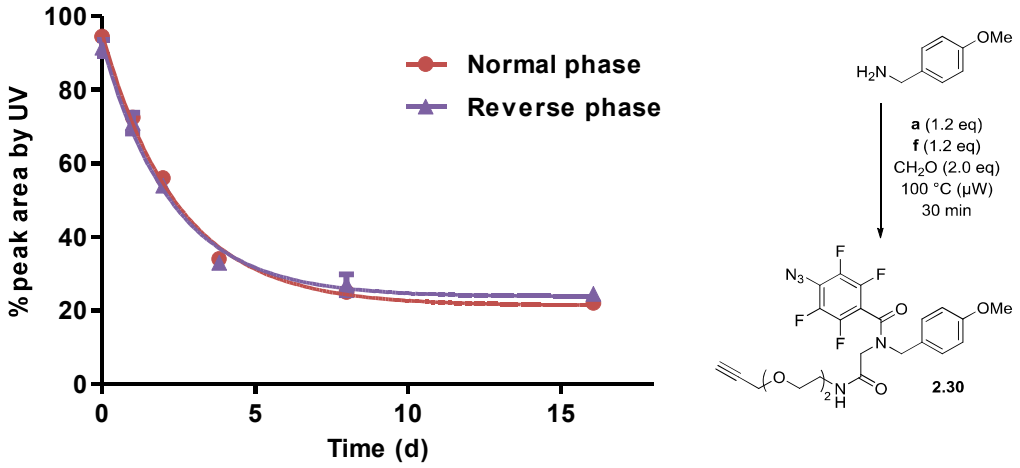


Figure 2.7 A simplified para-methoxy benzylamine analogue **2.30** was prepared and purified using normal and reverse phase chromatography. Both samples of isolated product were stored as solids at room temperature and showed identical degradation profiles.

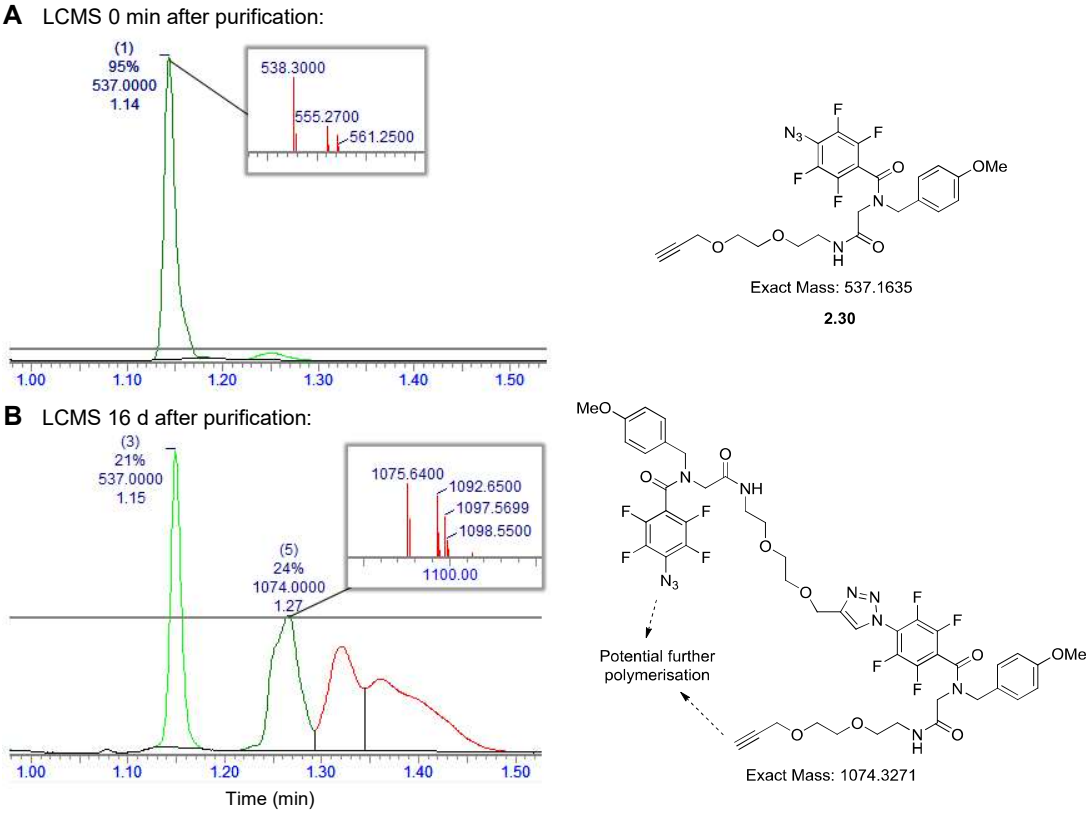


Figure 2.8 A) LCMS UV chromatogram of **2.30** immediately after purification (0 min timepoint in Figure **2.7**). B) LCMS UV chromatogram of **2.30** after 16 d stored as a solid at rt. $[M+H]^+$ 1075.6 Da was indicative of dimerisation.

2.3 Investigations into the effect of linker-length on photocrosslinking efficiency

It was postulated that variations in the vector of the photoreactive group and its linker length from the active site would have an influence on labelling efficiency. It was anticipated that an optimal linker length would constrain the photoreactive group close to the surface of the protein to favour crosslinking events over quenching of the reactive intermediate by solvent, while also being long enough so that binding of the affinity function would not be disrupted. To investigate the effect of linker length on photocrosslinking efficiency, a series of photoaffinity probes were designed and synthesised. A more potent pan-BET inhibitor was chosen as the affinity function, and amine derivatives with three different linker lengths were synthesised. These were employed with the five commonly used photoreactive groups previously discussed in a five-by-three array to obtain 15 PAL probes, where each photoreactive group was positioned at three different lengths from the affinity function. The combined impact of linker length and photoreactive group on physicochemical properties, cellular potency and crosslinking efficiencies was examined, along with crystallography and LC-MSMS studies to inform on the requirements for optimal PAL probes.

2.3.1 Selection of a potent BET inhibitor as a proof-of-concept affinity function

A potent pan-BET inhibitor **2.31** was chosen as the affinity function.[196,197] The X-ray crystal structure of **2.31** co-crystallised with BRD4 BD1 is shown in Figure 2.9 (Crystallography performed by Dr Chun-wa Chung). The pyridone warhead of **2.31** forms a hydrogen bond with N140 and with one of the four conserved water molecules within the BRD4 BD1 active site. The benzimidazole core forms a hydrogen bond with another water molecule and forms a CH–π interaction with a hydrogen of L92. The pendant tetrahydropyran occupies a hydrophobic region known as the WPF shelf (W81, P82 and F83). The 7-position of the benzimidazole core was identified as an appropriate solvent-exposed vector that could be elaborated to install primary amines to allow PAL probes to be synthesised using the Ugi protocol developed in this work. From this vector, three pendant amines were synthesised with a short (**2.32**), medium (**2.33**) and long (**2.34**) linker length respectively.

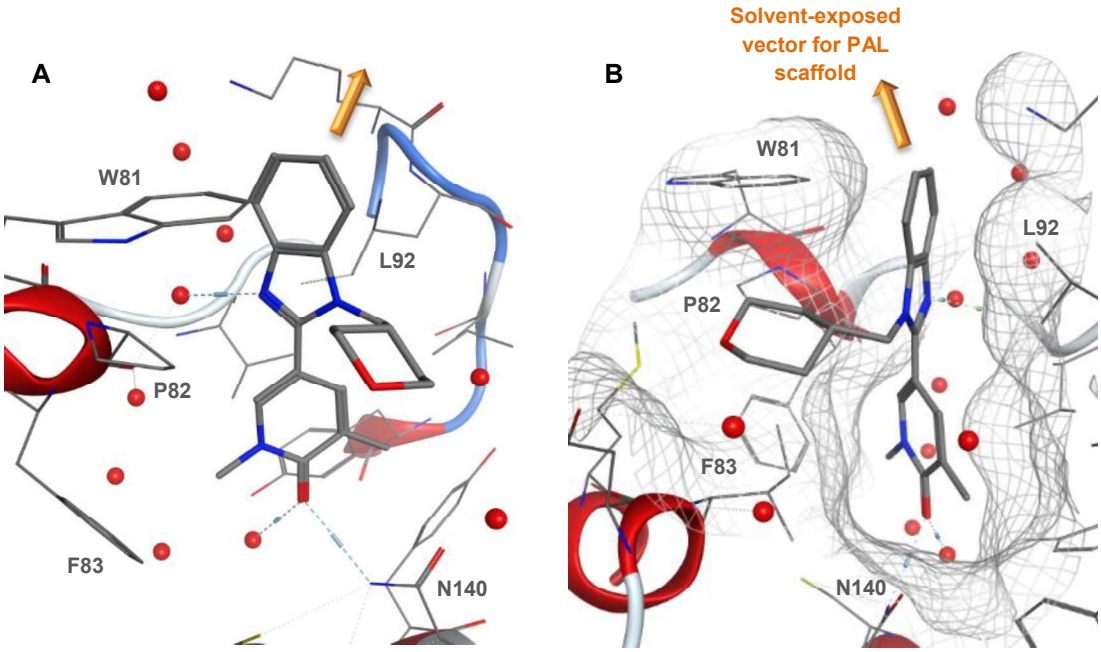


Figure 2.9 X-ray crystal structure of affinity function **2.31** co-crystallised with BRD4 BD1 (performed by Dr Chun-wa Chung). The carbonyl of the pyridone warhead forms a hydrogen bond with N140 and with a conserved water molecule. A) The benzimidazole core forms a CH–π interaction with L92 and a hydrogen bond with a water molecule through the unsubstituted nitrogen. B) The 7-position of the benzimidazole core was selected as the vector for further elaboration to install a primary amine and subsequently the PAL scaffold.

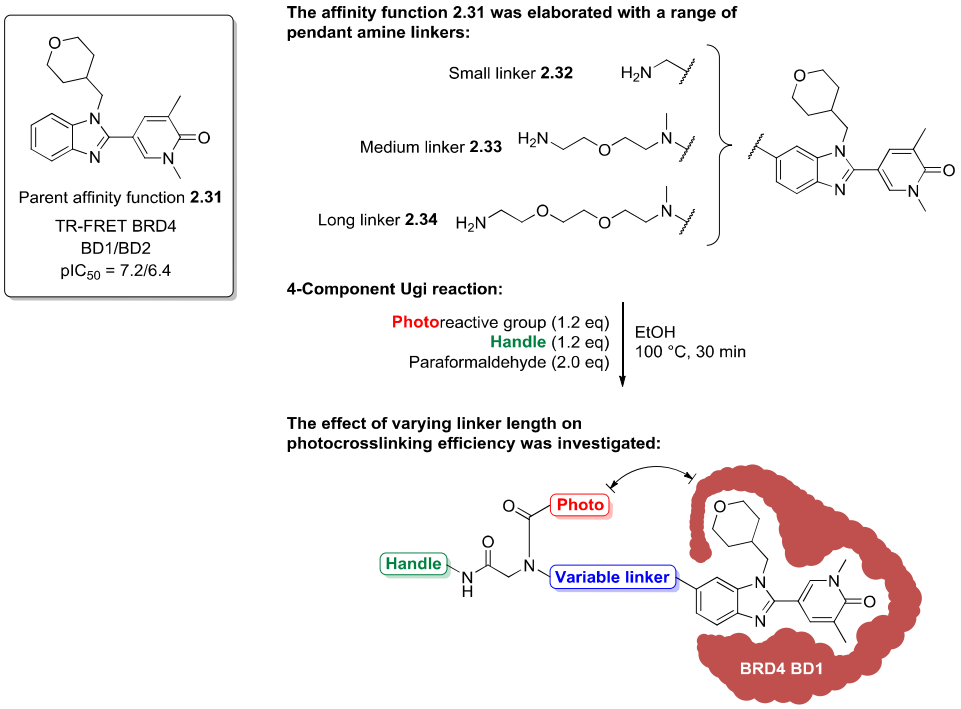
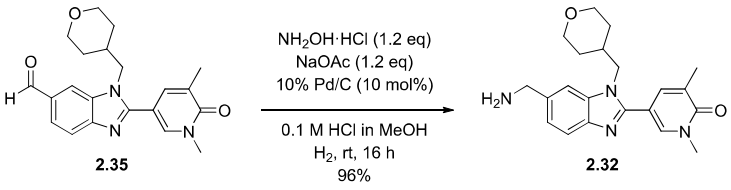


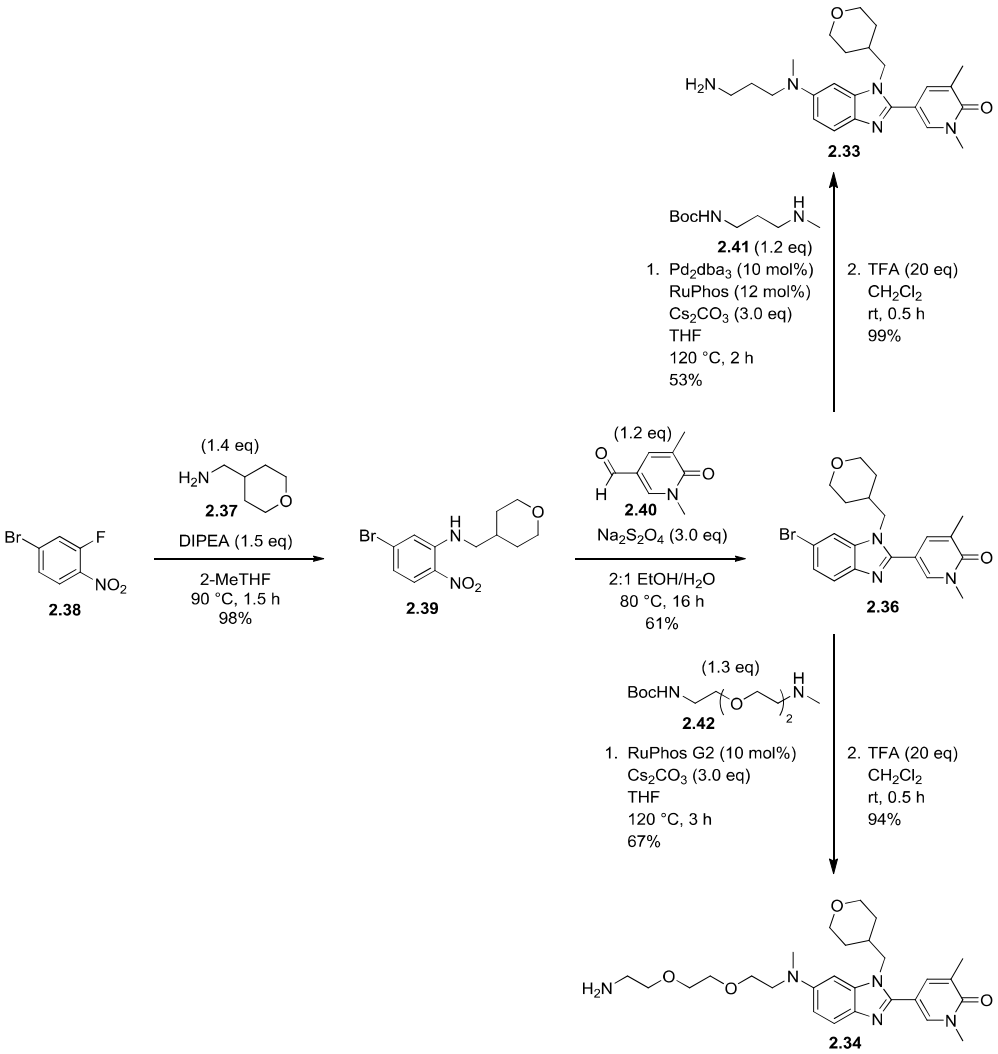
Figure 2.10 Three amine derivatives of the parent affinity function **2.31** containing short (**2.32**), medium (**2.33**) and long (**2.34**) linkers were synthesised. PAL probes were synthesised by the Ugi protocol and used to investigate the relationship between linker length and photocrosslinking efficiency.

2.3.2 Synthesis of the amine affinity function with three different linker lengths

To synthesise the short linker affinity function **2.32**, aldehyde **2.35** (obtained from the GSK compound collection) was condensed with hydroxylamine and the resulting imine was reduced by hydrogenation under acidic conditions (Scheme 2.5). The medium and long linker affinity functions (**2.33** and **2.34** respectively) were synthesised as shown in Scheme 2.6.

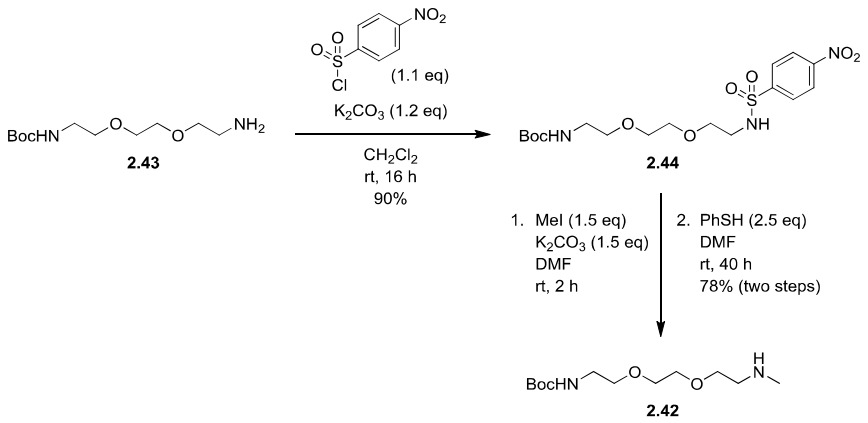


Scheme 2.5 Synthesis of short linker affinity function **2.32**.



Scheme 2.6 Synthesis of medium (**2.33**) and long linker (**2.34**) amine affinity functions.

The common intermediate **2.36** was synthesised by S_NAr reaction of the tetrahydropyran WPF shelf group **2.37** with **2.38**, followed by a reductive cyclisation with the pyridone warhead aldehyde **2.40** in moderate yield (60% over two steps). Buchwald-Hartwig coupling was then used to install the medium (**2.41**) and long (**2.42**) amine linkers. Subsequent Boc deprotection afforded **2.33** and **2.34** respectively. To prevent any unwanted condensation of formaldehyde on the aniline nitrogen during the Ugi reaction and to obtain more desirable physicochemical properties (fewer hydrogen bond donors), *N*-methylated linkers were used. The medium linker **2.41** was commercially available, and the longer linker **2.42** was synthesised following Scheme 2.7. Primary amine **2.43** was activated by forming a Fukuyama sulfonamide **2.44**, which was methylated using methyl iodide. Subsequent deprotection produced **2.42** in good overall yield (70%).



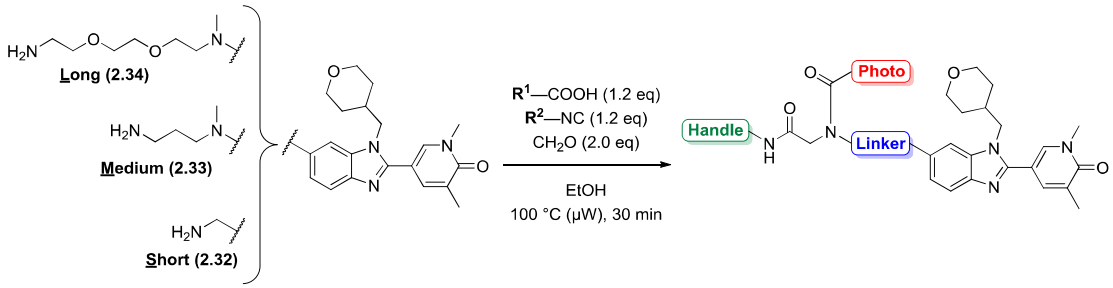
Scheme 2.7 Synthesis of *N*-methylated long linker **2.42**.

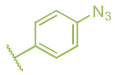
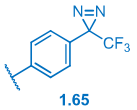
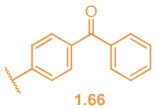
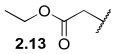
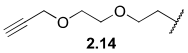
2.3.3 Using the Ugi protocol to synthesise 15 PAL probes *via* parallel synthesis

The three amine affinity functions (**2.32–2.34**) were combined with the five photoreactive groups (**1.16**, **1.63–1.66**) *via* parallel synthesis using the Ugi protocol to access 15 photoaffinity probes (**2.45–2.59**) (Table 2.2). The reactions were heated with microwave irradiation (100 °C, 30 min), and each reaction mixture was then purified directly by automated reverse-phase purification. Moderate synthetic yields were obtained (11–61%), further demonstrating the tolerance of these conditions for various amine affinity functions. The alkyne handle (**2.14**) was incorporated into all probes not

containing an azide functionality, where ethyl-ester isonitrile (**2.13**) was employed to avoid unwanted cycloaddition reactions (discussed in Section 2.2.5).

Table 2.2 Ugi reaction protocol used to synthesise 15 PAL probes by parallel synthesis. Five photoreactive groups (**1.16**, **1.63–1.66**) were coupled with three linker lengths (**2.32–2.34**).



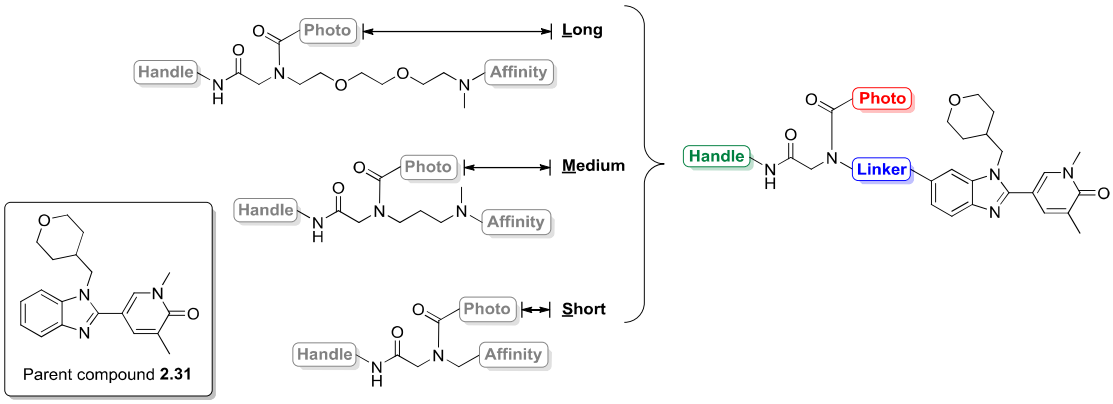
Photoreactive group	 1.16	 1.63	 1.64	 1.65	 1.66
S (2.32)	**2.45** (61%)	**2.46** (31%)	**2.47** (15%)	**2.48** (19%)	**2.49** (37%)
M (2.33)	**2.50** (60%)	**2.51** (44%)	**2.52** (33%)	**2.53** (11%)	**2.54** (47%)
L (2.34)	**2.55** (53%)	**2.56** (41%)	**2.57** (39%)	**2.58** (19%)	**2.59** (52%)
Handle	 2.13		 2.14		

2.3.4 Physicochemical properties and biochemical and cellular potencies of the 15 PAL probes

To establish how the various photoreactive groups might impact the cellular behaviour of the probes, the physicochemical properties of each probe at pH 7.4 were measured (Table 2.3, physicochemical properties were measured by Iain Reid, Ferdausi Mazumder, Terrence Johnson and Robert Armstrong). Good aqueous solubility was maintained (115–857 μM) upon addition of the PAL scaffold to the parent compound **2.31** (sol = 391 μM), while permeability was reduced. All probes showed an increase in chromLogD relative to **2.31**, which was most pronounced for the probes containing aromatic photoreactive groups. This increase in lipophilic character was reflected in an increase in percentage-binding to human serum albumin (HSA), which may lead to non-specific binding in whole cells or lysates. Overall, probes containing the alkyl diazirine (**2.47**, **2.52** and **2.57**) showed the least deviation in properties from parent **2.27**. Generally, introduction of the Ugi scaffolds showed a moderate but acceptable

effect on physicochemical properties, supporting its application as a general approach for the synthesis of PAL probes. Satisfyingly, all 15 probes retained pan-BET activity, showing similar levels of binding to BRD4 BD1 and BD2 (TR-FRET, $pIC_{50} \leq 0.6$ difference between each domain for all 15 probes) with a *ca*. 5 to 10-fold increase in biochemical potency over the parent compound **2.31**, indicating no disfavoured clashing of the Ugi scaffold with either domain. To assess cellular activity, the probes were screened in a human whole blood MCP-1 assay. The monocytic population within the blood was challenged with lipopolysaccharide (LPS), which causes downstream production of the pro-inflammatory cytokine MCP-1.[198] BET targeting compounds (such as I-BET762) have been shown to inhibit the release of this cytokine, an important component involved in many inflammatory diseases.[199] All 15 probes were found to have slightly reduced cellular potency (pIC_{50} = 5.9–6.5), compared to parent **2.31** (pIC_{50} = 6.9). This was likely due to a balance between increased target affinity (TR-FRET biochemical potencies) and reduced permeability, as suggested from the artificial membrane permeability (AMP).

Table 2.3 Physicochemical properties and potencies of 15 PAL probes (**2.45–2.59**) containing five photoreactive groups and three amine affinity functions with varied linker lengths. Physicochemical properties were measured by Iain Reid, Ferdausi Mazumder, Terrence Johnson and Robert Armstrong. TR-FRET assays and the MCP-1 whole blood assay were performed by Alex Phillipou.



Photoreactive group															
Linker length Compound	**2.31**	**S 2.45**	**M 2.50**	**L 2.55**	**S 2.46**	**M 2.51**	**L 2.56**	**S 2.47**	**M 2.52**	**L 2.57**	**S 2.48**	**M 2.53**	**L 2.58**	**S 2.49**	**M 2.54**	**L 2.59**
pIC_{50} BD1[a]	7.2	7.8	7.9	7.8	7.8	7.8	7.7	7.8	7.8	7.8	8.0	7.9	7.6	8.1	7.8	8.3
pIC_{50} BD2[b]	6.4	7.3	7.4	7.3	7.2	7.4	7.2	7.3	7.4	7.3	7.5	7.7	7.2	7.8	7.6	7.8
pIC_{50} MCP-1[c]	6.9	6.4	6.2	6.3	6.4	6.1	6.1	6.5	6.2	5.9	6.3	6.2	6.3	6.4	6.2	6.5
HSA (%)[d]	54	86	85	87	79	83	81	42	76	73	90	91	90	89	88	87
ChromLogD	2.7	4.4	4.5	4.8	3.6	3.8	4.1	2.9	3.5	3.8	4.6	4.8	5.0	4.0	4.1	4.4
Solubility (μM)[e]	452	293	115	244	254	232	231	519	398	444	379	422	491	303	453	857
AMP (nm/sec)[f]	223	98	27	76	23	19	21	<3	<3	<3	52	37	35	10	<3	7.9

[a] TR-FRET assay with recombinant BRD4 BD1. [b] TR-FRET assay with recombinant BRD4 BD2. [c] Assay for the inhibition of MCP-1 cytokine production from monocytes in human whole blood. [d] Human Serum Albumin binding. [e] Aqueous solubility measured by charged aerosol detection (CAD). [f] Artificial Membrane Permeability. ChromLogD, AMP and aqueous solubility were measured at pH 7.4.

2.3.5 Photocrosslinking timecourses of potent BET PAL probes at 302 nm irradiation

Timecourse experiments were conducted with BRD4 BD1 to assess the effect of the three linker-lengths on photocrosslinking yield. The 15 probes (6 μM) were incubated with BRD4 BD1 (3 μM) and irradiated (302 nm) over a timecourse (Figure 2.11). Comparing the photoreactive groups, the aryl azide probes gave the highest levels of labelling and rate of photoactivation for all three linker lengths. Probe **2.45** gave the highest levels of labelling (*ca.* 80%) and showed an extremely fast rate of photoactivation, reaching full activation before the first irradiation timepoint (<1 min). A more detailed timecourse with shorter irradiation times was conducted (Figure 2.12). Probe **2.45** showed full photoactivation within 15 seconds of irradiation under these conditions. Applications exploiting the high level and fast rate of BRD4 BD1 labelling

by **2.45** were explored further and are presented in Chapter 3. However, the application of aryl azides in cell-based PAL studies were limited due to their observed incompatibility with alkyne handles. Of the remaining three alkyne-compatible photoreactive groups, benzophenone gave the highest crosslinking yields. Interestingly, these probes showed high crosslinking yields for the short (**2.49**) and long linker (**2.59**) (*ca.* 40% and 60% respectively), while the medium linker (**2.54**) gave poor crosslinking (<5%). The aryl and alkyl diazirines showed moderate levels of photolabelling, best exemplified by the short linker probe **2.49** (15%). The alkyl diazirines were slower to photoactivate, requiring *ca*. 20 min irradiation versus *ca*. 2 min for the aryl CF_3 diazirines. Although the alkyl diazirine probes had the most favourable physicochemical profiles, their slow activation and low levels of photolabelling may lead to inefficient capture of target proteins in live cells and lysates.

Overall, the short linker probes (**2.45–2.49**) gave the highest levels of photolabelling, with an average labelling of 39% after 10 min of irradiation (*versus* 10% and 17% for the medium (**2.50–2.54**) and long linker (**2.55–2.59**) probes respectively). The large reduction in photocrosslinking yield between the short and medium linker was surprising, since the linker length had only increased by three atoms. The long linker probes showed the lowest levels of photocrosslinking, except for the benzophenone probe **2.59**. Investigations into the high levels of labelling by **2.59** are presented in Section 2.3.8.

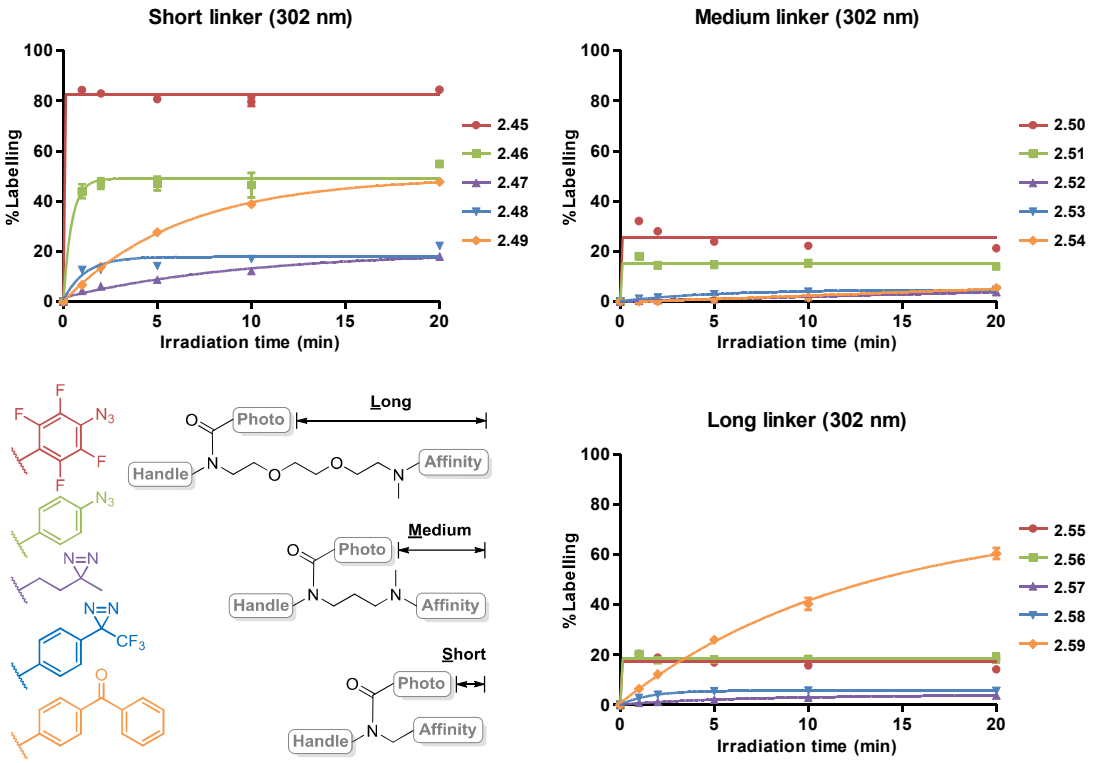


Figure 2.11 Photocrosslinking timecourses conducted under 302 nm irradiation with recombinant BRD4 BD1 (3 µM) and 15 PAL probes (6 µM) grouped by linker length. Photocrosslinking yields were determined by intact protein LCMS and calculated using Equation 2.1.

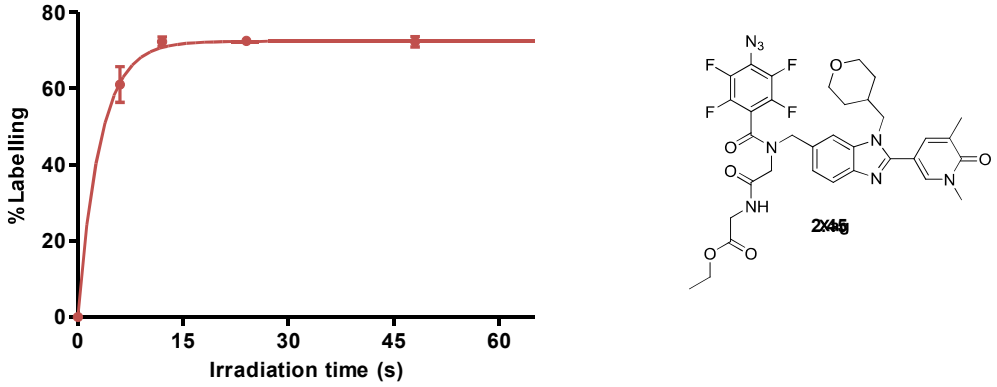


Figure 2.12 Short photocrosslinking timecourse for tetrafluoroaryl azide probe **2.45**. Probe **2.45** (10 µM) was incubated with recombinant BRD4 BD1 (1 µM) on ice and irradiated for the given length of time before determining the percentage labelling by intact protein LCMS analysis and Equation 2.1.

2.3.6 Demonstrating specific PAL labelling to the active site of BRD4 BD1

To demonstrate that the observed labelling for all 15 probes was due to specific binding in the active site of BRD4 BD1, a mixture of BRD4 BD1 (3 μM) and PAL probe (6 μM) was irradiated (302 nm, 10 min) with and without an excess of competitive ligand **2.60** (100 μM) (Figure 2.13). Without competition (red), the mass adduct corresponding to the photoactivated probe was observed. For all probes, no labelling was observed in the presence of excess **2.60** (blue). This demonstrated that the observed labelling was due to the PAL probe occupying the same binding site as the parent affinity function (*i.e.* the acetyl lysine binding pocket) prior to irradiation. This was consistent with the retained biochemical potency of the PAL probes *vs* the parent affinity function **2.31**.

These photoaffinity labelling experiments with recombinant BRD4 BD1 demonstrated that photocrosslinking yields are highly dependent on the choice of photoreactive group, irradiation wavelength and linker length. X-ray crystallographic studies with BRD4 BD1 and selected probes were pursued, with the aim of identifying how the local protein environment surrounding the photoreactive group may influence the photocrosslinking yields observed in these experiments.

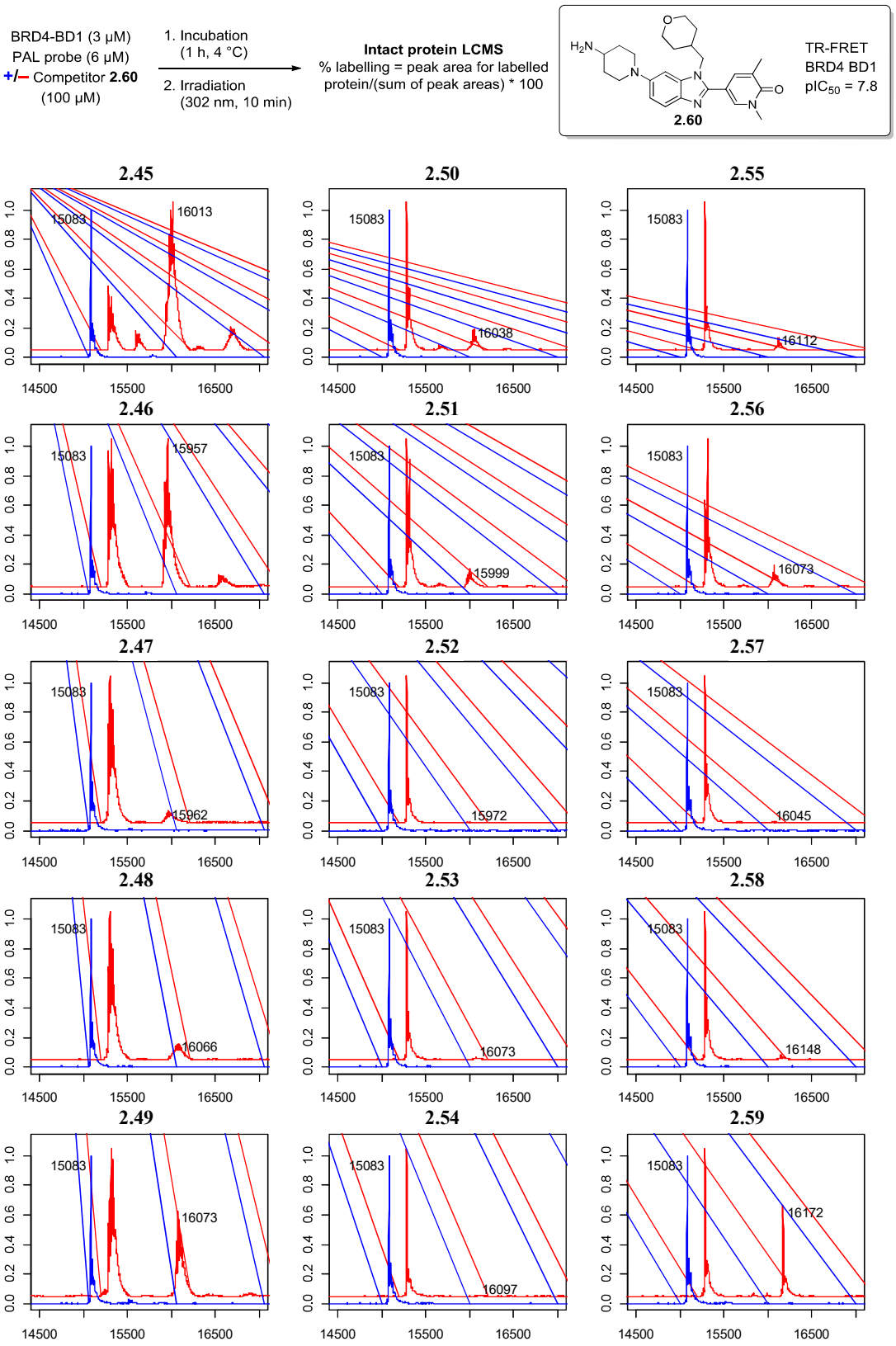


Figure 2.13 Deconvoluted LCMS spectra following irradiation of 15 PAL probes (6 µM) with BRD4 BD1 (3 µM) with (blue) and without (red) competitor compound **2.60** (100 µM) present. Photocrosslinking was inhibited in the presence of competitor, indicating that labelling was specific due to the PAL probe occupying the same binding site as the parent affinity function.

2.3.7 X-ray crystallography of selected PAL probes with BRD4 BD1

To better understand the relationship between linker length, photoreactive group and photocrosslinking efficiency, selected probes were co-crystallised with BRD4 BD1. For the probes that showed the highest levels of labelling (**2.45** and **2.46**), co-crystal structures with BRD4 BD1 were obtained (Figure 2.14, X-ray crystallography performed by Dr Chun-wa Chung).

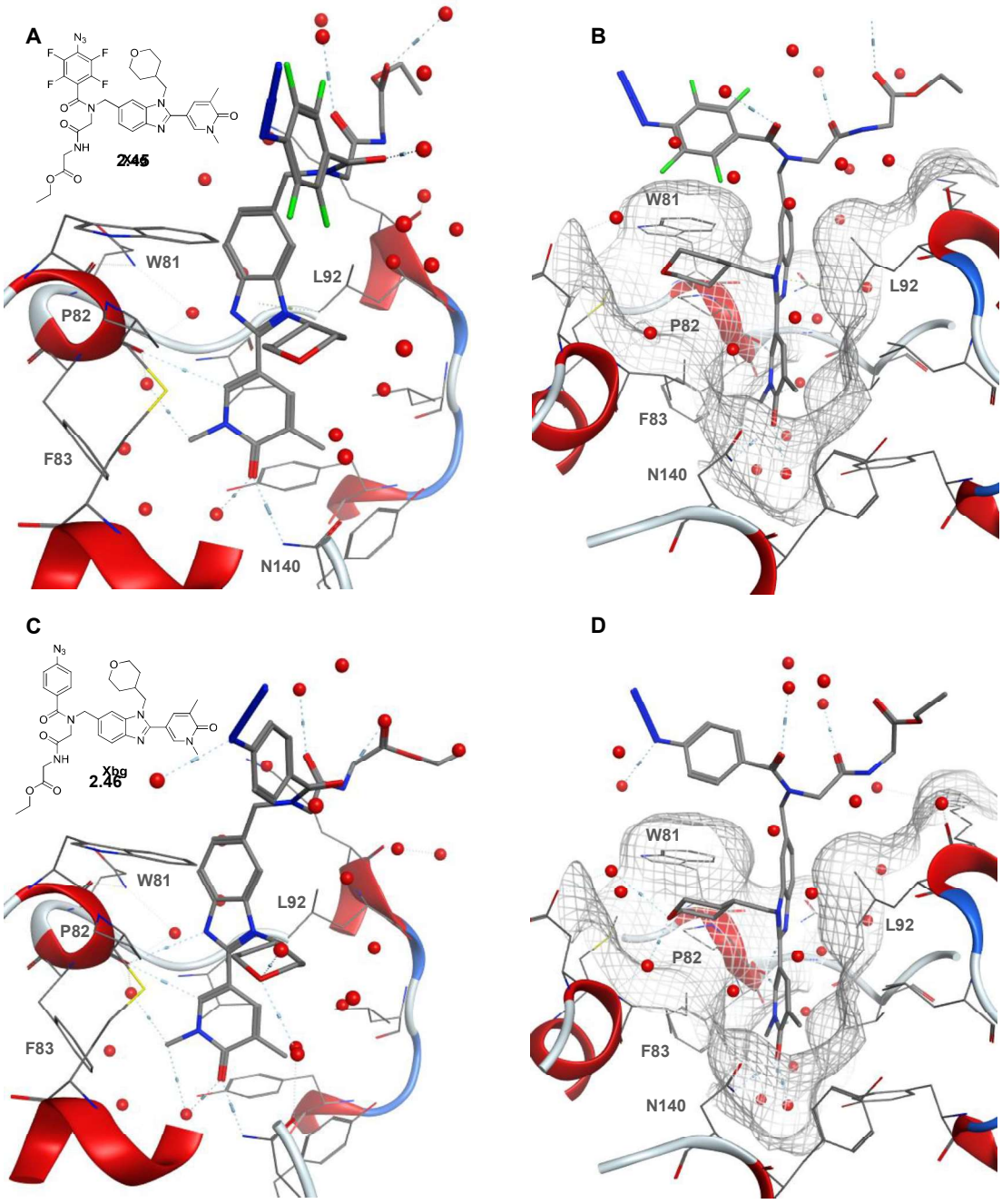


Figure 2.14 Co-crystal structures of **2.45** (A, B) and **2.46** (C, D) with BRD4 BD1. Both probes share a similar binding mode, and the key interactions of the affinity function were conserved (See Figure 2.9). The short linker projects the Ugi scaffold out of the binding pocket so that binding by the affinity function is not disturbed, and the photoreactive groups are retained close to the surface of the protein. Both photoreactive groups occupy the same space proximal to W82 of the WPF shelf. X-ray crystallography was performed by Dr Chun-wa Chung.

In both cases, the interactions of the affinity function were conserved with those of the parent **2.31** (see Figure 2.9) and both aryl azide photoreactive groups were orientated proximal to W82 of the WPF shelf. Despite these similar binding modes, **2.45** produced higher levels of labelling *versus* **2.46** (80% *vs* 50% respectively), which may be explained through the lifetimes of their reactive intermediates. Previous studies have shown that upon irradiation, the tetrafluoroaryl azide photoreactive group **1.16** produces a longer-lived singlet nitrene, and also that the re-arranged dehydroazepine intermediate is more electrophilic than the non-fluorinated analogue produced from **1.63**.[98]

A co-crystal structure of the short linker trifluoromethyl aryl diazirine probe **2.48** with BRD4 BD1 was also obtained. The structure captured the photoreactive group in two conformations (Figure 2.15, X-ray crystallography performed by Dr Chun-wa Chung). In the first conformation, the trifluoromethyl aryl diazirine was found proximal to W82, similar to the binding mode of the two aryl azide photoreactive groups (Figure 2.14). The second conformation positioned the photoreactive group on the opposite side of the acetyl lysine site, close to L92. In this second conformation, the photoreactive group may be in a sub-optimal position to react with neighbouring residues, contributing to the low photocrosslinking yields observed for **2.44** in the photocrosslinking timecourse studies. The three crystal structures obtained for probes **2.45**, **2.46** and **2.48** illustrate that the short linker projects the PAL Ugi scaffold far enough out of the binding pocket as to not disrupt binding of the affinity function, while still retaining the photoreactive groups close to the surface of the protein.

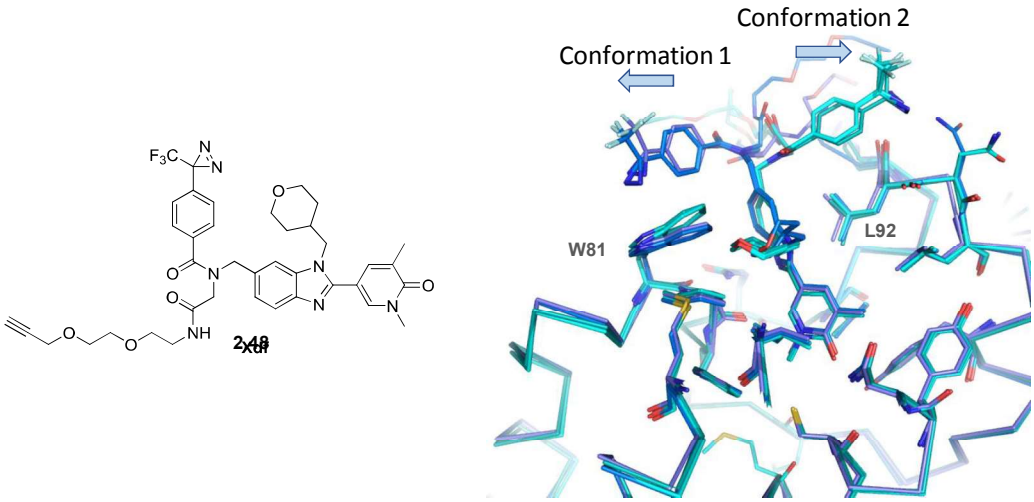


Figure 2.15 Co-crystal structure of probe **2.48** in the acetyl lysine binding pocket of BRD4 BD1. The trifluoromethyl aryl diazirine photoreactive group occupies two different conformations. One conformation showed the photoreactive group proximal to W81, which was similar to the binding mode shown by the aryl azide photoreactive groups. The second conformation placed the photoreactive group proximal to L92. X-ray crystallography was performed by Dr Chun-wa Chung.

The benzophenone probe **2.59** was the only long linker probe to give high levels of labelling (*ca*. 60%) in the photocrosslinking timecourse experiments with BRD4 BD1. It was proposed that the long and flexible PEG linker may allow for the hydrophobic benzophenone (and the diradical excited state produced upon irradiation) to sample transient interactions and crosslink to residues distal from the acetyl-lysine pocket. A co-crystal structure of **2.59** in BRD4 BD1 was obtained and is shown in Figure 2.16 (X-ray crystallography performed by Dr Chun-wa Chung). The benzophenone photoreactive group was observed in two different conformations within the asymmetric unit of this crystal structure, highlighting its high flexibility. One conformation places the benzophenone in an energetically unfavourable out-of-plane conformation, close to the protein surface and forming a hydrogen bond with W81. The second conformation extends the benzophenone into solvent, where more relaxed dihedral angles were adopted. The first may represent a conformation that leads to protein labelling, and the second may be expected to lead to reaction with water or buffer components. Both are likely to be two of many conformations dynamically sampled in solution. Benzophenone has been shown to label methionine residues regioselectively (see Section 1.2.2.3).[118,119] The first conformation positions the benzophenone close (6.7 Å) to M149, which may account for the high levels of

photocrosslinking observed (*ca*. 60%) in solution. To find the residue that was covalently modified by **2.55** in solution, tryptic digestion and LC-MSMS experiments were performed.

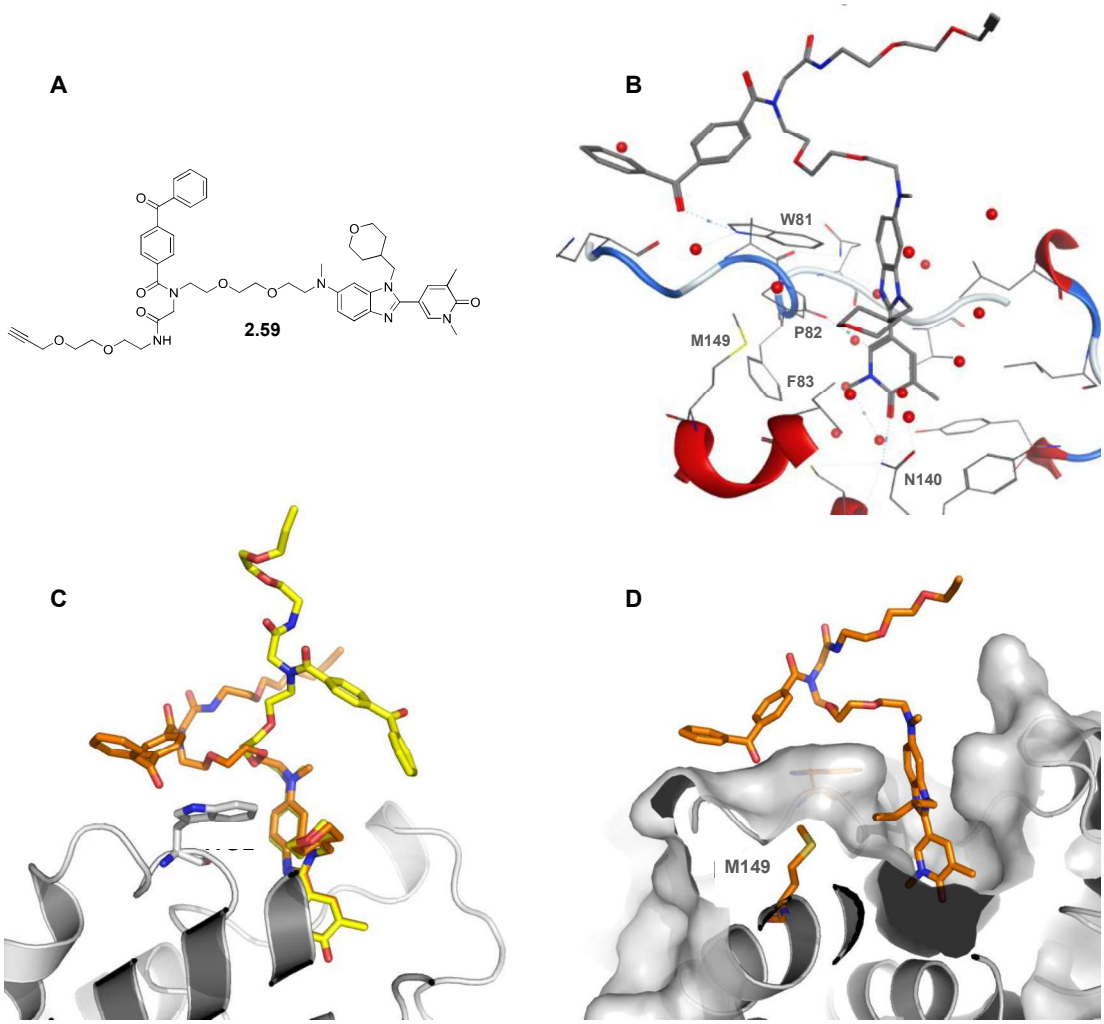


Figure 2.16 (A, B) Co-crystal structure of probe **2.59** with BRD4 BD1. The affinity function showed the same conserved interactions as parent **2.31** (C, D) The benzophenone of chain A (orange) was observed adjacent to W81 of the WPF shelf and in close proximity (6.7 Å) to M149. Benzophenone of chain B was directed into solvent (yellow). X-ray crystallography was performed by Dr Chun-wa Chung.

2.3.8 Identifying the site of photocrosslinking with probe 2.59 to BRD4 BD1

LC-MSMS analysis was used to identify the site of covalent modification by **2.59** (Figure 2.17). BRD4 BD1 (10 µM) was incubated with **2.59** (10 µM) and irradiated (302 nm). A shorter irradiation time (6 min) was chosen to minimise photo-damage to the protein, while still achieving a moderate photocrosslinking yield. A non-irradiated

control was also prepared. The samples were analysed by intact protein LCMS to calculate a photocrosslinking yield (*ca.* 40%) and confirm the correct mass for covalent modification with **2.59** (Figure 2.18). To identify the site of modification, the samples were purified by SDS-PAGE and the protein bands corresponding to BRD4 BD1 were excised and digested with a combination of trypsin and endopeptidase LysC. The digests were analysed by LC-MSMS using a data-dependent acquisition (DDA) method. Proteolytic digestion and LC-MSMS analysis was conducted by Ken Fantom.

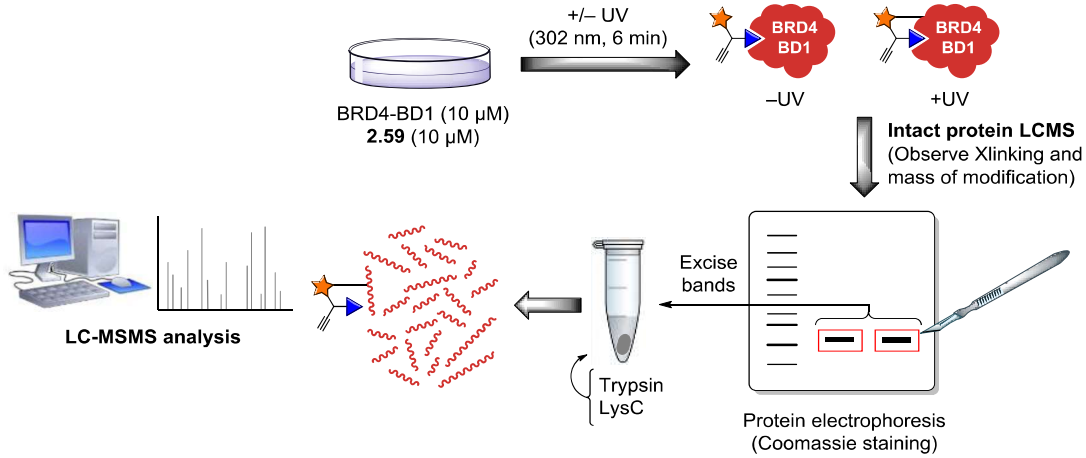


Figure 2.17 LC-MSMS analysis workflow for identifying the site of covalent modification by probe **2.59** on BRD4 BD1. Protein gel electrophoresis and LC-MSMS analysis was performed by Ken Fantom.

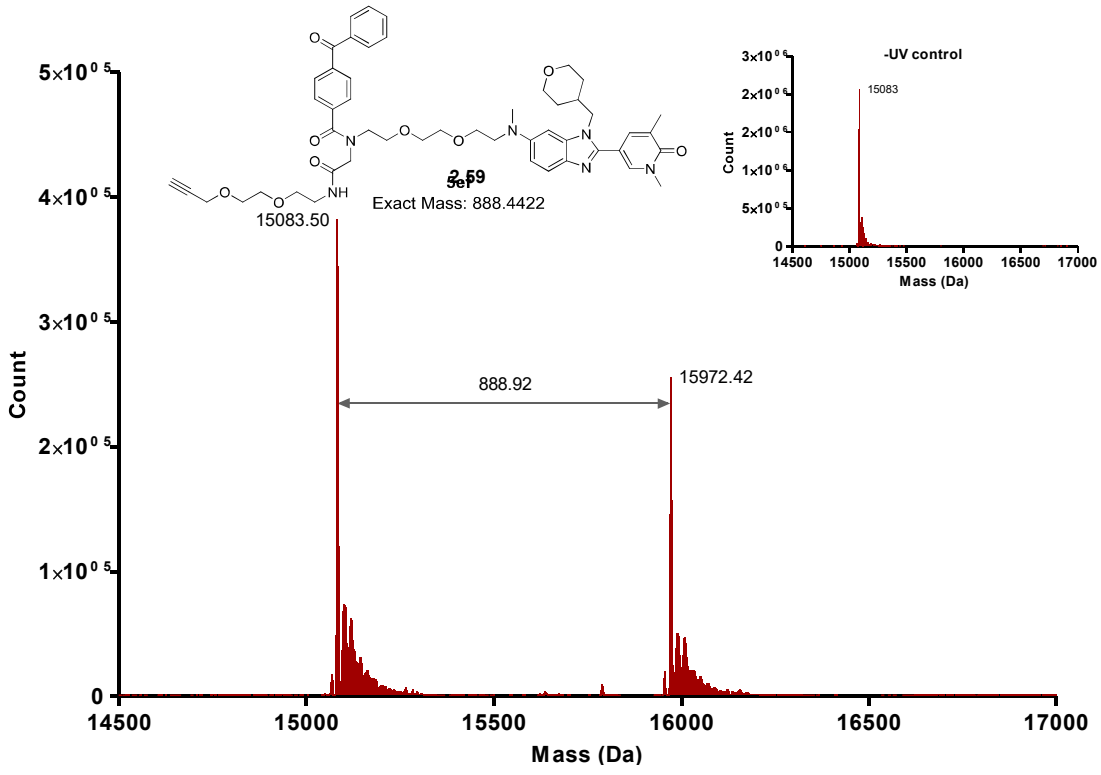


Figure 2.18 Intact protein LCMS analysis of BRD4 BD1 (10 µM) irradiated (302 nm, 6 min) with **2.59** (10 µM). No labelling was observed without UV irradiation.

Uninterpreted MSMS spectra were matched against the BRD4 peptide sequence using Mascot and allowing for **2.59** (888.4422 Da) as a variable modification on any residue. High sequence coverage (86%) was achieved for the +UV sample. The 3+ ion for the BRD4 peptide PGDDIVLMAEALEK (142–156) + **2.59** was identified (*m/z* 797.0735) (Figure 2.19). To uniquely assign the site of modification within the peptide sequence, the sample was analysed using a targeted parallel reaction monitoring (PRM) method, where only the precursor ions of *m/z* 797.0735 (3+ ion of **2.59**-modified PGDDIVLMAEALEK) were selected for fragmentation. This PRM analysis robustly identified **2.59** modification on M149 (Figure 2.20).

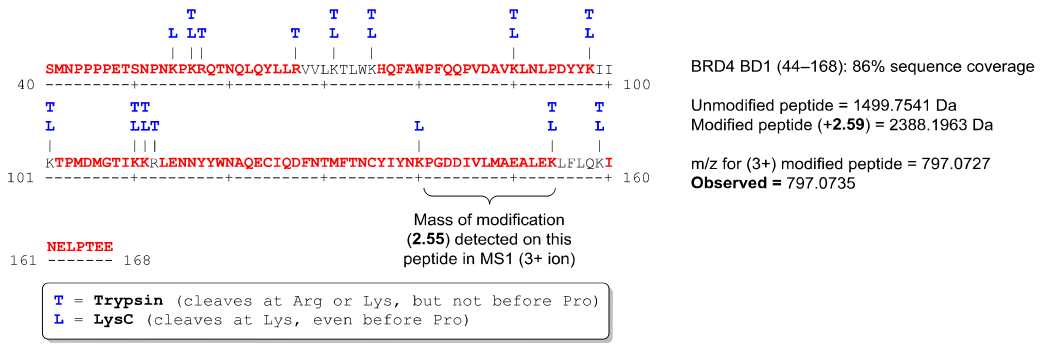


Figure 2.19 The amino acid sequence of BRD4 BD1 (44-168) recombinant protein used for tryptic digestion and LC-MSMS analysis. Peptides in red were identified by LC-MSMS analysis, giving 86% sequence coverage. A Mascot search (allowing for +**2.59** as a variable modification) identified a 3+ ion (*m/z* 797.0735) corresponding to the modified peptide PGDDIVLMAEALEK+**2.59**.

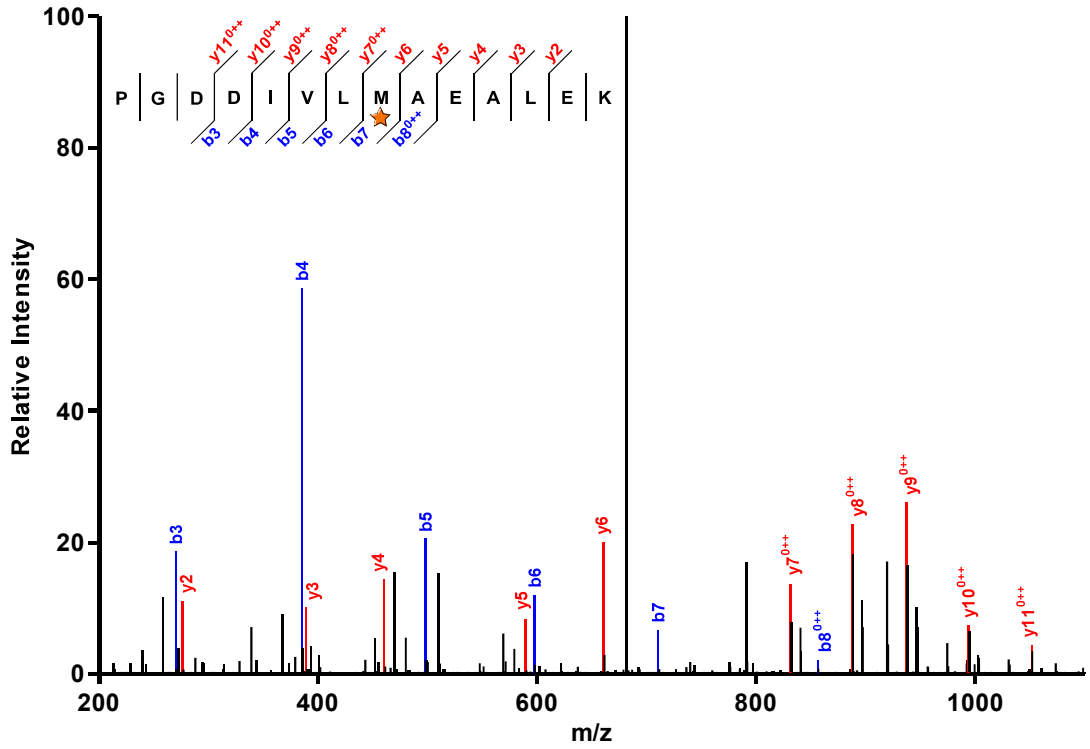


Figure 2.20 MS2 spectra for the 3+ ion (*m/z* 797.0735) corresponding to modified peptide PGDDIVLMAEALEK+**2.59** from a parallel reaction monitoring (PRM) LC-MSMS method. Good coverage of b- and y-ions were observed, indicating modification (**2.59**) of M149.

These findings demonstrate that constraining the photoreactive group close to the surface of the protein *via* a short linker is not the only effective approach for achieving efficient labelling. In this case, a long linker allowed benzophenone to regioselectively react with a distal methionine, resulting in moderate levels of labelling. However, in the absence of any structural information regarding the active site of the protein, this moderate level of labelling would be impossible to predict. Furthermore, if this strategy was used to identify novel binding sites on new proteins, any labelled methionine residues may be remote from the true binding site occupied by the affinity function, hindering accurate determination of the true binding site.[115] Although probe **2.59** used in this work achieved high levels of labelling, the regioselectivity for methionine shown by benzophenone is antithetic to an ideal photoreactive group. Ideal photoreactive groups should label neighbouring residues without bias and not require positioning close to preferred residues to achieve high levels of labelling.[115]

2.3.9 Comparison of protein oxidation *versus* linker length

Of the 15 PAL probes examined in the photocrosslinking timecourse experiments, some probes caused oxidation of the protein, characterised by multiple +16 Da species. For example, the peaks for both unmodified and labelled protein observed with the short linker probes **2.45–2.49** (Figure 2.13, left column of plots) were relatively broad (*ca*. 50 Da) *versus* the narrow peaks observed with the medium and long linker probes after irradiation (302 nm, 10 min). An expanded view of the unmodified protein peak for the short *vs* long linker probes is shown in Figure 2.21. The short linker probes produced extensive oxidation of the protein, which appeared to be independent of photocrosslinking yield. One postulation for these high levels of oxidation is that reactive intermediates or reactive oxygen species (ROS) are formed from the initial excited photoreactive group during irradiation, which are produced close to the surface of the protein by the short linker probes. These reactive species may cause protein oxidation for the short linker probes but be quenched by solvent or buffer components for the medium and long linker probes. Despite these speciation events, the overall photocrosslinking yields could be determined using intact protein LCMS. However, the extent of protein oxidation induced by the PAL probes may need to be addressed in other PAL applications, such as those involving proteolytic digestion and analysis by LC-MSMS (*e.g.* photocrosslinking site determination or chemoproteomics). In these experiments, any oxidative modifications to the protein would need to be recognised by the data processing algorithm. Therefore, the levels of protein oxidation induced by the PAL probe may be another parameter to consider when developing probes.

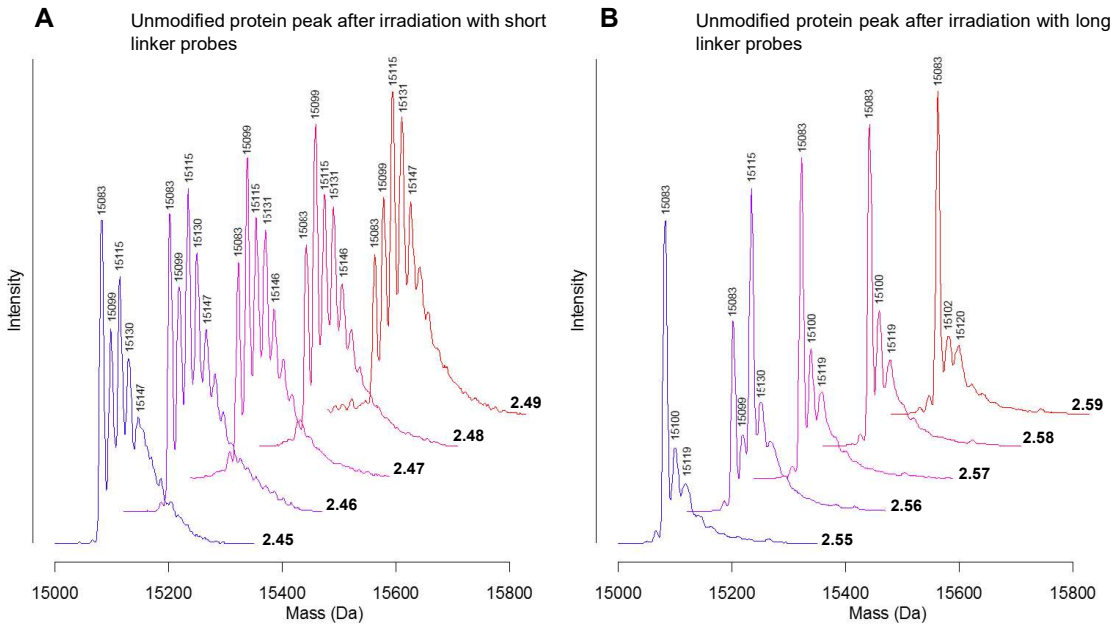


Figure 2.21 Unmodified protein peak in the deconvoluted spectra following irradiation (302 nm, 10 min) with the short (A) and long (B) linker probes. The short linker probes showed extensive oxidation (+16 Da species), possibly due to the photoreactive group (and the reactive intermediates produced upon excitation) being produced close to the surface of the protein.

2.3.10 Photocrosslinking timecourses of potent BET PAL probes at 365 nm irradiation

Photocrosslinking timecourse studies for all 15 probes were also carried out under 365 nm irradiation (Figure 2.22). At this wavelength, poor photocrosslinking yields were observed for all short linker probes. The medium and long linker aryl azides (**2.51** and **2.56**) afforded high levels of labelling (*ca.* 80%) and tetrafluoroaryl azides (**2.50** and **2.55**) afforded moderate levels of labelling (*ca*. 20%). Their rate of photoactivation was slower (*ca*. 5 min for complete activation) relative to their activation rate at 302 nm (<1 min). This inversion of labelling efficiency between the aryl azides at 365 *vs* 302 nm further demonstrated the challenges in predicting labelling efficiencies. Interestingly, the medium and long linker aryl azides **2.51** and **2.56** appeared to show a much faster rate of photoactivation *versus* the short linker aryl azide **2.46**, suggesting that the linker length can influence the photoexcitation of the aryl azide, perhaps by changing the electronic environment of the photoreactive group. The diazirine and benzophenone probes gave low levels of photolabelling (<10% after 80 min). Despite

these photoreactive groups having a reported λ_{max} *ca*. 350–360 nm, the 365 nm lamp used in this work may not have a strong enough irradiance (7.6 mW cm^{-2}) to achieve fast photoactivation rates (see Section 2.2.4).[39,62,100,115]

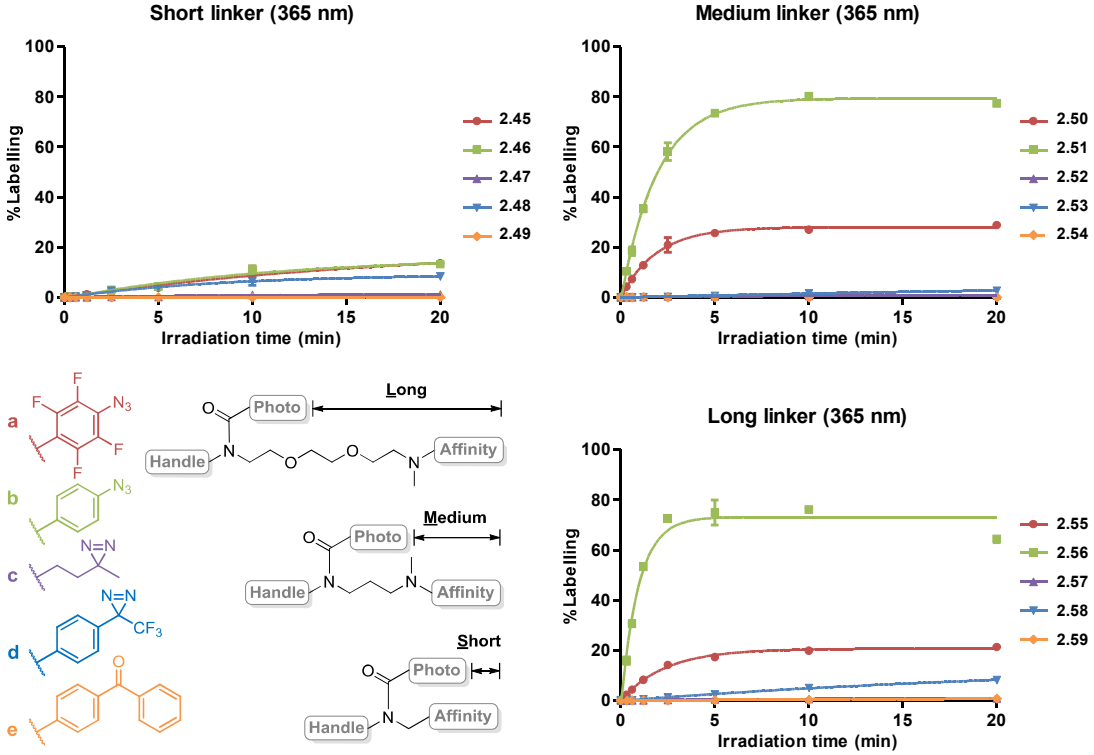


Figure 2.22 Photocrosslinking timecourses following irradiation (365 nm) with recombinant BRD4 BD1 (3 µM) and 15 PAL probes (6 µM), grouped by linker length.

2.3.11 Comparison of photodegradation to BRD4 BD1 by 302 nm *versus* 365 nm irradiation

The overall signal (sum of unmodified and single labelled protein peak area in the deconvoluted spectra) decreased with increasing irradiation time for 302 nm but remained approximately constant for 365 nm irradiation. This was indicative of photodegradation of the protein by 302 nm irradiation. To assess the photostability of BRD4 BD1 under both lamps used in this work, a timecourse with BRD4 BD1 with no PAL probe present was conducted with 302 nm and 365 nm irradiation (Figure 2.23).

The peak area for the protein was plotted against irradiation time. BRD4 BD1 was stable under 365 nm irradiation, however 302 nm irradiation degraded the protein with

a half-life of approximately 5 minutes. The mechanism of the photodegradation was unclear, as no other prominent peaks were observed in the intact mass spectrum (600-3200 Da) or in the deconvoluted spectra (>4000 Da). This highlighted that the choice of irradiation wavelength should not only be based on achieving high PAL probe activation rates, but also on the photostability of the protein, as other proteins may show a higher or lower photosensitivity than BRD4 BD1.

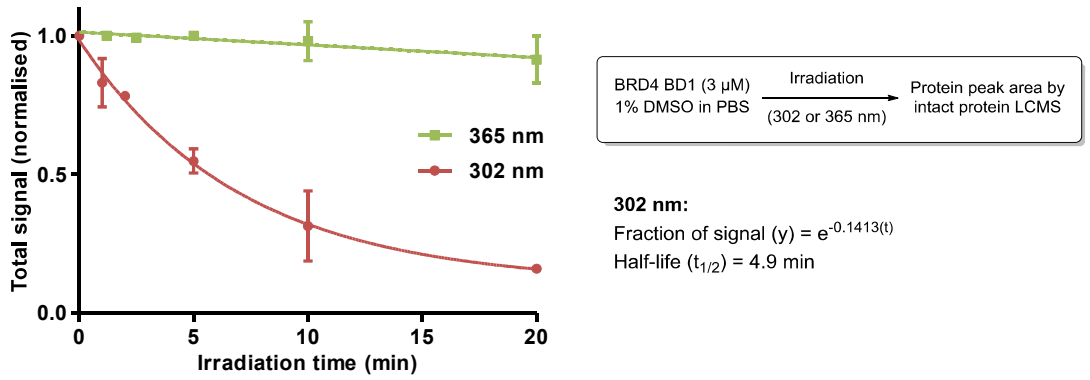


Figure 2.23 Photodegradation of BRD4 BD1 deconvoluted peak area with increasing irradiation time at 302 nm and 365 nm. The deconvoluted protein peak area for unmodified BD1 was obtained by intact protein LCMS. Appreciable photodegradation was observed with 302 nm irradiation, with half the signal being lost after *ca*. 5 min of irradiation.

2.3.12 Future work to aid binding site identification studies

Identification of the site of photocrosslinking can be challenging, and is further confounded if the labelling efficiency is low.[26] A single labelling event may be a summed result of photocrosslinking to numerous different residues. Furthermore, any photoinduced oxidations or other modifications to the peptides reduces the observable signal for each species, and unless the search algorithms can account for these modifications, they will not be identified.

Due to the difficulties in identifying the site of covalent modification(s) by PAL, a number of strategies have been recently developed to aid identification of labelled peptides.[200-202] One approach is to use a mixed isotope strategy, where two structurally identical probes are synthesised, but differ by their exact mass by the incorporation of stable isotopes within the second probe (Figure 2.24).[200,203] This approach has been coined Stable Isotope Labelled Inhibitors for Crosslinking (SILIC).[204]

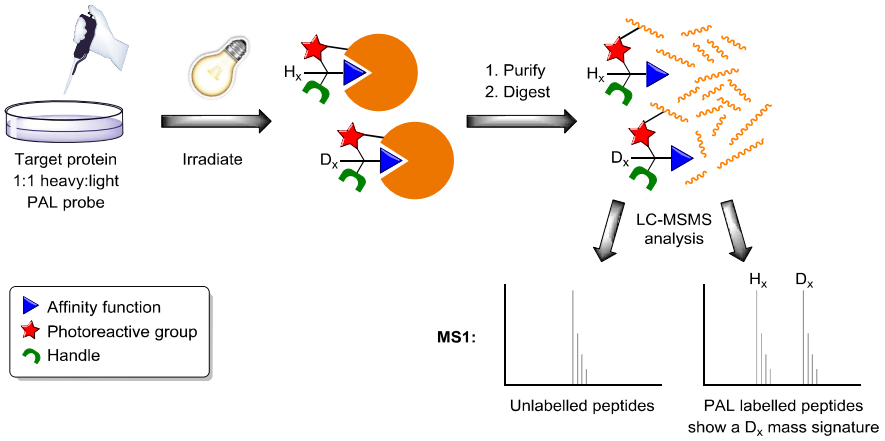


Figure 2.24 Stable Isotope Labelled Inhibitors for Crosslinking (SILIC) for identifying photolabelled peptides in the MS1 spectra of LC-MSMS analysis. The target protein is labelled with a 1:1 mixture of PAL probe containing heavy and light isotopes. Labelled peptides show a characteristic isotope pattern reflective of the mass difference between the heavy and light probes.

The Ugi protocol described in this work may be well suited to this approach, as deuterated paraformaldehyde (*ca*. £50/g) could be used to synthesise D_2-PAL probe analogues (Figure 2.25).

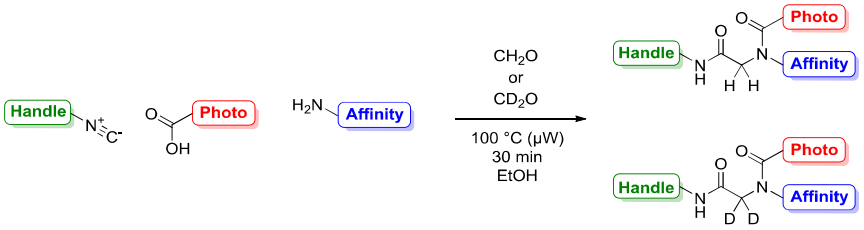


Figure 2.25 The Ugi protocol described in this work could be used with deuterated paraformaldehyde to access probes suitable for SILIC studies to find sites of covalent modification by PAL.

2.4 Conclusions

This work describes the development of a one-step protocol for the synthesis of PAL probes using the Ugi four-component reaction. This protocol is compatible with an array of affinity functions, photoreactive groups and bio-orthogonal handles. The protocol is applicable to parallel synthesis, which was used to generate 15 PAL probes containing five different photoreactive groups at three different linker lengths from the affinity function. These probes were used to investigate the relationship between photoreactive group, linker length and irradiation wavelength on photocrosslinking

efficiency with BRD4 BD1. For the probes that showed high levels of photocrosslinking, X-ray crystallography and LC-MSMS analysis were used to inform on the factors that determine efficient labelling at the residue level.

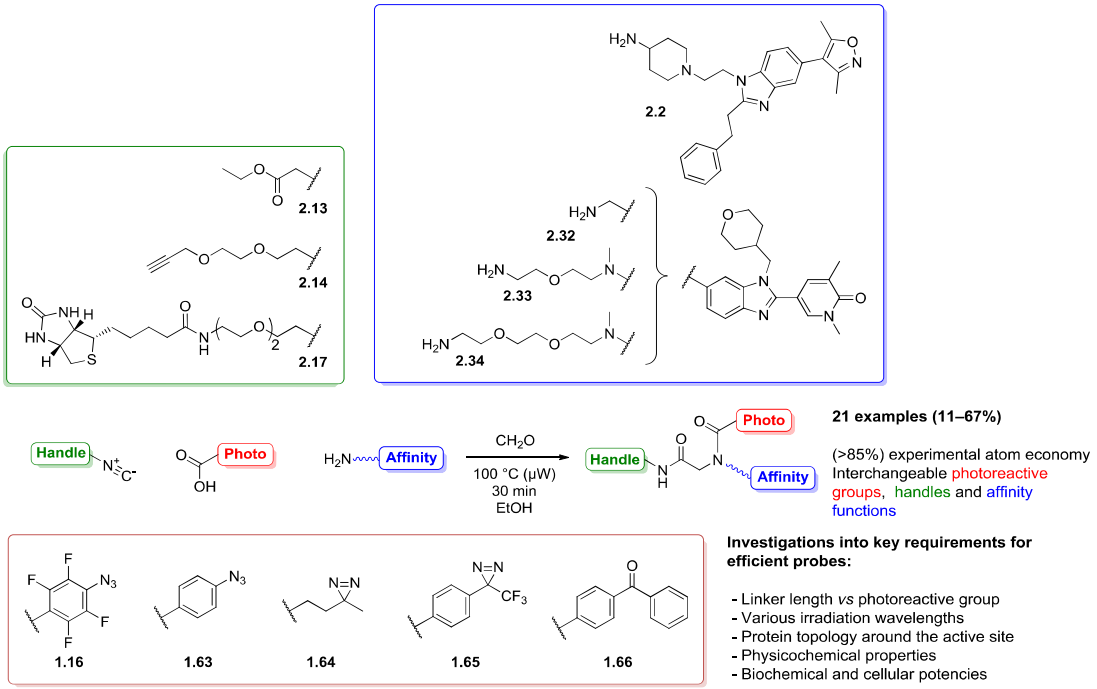


Figure 2.26 The Ugi four-component reaction protocol for the synthesis of an array of PAL probes in one step. The reaction showed good synthetic tolerance for a range of affinity functions, photoreactive groups and handle moieties. The protocol was used to synthesise 15 PAL probes in parallel that contained various photoreactive groups at short, medium or long linker lengths. These probes were used to investigate the relationship between photoreactive group, linker length and irradiation wavelength on photocrosslinking efficiency.

A comparison of two irradiation wavelengths (302 nm *vs* 365 nm) revealed that 302 nm gave faster rates of PAL probe activation and higher levels of labelling despite the 365 nm lamp having twice the irradiance of the 302 nm lamp (7.55 mW cm^{-2} *vs* 3.78 mW cm^{-2} respectively). However, irradiation at 302 nm also induced photodegradation of the protein, whereas the protein was stable under 365 nm irradiation.

Of the five photoreactive groups compared in this work, the aryl azides showed the highest and fastest levels of labelling, particularly the short linker tetrafluoroaryl azide **2.45**. A significant finding from these studies was the observed incompatibility of the aryl azide photoreactive groups with the alkyne handle. The alkyne handle is often

used to enable PAL pulldown experiments in lysates and live cells, and therefore represents a significant limitation in the application of aryl azide PAL probes.

Of the remaining three alkyne-compatible photoreactive groups, the trifluoromethyl aryl diazirine and benzophenone probes gave low to moderate levels of photocrosslinking. The alkyl diazirine probes gave the lowest photocrosslinking yields to BRD4 BD1 used in these studies.

Comparing the short, medium and long linker lengths, the highest levels of labelling were achieved with the short linker probes. X-ray crystallography of selected short linker probes showed that the short linker length placed the photoreactive group at an optimal position on the surface of the protein just outside of the acetyl lysine binding pocket. However, these short linker probes also gave extensive levels of protein oxidation compared to the medium or long linker probes, likely due the production of oxidative reactive intermediates in a high local concentration at the surface of the protein. It was found that high levels of labelling could also be achieved with a long linker benzophenone probe (**2.59**). Tryptic digestion and LC-MSMS analysis was used to determine that this high level of labelling was due to the benzophenone regioselectively reacting with a preferred methionine residue distal from the active site.

For all three alkyne-compatible photoreactive groups, acceptable levels of biochemical potency, membrane permeability and cellular activity were maintained upon addition of the Ugi scaffold to the parent affinity function, supporting the application of Ugi-derived PAL probes to studies in live cells. The alkyl diazirine probes showed the smallest deviation in physicochemical properties relative to the parent affinity function, suggesting that these probes may best replicate the interactions of the parent compound in cells. However, due to the low photocrosslinking yields observed for the alkyl diazirine probes, these probes may fail to successfully covalently capture these interactions.

The studies presented in this Chapter indicate that efficient labelling is highly dependent on a choice of photoreactive group, linker length, irradiation wavelength, and on the protein topology surrounding the active site. Therefore, the design of PAL probes with guaranteed high levels of labelling is non-trivial. Furthermore, if the probe is to be used in lysates or whole cells, appropriate physicochemical properties must

also be considered. Hence, this Ugi protocol is an ideal approach to access a wide variety of PAL probes in parallel, to rapidly sample for high levels of labelling and acceptable physicochemical profiles. Further applications of these PAL probes with recombinant protein are discussed in Chapter 3, and the performance of Ugi-derived PAL probes in live-cell MS-based proteomics is discussed in Chapter 4.

Chapter 3 Photoaffinity biochemical assays

3.1 Biochemical screening in target-based drug discovery

Advances in genomics and recombinant protein technology over the last three decades have allowed target-based drug discovery to become the main approach adopted by pharmaceutical companies to discover new small molecule therapeutics for patients.[205,206] Genes, gene products or the components of signalling pathways that are involved in a particular disease are identified through genetic analysis or biological observations.[2] Recombinant protein technologies allow for the key target proteins involved in the disease to be expressed and purified for use in biochemical or biophysical screening methods to identify small molecule ligands.[207-209] Advances in automation and the miniaturisation of assay formats have also aided the development of high throughput assay platforms, which are used for both hit identification (high throughput screening) and for optimisation of the affinity of hit and lead compounds for the target protein of interest during optimisation cycles.[210-212] Fluorescence-based assays have become the most widely used assay format for lead optimisation cycles due to their applicability to miniaturisation, speed of analysis, sensitivity, and dynamic range.[213] These include fluorescence resonance energy transfer (FRET), AlphaScreen™, and fluorescence polarisation (FP) assay methods which are briefly described in Sections 3.1.1–3.1.3 respectively.[214-219]

3.1.1 Fluorescent resonance energy transfer assays

Fluorescent resonance energy transfer (FRET) assays operate on the principle that energy can be non-radiatively transferred between a donor fluorophore and acceptor fluorophore that are in close proximity to each other (Figure 3.1).[216] Chelated lanthanides such as europium(III) or terbium(II) are most commonly used as the donor fluorophore, which is conjugated to an antibody that recognises a purification or fusion tag present on the target protein. A tool molecule containing a long-lived fluorescent dye is also used. Binding of the tool molecule to the active site of the target protein brings the donor and acceptor fluorophores into close proximity. Upon excitation (337 nm), the excited Eu^{3+} complex transfers energy to the acceptor fluorophore by FRET.

The acceptor fluorophore then emits light at a longer wavelength, which can be detected. The efficiency of FRET is inversely proportional to the sixth power of the distance between the fluorophores.[220] Therefore, emission from the fluorescent tool compound (665 nm) can be reduced by displacement of the tool compound from the binding site by a competitive ligand in a dose-response relationship. In time resolved-FRET (TR-FRET), reading of the emission wavelength is usually delayed (50–150 μs) after initial excitation, so that any short-lived background fluorescence has subsided.

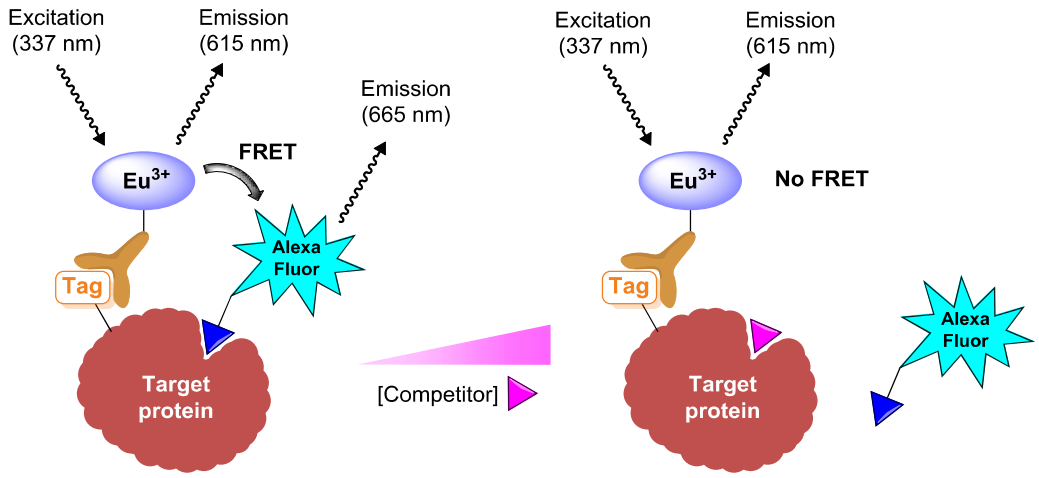


Figure 3.1 Schematic illustrating the principle behind fluorescence resonance energy transfer (FRET) assays.[216] When a donor-acceptor fluorophore pair are brought into close proximity, the excited donor can transfer energy *via* a radiationless process to the acceptor fluorophore. The intensity of the fluorescent emission from the acceptor fluorophore is directly proportional to the concentration of donor–acceptor pairs in proximity, which can be reduced in a dose-dependent manner by a competitive ligand.

3.1.2 Bead-based AlphaScreen™ assay platform

AlphaScreen™ is a bead-based technology where streptavidin-coated donor beads are irradiated with 680 nm light. The donor beads contain a photosensitiser which produces singlet oxygen upon 680 nm irradiation. In aqueous biochemical assay buffers, singlet oxygen has a very short lifetime (*ca*. 4 μs) and can only diffuse *ca*. 200 nm from the donor bead. The acceptor beads contain a chemiluminescent compound which produces 370 nm light upon reaction with singlet oxygen. The energy of this emission is transferred by FRET to fluorophores also contained within the acceptor

which emit at a longer wavelength (520–620 nm). In competition-based assays, a biotinylated tool compound is immobilised on the streptavidin-coated donor beads. The acceptor beads are coated with an antibody for a purification tag present on the target protein. Ligand binding to the target protein brings donor and acceptor beads into close proximity, where the singlet oxygen can transfer and produce a fluorescent signal. In the presence of competitor compound, the tool compound is displaced from the active site, decreasing the observed fluorescent signal in a concentration dependent manner. One disadvantage of AlphaScreen™ assays is that the reagents are sensitive to ambient light and are relatively expensive compared to those required for other assay formats. Also, the reactive singlet oxygen produced from the donor bead can be scavenged by competitor compounds, producing false positives.[214]

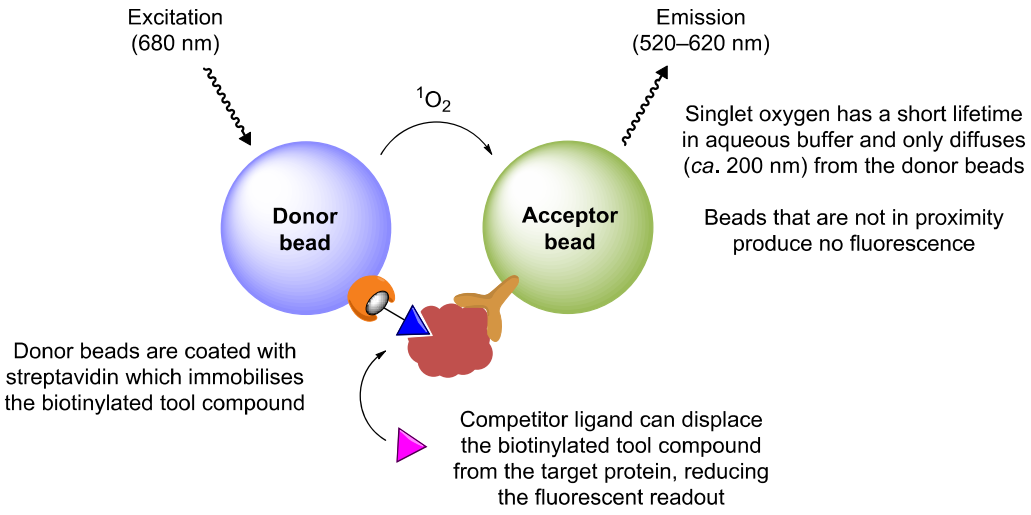


Figure 3.2 The AlphaScreen™ bead-based assay relies on a proximity-induced fluorescence readout. Donor beads are coated with streptavidin which captures a biotinylated tool compound that has affinity for the target protein.[214] The target protein is captured by acceptor beads that are coated with antibodies that recognise purification or fusion tags. Excitation of the donor beads produces a high local concentration of singlet oxygen, which reacts with the acceptor beads to produce a chemiluminescent signal that excites fluorophores also present in the acceptor bead by FRET. Only acceptor beads in the proximity of donor beads produce a fluorescent signal, which can be reduced in a dose-dependent manner by a competitive ligand.

3.1.3 Fluorescence polarisation assays

Fluorescence polarisation (FP) biochemical assays operate on the principle that if a fluorophore is excited with one plane of polarised light, the polarisation of the

fluorescent emission will be distorted due the molecular rotation of the fluorophore in solution between the excitation and emission events (Figure 3.3).[218] The sample is irradiated at the excitation wavelength of the fluorophore with a single plane of polarised light. The change in polarisation of the emitted light is measured by comparing the intensity of fluorescence in a parallel and perpendicular plane relative to the initial excitation polarisation plane.[219] If the fluorescent ligand is bound to the target protein, the complex rotates relatively slowly by Brownian motion within the excitation–emission timeframe, and a large fraction of the light remains polarised in the parallel plane ($I_{parallel} > I_{perpendicular}$). A competitive ligand can displace the fluorescent tool compound from the protein, allowing it to freely rotate in solution at a much faster rate, resulting in a high degree of depolarisation ($I_{parallel} \approx I_{perpendicular}$). FP assays are more operationally simple compared to proximity-based TR-FRET or AlphaScreen™ assays. However, FP assays usually have a lower dynamic range and are less sensitive due to increased background fluorescence.[221] High concentrations of competitive ligand can also aggregate with the fluorescent tool compound which can lead to false negatives.[218]

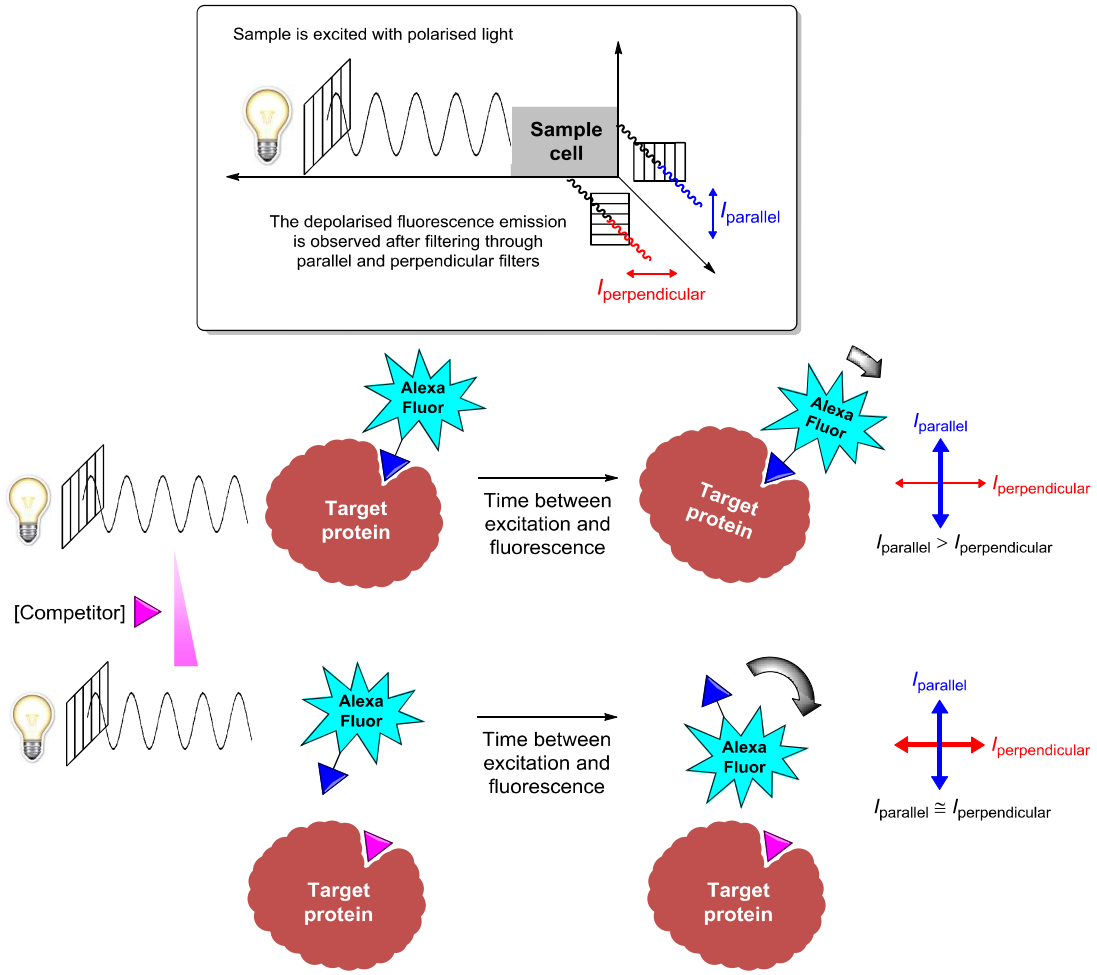


Figure 3.3 Schematic illustrating the principle of fluorescence polarisation (FP).[218] A fluorescent tool compound is excited by a single plane of polarised light. Within the timeframe between excitation and emission, the compound can rotate freely in solution. Due to this relatively fast rotation, the fluorescent emission is largely depolarised. If the tool molecule is bound to the target protein, the complex rotates at a slower rate in solution. Therefore, the extent of depolarisation of the fluorescent emission will be less under these conditions. A competitive ligand can compete with the tool molecule for the binding site, thus the degree of depolarisation can be used to report on the affinity of the competitive ligand.

One caveat to each of these three fluorescence-based biochemical assay platforms is that only one recombinant protein can be screened per experiment. For each new target protein of interest, a new assay needs to be developed and optimised. A biochemical method that is capable of screening multiple proteins at once would allow for the simultaneous optimisation of potencies against various on- and off-targets in a single assay. This would be highly beneficial to the lead optimisation stage of small molecule drug discovery, as more informative assays may help to reduce candidate cycle times.[9] Photoaffinity probes are used to report on non-covalent interactions formed by the affinity function with target proteins and can be displaced from the binding site by an

excess of competitive ligand prior to irradiation (see Section 2.3.6). Thus, the PAL approach may be amenable to biochemical screening (Figure 3.4), where PAL probes could be used to inform on the affinity of competitive ligands under certain assumptions:

- Excitation and formation of the reactive species does not affect the binding of the affinity function or the equilibrium between the compounds and the active site.
- The excitation and formation of the reactive species occurs at the same rate for when the probe is bound to the active site and when it is free in solution.

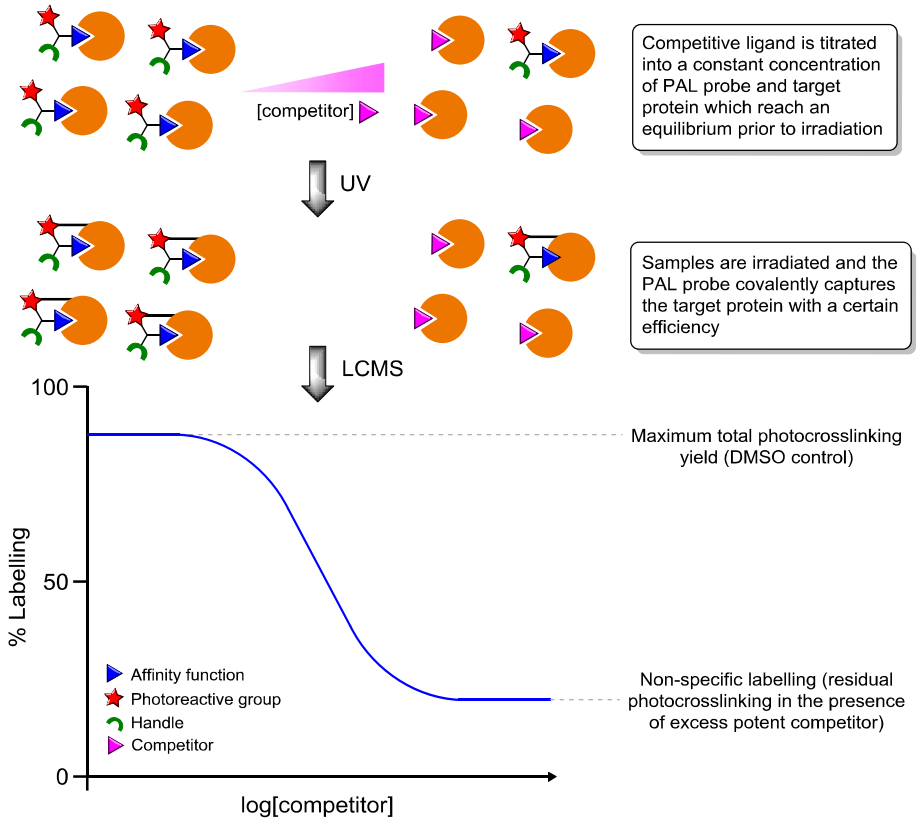


Figure 3.4 Application of photoaffinity labelling to enable a biochemical displacement assay. The PAL probe can be displaced by a competitor prior to UV irradiation. This equilibrium between the target protein, PAL probe and competitor is captured during the irradiation event. The reduction in covalent modification due to the displacement of the PAL probe by a concentration gradient of competitor can be measured by intact protein LCMS and used to inform on the relative potency of different ligands.

Under these two conditions, the percentage of specific photocrosslinking is directly proportional to the fractional occupancy of PAL probe present in the active site at equilibrium. The percentage photocrosslinking can be determined by intact protein

LCMS (see Sections 1.2.4.1 and 2.2.3). This readout may allow for multiple proteins to be screened within the same assay, provided that the proteins are of different masses. This offers a unique advantage over current fluorescence-based biochemical screening methods, which can only be used to study the binding of ligands to a single protein. This work involved the development of a dual-domain PAL displacement assay, using BRD4 BD1 and BD2 as proof-of-concept target proteins. To develop a working PAL displacement assay, several optimisation studies were performed. A series of known BET inhibitors were then used to validate the assay, and the performance of the assay was benchmarked by comparison to TR-FRET BD1 and BD2 assays.

3.2 PAL displacement assay optimisation

To establish a working displacement assay, several key parameters had to be determined. These are presented below in Table 3.1.

Table 3.1 Parameters to be determined to establish a working PAL displacement assay

Parameter	Requirements
Optimal PAL probe	High % total labelling to BD1 and BD2. Low % non-specific labelling to BD1 and BD2.
Non-specific binding	Reduce non-specific labelling or establish a workflow to account for it.
Irradiation time	Optimise to enable activation of PAL probe to label BD1 and BD2, while limiting photodegradation of the protein or competitor compounds.
Active protein concentration	Establish the active protein concentration of BD1 and BD2 so that protein concentration can be accurately set, and the lower limit of competitor concentration determined.
Multiple domain LCMS	Establish whether crosslinking can be achieved to both domains in the same sample well and if this % labelling can be determined accurately within the same LCMS injection.

3.2.1 Choosing an optimal probe for the PAL displacement assay

The first parameter to explore in the optimisation studies towards a working PAL displacement assay was choosing an optimal PAL probe. To find which probe had the highest levels of photocrosslinking to both BD1 and BD2, the 15 probes **2.45–2.59**

synthesised in Chapter 2 (Section 2.3.3) (5 µM) were incubated with BD1 or BD2 (1 µM) and irradiated for 10 min (Figure 3.5).

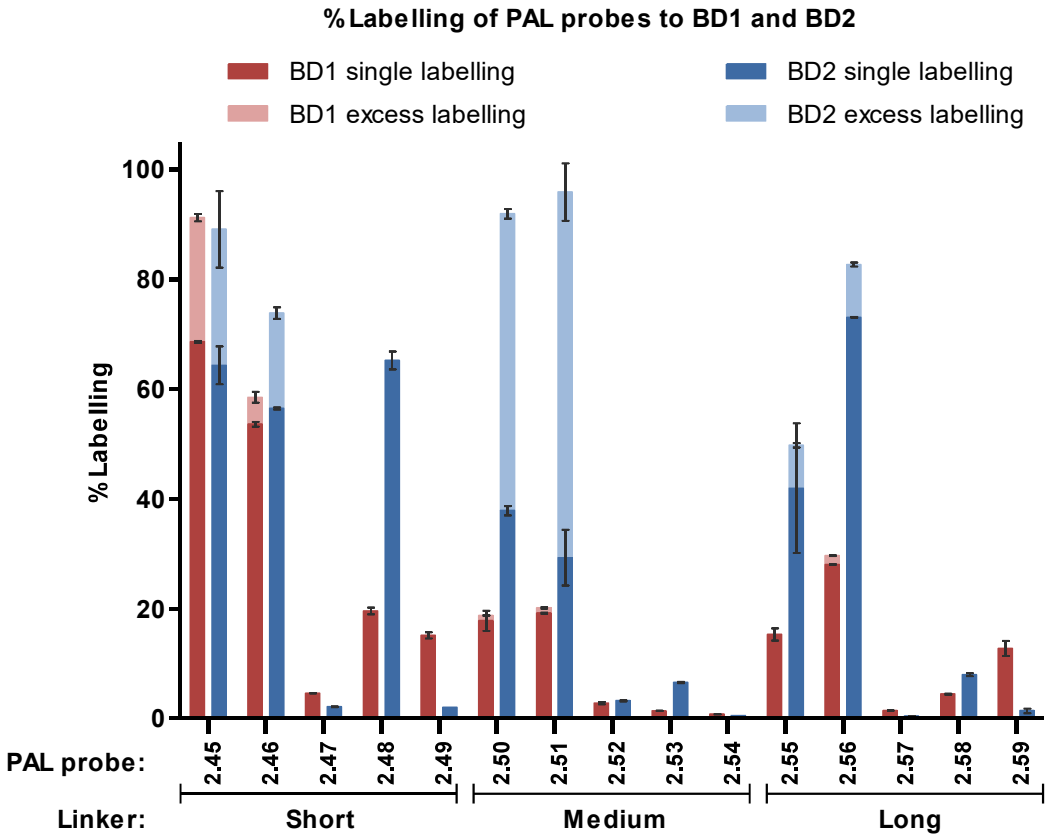


Figure 3.5 Photocrosslinking yields of 15 probes (5 µM) to BRD4 BD1 and BD2 (1 µM each) after irradiation (302 nm, 10 min) determined by intact protein LCMS. Percentage single labelling refers to the deconvoluted peak area for a single covalent modification of PAL probe to the protein. Excess labelling refers to the summed deconvoluted peak areas for any additional covalent modifications by the PAL probe. Values are averages of two replicates.

Of the 15 PAL probes, the aryl azide probes (**2.45**, **2.46**, **2.50**, **2.51**, **2.55** and **2.56**) showed the highest levels of single labelling *versus* the other photoreactive groups, however this was accompanied by moderate-to-high levels of excess labelling. This was more pronounced with BD2, with medium linker aryl azide probes **2.50** and **2.51** giving high levels of excess labelling (*ca*. 50% and 65% respectively). Tetrafluoroaryl azide probe **2.45** gave the highest average levels of photolabelling to BD1 and BD2. This probe also showed the fastest photoactivation rate when labelling BD1 in the earlier timecourse studies (Chapter 2, Figure 2.12). Thus, probe **2.45** was selected for

further studies. Attention then turned to reducing the moderate levels of excess labelling given by **2.45** (*ca*. 20%) to both domains.

3.2.2 Experiments to reduce non-specific labelling of PAL probe 2.45

Surfactants are often used in biochemical assays to reduce non-specific binding.[222] However, most surfactants are not compatible with LCMS, due to potential column damage and ionisation suppression of the intact protein.[223,224] Low MS-compatible concentrations of 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) were tested (Figure 3.6). In an attempt to reduce any non-specific binding, the experiment was performed with a low probe (2 µM) and protein (0.2 µM) concentration, which was just within the limit of detection. The experiment was also carried out with a relatively high concentration of DMSO (3%). Unfortunately, the addition of CHAPS (10 µM or 50 µM) showed no alleviation of non-specific binding for probe **2.45** to BD1 compared to a PBS-only control.

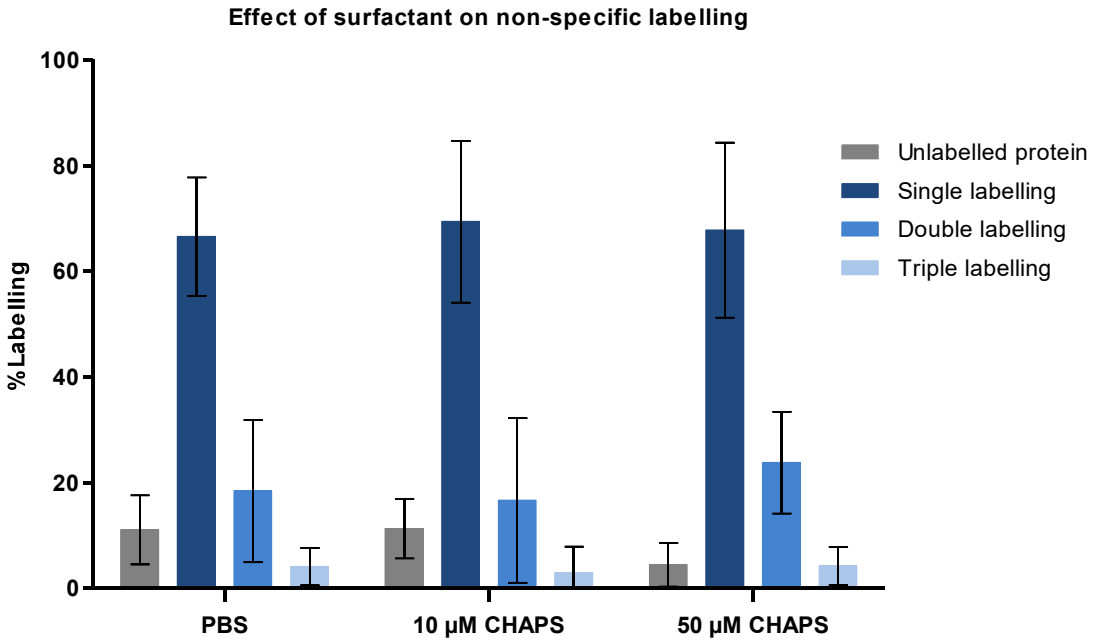


Figure 3.6 The addition of CHAPS surfactant (10 µM or 50 µM) had no effect on reducing non-specific labelling of probe **2.45** (2 µM) to BRD4 BD1 (0.2 µM). Samples were irradiated at 302 nm for 2 min. Experiment performed in triplicate.

Although the level of non-specific binding for probe **2.45** was not reduced by the addition of CHAPS, it was desired to progress exploration of the assay with *ca.* 20%

excess labelling. One advantage of having a MS-based readout is that the percentage of single and double (or higher) labelled species can be accurately determined. These peak ratios can be used in combination with experimentally determined non-specific labelling to calculate the specific labelling of the PAL probe.

3.2.3 Timecourse of PAL probe 2.45 with BRD4 BD1 and BD2

The photocrosslinking timecourse studies presented in Chapter 2 (Figure 2.12) showed that probe **2.45** gave fast rates of photolabelling to BD1 under 302 nm irradiation (full photoactivation in <1 min). The rate of labelling to both BRD4 BD1 and BD2 domains by PAL probe **2.45** was assessed (Figure 3.7). **2.45** (5 μM) was incubated with each domain (1 μM) and irradiated over a timecourse. The percentage labelling was determined by intact protein LCMS. Probe **2.45** labelled BD1 and BD2 with approximately equal rate, and full photoactivation was reached within 1 min for both proteins. As outlined in the requirements for a PAL displacement assay, the shortest possible irradiation time was required to minimise photodamage to the protein (as demonstrated in Figure 2.23) and limit any potential photoactivation of the competitor compounds. To compromise between achieving full photoactivation of **2.45** and having a minimum UV exposure time, 0.6 min was chosen as a suitable irradiation time for future experiments. The deconvoluted spectra for the timecourse involving BD1 are shown in Figure 3.8 to illustrate the observation of single and double labelled species. Interestingly, the highest peak in the double labelled population was an oxidised species (+32 Da), whereas the highest peak in the single labelled population corresponded to the addition of probe only (+698 Da).[225] One explanation of oxidation is that during the irradiation event, the nitrene formed on the probe can react with water and/or buffer components, leading to the formation of reactive oxygen species (ROS).[95] The doubly labelled species would have a high local concentration of PAL probe, which may result in an accompanying high local concentration of ROS.

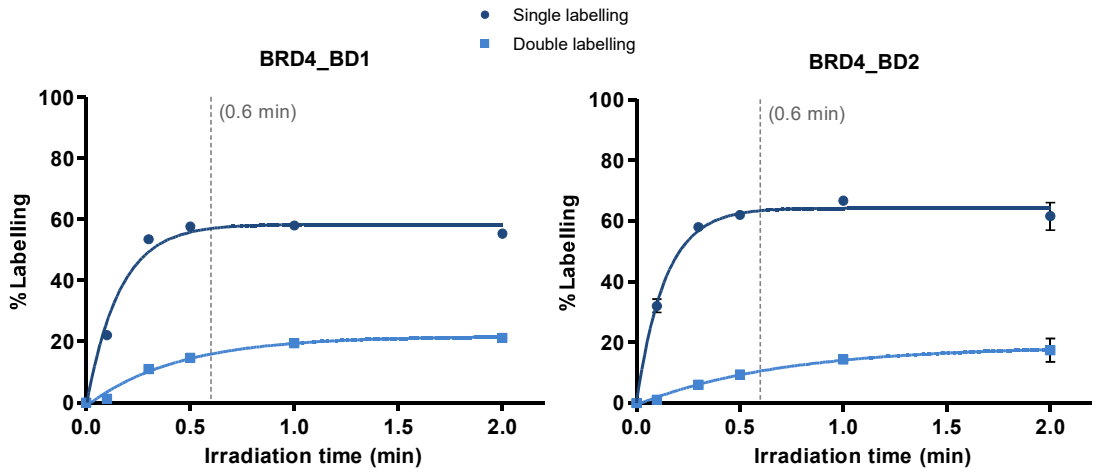


Figure 3.7 Irradiation timecourse of BRD4 BD1 (1 µM) with **2.45** (5 µM) (left). Irradiation timecourse of BRD4 BD2 (1 µM) with **2.45** (5 µM) (right). 0.6 min was chosen as the optimal irradiation time for future experiments involving PAL probe **2.45**.

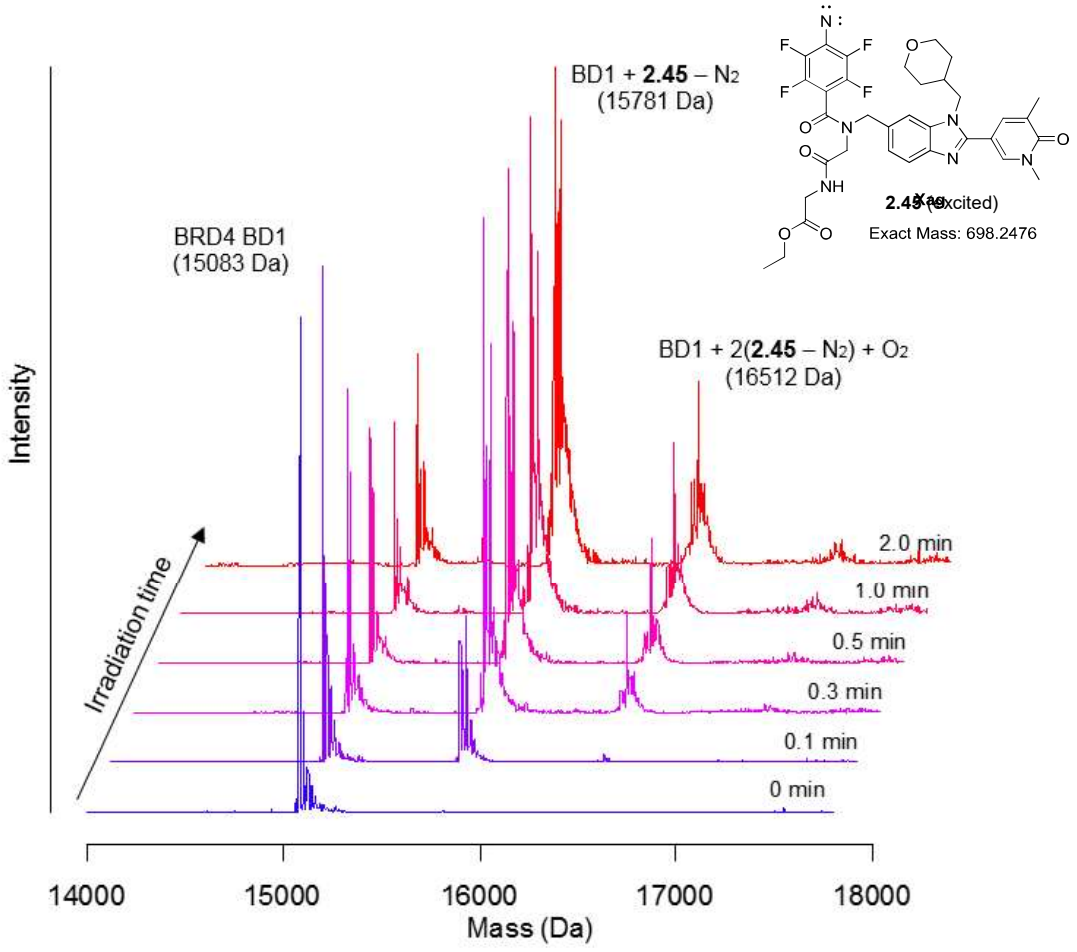


Figure 3.8 Photocrosslinking timecourse of probe **2.45** (5 µM) with BRD4 BD1 (5 µM) at increasing irradiation times (302 nm). The mass of the single labelled species corresponds to (BD1 + probe – N_2) Da, and the mass of the double labelled species corresponds to (BD1 + 2(probe – N_2) + 32) Da. 0.6 min was chosen as the optimal irradiation time.

3.2.4 Experiments to determine the active concentration of BRD4 BD1 and BD2

To determine the lower limit for competitor concentration allowed in the PAL displacement assay, it was necessary to determine the active concentration of protein. The active protein concentration is the concentration of correctly folded protein that has an active site available for specific binding by the PAL probe.[226] For this experiment, BD1 (*ca*. 1 µM total protein concentration estimated using a NanoDrop absorbance assay) and BD2 (*ca*. 2 µM) were irradiated with **2.45** over a range of concentrations below and above the expected protein concentration (Figure 3.9).

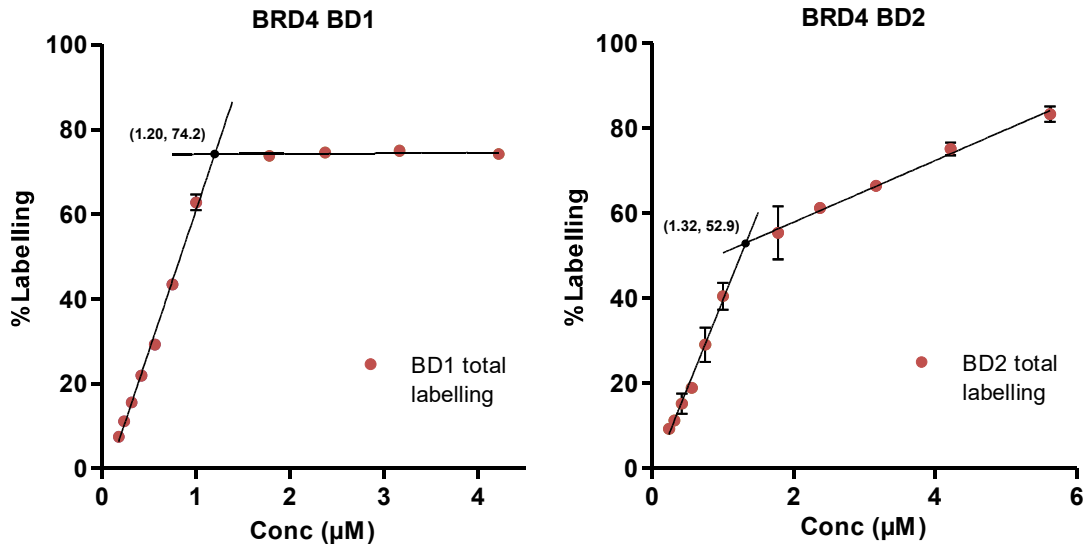


Figure 3.9 Experiment to determine the active concentrations of BRD4 BD1 and BD2. BD1 total concentration was estimated as 1 µM and the active protein concentration was found to be 1.20 µM. BD2 total concentration was estimated as 2 µM and the active protein concentration was found to be 1.32 µM. These values were used to accurately adjust protein amounts to 1.0 µM active concentration in the displacement assay.

Based on the potency of **2.45** in the BRD4 BD1 and BD2 TR-FRET assays presented in Section 2.3.4 (pIC_{50} = 7.8 and 7.3 respectively), the dissociation constant (K_d) of **2.45** was estimated to be in the nanomolar range. Therefore at 1–2 µM of protein, any **2.45** present was assumed to fully bind to the active site. As the concentration of **2.45** increases from zero, the probe binds fully to the active site to produce a linear relationship between percentage labelling and concentration. A point is reached where the concentration of **2.45** is equal to the concentration of active protein, and there are

no more sites available for specific binding. For concentrations of **2.45** past this point, the slope should be flat since no more photocrosslinking can occur *via* specific binding. Any additional slope after this point is due to non-specific binding. The point at which these two slopes intersect gives the active concentration of protein. For BD1, the active concentration was higher than expected and found to be 1.2 µM (for a sample that had an estimated total concentration of 1 µM using a NanoDrop absorbance assay). For BD2, the active concentration was found to be 1.32 µM (from a sample that had an estimated total concentration of 2 µM). This finding allowed for both protein concentrations to be accurately set to 1.0 µM each for the displacement assay.

3.2.5 Combining BRD4 BD1 and BD2 in the same photocrosslinking experiment

An advantage of using a PAL displacement assay with an intact protein LCMS readout is that the binding affinity of a competitor compound to multiple domains can be determined within the same experiment. To investigate whether PAL labelling to BRD4 BD1 and BD2 could be conducted and analysed from the same sample well, both proteins were combined (1 µM final active concentration each) and incubated with PAL probe **2.45** (5 µM). The sample was then analysed using intact protein LCMS before and after irradiation (302 nm, 0.6 min) (Figure 3.10). The two proteins were resolved by the LC method (7.7 min/sample), with BRD4 BD2 eluting first (*ca*. 2.55 min), followed by BD1 (*ca*. 2.85 min). This allowed both peaks to be deconvoluted separately. Interestingly, although both proteins show a similar total ion count (TIC) intensity (1.2–1.4 × 10^6), prior to irradiation, the deconvoluted peak intensity for BD2 was approximately 4.5-fold weaker than for BD1. This difference was even more pronounced after irradiation, with crosslinked BD2 showing a 7.5-fold weaker deconvoluted intensity than for BD1. This suggests that the recombinant BD2 domain may be less ionisable than the BD1 domain, and/or more unstable under UV irradiation (302 nm). Due to the lower intensities observed for BD2, it was anticipated that percentage labelling to BD2 in the displacement assay would appear with higher noise and standard deviation than for the percentage labelling to BD1. However, this experiment demonstrated that **2.45** could label both BD1 and BD2 within the same sample well, and that both proteins could be detected by MS with sufficient intensity

to determine percentage total labelling. After satisfying the optimisation requirements outlined in Section 3.2, the practical workflow of the assay was addressed.

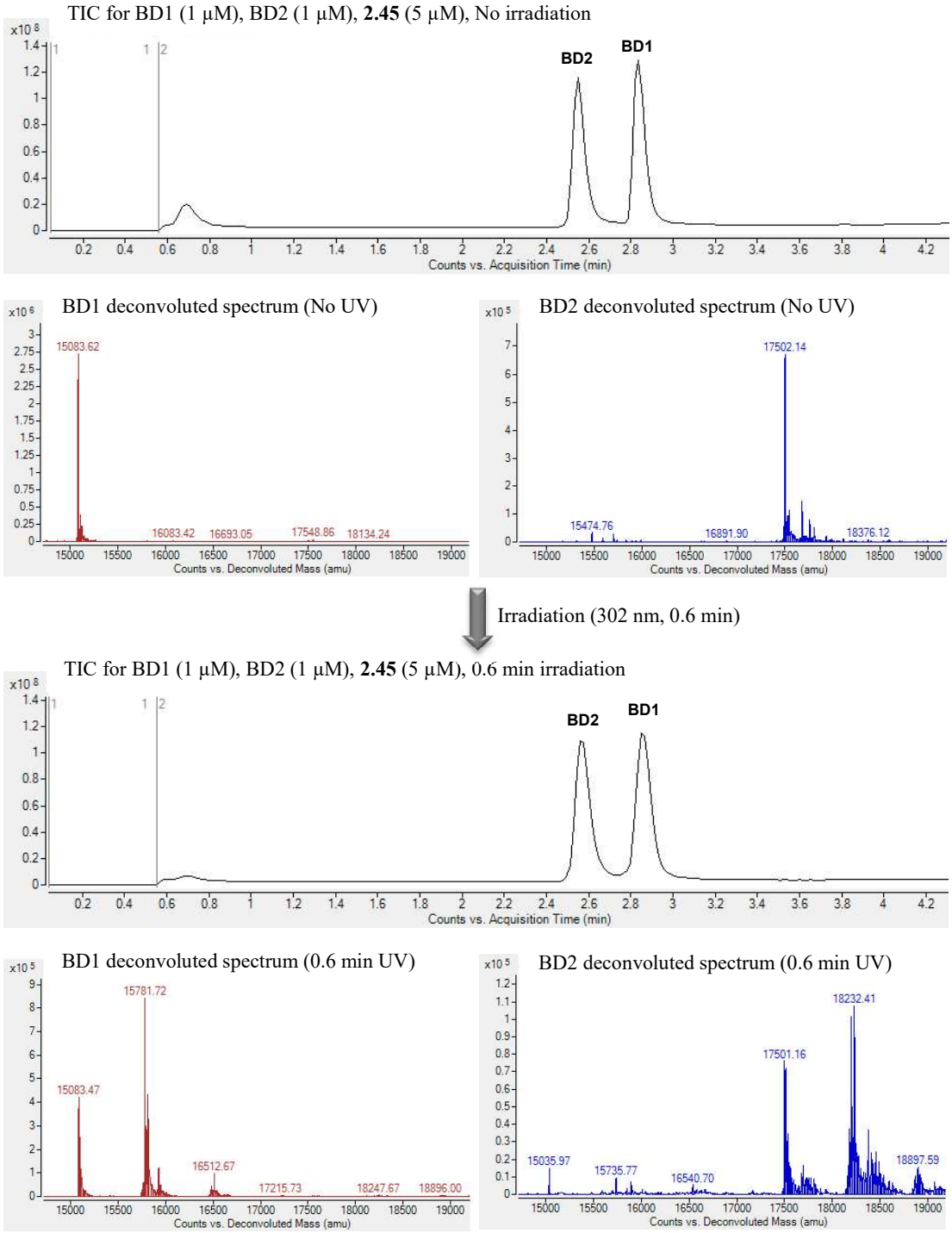


Figure 3.10 Investigating labelling of BRD4 BD1 and BD2 domains by PAL probe **2.45** within the same sample using intact protein LCMS. Both proteins were resolved by chromatography, and the same labelling profiles were observed for each domain as was observed in single-domain experiments.

3.2.6 Establishing a PAL displacement assay platform using intact protein LCMS analysis

The assay workflow was optimised so that only a single assay plate was required from compound dispensing through to analysis by intact protein LCMS (Figure 3.11). The compounds to be screened were acoustically dispensed to Greiner 384 low-volume microplates (150 nL in DMSO per well) using a Labcyte Echo555 acoustic dispenser. A solution of protein (BRD4 BD1 and BD2, 1 µM each) and probe **2.45** (5 µM, 0.1% DMSO) was added to each well (15 µL) by multi-channel pipette. The plate was then allowed to stand on ice for 1 h to reach equilibrium before irradiation (302 nm, 0.6 min). After irradiation, the plate was sealed with adhesive film and sampled directly by the intact protein LCMS instrument (7.7 min per sample, 10 µL injection). For data analysis, each protein peak in the TIC was extracted and deconvoluted, and this data was exported as a comma separated values (csv) file containing mass (Da) *vs* intensity. RStudio was used to calculate the percentage total labelling to BD1 and BD2, and these values were exported as a csv file for interpretation in Graphpad Prism. This workflow was used for a proof-of-concept study to investigate whether PAL probe **2.45** could inform on the BD1 and BD2 affinities of a large test set of competitor compounds.

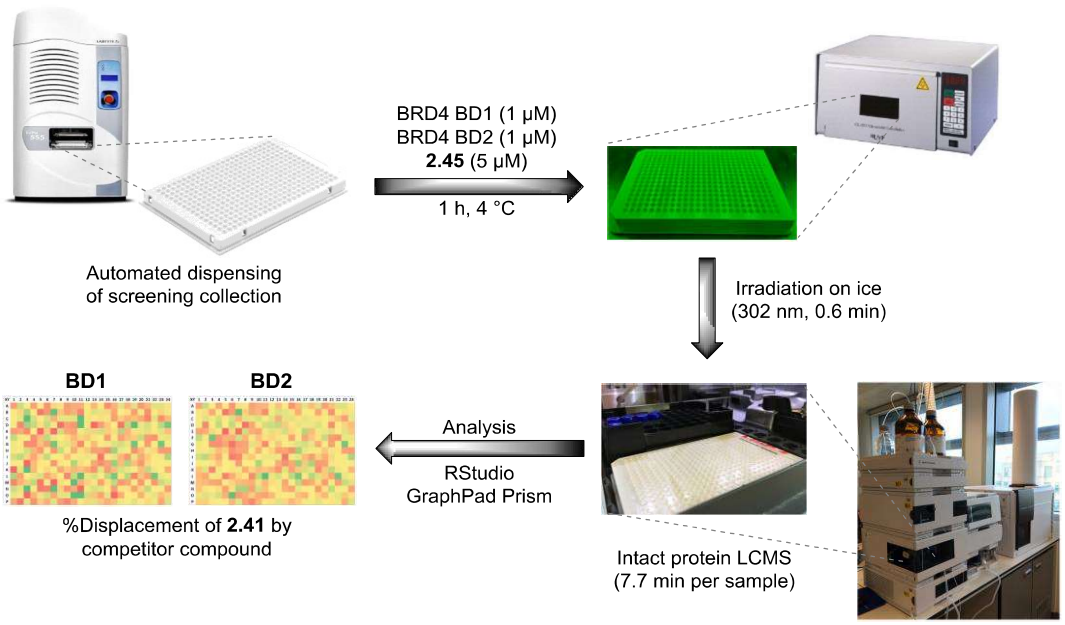


Figure 3.11 Optimised PAL displacement assay workflow. Competitor compounds were acoustically dispensed into 384-well low-volume plates (150 nL). A solution of BRD4 BD1 and BD2 protein (1 µM) and **2.45** (5 µM) was added using a multi-channel pipette (15 µL). After 1 h equilibration on ice, the plates were irradiated (302 nm, 0.6 min) and sealed before being analysed directly by intact protein LCMS (10 µL injection, 7.7 min per sample). The BD1 and BD2 protein peaks were deconvoluted and this data was exported as a csv file. RStudio is used to calculate percentage total labelling to BD1 and BD2. These values were then interpreted with GraphPad Prism.

3.3 Proof-of-concept studies for a dual-domain PAL displacement assay

3.3.1 Choosing a test set of compounds for the dual-domain PAL displacement assay

To obtain a test set of compounds for a proof-of-concept study, the chemical database maintained by the European Molecular Biology Laboratory (CHEMBL) was searched for compounds that had any BET activity annotated in their chemical compound profile. The targets “BRD4”, “BRD3”, “BRD2” and “BRDT” were searched and the resulting lists of compounds (in SMILES format) were exported as csv files (Figure 3.12). The lists were combined, and any duplicates were removed which gave 2109 compounds. This SMILES list of compounds was then searched for compounds that had an assigned GSK registration number (387 compounds). Of these, 264 compounds were available to use as 10 mM stock solutions in DMSO. This set of compounds was

labelled the "CHEMBL set". The SMILES notation and GSK registration numbers for the 264 compounds of the CHEMBL set are shown in Table 7.12.

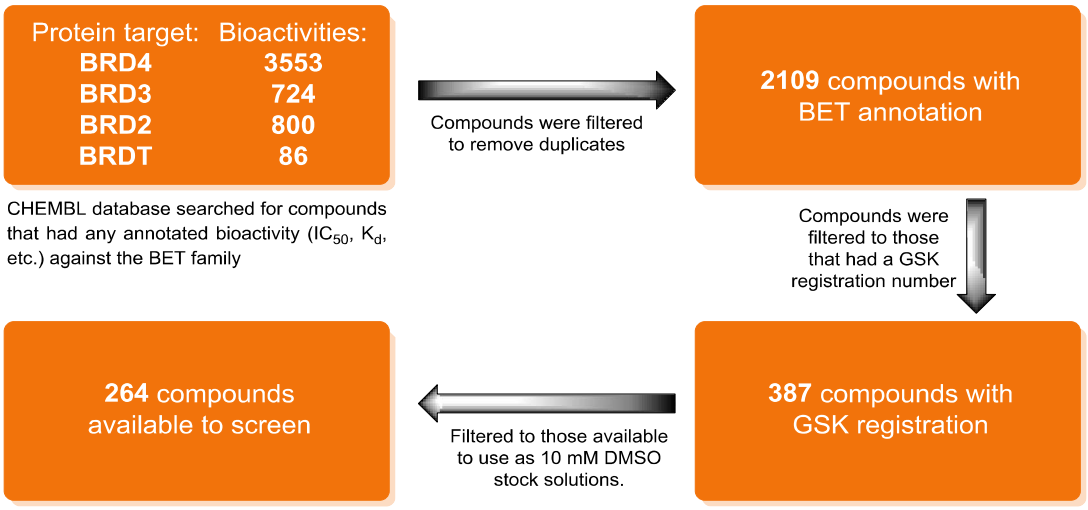


Figure 3.12 Selection of compounds for the displacement assay test set. Compounds with BET annotation were filtered to those that were present in the GSK compound collection and available as 10 mM stock solutions in DMSO. SMILES codes for the 264 compounds screened are given in Table 7.12.

3.3.2 Single-shot screening of the CHEMBL set by PAL displacement assay

The CHEMBL set of compounds (264) were screened at a single concentration (100 μM) in duplicate following the optimised assay workflow described in Section 3.2.6. The aim of this preliminary screen was to rank the compounds by the amount they displaced the PAL probe **2.45**, so that the most potent compounds could be selected and screened in a full-curve dose-response experiment. The percentage labelling of **2.45** to both domains in the presence of the various competitor compounds was calculated using Equation 3.1. The lowest percentage labelling observed (from the most potent compound at each domain) was used as an estimate for non-specific labelling. This value was subtracted from the percentage labelling obtained for the remaining compounds. These values were then normalised to the highest levels of labelling observed (DMSO control) using Equation 3.2. This value for percentage normalised labelling was subtracted from 100 to give a percentage displacement of the PAL probe **2.45** from both domains by each competitor compound (Equation 3.3). The values for percentage displacement were then averaged over the two replicates. These values are given in Table 7.12.

Equation 3.1

$$\%\ Labelling = \frac{peak\ height\ for\ single\ labelled\ protein}{(peak\ height\ for\ unlabelled\ protein + single\ labelled\ protein)} \times 100$$

Equation 3.2

$$\%\ normalised\ labelling = \frac{\%\ labelling - A}{B - A} \times 100$$

$A = lowest\ \%\ labelling\ observed\ (most\ potent\ competitor)$

$B = highest\ \%\ labelling\ observed\ (DMSO\ control)$

Equation 3.3

$$\%\ displacement = 100 - \%\ normalised\ labelling$$

To assess the reproducibility of the single-shot assay, the values for percentage displacement obtained from both replicates were plotted against each other (Figure 3.13). The displacement of PAL probe **2.45** from BD1 was found to be more reproducible than for BD2, showing a tighter correlation between each replicate (R^2 = 0.95 and 0.78 respectively). This may be due to BD2 having a 7.5-fold lower signal than BD1 in the intact protein LCMS analysis (determined in Section 3.2.5).

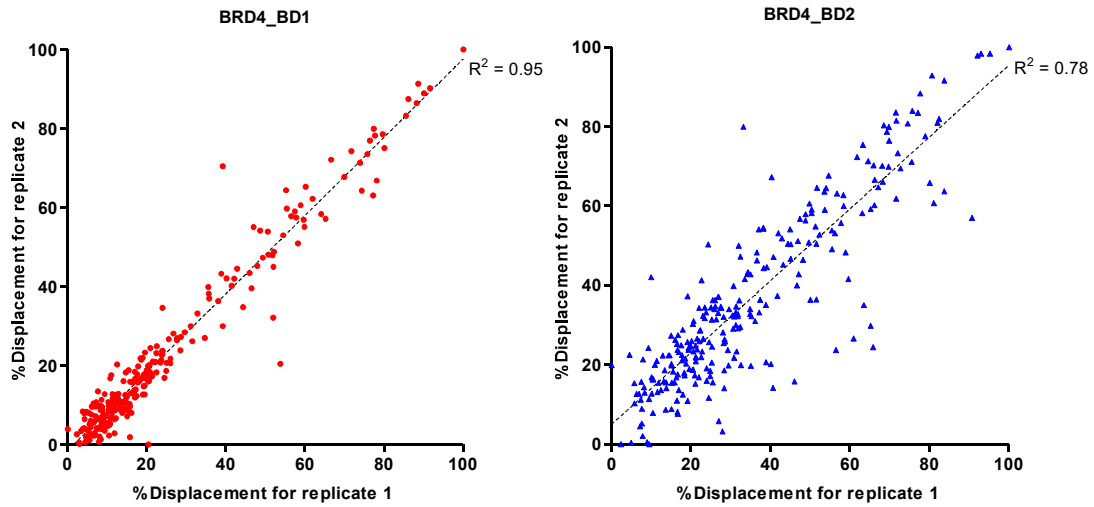


Figure 3.13 Percentage displacement values from each replicate were plotted for both BD1 and BD2 domains. Percentage displacement was more reproducible for BD1 (R^2 = 0.95) than for BD2 (R^2 = 0.78).

The average percentage displacement values for BRD4 BD1 and BD2 were plotted (Figure 3.14). Compounds that appeared to be BD1 or BD2 selective were identified (*e.g.* **F02** and **C01**, respectively), along with potent competitors at both domains (*e.g.* **A02**, **H01** and **K01**). From the 264 compounds screened, 44 compounds showed ≥60% displacement at either or both domains (Figure 3.14, highlighted as dark red data points). These compounds were selected for screening by full-curve dose-response. The high concentration (100 μM) used in this single shot experiment would be the highest concentration used in the dose-response follow-up studies. Therefore 60% displacement was chosen as the selection parameter, as the compounds selected would need to pass the IC_{50} (concentration of competitor that displaces the PAL probe by 50%) to give acceptable dose-response curves.

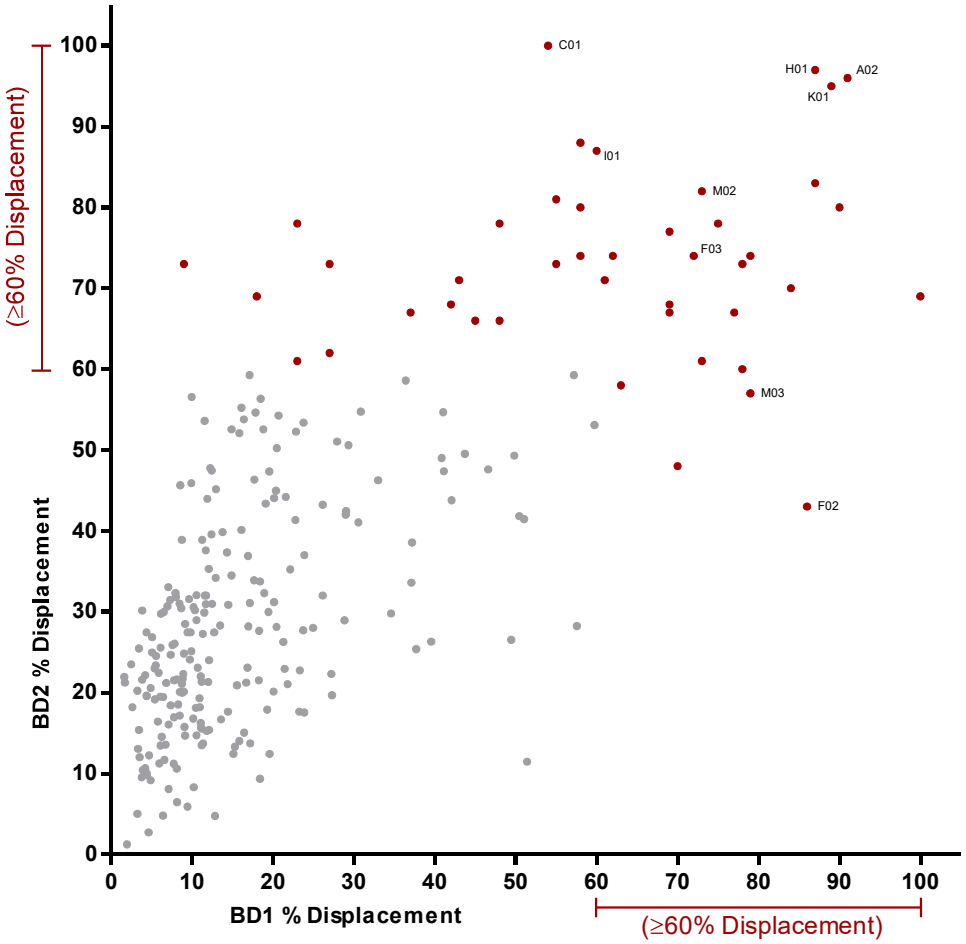


Figure 3.14 Single concentration photoaffinity displacement assay with BRD4 BD1 and BD2. The percentage displacement of PAL probe **2.45** (5 μM) by a single concentration (100 μM) of competitor compound from the CHEMBL set is shown as an average of two replicates. Compounds that gave ≥60% displacement (red) were chosen for full-curve dose-response studies. The dose-response curves for the labelled compounds are shown in Figure 3.15. The dose-response curves for the remaining compounds are shown in Figure 7.1 to Figure 7.4.

3.3.3 Full curve dose-response PAL displacement assay with selected compounds from the CHEMBL set

The 44 competitor compounds selected from the preliminary single-shot displacement screening were dispensed in an 8-point serial dilution (1 in 2) with a final concentration ranging from 100 µM to 0.78 µM. A DMSO control was included as the ninth data point. The compounds were screened in duplicate following the optimised assay workflow developed in Section 3.2.6. However, an increased irradiation time was used (302 nm, 1 min) to ensure the full photoactivation of probe **2.45**. Intact protein LCMS analysis was also further optimised to reduce analysis time to 5.75 min/sample by introducing a higher flow rate (1.2 mL min^{-1}) during the column wash stage (see Section 0). Percentage total labelling was calculated using Equation 3.4.

Equation 3.4

$$\%\ labelling = \frac{B + 2C}{A + B + 2C} \times 100$$

where:

A = peak height for unlabelled protein

B = peak height for single labelled protein

C = peak height for double labelled protein

Example full-curve plots are shown in Figure 3.15 for the nine competitor compounds that were labelled in the single-shot plot (Figure 3.14). The full-curve plots for the remaining 35 compounds are shown in Figure 7.1 to Figure 7.4.

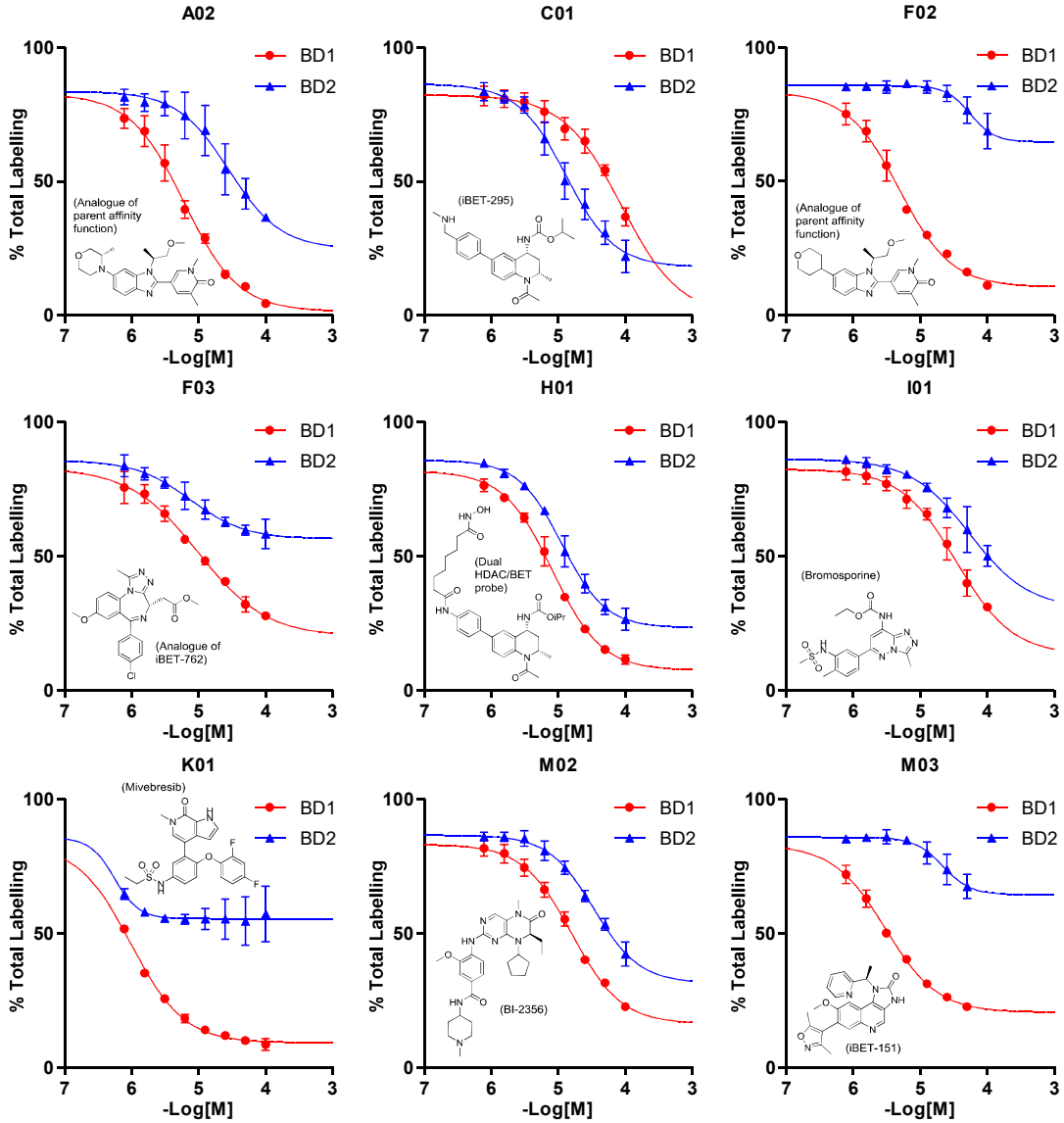


Figure 3.15 Selected dose-response curves for known BET inhibitors (compounds labelled in Figure 3.14) as determined by PAL displacement assay. **F02** and **M03** showed selectivity for BRD4 BD1. **C01** showed selectivity for BRD4 BD2. The data was fitted with a four-parameter non-linear regression with a bottom constraint >0.0 (GraphPad Prism). The dose-response curves for all 44 compounds screened are shown in Figure 7.1 to Figure 7.4.

The dose-response data was fitted with a four-parameter non-linear regression (Graphpad prism) which gave good fitting and allowed PAL pIC_{50} values to be accurately determined for nearly all competitor compounds for both proteins.

Compound **C01** (i-BET295) showed selectivity for BD2, and compounds **F02** and **M03** (iBET-151) showed selectivity for BD1.[227,228] **K01** (ABBV-075/Mivebresib) was highly potent against both BD1 and BD2.[229] The dual-target ligands such as **H01** (dual HDAC/BET probe) and **M02** (BI-2356, dual PLK1/BRD4 inhibitor) gave

well-defined curves and were equipotent at BD1 and BD2.[230,231] The full curves for the remaining competitor compounds are shown in Figure 7.1 to Figure 7.4.

Seven compounds showed additional protein modification with masses corresponding to covalent labelling of the protein by the competitor compound itself, five of which contained a sulfonamide. These are shown in Figure 3.16. Covalent competitor labelling was up to 20-fold more intense with BD2 than with BD1. The BD2 spectra for three compounds that showed the highest levels of competitor labelling are shown in Figure 3.17. The dual JAK2/BRD4 inhibitor (TG-101209) **E01** showed covalent labelling (*ca*. 50%) to both the unmodified protein and the probe-labelled protein species.[231] Fragment-derived BET inhibitor **J01** showed similar high levels of competitor labelling, and increased levels of oxidation (multiple +16 Da species).[232] All the competitor labelling compounds that contained a sulfonamide labelled the protein with the exact mass of the competitor. Compound **O02**, which did not contain a sulfonamide, showed modification of both the unlabelled and probe labelled species (*ca*. 30%), however the mass addition corresponded to (competitor + 15 Da), suggesting additional oxidation. For all seven compounds, competitor labelling was concentration-dependent. It is also unclear whether all sulfonamides are liable to covalently modify the target protein, as no labelling was observed for two other sulfonamide-containing competitors **I01** (Bromosporine) and **K02** (see Figure 3.15 and Figure 7.3 respectively). One area of future work would be to investigate whether these covalent labelling events occur in the absence of UV irradiation.

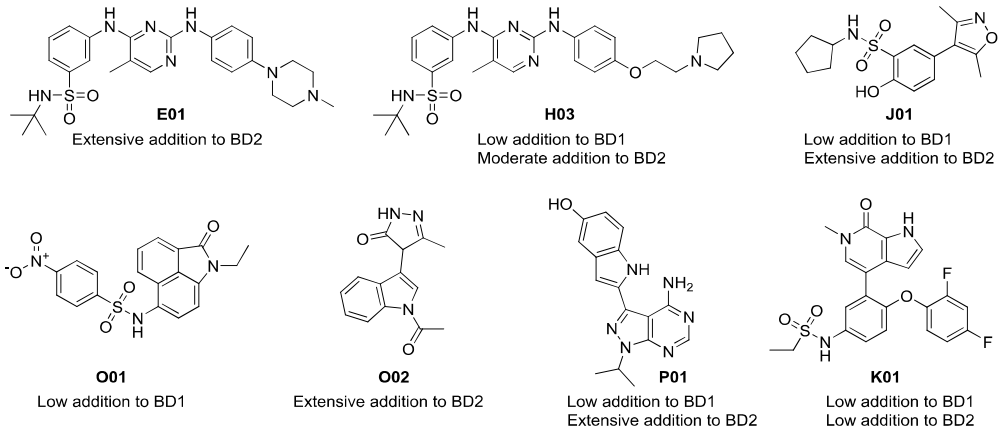


Figure 3.16 Seven compounds from the dose-response follow-up set that showed covalent modification of the target protein(s). Five out of the seven compounds contained sulfonamides, and the covalent adducts observed in these samples corresponded to the exact masses of each competitor.

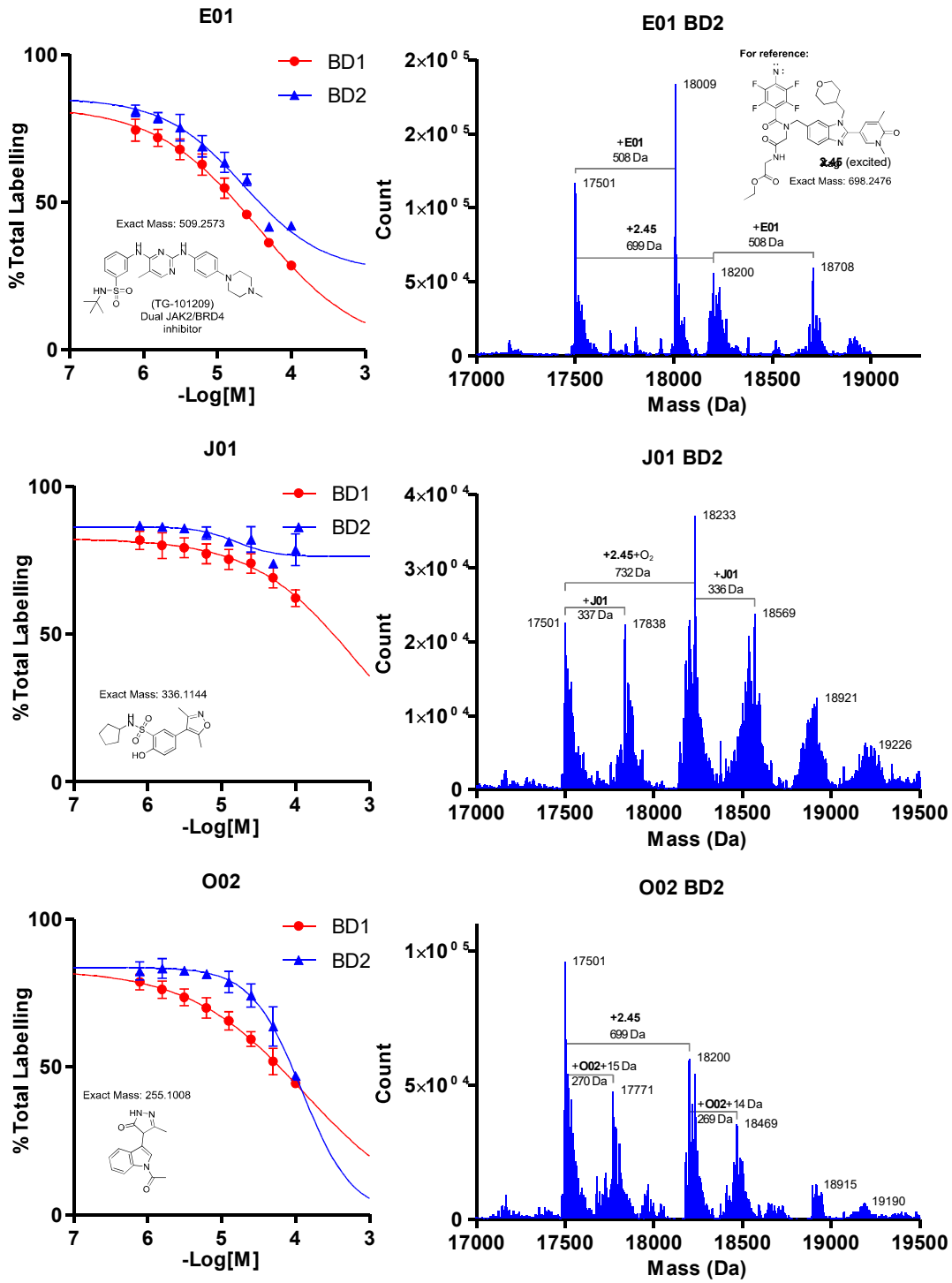


Figure 3.17 Three compounds that showed the highest levels of undesired covalent competitor labelling to the target protein. Labelling was *ca*. 20-fold more intense to BD2 (shown right) than to BD1. For the competitors containing a sulfonamide, adducts with the exact mass of the competitor were observed. Competitor **O02** showed mass additions of (competitor + 15 Da).

To compare the pIC_{50} values derived from the PAL displacement assay to those obtained in an orthogonal biochemical assay, the set of 44 compounds were screened

in BRD4 BD1 and BD2 TR-FRET assays (performed by Alex Phillipou). The pIC_{50} values obtained from each assay (PAL *vs* TR-FRET) were plotted for each protein (Figure 3.18). A good correlation was observed between both assays for both domains, except for two outliers; **A01** and **F01**. Both compounds were highly potent in the PAL displacement assay but were both below the lower limit (pIC50 <4.3) of the TR-FRET assays for both BD1 and BD2. These compounds share a common 1,3-dimethyl benzimidazolone core and are analogues of the BRPF1 chemical probe GSK6853.[233,234] No covalent competitor labelling to either BD1 or BD2 was observed in the deconvoluted spectra for both compounds. **A01** and **F01** showed standard binding curves (Hill slope = *ca*. 1) to BD2, with a Hill slope of 1.09 and 0.94 respectively. However, Hill slopes of less than 1 were obtained with BD1 (0.88 and 0.66 respectively), which indicate that the percentage labelling by PAL probe **2.45** could have been reduced by a mechanism other than independent competitive binding. One potential reason for a non-competitive reduction in photocrosslinking by PAL probe **2.45** with increasing concentrations of **A01** or **F01** may be due to these compounds strongly absorbing UV light at *ca*. 302 nm. Future experiments could be conducted to measure the UV absorbance of these compounds, and to monitor for changes in the rate of photoactivation of **2.45** (by observing the mass of the intact probe rather than photocrosslinking to recombinant protein) in the presence of a range of concentrations of **A01** or **F01**.

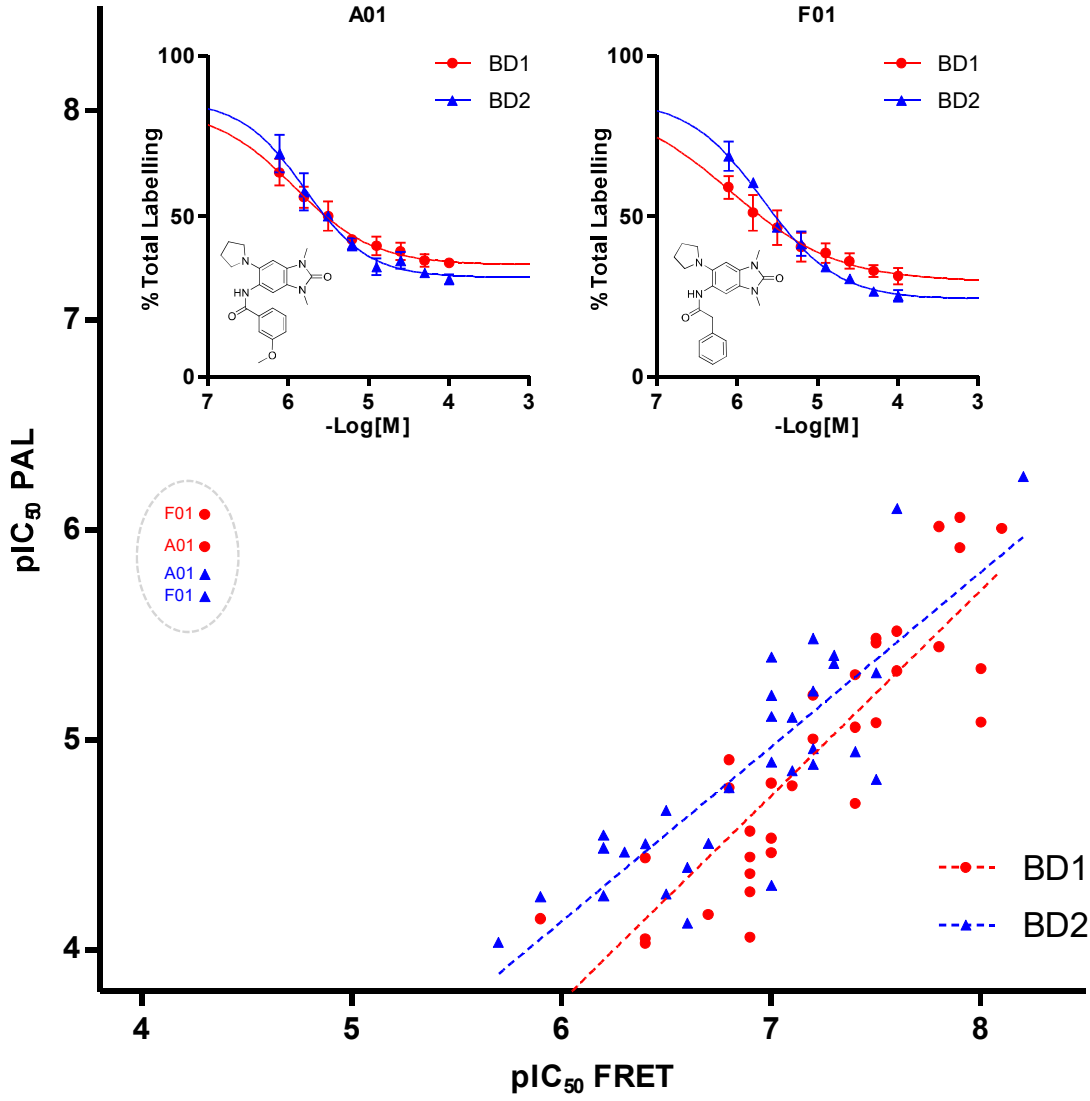


Figure 3.18 Comparison of the pIC_{50} values obtained from the PAL displacement assay and from a TR-FRET biochemical assay. Good correlation was observed between the two assays, except for two structurally similar compounds **A01** and **F01** (PAL dose-response curves inset) which appeared highly potent in the PAL displacement assay but were below the limit of the TR-FRET assay for both BD1 and BD2.

3.3.4 Deriving pK_i values for each competitor compound using the Munson-Rodbard correction to the Cheng-Prusoff equation

The remaining compounds showed good correlation between both assays, however the range of pIC_{50} values from the PAL displacement assay were shifted lower than that of the TR-FRET assay. The Munson-Rodbard correction to the Cheng-Prusoff equation was used to normalise the range of the PAL assay to that of the TR-FRET

assay so they could be better compared (Equation 3.5).[235] The pIC_{50} values of PAL probe **2.45** in the TR-FRET assays (BD1/BD2 pIC_{50} = 7.8/7.3 respectively) were used as a surrogate for the K_d value of the PAL probe at each domain. As 5 µM of PAL probe **2.45** (p*) was used with 1 µM of BD1 and BD2, the initial bound/free ratio of PAL probe (y_0) was entered as 2/3. The K_i values produced using Equation 3.5 were transformed to pK_i using Equation 3.6.

Equation 3.5

$$K_i = \frac{IC_{50}}{1 + \frac{p^*(y_0 + 2)}{2K_d(y_0 + 1)} + y_0} + K_d\left(\frac{y_0}{y_0 + 2}\right)$$

where:

IC_{50} = value obtained in the PAL displacement assay

p* = Initial concentration of PAL probe **2.45**

K_d = IC_{50} value for PAL probe obtained in the TR-FRET assays

y_0 = initial ratio of bound/free PAL probe

Equation 3.6

$$pK_i = -log_{10}K_i$$

The derived pK_i values from the PAL displacement assay were plotted against the pIC_{50} values obtained from the TR-FRET assays for both domains (Figure 3.19). The compounds showed good correlation between the assays for both BD1 and BD2 (R^2 = 0.71 and 0.76 respectively). This demonstrated that the PAL displacement assay developed in this work was a viable biochemical screening method that can be used to inform on the relative affinities of competitor compounds to two target proteins within the same experiment.

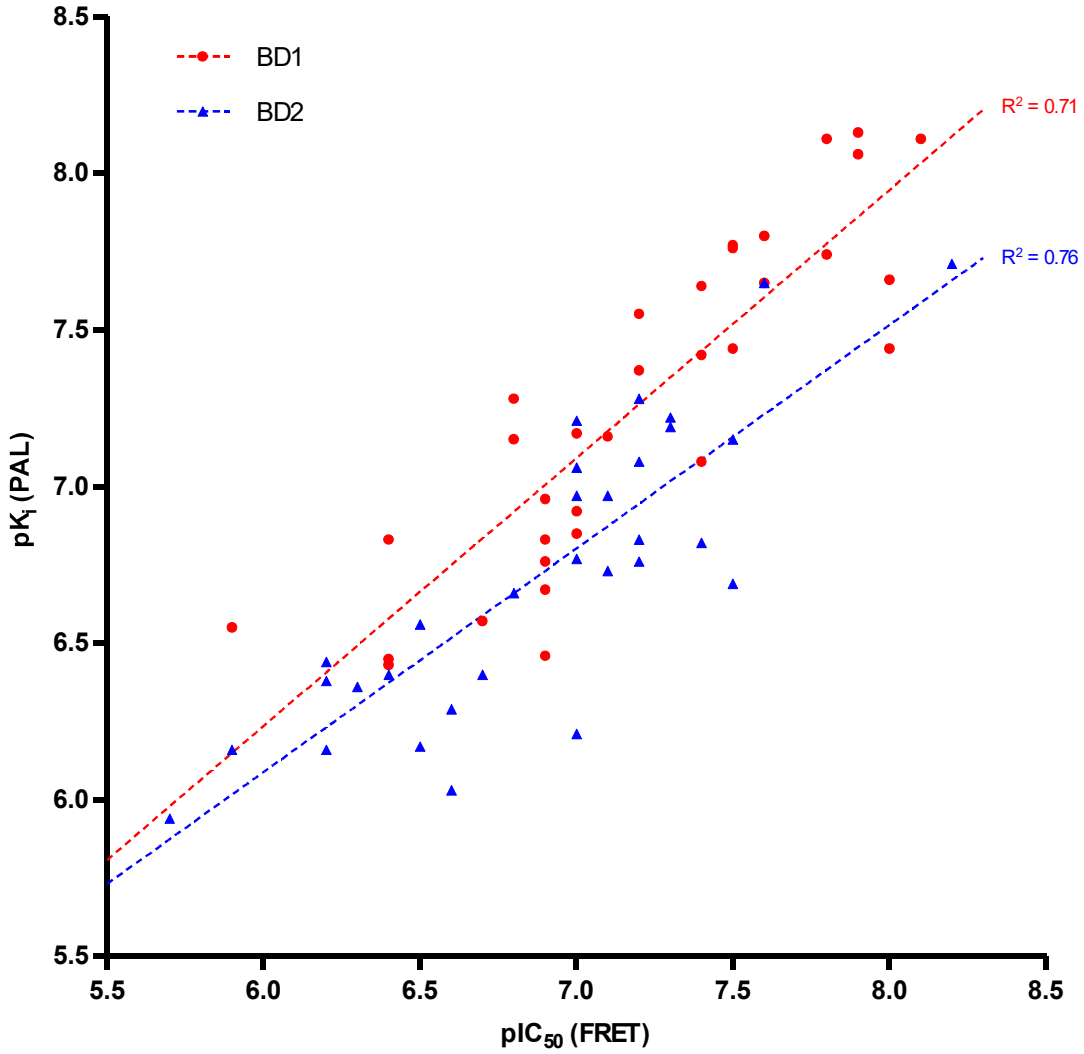


Figure 3.19 Comparison of the derived pK_i values from the PAL displacement assay and the pIC_{50} values obtained in the TR-FRET assay for both domains. The pK_i values were obtained by transforming the IC_{50} values from the PAL displacement assay using the Munson-Rodbard correction to the Cheng-Prusoff equation.[235] A good correlation was observed between the PAL displacement assay and the TR-FRET assay for both BD1 and BD2 domains (R^2 = 0.71 and 0.76 respectively).

The derived pK_i values from the PAL displacement assay for both domains were plotted to explore the BD1/BD2 selectivity of the competitor compounds (Figure 3.20). **C01**, **C02**, **I03** and **L02** were identified as BD2 selective (≤0.4-fold), which was also observed in the TR-FRET assays (Table 3.2). **L01** and **G02** (iBET-295) were also identified as highly potent selective BD2 ligands and were above the upper limit of the PAL displacement assay for BD2. Competitors **F02**, **A02**, **M03** (iBET-151), **M02**, **P02**, **B02**, **E03** were identified as BD1 selective, which also agreed with the selectivity observed in the TR-FRET assays (Table 3.2).

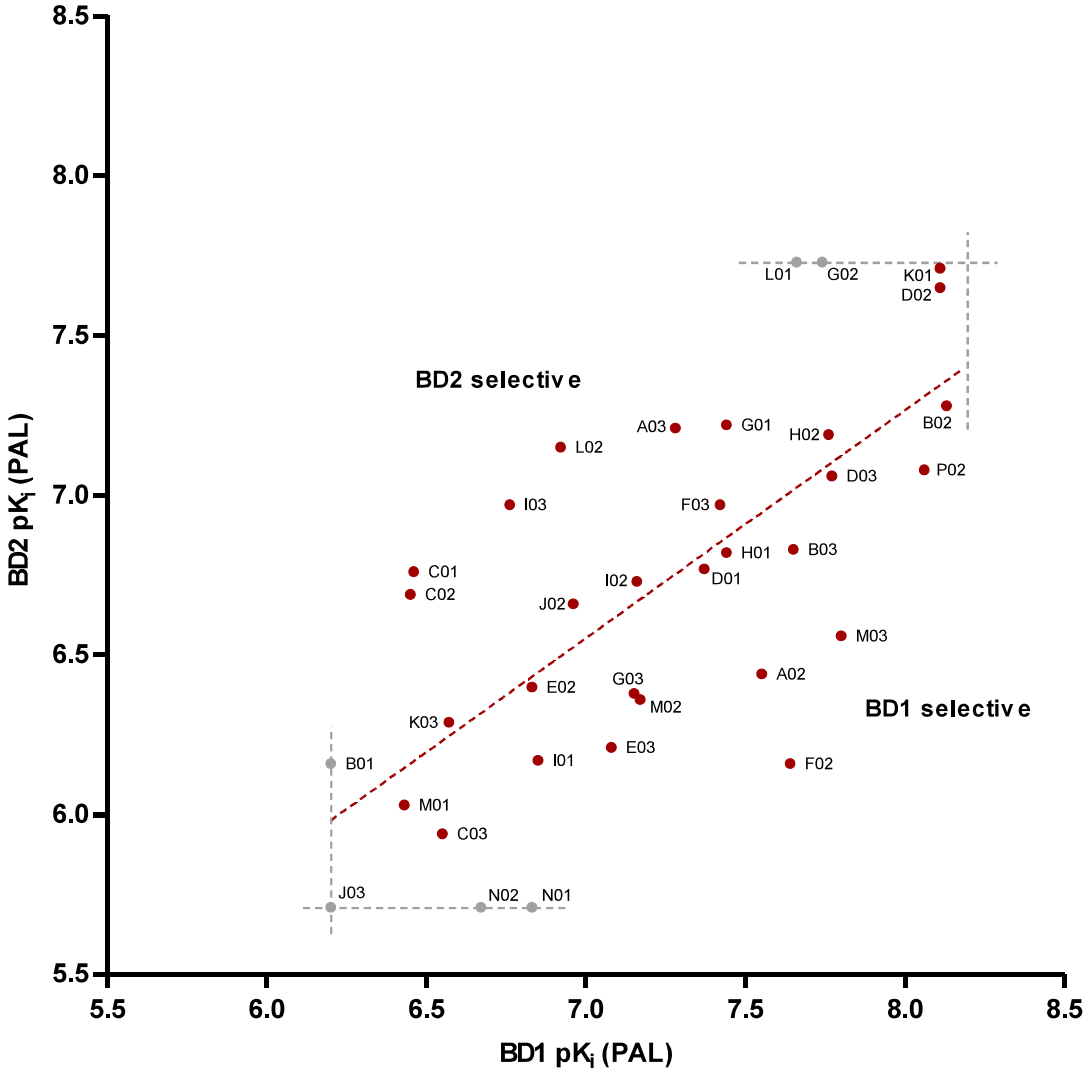


Figure 3.20 Plotting the pK_i values for BD1 and BD2 derived from the IC_{50} values obtained in the PAL displacement assay. The lower and upper limits of the assay are shown by dashed grey lines. The BD1/BD2 selectivity for each compound agreed with the selectivity observed in TR-FRET assays for BD1 and BD2 (Table 3.2).

These experiments demonstrated that the dual-domain PAL displacement assay could be successfully used to determine the selectivity and potency of an array of competitor compounds for BRD4 BD1 and BD2, with a dynamic range of two log units. These values showed good agreement with the relative potencies and BD1/BD2 selectivity observed in TR-FRET assays for both domains. Both data sets are compared in Table 3.2. The 44 compounds assayed were structurally diverse and included a range of well-known BET inhibitors such as iBET-295 (**C01**), TG-101209 (**E01**), iBET-726 (**G02**), bromosporine (**I01**), Mivebresib (**K01**), BI-2356 (**M02**) and iBET-151 (**M03**).

By benchmarking the PAL displacement assay against orthogonal TR-FRET assays for BD1 and BD2, this proof-of-concept study successfully demonstrated that a dual-domain PAL displacement assay could be used to inform on the relative affinities and selectivity of competitors to both BD1 and BD2 domains within the same experiment.

Table 3.2 The TR-FRET, single-shot PAL displacement, and full-curve PAL displacement data for the 44 competitor compounds screened against BRD4 BD1 and BD2 in this proof-of-concept work.

Competitor compound	Compound details	Competitor labelling	BD1 TR-FRET (pIC_{50})	BD2 TR-FRET (pIC_{50})	% Disp. BD1 (single-shot)	% Disp. BD2 (single-shot)	BD1 PAL (full-curve) (pIC_{50})	BD2 PAL (full-curve) (pIC_{50})	BD1 PAL (full-curve) (pK_i)	BD2 PAL (full-curve) (pK_i)
A01	BRPF1 probe analogue[233,234]		<4.3	<4.3	58	88	5.9	5.8	8.1	7.5
A02	Analogue of PAL probe affinity function[196,197]		7.2	6.2	91	96	5.2	4.5	7.6	6.4
A03	Analogue of iBET-762[236]		6.8	7.0	42	68	4.9	5.4	7.3	7.2
B01	Analogue of iBET-762[236]		5.5	5.9	18	69	<3.8	4.3	<6.2	6.2
B02	Analogue of PAL probe affinity function[196,197]		7.9	7.2	100	69	6.1	5.5	8.1	7.3
B03	Analogue of PAL probe affinity function[196,197]		7.6	7.2	84	70	5.3	5	7.7	6.8
C01	iBET-295[227]		6.9	7.2	54	100	4.1	4.9	6.5	6.8
C02			6.4	7.5	23	78	4.1	4.8	6.5	6.7
C03	Dual HDAC/BET probe[230]		5.9	5.7	27	73	4.1	4	6.5	5.9
D01			7.2	7.0	75	78	5	4.9	7.4	6.8
D02	Analogue of PAL probe affinity function[196,197]		7.8	7.6	87	83	6	6.1	8.1	7.7
D03	Analogue of iBET-151[228]		7.5	7.0	78	73	5.5	5.2	7.8	7.1
E01	TG-101209 (Dual JAK2/BRD4 inhibitor)[183,231]	Y	5.9	5.6	58	80	4.4	4.6	6.8	6.5
E02			6.9	6.7	69	68	4.4	4.5	6.8	6.4
E03			7.4	7.0	73	61	4.7	4.3	7.1	6.2
F01	BRPF1 probe analogue[233,234]		<4.3	<4.3	55	81	6.1	5.7	8.1	7.4
F02	Analogue of PAL probe affinity function[196,197]		7.4	6.2	86	43	5.3	4.3	7.6	6.2
F03	Analogue of iBET-762[236]		7.4	7.0	72	74	5.1	5.1	7.4	7.0
G01	Analogue of iBET-762[236]		7.5	7.3	63	58	5.1	5.4	7.4	7.2
G02	iBET-726[227]		7.8	8.2	90	80	5.4	>6.3	7.7	>7.7
G03	Analogue of iBET-151[228]		6.8	6.2	79	74	4.8	4.5	7.2	6.4
H01	Dual HDAC/BET probe[230]		8.0	7.4	87	97	5.1	4.9	7.4	6.8
H02	Analogue of iBET-151[228]		7.5	7.3	78	60	5.5	5.4	7.8	7.2
H03	Dual kinase/BRD4 inhibitor[183,231]	Y	6.4	6.0	48	78	3.9	<3.8	6.3	<5.7
I01	Bromosporine[237]		7.0	6.5	60	87	4.5	4.3	6.9	6.2
I02			7.1	7.1	55	73	4.8	4.9	7.2	6.7
I03	Analogue of iBET-762[236]		6.9	7.1	62	74	4.4	5.1	6.8	7.0
J01		Y	5.8	5.7	27	62	<3.8	4.8	<6.2	6.7
J02			6.9	6.8	69	67	4.6	4.8	7	6.7
J03			5.9	5.9	43	71	<3.8	<3.8	<6.2	<5.7
K01	ABBV-075/Mivebresib[229]	Y	8.1	8.2	89	95	6	6.3	8.1	7.7
K02			5.3	6.4	9	73	NA	4.5	–	6.4
K03			6.7	6.6	48	66	4.2	4.4	6.6	6.3
L01	Analogue of iBET-726[227]		8.0	8.5	69	77	5.3	>6.3	7.7	7.7
L02	Analogue of iBET-762[236]		7.0	7.5	61	71	4.5	5.3	6.9	7.1
L03	DMSO control		–	–	0	0	–	–	–	–
M01			6.4	6.6	29	42	4	4.1	6.4	6.0
M02	BI-2356 (Dual PLK1/BRD4 inhibitor)[231,238]		7.0	6.3	73	82	4.8	4.5	7.2	6.4
M03	iBET-151[228]		7.6	6.5	79	57	5.5	4.7	7.8	6.6
N01	Analogue of iBET-151[228]		6.4	5.8	70	48	4.4	<3.8	6.8	<5.7
N02	Analogue of iBET-762[236]		6.9	6.9	58	74	4.3	<3.8	6.7	<5.7
O01	Benzo[cd]indol-2(1H)-one BET inhibitor[239]	Y	5.6	4.5	23	61	4.3	<3.8	6.7	<5.7
O02		Y	NA	NA	45	66	3.9	3.9	6.3	5.8
P01		Y	4.8	4.8	37	67	<3.8	3.9	<6.2	5.8
P02	Analogue of iBET-151[228]		7.9	7.2	77	67	5.9	5.2	8.1	7.1

3.4 Conclusions and future work

In conclusion, a novel biochemical screening platform was established in this work. The assay used a PAL probe as a reporter compound, and the relative affinities of competitor compounds could be measured by the extent of PAL probe displacement from the target protein. Recombinant BRD4 BD1 and BD2 were used as two proof-of-concept target proteins. One key advantage of this new screening method is the ability to measure affinities of competitor compounds to multiple proteins within the same sample.

PAL probe **2.45** was chosen as the reporter probe due to the high photocrosslinking yields observed with both proteins, and for its fast photoactivation rate (302 nm, ≤1 min). A variety of practical aspects to the assay platform were optimised to give the workflow shown in Figure 3.21. A test set of compounds that had known BET activity were chosen from the CHEMBL database and were dispensed to 384-well plates using automated acoustic dispensing. A solution of PAL probe **2.45** (5 μM) and BRD4 BD1 and BD2 (1 μM each) was then added. The plates were equilibrated on ice and then irradiated. The extent of photocrosslinking by **2.45** to both domains was then analysed directly using intact protein LCMS.

Initially, the test set of compounds (264) were screened at a single high concentration (100 μM). Compounds that showed ≥60% displacement of PAL probe **2.45** from either domain were followed up with full-curve 9-point dose-response experiments to obtain pIC_{50} values for BD1 and BD2. These pIC_{50} values were transformed using the Munson-Rodbard correction to the Cheng-Prusoff equation to produce pK_i values for each competitor, which showed good correlation with TR-FRET pIC_{50} values for both BD1 and BD2 (R^2 = 0.71 and 0.76 respectively). These pK_i values obtained from the dual-domain PAL displacement assay could be compared against each other to inform on the BD1 or BD2 selectivity of the competitor compounds. The BD1/BD2 selectivity observed in the PAL displacement assay agreed with the selectivity observed in the TR-FRET BD1 and BD2 assays.

One unresolved issue using this approach is that a subset of compounds (7 out of 44) showed concentration-dependent covalent labelling of the target protein. In these cases, labelling of BD2 was *ca*. 20-fold higher than for BD1. Five of these compounds contained a sulfonamide moiety. Future work should involve investigating the

mechanisms for these undesired covalent labelling events by the competitor compound. For example, whether this covalent labelling occurs in the absence of UV irradiation should be investigated.

When compared to the TR-FRET assays, all compounds showed good correlation except for two compounds (**A01** and **F01**) that contained a 1,3-dimethyl benzimidazolone core structure. Future work is required to investigate why these compounds appeared as false positives in the PAL displacement assay, or as false negatives in the TR-FRET assay. The assay platform may also be developed further to increase the number of protein domains per sample to obtain more information without extending assay time. For example, BET isoform selectivity could be analysed using recombinant BRD2 BD1, BRD3 BD1 and BRD4 BD1. Another use of this assay platform could be to screen against a desired on-target and undesired off-target protein simultaneously within the same experiment. Despite being optimised to 5.75 min/sample, analysis by intact protein LCMS is still the rate-limiting step in this assay workflow. Faster sampling technologies such as Rapidfire MS (*ca*. 8 s per sample) or RapiFlex MALDI MS (*ca.* 0.3 s per sample) may be applicable to increase the throughput of this new assay platform.[240,241]

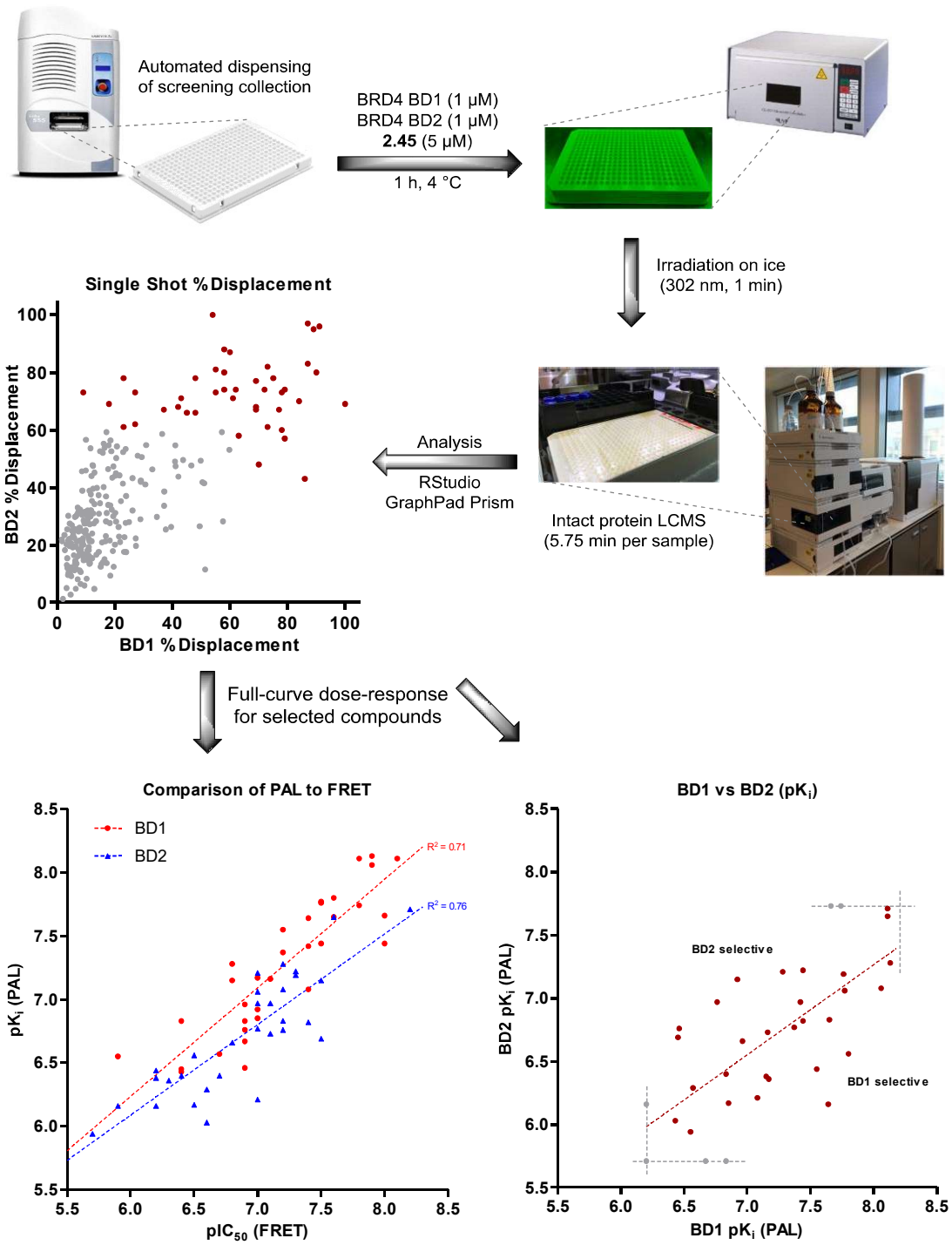


Figure 3.21 Dual-domain PAL displacement assay workflow developed in this work. As a proof-of-concept, 264 compounds with known BET activity were screened against BRD4 BD1 and BD2, using PAL probe **2.45** as the reporter compound. Compounds that showed ≥60% displacement of **2.45** were screened in follow-up dose response experiments to give pIC_{50} values for each competitor to BD1 and BD2. The relative affinities obtained from the PAL displacement assay showed good correlation to those obtained in orthogonal BD1 and BD2 TR-FRET assays, and the data was used to inform on the BD1/BD2 selectivity of each competitor compound.

Chapter 4 Application of Ugi-derived PAL probes in photoaffinity-based proteomic profiling

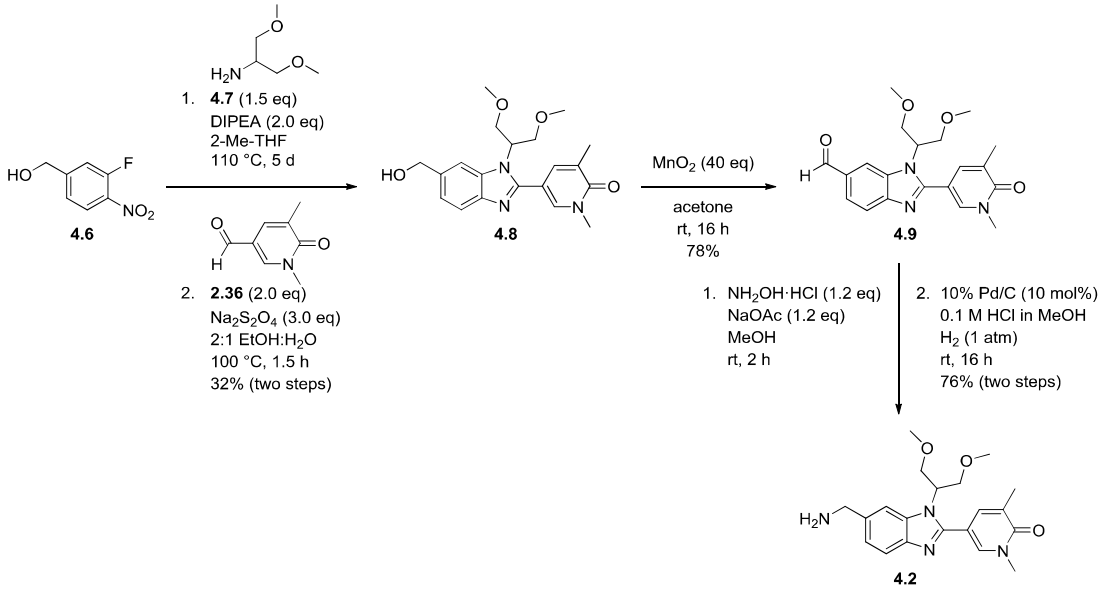
4.1 Introduction

MS-based chemoproteomics is one of the most powerful approaches used to identify the on- and off-target proteins engaged by a compound of interest.[8,24,25,28,80] Of the three capture-based protein profiling methods discussed in Section 1.2.1 (affinity-, activity- and photoaffinity-based), photoaffinity labelling offers the potential to capture the most relevant non-covalent interactions formed by the compound of interest in live cells.[21,26,39] To demonstrate the utility of PAL probes that could be synthesised from an amine-functionalised compound of interest in a single step using the Ugi protocol developed in Chapter 2, these probes were used directly in live cells to profile the on- and off-targets of the compound of interest. This straight-forward and generic approach to photoaffinity-based proteomic profiling would be highly applicable to lead optimisation and pre-clinical stages of small molecule drug discovery programmes.[80,242] The abilities of three different photoreactive groups to enrich the BET family of proteins from HL-60 cells were compared. The BET family of proteins was robustly identified by two of the three photoreactive groups, which demonstrated two key points. Firstly, that PAL probes synthesised using the Ugi protocol developed in this work could be used directly in live cells to find the on and off-targets of the parent affinity function. Secondly, that a range of PAL probes should be synthesised and trialled in the photoaffinity-based profiling experiments to obtain a more informative profile of the engaged PAL probes.

4.2 Synthesis of probes with increased permeability and cellular potency

Photoaffinity-based proteomic profiling in live cells typically requires the use of a clickable alkyne handle to allow for cell permeability.[17] The less membrane-permeable biotin enrichment handle is then installed using copper-catalysed azide-alkyne cycloaddition (CuAAC) chemistry after cell lysis.[35] The aryl azide photoreactive

groups **1.16** and **1.63** were found to be incompatible with the alkyne handle **2.14** (Section 2.2.5). Therefore, these probes were not suitable for progression to studies in live cells. These studies focused on the alkyne-compatible alkyl diazirine **1.64**, trifluoromethyl aryl diazirine **1.65** and benzophenone **1.66** photoreactive groups. In order to maximise the cellular activity of the PAL probes, an alternative affinity function **4.1** was used, which had improved cellular activity in the MCP-1 whole blood assay compared to the parent affinity function **2.31** used in Chapter 2 (pIC_{50} = 7.3 *vs.* 6.9 respectively). Therefore, a short linker amine derivative (**4.2**) of **4.1** was synthesised (Scheme 4.1). A S_NAr reaction of starting material **4.6** with amine **4.7** afforded an intermediate which underwent a reductive cyclisation with the pyridone aldehyde **2.36** to afford **4.8**. Manganese dioxide was used to oxidise the benzylic alcohol **4.8** to the corresponding aldehyde **4.9**, which was condensed with hydroxylamine. The resulting oxime was reduced by hydrogenation over palladium to give the desired amine affinity function **4.2**.

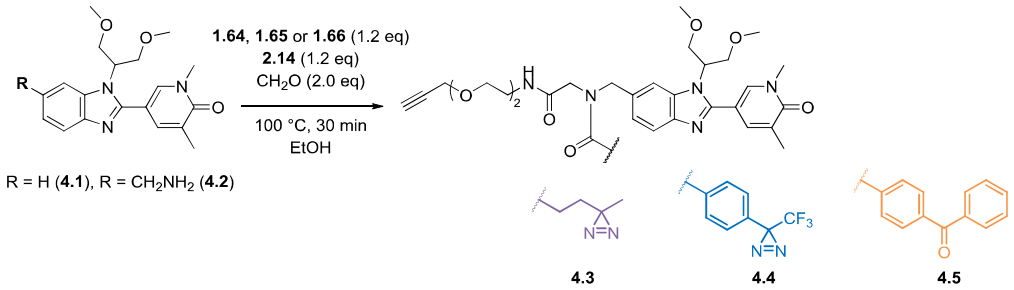


Scheme 4.1 Synthesis of **4.2**, a more cell-permeable analogue of **2.28** with a dimethoxy WPF shelf group in place of the tetrahydropyran group.

Using the Ugi protocol, PAL probes **4.3**, **4.4** and **4.5** were synthesised in good yields (53–70%) from **4.2** and photoreactive groups **1.64**, **1.65** and **1.66** respectively (Table 4.1). Importantly, these probes showed improved cellular potency (pIC_{50} = 7.2, 6.9 and 6.9) compared to their analogues (**2.43**, **2.44** and **2.45**) used in Chapter 2 (pIC_{50} = 6.5,

6.3 and 6.4 respectively, Table 2.3, page 74). Physicochemical properties of the three probes **4.3**, **4.4** and **4.5** were measured (by Iain Reid, Ferdausi Mazumder, Terrence Johnson and Robert Armstrong) to compare the effects of each photoreactive group. The alkyl diazirine probe **4.3** gave the lowest perturbation of physicochemical properties relative to the parent **4.1**, which was consistent with previous results. Alkyl diazirine **4.3** also had the lowest human serum albumin (HSA) binding (39%), compared to the higher binding observed for probes **4.4** and **4.5** (89% and 88% respectively). This higher protein binding to HSA may be indicative of higher levels of non-specific enrichment of off-target proteins by probes **4.4** and **4.5** when used in lysates or live cells.

Table 4.1 Physicochemical properties and potencies of the Ugi-derived PAL probes used in lysate and live cell chemoproteomics experiments. Measurements made by Iain Reid, Ferdausi Mazumder, Terrence Johnson and Robert Armstrong.



Compound	**4.1**[h]	**4.3**	**4.4**	**4.5**
Synthetic yield (%)[a]	–	53	53	70
pIC_{50} BD1[b]	7.4	8.0	8.1	8.0
pIC_{50} BD2[c]	6.2	7.2	7.3	7.4
pIC_{50} MCP-1[d]	7.3	7.2	6.9	6.9
ChromLogD	3.0	3.4	4.9	4.2
Solubility (μM)[e]	529	493	519	489
AMP (nm/sec)[f]	308	8.5	100	19
HSA (%)[g]	36	39	89	88

[a] Isolated yields. [b] TR-FRET assay with recombinant BRD4 BD1. [c] TR-FRET assay with recombinant BRD4 BD2. [d] Assay for the inhibition of MCP-1 cytokine production from monocytes in human whole blood. [e] Solubility data measured by charged aerosol detection (CAD) (μM). [f] Artificial membrane permeability (nm/sec). [g] Percentage binding to human serum albumin. [h] Parent compound **4.1** included for comparison. ChromLogD, AMP and aqueous solubility were measured at pH 7.4.

The full length endogenous BET proteins contain two bromodomains, BD1 and BD2, both of which may be enriched by PAL capture. Therefore, the crosslinking yields of

all three probes **4.3**–**4.5** were determined with both domains using recombinant BRD4 BD1 and BD2 (Figure 4.1).

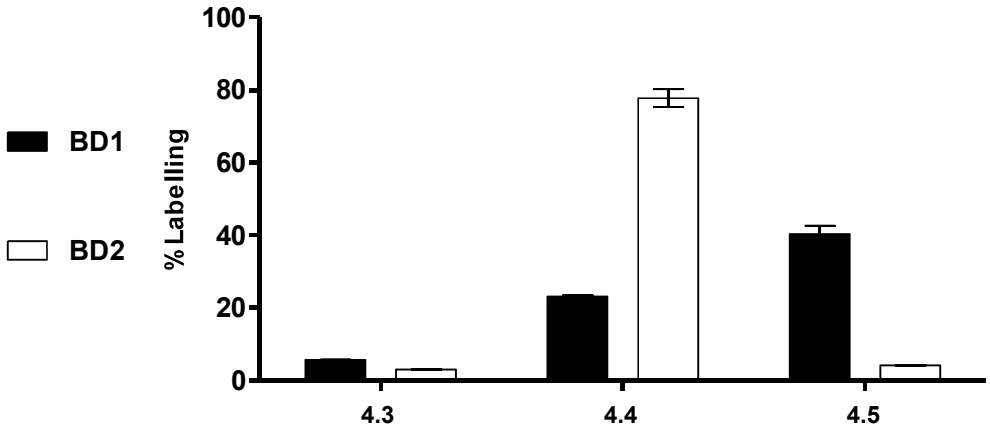


Figure 4.1 Photocrosslinking yields (302 nm, 10 min) for probes **4.3**, **4.4** and **4.5** (5 μM) with recombinant BRD4 BD1 and BD2 (1 μM). Experiment performed in duplicate.

The trifluoromethyl aryl diazirine probe **4.4** gave the highest average photolabelling across both domains (*ca.* 50%), followed by the benzophenone probe **4.5** (*ca.* 22%). The alkyl diazirine probe **4.3** gave poor levels of labelling at both domains (*ca.* 4%). Although probe **4.3** showed the most desirable physicochemical profile, these low levels of labelling to recombinant protein suggested that this probe may fail to capture target proteins in lysates or live cells. Despite the high sequence homology between BD1 and BD2, **4.4** and **4.5** showed high variations in crosslinking efficiencies, demonstrating that subtle changes in protein topology surrounding the active site can have a significant impact on labelling efficiency. After profiling the physicochemical properties, cellular potency and crosslinking efficiencies of the three PAL probes, attention was turned to developing a CuAAC ligation protocol to enable their application in live cells.

4.3 Optimisation studies towards photoaffinity-based proteomic pulldown experiments in lysates and live cells

Protocols for affinity- and activity-based profiling in live cells had been previously developed at GSK (Cellzome) using copper-free click reactions, such as the inverse electron demand Diels-Alder (IEDDA) reaction between *trans*-cyclooctene (TCO) and

tetrazine (Figure 4.2).[34,79,243] In this work, these workflows were modified to incorporate UV irradiation and CuAAC ligation steps, so that the alkyne PAL probes synthesised using the Ugi protocol could be employed. Various modifications from CuAAC-enabled workflows found in the literature were evaluated to develop an optimised protocol.[18,21,35,61]

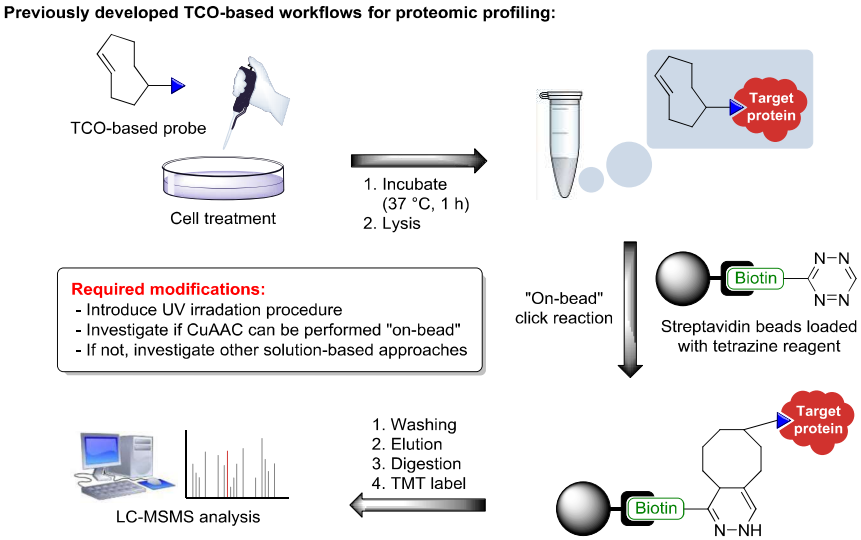


Figure 4.2 Affinity-based proteomic profiling procedure developed by Cellzome. This procedure was adopted and modified to produce a CuAAC-enabled photoaffinity-based proteomic profiling workflow.

4.3.1 Optimisation of CuAAC conditions with recombinant protein

CuAAC conditions reported by Cravatt and co-workers were used as an initial starting point for the optimisation studies.[18] The amount of azide component required for an efficient CuAAC reaction rate at room temperature was investigated using these conditions with recombinant BRD4 BD1 labelled with PAL probe **4.5** (Figure 4.3).

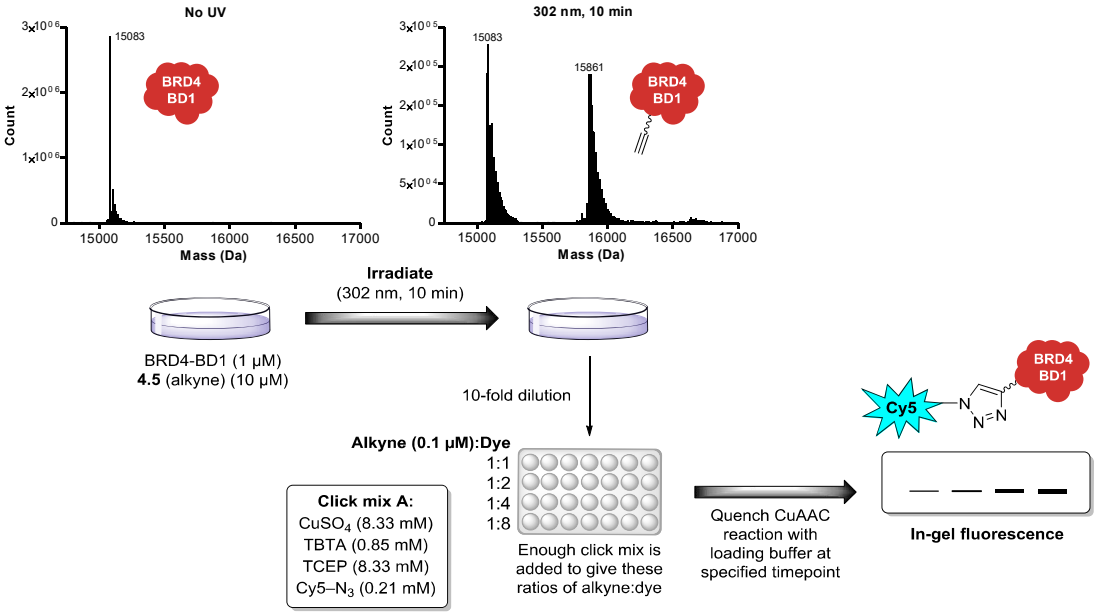


Figure 4.3 Timecourse experiment to investigate the rate of labelling with increasing loadings of click mix A. BRD4 BD1 (1 μM) was incubated with **4.5** (10 μM) and irradiated (302 nm, 10 min). The irradiated samples were diluted 10-fold (giving an alkyne concentration of 0.1 μM). The appropriate amount of click mix A was added to give 1:1, 1:2, 1:4 and 1:8 ratios of alkyne:Cy5 azide. The CuAAC reaction proceeded until quenched by the addition of loading buffer and these samples were analysed by in-gel fluorescence.

A sample of BRD4 BD1 (1 μM) photolabelled with **4.5** (10 μM) was prepared (Figure 4.3). This sample was diluted 10-fold to give a total alkyne concentration of 1 μM, and then split into four samples. Click mix A (containing 0.21 mM Cy5-azide) was added to each sample to give a final alkyne:azide ratio of 1:1, 1:2, 1:4 and 1:8 respectively. These click reactions proceeded in parallel at room temperature until quenched by the addition of loading buffer containing 1 M dithiothreitol (DTT).[39] These samples were subjected to polyacrylamide gel electrophoresis (PAGE) and any fluorescently labelled protein was detected and quantified by fluorescent imaging (λ_{ex} 680 nm) (Figure 4.4, left). For the zero-minute timepoint, the loading buffer was added prior to the click mix. In this sample, no fluorescence was observed, demonstrating that the loading buffer inhibited the reaction and acted as a robust quench. No fluorescence was observed for a DMSO control containing no PAL probe **4.5**. The fluorescence from each band was measured and plotted against click reaction time. The points were fitted with non-linear regression (one-phase decay, GraphPad Prism). The predicted

maximum (Y_{max}) values were used to normalise the data to 100% reaction completion (Figure 4.4, right).

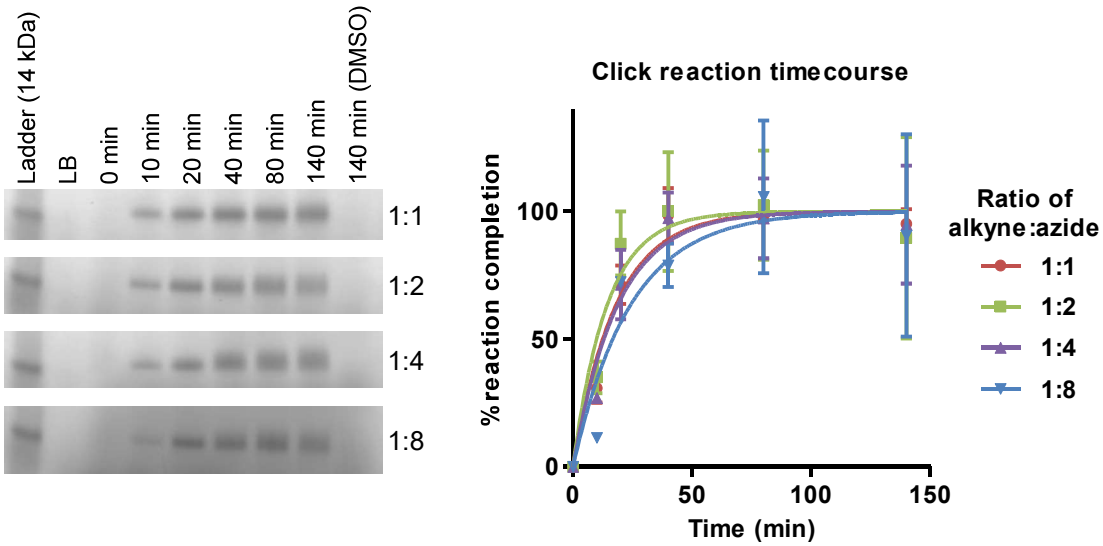


Figure 4.4 In-gel fluorescence of click reaction timecourse with increasing concentrations of click mix A (left). At the 0 min timepoint, the loading buffer was added prior to the click mix, demonstrating quenching. For 140 min (DMSO), the click reaction was performed with protein that was irradiated with DMSO control (no probe present). LB = loading buffer. Absolute fluorescence was measured from each band (n=2 gels) and the background from the surrounding gel area was subtracted using Image Studio. These values were plotted (GraphPad Prism) and fitted with a non-linear regression (one-phase decay) to get a predicted maximum (Y_{max}) fluorescent labelling (data not shown) which was used to normalise the data to 100% reaction completion (right).

There was no significant difference in rate observed with increasing concentrations of click mix, and maximum labelling was achieved after *ca*. 80 min. Greater than 95% maximum labelling was achieved after 1 h, and so this reaction time was chosen for future experiments as a compromise between achieving high levels of ligation and minimising the amount of time that the protein was incubated at room temperature.

4.3.2 Investigating the compatibility of CuAAC ligation with components of the lysis buffer

The compatibility of the CuAAC conditions with the lysis buffer used in the previously reported TCO-based proteomic workflows was investigated.[34,79,243] Initially, lysis buffer containing Tris-HCl was used, however preliminary experiments identified that this buffer inhibited the CuAAC reaction. Tris buffers have been reported to slow CuAAC reactions due to competing with the TBTA ligand for copper(I) or by forming

stable copper(II) chelates.[128,129,244] The lysis buffer was modified by using 50 mM HEPES buffer in place of 50 mM Tris-HCl. The buffer components are listed in Table 4.2. Each component was made up in 50 mM HEPES. **4.5** (10 µM) was incubated with BRD4 BD1 (1 µM) in 50 mM HEPES and irradiated (302 nm, 10 min) (Figure 4.5). The irradiated sample was then diluted (1 in 10) in each pre-prepared buffer component. A positive control sample in PBS buffer was also prepared. Click mix A was added to each sample and after 1 h at room temperature, the reaction was quenched by the addition of loading buffer. The samples were then analysed using in-gel fluorescence (Figure 4.6).

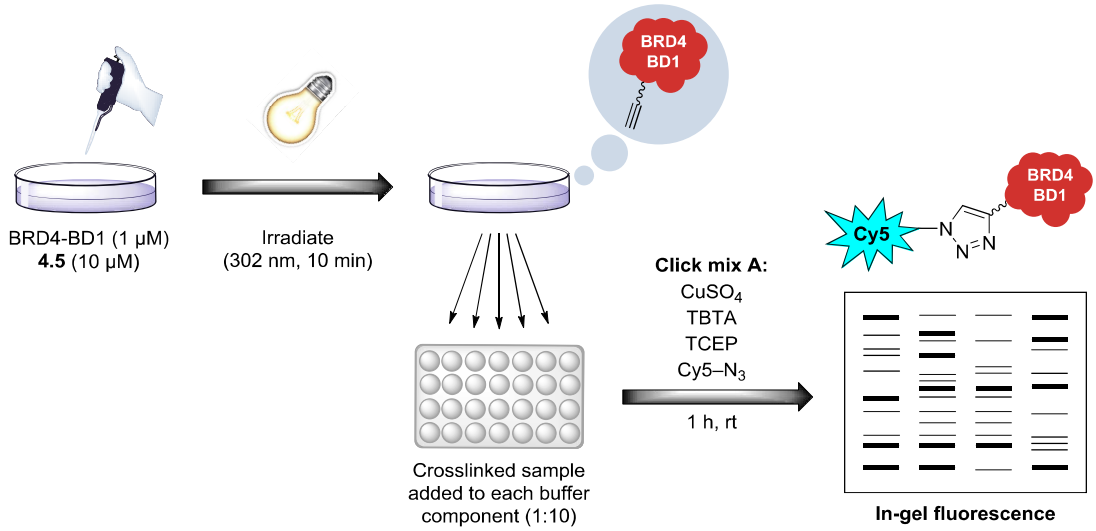


Figure 4.5 Investigation of the effects of the components of the lysis buffer on CuAAC ligation. A batch of crosslinked BRD4 BD1 (1 µM with probe **4.5** (10 µM)) was prepared in 50 mM HEPES and added to each individual buffer component listed in Table 4.2. An irradiated sample in PBS buffer was also prepared as a positive control. Click mix A was added (alkyne:azide = 1:4) and the samples were incubated at rt for 1 h before being quenched with loading buffer and analysed using in-gel fluorescence.

Good labelling was observed for all lanes, except for 1 mM sodium orthovanadate and 0.2% SDS (lanes 6 and 9 respectively, Figure 4.6). SDS and other ionic surfactants have been reported to slow or inhibit CuAAC ligation in the literature.[35] Sodium orthovanadate is used in lysis buffers to inhibit protein phosphotyrosyl phosphatases (PTPs) and avoid dephosphorylation of proteins after lysis.[32,245] Since this work did not involve the study of phosphorylation states and PAL capture would be performed prior to cell lysis, this component was removed. Interestingly, higher levels of fluorescence were observed in HEPES *versus* PBS (lanes 2 and 11 respectively),

suggesting that PBS had a quenching effect on either the photocrosslinking or CuAAC reaction.

Table 4.2 Components of the buffer were made up in 50 mM HEPES

Buffer component
50 mM HEPES
5% glycerol
1.5 mM $MgCl_2$
150 mM NaCl
1 mM Na_3VO_4
25 mM NaF
0.4% IGEPAL CA-630
0.2% SDS

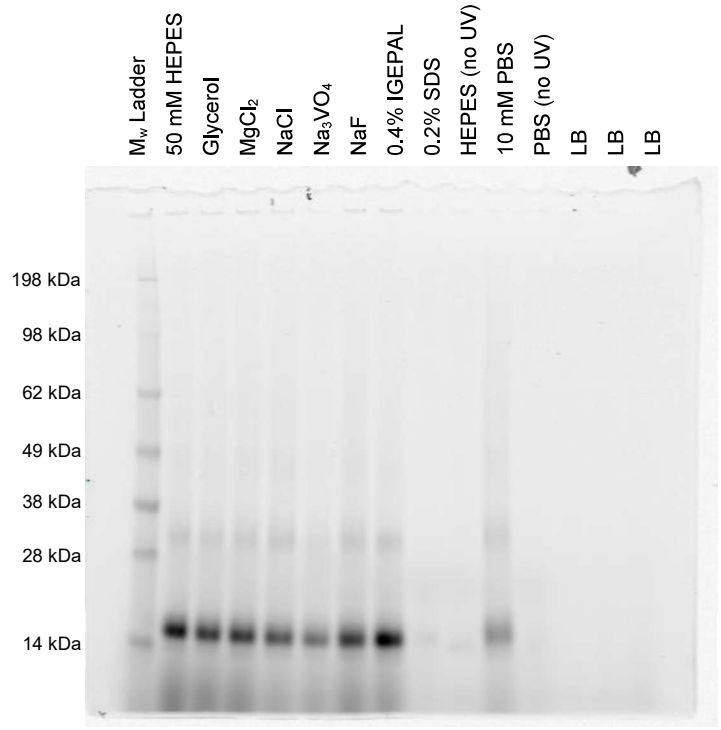


Figure 4.6 In-gel fluorescence of crosslinked BRD4 BD1 conjugated to Cy5 dye *via* CuAAC chemistry with buffer components made up in HEPES. 0.2% SDS completely inhibited the CuAAC reaction, and Na_3VO_4 showed partial inhibition. No fluorescence was observed without UV irradiation.

Satisfied that the CuAAC ligation was compatible with the individual components of the lysis buffer, the reaction was examined with the combined buffer in HL-60 cell lysate.

4.3.3 CuAAC ligation in HL-60 lysate

CuAAC ligation in HL-60 lysate was investigated using an in-gel fluorescence readout, as it enabled faster data turnaround (*ca*. few hours) compared to a full MS-based proteomic pulldown experiment (*ca*. 2 weeks).[39] It was anticipated that the quantities of endogenous BET proteins in HL-60 cell lysate would be too low for identification by in-gel fluorescence, so the lysate was spiked with recombinant BRD4 BD1 to give an observable signal. HL-60 cell lysate was prepared using the modified HEPES lysis buffer and diluted to 1 mg/mL. This lysate was then spiked with

recombinant BRD4 BD1 over a range of concentrations (0–0.8 μM). These mixtures were incubated with **4.5** (10 μM) and irradiated (302 nm, 10 min). CuAAC ligation was performed using Click mix B (alkyne:azide = 1:2, rt, 1 h), which used Cy5.5 azide (λ_{ex} 678 nm) in place of Cy5 azide (λ_{ex} 647 nm) to achieve higher sensitivity by being closer to the excitation wavelength of the gel imager laser (680 nm) (Figure 4.7).

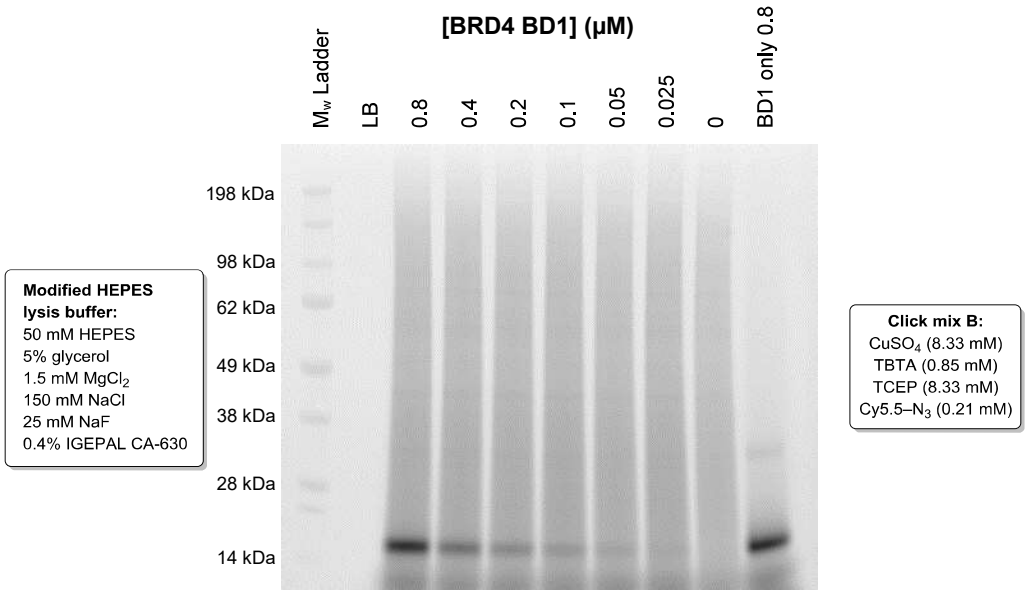


Figure 4.7 In-gel fluorescence for HL-60 lysate (1 mg/mL) spiked with increasing concentrations of recombinant BRD4 BD1 (0–0.8 μM). The samples were incubated with **4.5** (10 μM) for 30 min on ice before irradiation (302 nm, 10 min). Click mix B (Cy5.5 azide in place of Cy5 azide) was used (alkyne:azide = 1:2). LB = loading buffer.

No fluorescently labelled proteins were identified in HL-60 lysate alone, presumably due to low endogenous expression of any PAL captured proteins. However, fluorescent labelling of recombinant BRD4 BD1 at concentrations of 0.1 μM and above could be clearly observed. Furthermore, the click reaction was equally efficient in HL-60 lysate (1 mg/mL) as with recombinant BRD4 BD1 protein alone (lanes 3 and 10 respectively). This in-gel fluorescence protocol was subsequently used to compare the labelling efficiencies of the three probes **4.3**, **4.4** and **4.5** in HL-60 lysate.

4.3.4 Comparison of photolabelling efficiency in HL-60 lysate using in-gel fluorescence

The labelling efficiencies of probes **4.3**, **4.4** and **4.5** to recombinant BRD4 BD1 in HL-60 lysate were compared. Probes **4.3**, **4.4** and **4.5** (1 μM) were incubated with BD1 spiked HL-60 lysate (0.1 μM in 1 mg/mL) in the presence or absence of the BET inhibitor (+)-JQ1 (200 μM).[246] These samples were then irradiated at two different wavelengths (302 and 365 nm, 10 min). Click mix B was used to install Cy5.5 dye, and the samples were analysed by in-gel fluorescence (Figure 4.8). A *ca*. 10-fold increase in fluorescence for BRD4 BD1 labelling was observed for the samples irradiated with 302 nm *versus* 365 nm. This was consistent with the irradiation wavelength comparison conducted in Section 2.3, thus 302 nm was chosen as the irradiation wavelength for future lysate and cell PAL experiments. Importantly, labelling by the three probes was reduced in the presence of excess (+)-JQ1. This demonstrated that the observed photolabelling was due to specific engagement of the PAL probe in the active site of BRD4 BD1.

Interestingly, the alkyl diazirine probe **4.3** gave more fluorescence than the trifluoromethyl aryl diazirine **4.4**, which contradicted the relative labelling efficiencies obtained in PBS buffer and observed by intact protein LCMS in Section 4.2 (Figure 4.1, *ca*. 5% and 25% respectively). Changes in buffer composition have been shown to appreciably affect the amount of labelling observed in PAL experiments by affecting the lifetimes of the reactive intermediates in solution.[62,247,248] This may explain the differences in labelling efficiencies observed for the three probes in the modified HEPES lysis buffer *versus* PBS.

The fluorescent labelling of recombinant BRD4 BD1 was quantified using imaging software (Image Studio). The fluorescence from the remaining lysate in each lane was also measured (Figure 4.9, left). These values were taken as the specific and background levels of photolabelling respectively. For each probe, these values were normalised to the total fluorescence (specific + background) measured (Figure 4.9, right). **4.3** showed the highest signal-to-background ratio for the three probes. This reflected the favoured physicochemical profile of **4.3**, which showed the lowest percentage HSA binding and chromLogD of the three probes (Table 4.1). Therefore, probe **4.3** was chosen for further optimisation studies in HL-60 cell lysate.

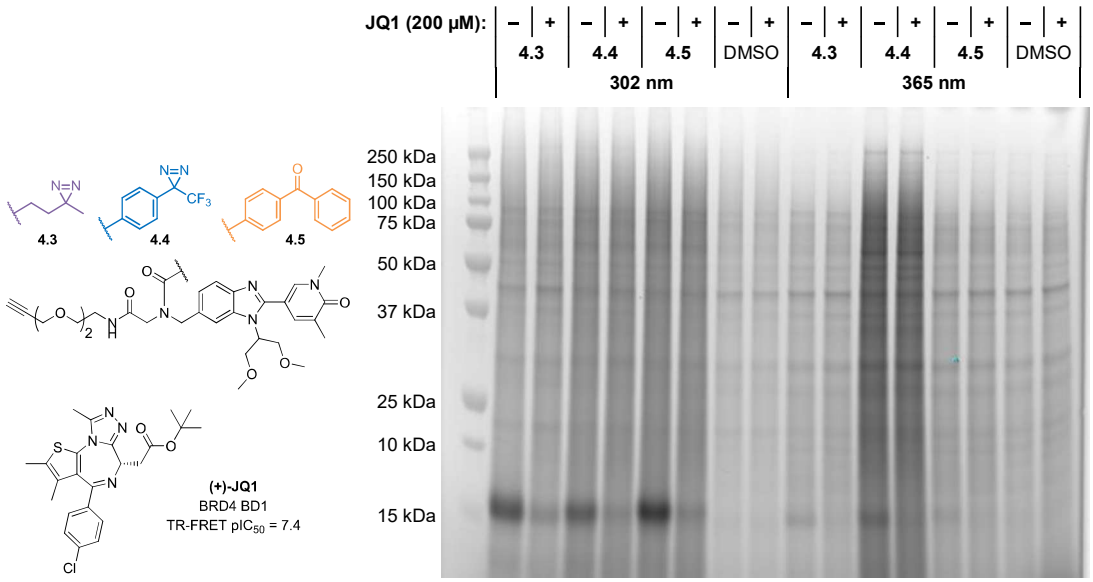


Figure 4.8 In-gel fluorescence of three probes (**4.3**, **4.4**, and **4.5**) after irradiation (302 or 365 nm) with BRD4 BD1 spiked HL-60 lysate (0.1 μM in 1 mg/mL) and CuAAC ligation with click mix B. Irradiation at 302 nm gave superior (>10-fold) levels of BRD4 BD1 labelling *vs* 365 nm. PAL labelling was reduced by an excess of (+)-JQ1 (200 μM). This demonstrated that PAL labelling was due to specific engagement of the PAL probe in the binding site of BRD4 BD1.

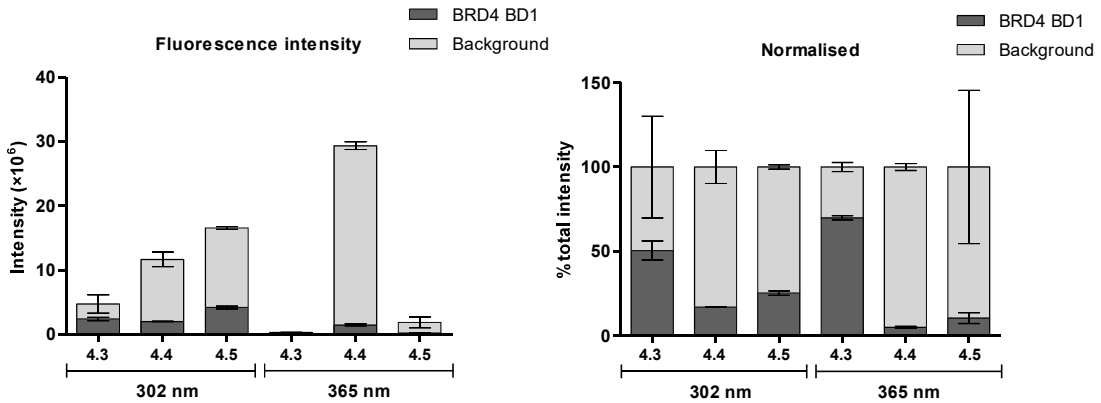


Figure 4.9 The absolute fluorescence intensity was measured (using Image Studio) for recombinant BRD4 BD1 labelling and for the remaining lysate and averaged over two duplicate gels. **4.3** gave the highest ratio of specific-to-background labelling.

4.3.5 Investigating "on-bead" CuAAC ligation

In the previously reported TCO-based pulldown workflows, biotin ligation was performed "on-bead", where biotin-tetrazine was immobilised on neutravidin beads and then introduced to the lysate containing the TCO probe (Figure 4.2).[34,79,243] Performing the biotin ligation on-beads offered a simplified pulldown workflow,

avoiding the need for a protein precipitation step which commonly follows CuAAC ligation reactions performed in solution.[21,34,61] The applicability of an "on-bead" CuAAC ligation was investigated in this work, where biotin-azide was immobilised on neutravidin beads and allowed to react with BRD4 BD1 spiked lysate that had been irradiated with PAL probe **4.3**.

Samples of BRD4 BD1 spiked HL-60 lysate (0.1 μM in 1 mg/mL) irradiated with probe **4.3** (1 μM) were prepared (Figure 4.10) and added to neutravidin beads that were pre-loaded with either biotin azide or biotin only (Alkyne:immobilised azide = 1:7). CuAAC reagents were added, which initiated $t_1 = 0$ for the first click reaction between any alkyne present and the immobilised biotin-azide. At specified timepoints, the two vials were gently centrifuged to pellet the beads, and a sample of the supernatant was taken before the vials were returned to mixing at rt. The sample of supernatant (which could no longer react with any immobilised biotin-azide) was added to click mix B with Cy5.5 azide (alkyne:azide = 1:10), which initiated $t_2 = 0$ for a second click reaction). When $t_2 = 1$ h for each sample, the second click reaction was quenched by the addition of loading buffer. If the CuAAC ligation proceeded with the immobilised biotin-azide in the first click reaction, there would be less free alkynylated BRD4 BD1 to react with Cy5.5 azide in the second click reaction and less fluorescence would be observed. The in-gel fluorescence results are shown in Figure 4.11. Unfortunately, no decrease in fluorescence was observed with increasing t_1, and no difference in fluorescence was observed between beads loaded with biotin-azide or biotin only. This suggested that the reaction with the immobilised biotin azide was too slow and an on-bead CuAAC ligation was impractical.[249] An alternative solution-based CuAAC ligation was therefore investigated.

The experiment was repeated with **4.3** (1 μM) crosslinked to recombinant BRD4 BD1 (0.1 μM) using either biotin-azide or biotin only in solution (Figure 4.12). CuAAC reagents were added (alkyne:azide = 1:7), and at various timepoints (t_1), samples were transferred to vials containing click mix B.

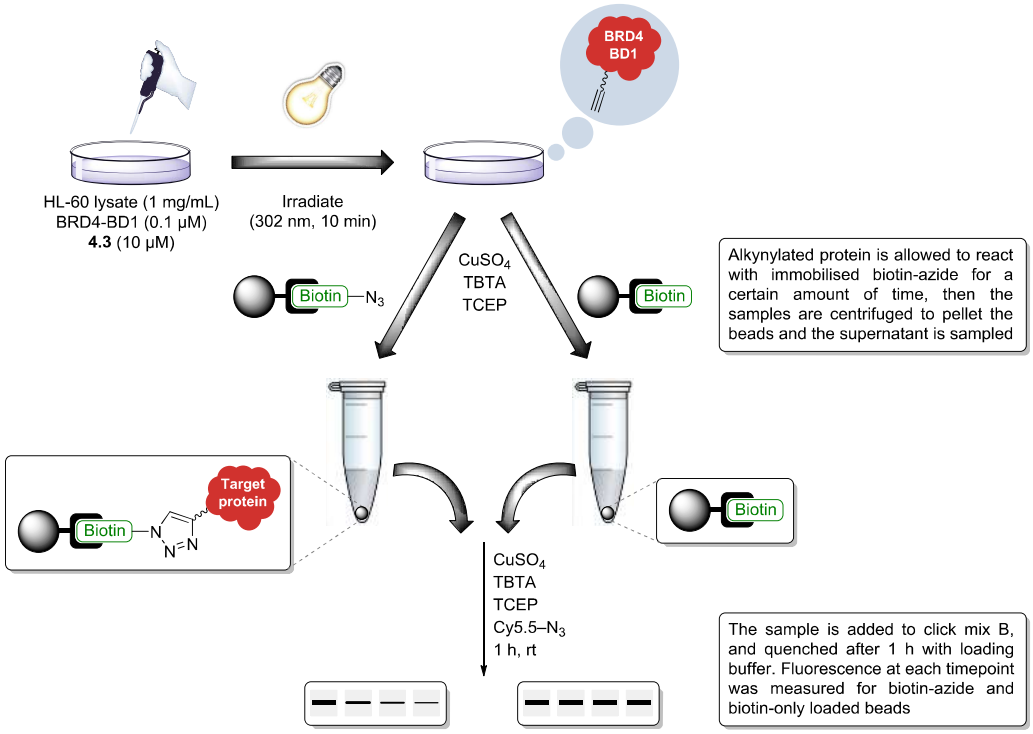


Figure 4.10 Schematic of CuAAC ligation performed with biotin-azide immobilised on neutravidin beads. Crosslinked samples of BRD4 BD1 spiked HL-60 lysate were added to neutravidin beads pre-loaded with biotin-azide or biotin only. CuAAC reagents were added to begin the click reaction (t_1=0). At each timepoint, the samples were centrifuged, and a sample of the supernatant was transferred to fresh click mix B (t_2=0). After t_2 = 1 h, these samples were quenched by the addition of loading buffer. CuAAC ligation to the immobilised biotin-azide depleted the free alkynylated protein in the supernatant, resulting in less fluorescence observed after the second click reaction.

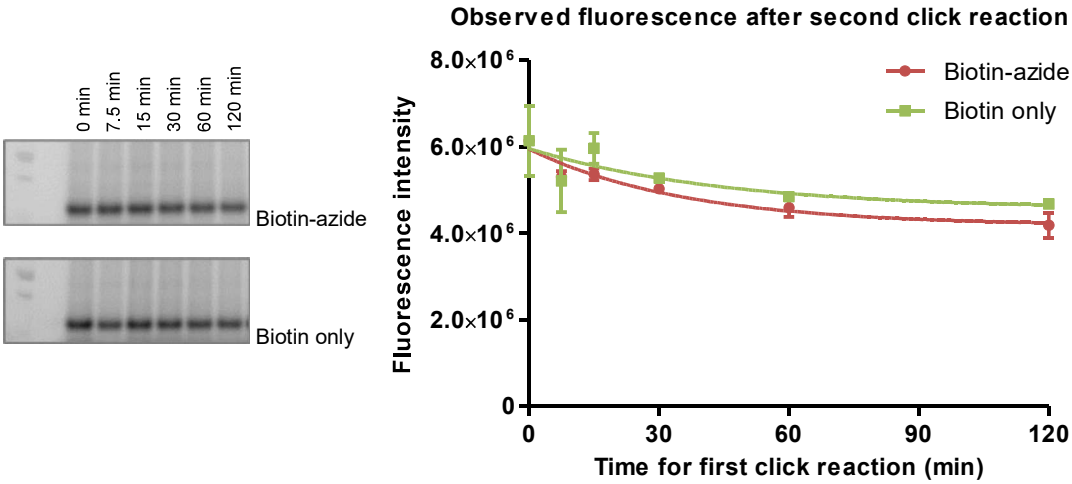


Figure 4.11 In-gel fluorescence analysis of samples after the second click reaction with Cy5.5 azide. No difference in fluorescence was observed between samples mixed with biotin-azide and biotin only beads. No significant decrease in fluorescence was observed with increasing t_1, demonstrating that the reaction between alkynylated BRD4 BD1 and the immobilised biotin-azide was too slow to be used in a pulldown workflow. Experiment performed in duplicate.

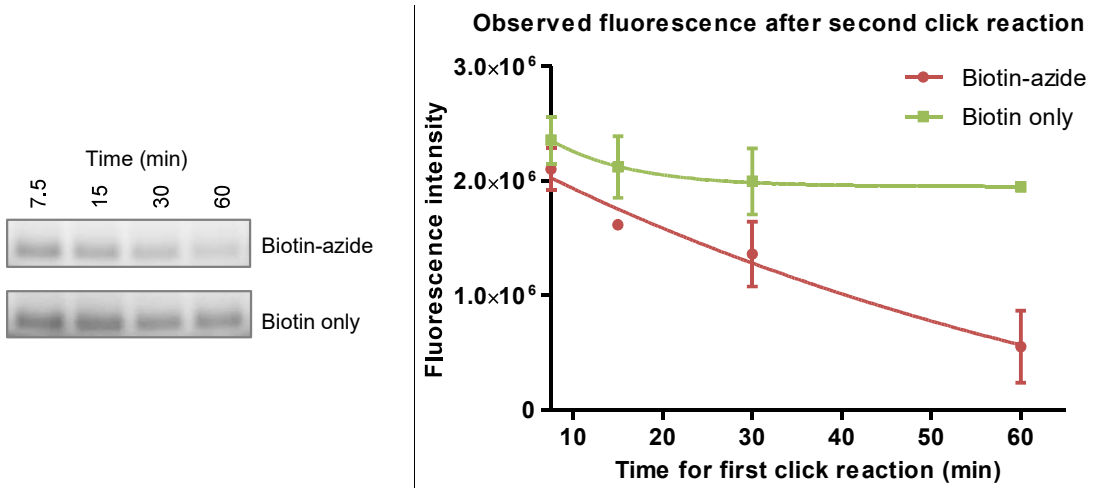


Figure 4.12 In-gel fluorescence of solution-based CuAAC ligation. BRD4 BD1 (0.1 μM) was irradiated with **4.3** (1 μM) and added to either biotin azide or biotin only in solution (1:7 alkyne:azide). CuAAC reagents were added, initiating $t_1 = 0$. At each timepoint, a sample was taken and added to click mix B (Cy5.5 azide:biotin azide = 5:1) and initiating $t_2 = 0$. At $t_2 = 1$ h, the click reactions were quenched by the addition of loading buffer. Experiment performed in duplicate.

As the biotin-azide was free in solution and could continue to react after sampling, a five-fold excess of Cy5.5 azide was used to compete for the remaining alkynylated BRD4 BD1. Pleasingly, a decrease in fluorescence was observed with increasing t_1. This demonstrated that the CuAAC ligation occurred with biotin-azide in solution, albeit at a slower rate than desired (<95% completion after 1 h). For these experiments, the ratio of alkyne:azide (1:7) was used to reflect the concentrations used in the reported TCO-based on-bead ligations.[34,79] However, in the CuAAC procedure reported by the Cravatt and co-workers, higher concentrations of biotin-azide were used (alkyne:azide = 1:45).[18] Moving forward with a solution-based CuAAC ligation, these higher concentrations of alkyne were chosen to maximise biotinylation of any PAL captured proteins.

4.3.6 Optimisation of a protein precipitation step

After establishing that an on-bead CuAAC ligation was infeasible and a solution-based alternative was required, the necessity for any protein precipitation clean-up step was examined. The Cravatt group reported the use of a chloroform-based protein precipitation to remove any small molecules after the biotin ligation.[18] For this work,

a more operationally simple, safer and greener method was desired. Therefore, an acetone-based precipitation reported by Taunton and co-workers was investigated, along with a control sample that had no protein precipitation step.[61]

A sample of BRD4 BD1 spiked HL-60 lysate (0.1 µM in 1 mg/mL) was incubated with **4.3** (1 µM) or DMSO control and irradiated (302 nm, 10 min) in duplicate (Figure 4.13). Biotin ligation was performed using the CuAAC conditions reported by Cravatt and co-workers (Click mix C, Figure 4.13). The samples were split into two workflows. In the first, the samples were incubated directly with neutravidin beads. For the second, the protein content of the sample was precipitated with acetone (4x sample volume, -25 °C) and pelleted by centrifugation. The supernatant containing the CuAAC reagents and excess biotin-azide was removed and the protein pellet was resolubilised and added to neutravidin beads. The beads from both workflows were then filtered and washed to remove any non-specifically bound proteins, and an on-bead trypsin/LysC digest was performed overnight. The peptide eluents from each sample were labelled with an individual TMT-reagent (8-plex). The samples were then combined and analysed by LC-MSMS.

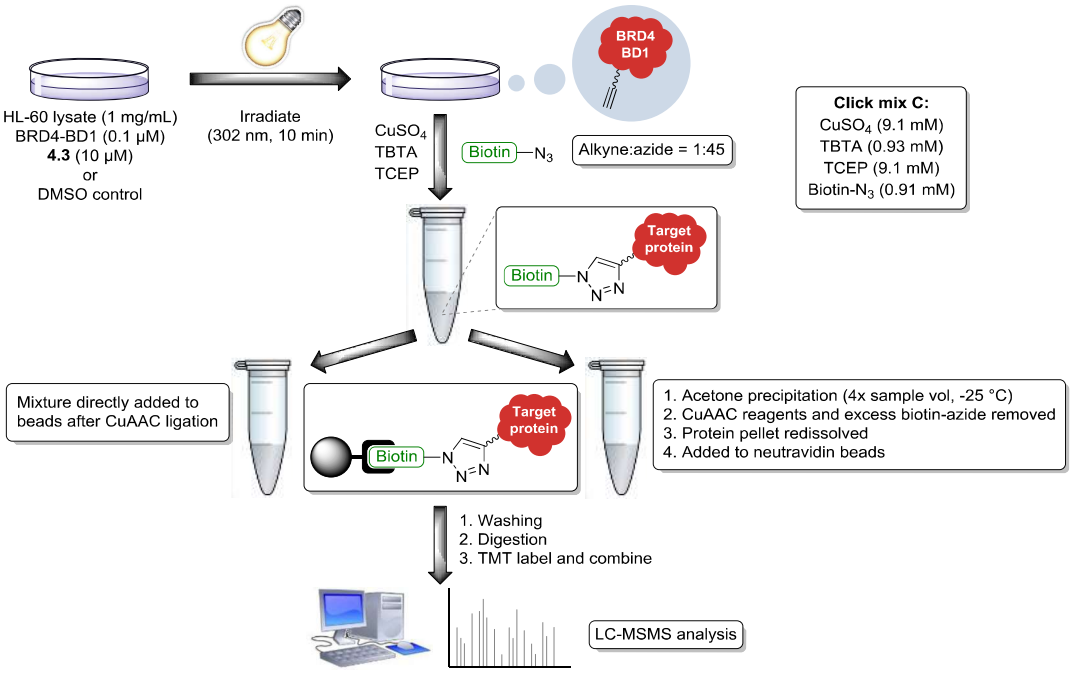


Figure 4.13 PAL workflow for determining the necessity of a protein precipitation clean-up step. Samples with no precipitation step were compared to samples that were precipitated with acetone.

It was observed that the average sum of the relative TMT intensities for the samples that had no precipitation were *ca*. three-fold less intense than those with an acetone

precipitation clean-up step (Figure 4.14, left). It was postulated that this reduction in protein abundance may have been due to the remaining free excess of biotin-azide (*ca*. 44 eq.) competing with biotinylated target proteins for the neutravidin sites on the beads. For the samples with an acetone precipitation step, the ratio of protein enriched by the presence of PAL probe *vs* DMSO (also referred to as the fold-change in enrichment) was calculated for both replicates, and these values were plotted on an XY scatter plot in a $\log_2$ scale (Figure 4.14 right). Gratifyingly, the spiked-in BRD4 BD1 was identified, and was enriched ≥8-fold by the presence of PAL probe *vs* DMSO in both replicates. For the samples that did not have a protein precipitation step, no proteins were significantly enriched (≥2-fold enrichment in both replicates). This demonstrated that a protein purification step was required, and an operationally simple acetone-based precipitation allowed for the PAL enrichment of the BRD4 BD1 target protein to be successfully observed. Furthermore, endogenous BRD3 was also identified and enriched ≥2-fold in both replicates. This demonstrated that this preliminary protocol involving CuAAC ligation with an acetone-based protein precipitation could be used to identify endogenous target proteins captured by PAL in cell lysate. Having established a pulldown protocol, the capture of endogenous proteins by the three Ugi-derived probes was investigated.

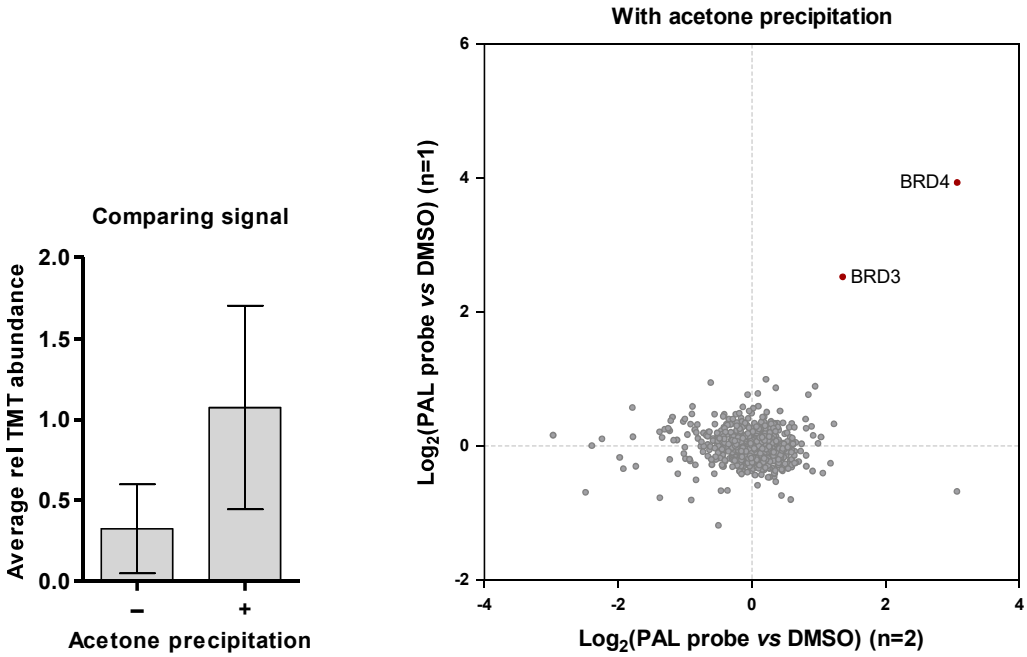


Figure 4.14 The average relative TMT abundance for the samples that had no protein precipitation was *ca.* 3-fold lower than for the samples that had an acetone precipitation clean-up step (performed in duplicate) (left). For the samples with no clean-up step, it was postulated that free biotin-azide (*ca.* 44 μM) may have competed with the biotinylated target proteins, reducing their abundance. For the samples that were cleaned up by acetone precipitation, the $\log_2$ of the ratio of protein identified by PAL probe *vs* DMSO control for each replicate were plotted against each other (710 proteins; ≥2 quantified unique spectra) (right). Spiked BRD4 BD1 was enriched ≥8-fold by PAL probe *vs* DMSO, and endogenous BRD3 was enriched ≥2-fold.

4.4 Photoaffinity-based proteomic profiling using Ugi-derived PAL probes

4.4.1 PAL capture of the BET family of proteins in HL-60 cell lysate

The three PAL probes **4.3**, **4.4** and **4.5** were employed in the established pulldown protocol using unmodified HL-60 lysate to compare the endogenous proteins captured by each photoreactive group (alkyl diazirine, trifluoromethyl aryl diazirine and benzophenone respectively). For each PAL probe, HL-60 lysate (2.2 mg/mL) was treated under three conditions in duplicate; PAL probe (1 μM), PAL probe + competition with (+)-JQ1 (200 μM), and DMSO control (Figure 4.15). These three conditions were used to distinguish between specifically and non-specifically enriched proteins.[43] The samples we incubated for 1 h on ice before irradiation (302 nm, 10 min). CuAAC ligation of biotin-azide was performed in solution, and the excess

reagents were removed with an acetone protein precipitation. The proteins were resolubilised and incubated with neutravidin beads (2 h, 4 °C). The beads were filtered and washed to remove any non-covalently bound proteins. The enriched proteins were then digested with a mixture of trypsin/LysC on the beads (rt, overnight). The resulting peptides were eluted, and each pair of duplicate samples of the three conditions (PAL probe, PAL probe + competition, DMSO) were labelled with a separate TMT reagent (6-plex). The samples were combined and analysed by LC-MSMS.

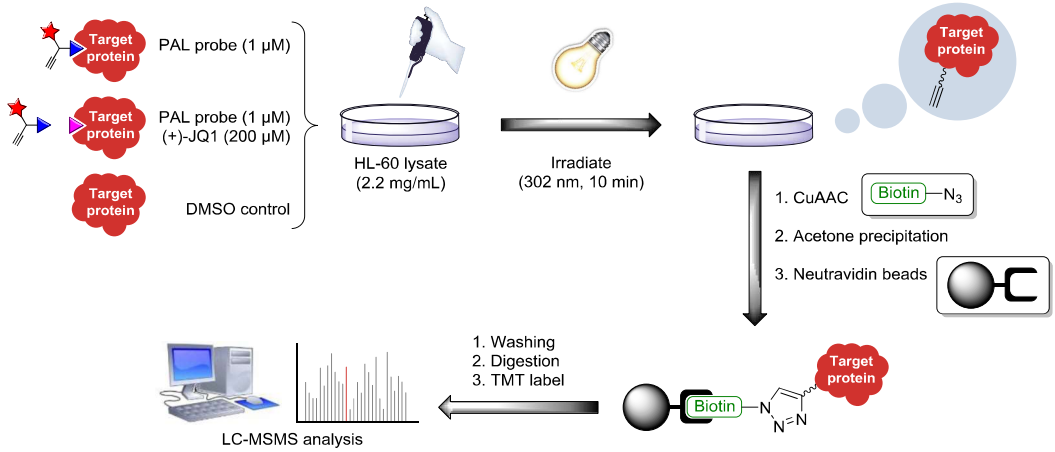
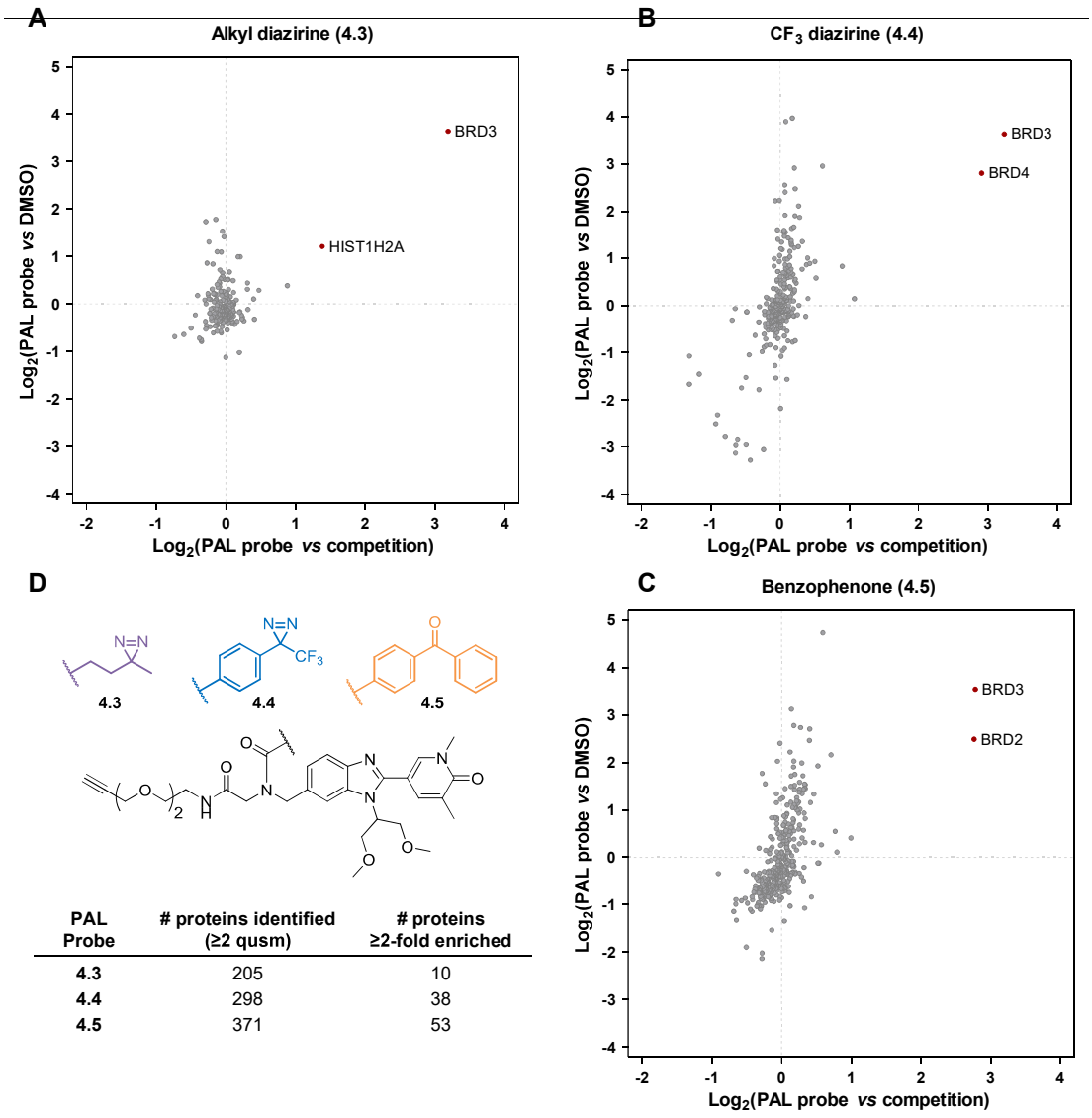


Figure 4.15 The three Ugi-derived probes **4.3**, **4.4** and **4.5** were employed in the established protocol with HL-60 lysate to identify and compare the endogenous proteins that were enriched by each photoreactive group. HL-60 lysate was treated under the three conditions shown in duplicate for each probe. Samples were incubated for 1 h on ice before irradiation. CuAAC ligation of biotin-azide was performed in solution and the excess reagents were removed using an acetone protein precipitation. The samples were incubated with neutravidin beads (2 h, 4 °C) and washed to remove any non-covalently bound proteins. On-bead digestion was performed, and the eluted peptides were labelled with a unique TMT reagent (6-plex) and analysed by LC-MSMS.

Experiments containing PAL probes **4.3**, **4.4** and **4.5** identified 205, 298 and 371 proteins (≥2 quantified unique spectra matches (qusm)) respectively (Figure 4.16). For each experiment, the ratio of protein enrichment for PAL probe *vs* DMSO control was plotted in a $\log_2$ scale on the y-axis. Proteins that were ≥2-fold enriched (≥1 on $\log_2$ scale) were assigned to be significantly enriched. Probes **4.3**, **4.4** and **4.5** significantly enriched 10, 38 and 53 proteins respectively (see Section 7.4.2.8, Table 7.20–Table 7.22, page 301). Interestingly, this correlated with their measured binding to HSA (Table 4.1). The ratio of protein enriched by PAL probe *vs* competition (PAL probe + (+)-JQ1) was plotted in a $\log_2$ scale on the x-axis. Increasing positive x values reported the extent that the PAL probe was displaced from the target protein by an excess of

(+)-JQ1 prior to irradiation. For alkyl diazirine **4.3**, endogenous BRD3 was enriched and competed ≥8-fold. Additionally, histone H2A (HISTH2A), which has no bromodomain, was enriched and competed ≥2-fold. There is no evidence to date of BRD3 directly interacting with H2A in cells, however recombinant BRD3 BD1 has been shown to have affinity for acetylated lysine regions (K35ac and K127ac) of H2A peptides *in vitro*.[180] PAL probes have been shown to proximally label protein binding partners/interactors through space, which could allow for the mapping of local protein environments.[250] Therefore, H2A may have been proximally labelled by **4.3** bound to BD1, through interaction with BD2 or *vice-versa*. For the trifluoromethyl aryl diazirine **4.4** and benzophenone **4.5** probes, only BET family proteins were significantly enriched *vs* DMSO and competed by excess (+)-JQ1. **4.4** competitively enriched BRD3 ≥8-fold, and BRD4 ≥6-fold. **4.5** competitively enriched BRD3 ≥6-fold and BRD2 ≥5-fold. It is worth mentioning that in each case, the missing BET family members (BRD2 and BRD3 respectively) were identified, but with only 1 qusm. This suggested that the BET family proteins were relatively close to the limit of detection under these experimental conditions, which could be due to many factors, such as low levels of endogenous expression, poor PAL capture or poor protein stability after lysis or during the pulldown. To increase the number of peptides for identification in future experiments, a higher initial protein concentration (>2.2 mg/mL) could be used, or the number of peptide identifications from two LC-MSMS sample injections could be combined to give greater analytical coverage. Competition by (+)-JQ1 demonstrated that the PAL probes enriched the BET family members through specific binding in the same active site as (+)-JQ1 (BD1 and/or BD2). For the proteins that were enriched but not competed by (+)-JQ1, it was unclear whether PAL capture to these proteins was through specific binding *via* the affinity function, or through non-specific binding from any part of the PAL probe. In future experiments, competition by excess parent compound would enable the identification of proteins captured through specific binding *via* the affinity function of the PAL probe, therefore informing on the on- and off-targets of the parent compound.



PAL Probe	# proteins identified (≥2 qusm)	# proteins ≥2-fold enriched
4.3	205	10
4.4	298	38
4.5	371	53

Figure 4.16 Log_2 proteomic plots for probes **4.3** (A), **4.4** (B) and **4.5** (C) used in HL-60 lysate. The ratio of protein enrichment for PAL probe *vs* DMSO control was plotted along the x-axis in a log_2 scale and represents the amount of protein enriched by PAL probe capture. The ratio of protein enrichment for PAL probe *vs* competition (PAL probe + (+)-JQ1) was plotted on the y-axis in log_2 scale and represents the extent that (+)-JQ1 displaced the PAL probe from the target protein prior to irradiation. D) **4.3**, **4.4** and **4.5** enriched (≥2-fold) 10, 38 and 53 proteins respectively.

4.4.2 PAL capture of the BET family of proteins in live HL-60 cells

To obtain a more biologically relevant representation of the proteins engaged by the three PAL probes, the experiment was repeated in live HL-60 cells. The cells were treated under the same three conditions that were used in cell lysate, however for competition, parent amine compound **4.2** (50 μM) was used in place of (+)-JQ1 (200 μM) for two reasons. Firstly, **4.3**, **4.4** and **4.5** were approximately 10-fold more potent

in the MCP-1 cellular assay (pIC_{50} = 7.1, 6.8 and 7.0 respectively) *vs* (+)-JQ1 (pIC_{50} = 6.2), therefore a high concentration of (+)-JQ1 (200 μM) would be required to displace the PAL probes, which has been shown to be cytotoxic. Secondly, by using the parent affinity function as the competitor, any non-competitively enriched proteins identified would likely have been captured due to binding through another part of the PAL probe other than the affinity function. Therefore, any competitively enriched proteins were more likely to be genuine targets of the parent compound, or closely neighbouring proteins thereof. The cells were treated under these three conditions and incubated for 1 h (Figure 4.17). The cells were irradiated (302 nm, 10 min) and lysed using the modified CuAAC-compatible lysis buffer. CuAAC ligation was used to install biotin, and the excess CuAAC reagents were removed using an acetone protein precipitation. The PAL captured proteins were then enriched and digested on neutravidin beads. The peptide eluents were TMT-labelled, and the combined sample was analysed by LC-MSMS.

In the experiments containing trifluoromethyl aryl diazirine **4.4** and benzophenone **4.5**, 135 and 95 proteins were ≥2-fold enriched by the presence of PAL probe *vs* DMSO control, respectively (Figure 4.18). These significantly enriched proteins are listed in Section 7.4.2.9, Table 7.24 and Table 7.25. For the proteins that were ≥2-fold enriched by **4.4**, only the BET family (BRD2, 3 and 4) were ≥2-fold competed by the presence of parent compound **4.2** (50-fold excess). This competitive enrichment demonstrated that PAL capture of the BET family was a result of specific target engagement *via* the affinity function of the PAL probe. For probe **4.5**, the BET family was enriched and competed ≥4-fold. **4.5** also ≥2-fold competitively enriched a nuclear zinc-finger domain (ZMYM1), and SUPT16H, which is a component of the FACT complex and a known BRD4 interactor.[51,251] Capture of these off-target proteins could have been a result of being in close proximity to one of the BET target proteins, enabling the BET-bound probe to crosslink to neighbouring interactors. Through proximity-induced crosslinking, PAL probes can identify protein binding partners, and the longer linker probes synthesised in Section 2.3.3 might prove particularly effective in sampling the local protein environment around the BET family in future experiments.[250]

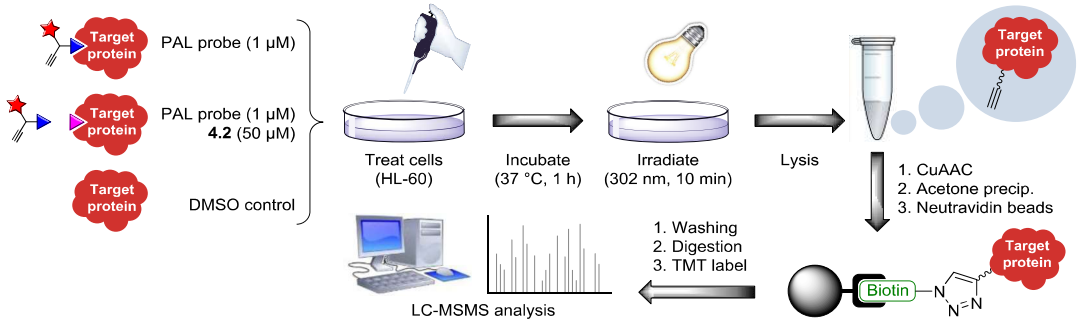


Figure 4.17 Modified CuAAC-enabled PAL workflow in HL-60 cells.

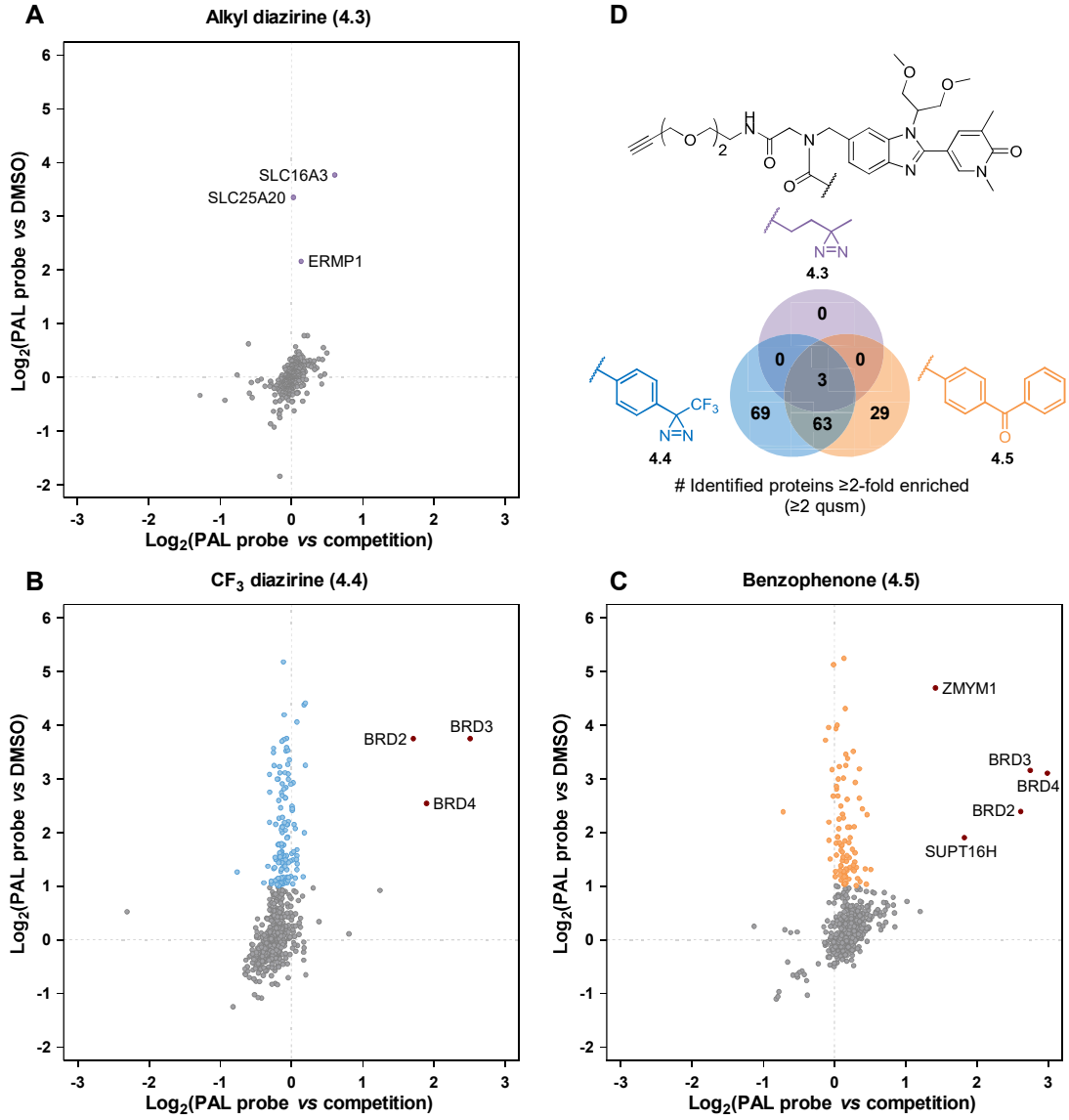


Figure 4.18 Log_2 proteomic plots for the proteins enriched by probes **4.3** (A), **4.4** (B) and **4.5** (C) from live HL-60 cells. The ratio of protein enrichment for PAL probe *vs* DMSO control was plotted on the y-axis in a log_2 scale and represents the amount of protein enriched by PAL probe capture. The ratio of protein enrichment for PAL probe *vs* competition (PAL probe + parent compound **4.2**) was plotted on the x-axis in log_2 scale and represents the extent that **4.2** displaced the PAL probe from the target protein prior to irradiation. D) **4.3**, **4.4** and **4.5** enriched (≥2-fold) 3, 135 and 95 proteins respectively.

The alkyl diazirine probe **4.3** did not afford robust identification of any BET family members, which was consistent with the low photocrosslinking yields observed with recombinant protein in Section 4.2. Two solute carrier proteins (SLC16A3 and SLC25A20) and an endoplasmic reticulum metallopeptidase (ERMP1) were the only proteins that were ≥2-fold enriched by **4.3** *vs* DMSO. These proteins were also enriched but not competed for in the experiments performed with **4.4** and **4.5**, which suggested that these proteins were enriched due to non-specific binding. This study demonstrated that while the alkyl diazirine photoreactive group is often favoured due to its small size, it is not always optimal for protein capture, and may have potential to be optimised further.[36]

For the 66 proteins that were enriched (≥2-fold) by both probes **4.4** and **4.5** but not competed by parent **4.2**, gene ontology (GO) term enrichment analysis was performed using DAVID bioinformatics resources using the total number of proteins identified (656) as a background. The biological process and molecular function GO terms were grouped with a statistical cut-off of P-value ≤ 0.05 (Figure 4.19). This analysis revealed a high representation of mitochondrial transport proteins, specifically the solute carrier (SLC) family of transporter proteins (Table 4.3).[252-254]

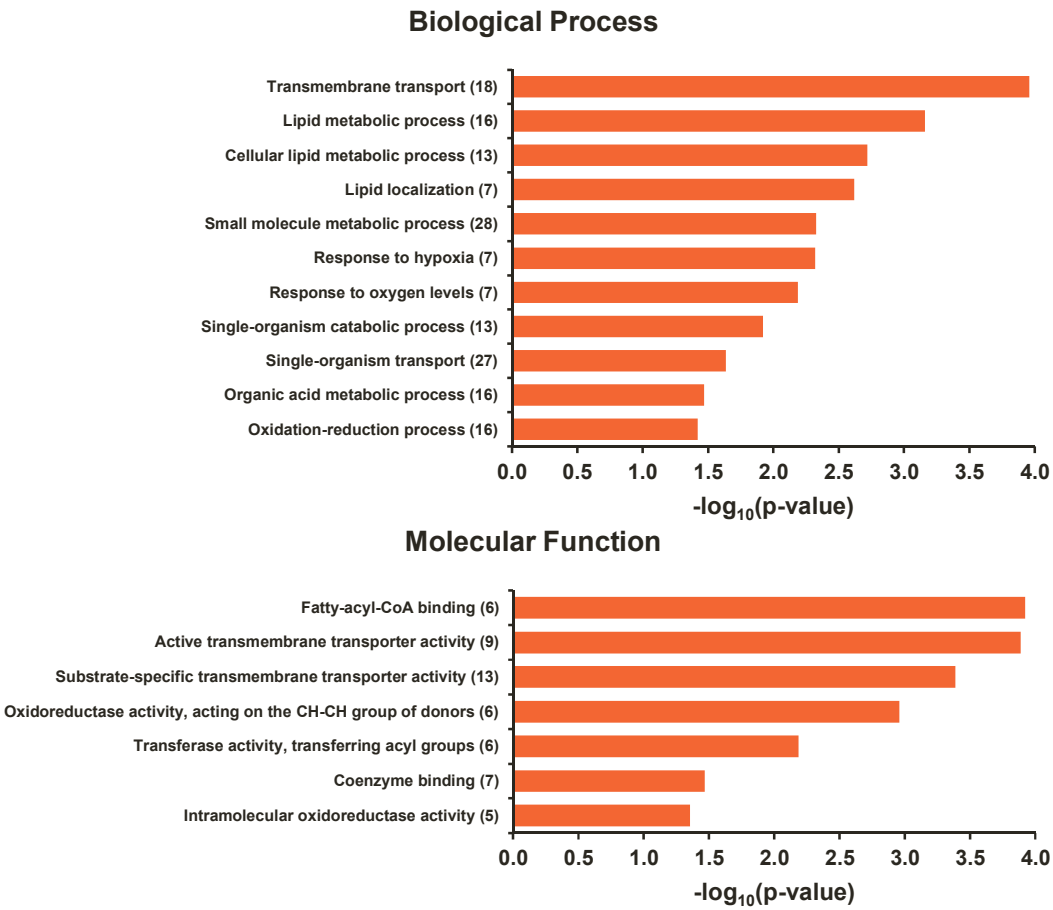


Figure 4.19 Gene ontology (GO) enrichment analysis for the 66 proteins that were ≥2-fold enriched by both **4.4** and **4.5** vs DMSO control. The number of genes grouped in each term is shown in parentheses. A statistical cut-off of p-value ≤ 0.05 was used. All identified proteins were used as a background list. Biological process (GOTERM_BP_3) and molecular function (GOTERM_MF_3) terms were obtained from DAVID bioinformatics resources (Version 6.8).[252,253]

Table 4.3 Genes grouped by the biological process GO term "transmembrane transport" in Figure 4.19 (top).

Gene name	Uniprot reference	Protein description	Log_2(PAL probe *vs* DMSO) Probe 4.4	Probe 4.5
ABCB10	Q9NRK6	ATP binding cassette subfamily B member 10	4.4	3.4
ATP5B	P06576	ATP synthase subunit beta, mitochondrial	1.3	1.3
ATP5C1	P36542	ATP synthase subunit gamma, mitochondrial	1.8	3.5
CPT1A	P50416	Carnitine palmitoyltransferase 1A	1.3	1.7
CPT2	P23786	Carnitine palmitoyltransferase 2	1.8	3.2
ERO1A	Q96HE7	Endoplasmic reticulum oxidoreductase 1 alpha	1.1	1.2
MFSD10	Q14728	Major facilitator superfamily domain containing 10	2.6	2.2
SCARB1	Q8WTV0	Scavenger receptor class B member 1	2.5	2.6
SLC16A1	P53985	Solute carrier family 16 member 1	2.2	2.2
SLC16A10	Q8TF71	Solute carrier family 16 member 10	4.4	2.4
SLC16A3	O15427	Solute carrier family 16 member 3	4.1	5.1
SLC16A7	O60669	Solute carrier family 16 member 7	3.5	4
SLC22A18	Q96BI1	Solute carrier family 22 member 18	3	5.2
SLC25A20	O43772	Solute carrier family 25 member 20	3.6	3.9
SLC25A3	Q00325	Solute carrier family 25 member 3	1.5	1.2
SLC29A1	Q99808	Solute carrier family 29 member 1 (Augustine blood group)	2.6	2.9

UBA52	P62987	Ubiquitin A-52 residue ribosomal protein fusion product 1	1.1	1.3
VDAC2	P45880	Voltage dependent anion channel 2	1.5	1.3

Non-competitive enrichment by both probes suggested that these proteins may have been captured by binding events through the moieties other than the affinity function present in both probes, such as the alkyne handle or core Ugi scaffold. Many of these proteins are involved in a range of solute or small molecule binding and transport, therefore it may be that these proteins recognise and bind to a wide variety of affinity functions.[254] The SLC25A20 gene encodes for the mitochondrial carnitine/acylcarnitine carrier (CAC) transporter protein, which is known to have multiple substrates and is involved in the translocation of acyl carnitine fatty acids into the mitochondrial matrix.[254-256] CAC deficiency in humans usually results in mortality in the neonatal period or during young infancy.[257] Interestingly, Cravatt and co-workers used PAL probe **4.10** to identify SLC25A20 as a key protein target of ingenol mebutate **4.11**, which is the active compound present in the topical drug Picato™, which used to treat actinic ketatosis.[67,258] The group also developed a selective tool molecule **4.12** for SLC25A20 from an initial photoaffinity fragment hit **4.13**, which enriched SLC25A20 from HEK293T cells over other photoaffinity fragments (see Section 5.1.1.1 for a more detailed discussion of this work).[18] The natural product derivative **4.11**, elaborated fragment **4.12**, and the three probes (**4.3**, **4.4** and **4.5**) used in this work have little or no obvious structural similarities (Figure 4.20), suggesting that SLC25A20 binds to a wide variety of ligands.

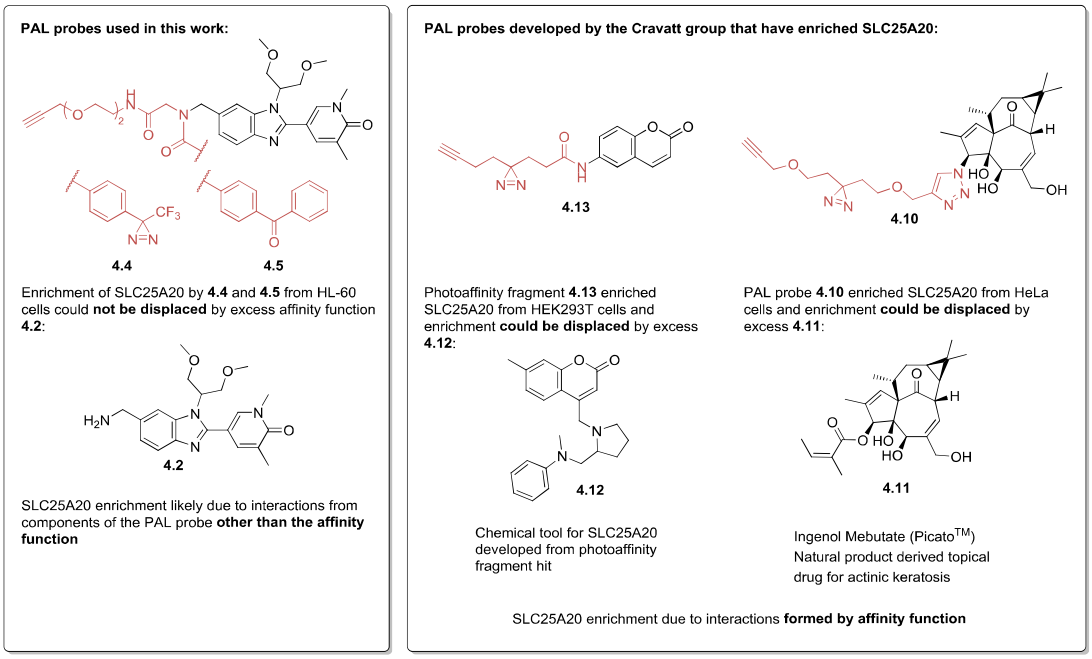


Figure 4.20 The two PAL probes **4.4** and **4.5** enriched a range of mitochondrial transporter proteins from HL-60 cells, however the enrichment of these proteins could not be competed by the excess of parent **4.2**, suggesting that the enrichment was due to interactions formed through components of the PAL probe other than the affinity function (Ugi scaffold, photoreactive groups, or alkyne handle). Recently, the Cravatt group has developed a chemical tool for SLC25A20 from a photoaffinity fragment hit **4.12** that selectively enriched SLC25A20 over other photoaffinity fragments.[18] The group also identified SLC25A20 as a key target of ingenol mebutate **4.11**, which is a natural product-derived marketed drug used to treat actinic ketatosis.[67] In these two cases, SLC25A20 enrichment by the respective PAL probe (**4.13** and **4.10**) could be competed by excess parent affinity function (**4.12** and **4.11** respectively). This demonstrated that SLC25A20 was specifically enriched by interactions formed by the affinity function in these two cases.

In future experiments, an Ugi-derived probe with no (or at least very minimal) affinity function could be used to obtain a background list of proteins engaged by the PAL scaffold. Additionally, Ugi probes with "dummy" photoreactive groups and handles (structural analogues with no ability to form reactive intermediates under UV light or engage in the CuAAC ligation respectively) could be synthesised and used as competitor compounds with the PAL probes used in this work.

4.5 Conclusions

A previously reported proteomic pulldown workflow was modified to produce a successful CuAAC-enabled photoaffinity-based workflow.[34,79] This allowed for the alkyne-containing probes synthesised using the one-step Ugi protocol developed in

Chapter 2 to be used in live-cell photoaffinity-based proteomic profiling experiments. The abilities of three alkyne-compatible photoreactive groups to capture the BET family of target proteins in HL-60 lysate and live cells were compared. These comparative target engagement studies showed that for a PAL probe to successfully report on the non-covalent interactions formed by the parent compound, several criteria must be satisfied. The PAL probe must maintain good cell permeability and potency at the same targets engaged by the parent compound, while being able to sufficiently label those proteins for successful identification. Choice of photoreactive group was also found to be critical. Two of the three probes (trifluoromethyl aryl diazirine **4.4** and benzophenone **4.5**) showed successful competitive enrichment of the BET family, however the minimalist alkyl diazirine **4.3** failed to robustly identify any BET family members in cells. *De novo* design of a single probe to satisfy these criteria is extremely challenging, and so it is advised to synthesise and trial a range of PAL probes to obtain a more complete picture of the true ligand-protein interactions formed by the parent compound of interest. The Ugi protocol reported in this work offers expedient access to a range of PAL probes from a compound of interest in one parallel step, and this work demonstrated that they can be used directly in photoaffinity-based proteomic profiling experiments to identify on- and off- targets in live cells.

4.5.1 Future work

After demonstrating that Ugi-derived PAL probes could be directly used in live-cell MS-based proteomic studies to report on the on- and off-targets of the parent compound, a number of possible future experiments were envisaged. Firstly, to investigate if this strategy is universal, PAL probes from other amine-functionalised compounds of interest (*e.g.* kinase inhibitors, known drug compounds, natural products or phenotypic hits) could be synthesised using the Ugi protocol and trialled in the photoaffinity proteomic profiling workflow developed here. Secondly, a serial dilution of competitor compound could be used to obtain a dose-response relationship between competitor concentration and the extent that the PAL probe is competed from the target proteins (analogous to the PAL displacement assay with recombinant protein

developed in Chapter 3). The power of this approach can be quickly recognised by imagining the use of a promiscuous PAL probe (*e.g.* broad-spectrum kinase inhibitor), where an inferred affinity (K_i) could be obtained for all the proteins that were specifically enriched in a single experiment. Similar approaches have been demonstrated with affinity-based and activity based proteomic profiling, however a photoaffinity-based approach has not yet been demonstrated.[32,61,259] Thirdly, the CuAAC ligation may be further optimised by using the conditions reported by Finn and co-workers using THPTA (**4.14**) as a ligand, or BTTAA (**4.15**) reported by Schumann and co-workers (Figure 4.21). These ligands have been reported to give superior ligation rates at lower temperatures *versus* the TBTA (**4.16**) ligand used in this work.[128,129,260]

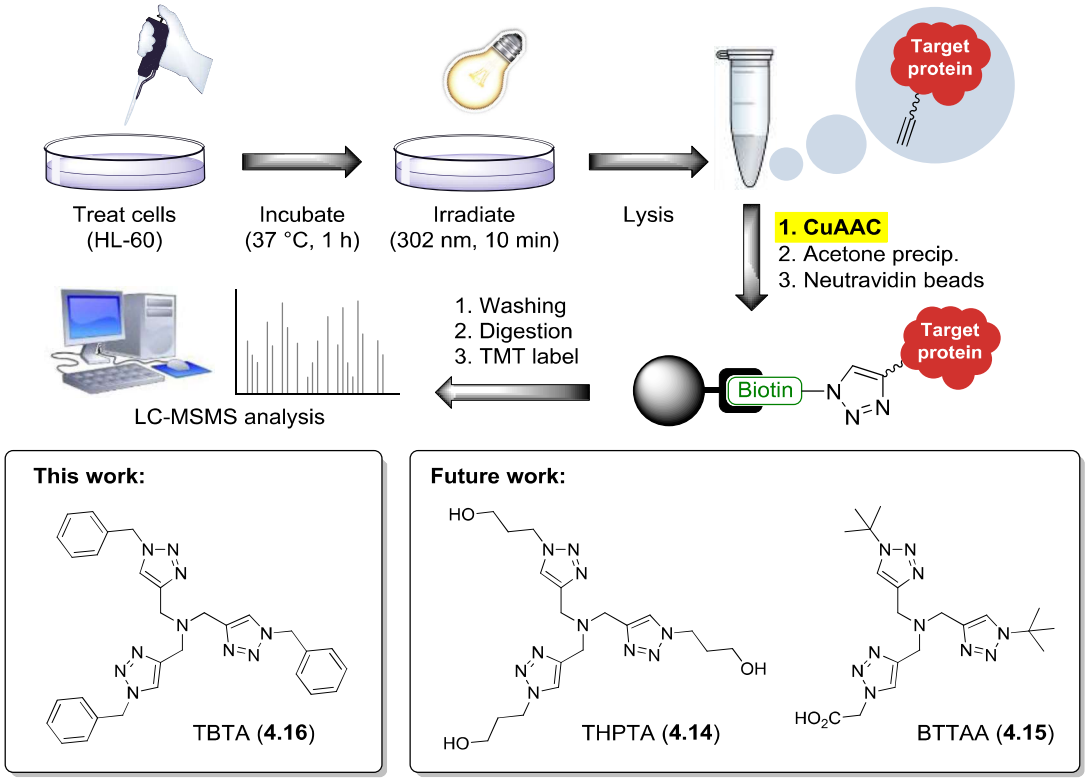


Figure 4.21 Future work to improve the efficiency of the CuAAC ligation step using THPTA or BTTAA ligands in place of the TBTA ligand used in this work.

Chapter 5 Photoaffinity Fragment-based Drug Discovery

5.1.1 Introduction to FBDD

In recent years, fragment-based drug discovery (FBDD) has become an attractive strategy to identify chemical starting points for a target-based drug discovery effort.[261-263] For a small molecule to be considered as a fragment, several criteria must be met. Analogous to Lipinski's 'Rule of Five' and Veber's rules for describing the drug-likeness of a compound, the 'Rule of Three' has been applied to describe fragments.[264,265] To be considered fragment-like, a compound must have a molecular weight <300 Da, have no more than 3 hydrogen bond donors or acceptors, a calculated logP ≤ 3, no more than 3 rotatable bonds and a polar surface area (PSA) ≤ 60 Å^2 (Figure 5.1).[266,267]

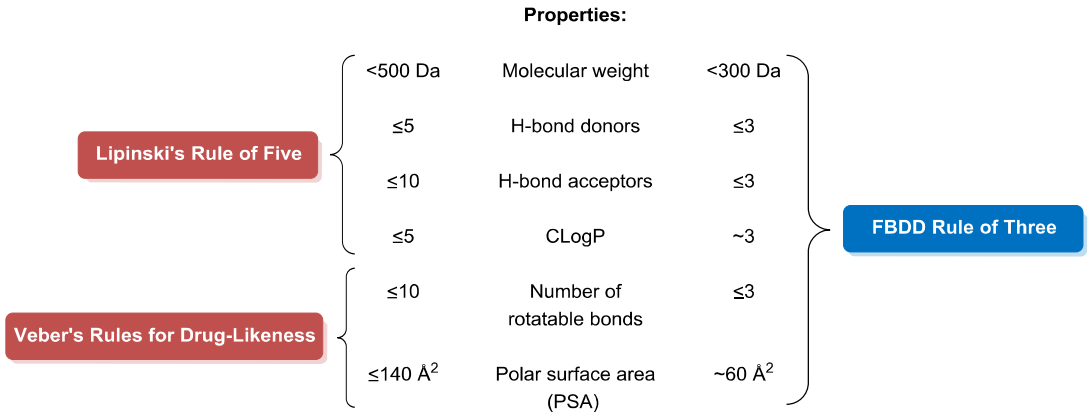


Figure 5.1 The Rule of Three used to describe small molecule fragments used in FBDD.

The number of possible drug-like molecules (those with M_w <500 Da) has been estimated to be at least 10^{30}, and may be as high as 10^{60}. Traditional HTS screening libraries (*ca*. 10^6 compounds with an average molecular weight *ca*. 400–500 Da) only cover a small fraction of this space ($10^6/10^{30} = 10^{-24}$). By imposing smaller fragment-like molecular weight limits (M_w < 300 Da), the number of possible molecules falls to approximately 10^9 possibilities. Therefore by screening *ca*. 10^3 molecules from a well-designed fragment library, a larger fraction of chemical space ($10^3/10^9 = 10^{-6}$) can be sampled, resulting in a higher hit rate.[268-271] This has been demonstrated in practice, where the hit rates from fragment-based NMR screening

were found to be 10–1000 times higher than those from traditional HTS screening efforts.[272]

Due to the small size of the fragment, there is inherently less chemical functionality to interact with the active site of the target protein, and so the binding affinities for hits are expected to be weak (K_d *ca.* 100–1000 μM).[267] However, this binding is relatively efficient, whereby a high fraction of the fragment's atoms are involved in the binding interaction. Ligand efficiency (LE) is a commonly used metric to assess the binding efficiency of a compound in proportion to its size, and is calculated by dividing the free energy of binding by the number of heavy atoms (non-hydrogen atoms) present in the compound.[271] One of the main appeals of FBDD is that by discovering fragments with high ligand efficiencies, the fragments can be rationally grown with the aid of structure-based drug design (SBDD) to give fully elaborated drug-like compounds with maintained high ligand efficiency (Figure 5.2).[268] Fragment linking can also be used if multiple fragment hits are identified within proximal sites of the binding pocket.[273]

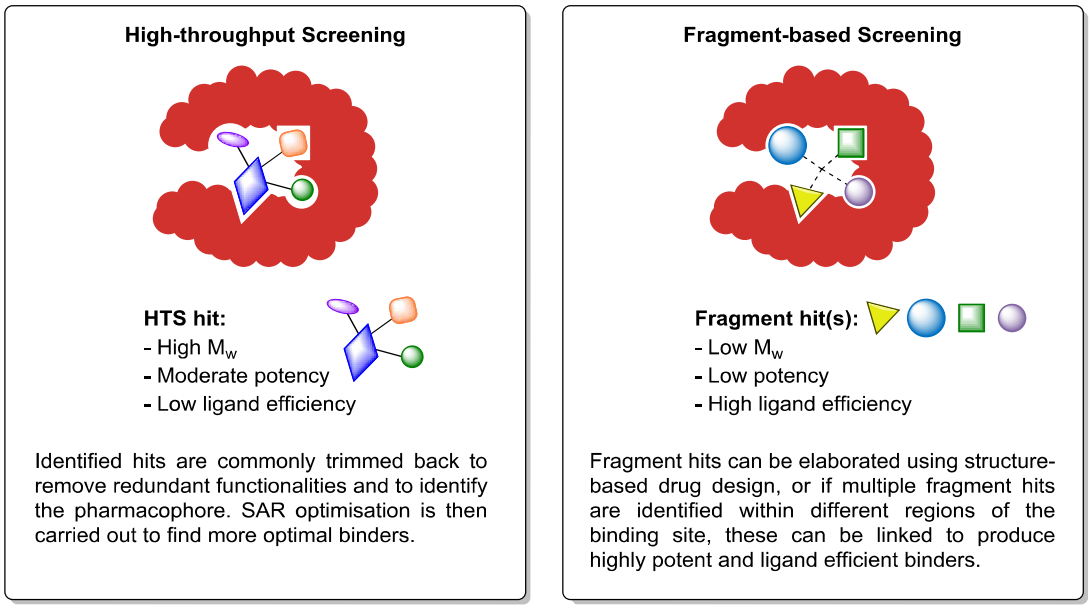


Figure 5.2 Comparison of high-throughput screening versus fragment-based screening. High-throughput screening is widely used to identify moderately potent hits that are useful starting points for SAR optimisation. However, these hits often contain redundant atoms that are not involved in key binding interactions and therefore contribute to low ligand efficiency. Fragment-based screening is used to identify weak but efficient small molecule binders (M_w <300 Da). These fragment hits are grown or linked with the aid of structure-based drug design to produce lead compounds (M_w *ca.* 500 Da) that are highly potent (*ca.* 1 nM) and ligand efficient.

One caveat to FBDD is that more sensitive analytical methods are required to detect the inherently weak binding of hit fragments to the target protein.[207] Additionally, the high (millimolar) concentrations of fragment required to measure the K_d values of hits can result in high levels of non-specific binding which must be accounted for to avoid false positives. Furthermore, issues with aqueous solubility can arise when using these high concentrations. The advantages and disadvantages of the most commonly used methods to measure fragment binding are summarised in Table 5.1.[269,271] Ligand observed NMR is a label-free technique, but requires a relatively large amount of soluble protein.[274,275] Surface plasmon resonance (SPR) requires lower amounts of protein and is more high-throughput, however the protein must first be immobilised on a surface while retaining the integrity of the active site.[276] X-ray crystallography produces accurate binding and structural information, however this method has the lowest throughput.[277] Conversely, high concentration biochemical screening can be performed with the highest level of throughput, but provides no structural binding information.[278] To compensate for the disadvantages or false positives arising from any one particular method, a range of methods are usually used in parallel.[261,269]

Table 5.1 Comparison of the analytical techniques commonly used to detect fragment binding.

Method	Sensitivity	Throughput (samples/day)	Structural insight
NMR (ligand observed)	low mM	Med (10^2)	Some
NMR (protein observed)	Low mM	Low (<10)	Yes
Native mass spectrometry	Low mM	High (10^3)	None
Surface plasmon resonance	Low mM	High (10^3)	None
X-ray crystallography	Mid mM	Low (10)	Yes
Thermal shift assay (TSA)	Low mM	High (10^3)	None
Biochemical screening	High µM	High (10^3)	None

Recently, chemical biology approaches to FBDD involving covalent fragment libraries have been developed.[273] These libraries typically contain electrophilic moieties and are screened against proteins containing nucleophilic residues in (or around) their active sites. The reactivity of the chosen covalent warhead is carefully designed, so that there is limited non-specific background reactivity with nucleophilic residues on the surface of the protein. Any covalent labelling is driven by an initial specific

reversible binding to the active site, which creates a high local concentration of electrophile near the active site's nucleophilic residue. This enables a selective covalent reaction to occur.[54,279] Therefore, any observed labelling is indicative of initial fragment binding. By covalently capturing the transient interaction between the fragment and the protein, more straight-forward detection methods than those listed in Table 5.1 can be used. For example, the covalent modification of the target protein by an electrophilic fragment can be observed by intact protein LCMS, or if the experiment is carried out in lysates or live cells, the captured target proteins can be identified by chemoproteomics.[77]

Johansson *et al.* recently identified a covalent tool molecule for the ubiquitin ligase HOIP from initial screening of *ca.* 100 covalent fragments.[79] An elaborated lead compound showed selective labelling of HOIP in live-cell MS-based proteomics, thus demonstrating that covalent fragment screening can be used to quickly develop covalent chemical probes, and potentially new covalent drug candidates.

Craven *et al.* reported a covalent fragment-based screening platform that used a fluorescence-based readout (Figure 5.3).[78] A 138-member fragment library that contained a variety of electrophilic warheads was incubated with either CDK2 protein or glutathione (GSH) (5 μM). At various timepoints, the thiol-reactive fluorogenic probe **5.1** was added in excess, which reacted with any remaining free thiol. By comparing the rates of protein labelling versus the intrinsic reactivity to thiols with GSH, the authors identified fragment **5.2** which labelled CDK2 with a 5-fold rate enhancement versus GSH. This demonstrated that this assay platform (coined quantitative irreversible tethering (qIT)) could be used to identify covalent fragment hits against protein targets that contained a surface-exposed cysteine.

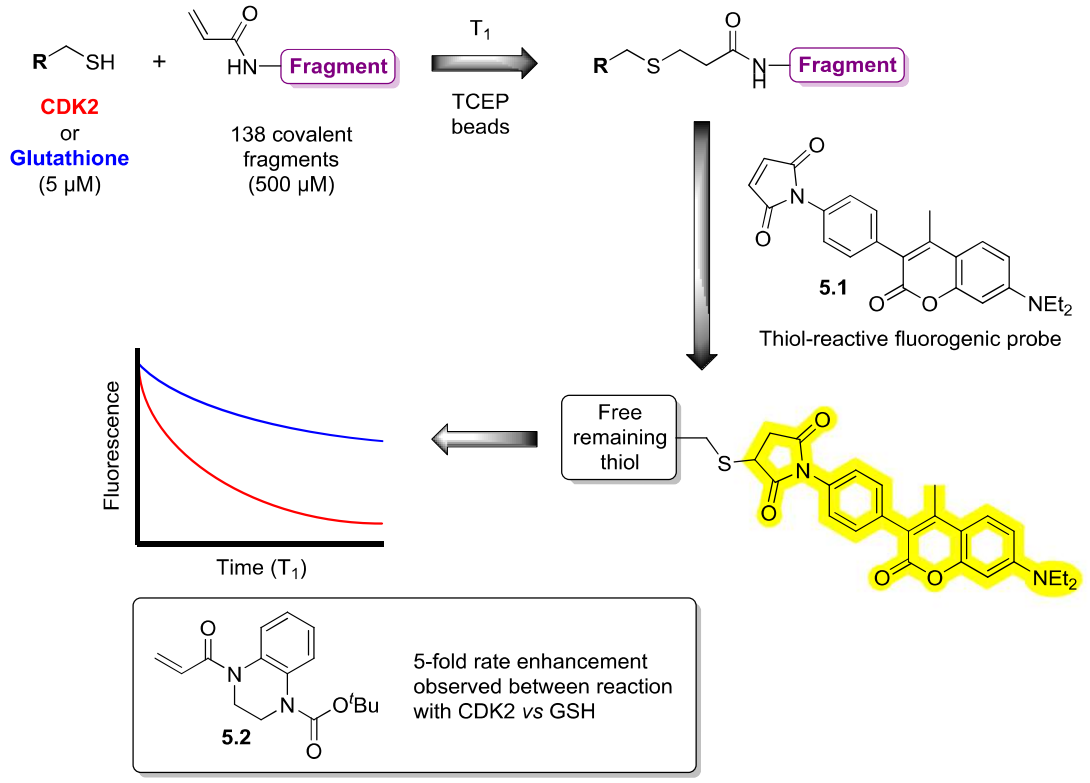


Figure 5.3 Quantitative irreversible tethering (qIT) assay platform developed by Craven *et al.*[78] A library of 138 covalent fragments (500 μM) were incubated with CDK2 protein or glutathione (5 μM). At various timepoints, excess **5.1** was added to react with any remaining free thiol and produce a fluorescent readout. The rate of reaction between the fragment and protein or GSH was compared and fragments that showed kinetic selectivity (>3-fold rate enhancement) for CDK2 were identified as hits (*e.g.* **5.2**).

One disadvantage of using residue-specific covalent modification is that only proteins containing those residues within or proximal to a binding site are applicable protein targets. Photoaffinity fragments offer a complementary approach, whereby the relatively unbiased nature of PAL capture versus electrophilic trapping can be used to target a wider range of protein families.

5.1.1.1 Application of PAL to FBDD

Cravatt and co-workers recently demonstrated that fully-functionalised photoaffinity fragments (FFFs) containing an affinity function, photoreactive group and bio-orthogonal handle could capture specific target proteins in live cells (Figure 5.4).[18] The group treated cells with 14 FFFs, and found varied proteomic interactions for each fragment *versus* an *N*-methyl amide control fragment **5.3**. Of the proteins that were

identified, prostaglandin reductase 2 (PTGR2) and the mitochondrial acylcarnitine transporter SLC25A20 were selected for follow up studies. Elaborated small molecule binders based on the initial hit fragments (**5.4** and **4.13**) were developed, resulting in the identification of selective chemical probes for both enzymes (**5.5** and **4.12** respectively), which had no previously reported chemical tools. This work demonstrated that selective tool compounds can be developed from initial screening of FFFs in live cells. One caveat to this approach is that it is target agnostic, whereby the target proteins of interest were chosen based on the levels of selectivity shown by the FFFs after the screen was performed. Screening FFFs in live cells with the aim of identifying fragment hits for a pre-determined target protein of interest would be impractical with current MS-based proteomic technologies. Thus, this approach has potential to be optimised further to suit the requirements of target-based drug discovery where the desired target protein is known.

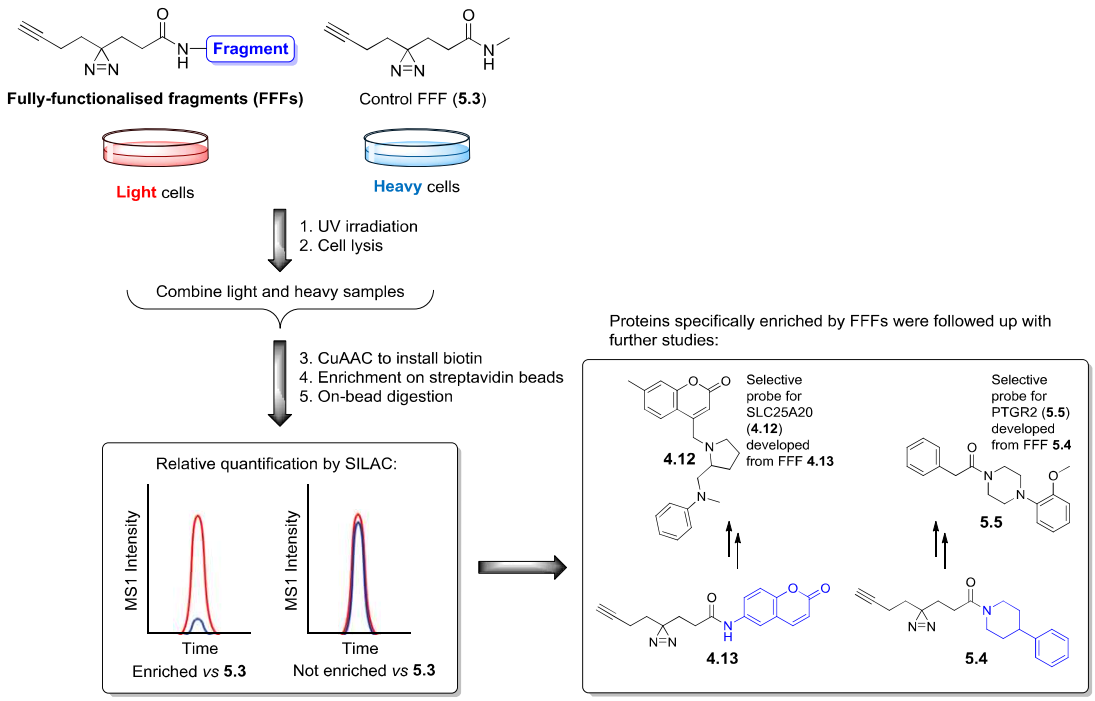


Figure 5.4 Fragment-based screening in live cells demonstrated by the Cravatt group. Fully-functionalised fragments (FFFs) were incubated in live cells and irradiated. MS-based proteomics with SILAC quantification was used to identify proteins enriched by FFFs versus a control FFF **5.3**. FFFs **4.13** and **5.4** showed moderately selective enrichment of SLC25A20 and PTGR2 respectively. Competitive screening of an elaborated small molecule library against these FFFs identified **4.12** and **5.5** as new selective chemical tools for SLC25A20 and PTGR2, demonstrating that new chemical probes could be discovered from initial photoaffinity-based fragment screening in live cells.

Inspired by this work, a more target-based approach was envisaged where a library of simplified photoaffinity fragments could be screened by irradiation in the presence of a recombinant target protein of interest (Figure 5.5). Any covalent modification to the target protein could be observed by intact protein LCMS, thus reporting on the affinity of the fragment to the target protein prior to irradiation. As this workflow would not involve a proteomic pulldown step, a pendant alkyne was not necessary, allowing the photoaffinity fragment to retain more fragment-like properties *versus* the fully-functionalised fragments (FFFs). These simplified photoaffinity fragments were termed "Photoaffinity Bits" or "PhABits". This approach offers a new methodology to find fragment hits, complementary to the methods introduced in Table 5.1.

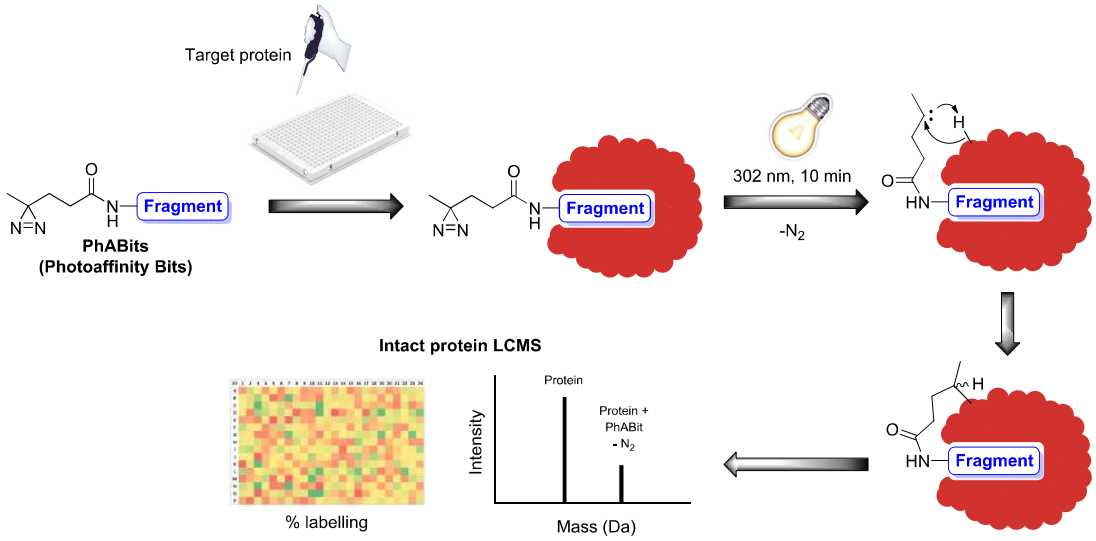


Figure 5.5 Proposed workflow for PhABit screening with recombinant proteins of interest. PhABits are incubated with recombinant protein (single compound per well) and irradiated (302 nm, 10 min). Covalent modification is observed by intact protein LCMS. The high-throughput intact protein LCMS workflow developed in Chapter 3 (Section 3.2.6) is applicable to this work (384-well format, 5.75 min/sample).

5.2 Proof-of-concept studies for PhABit screening

In a preliminary study, ten BET-targeting PhABits were synthesised and their labelling efficiency to BRD4 BD1 was examined by intact protein LCMS. BRD4 BD1 was chosen as a proof-of-concept protein so that any PhABit hits which showed potency by PhABit screening could be cross-validated in a TR-FRET assay.

5.2.1 Synthesis of PhABits

Ten amine-containing fragments containing known BET binding motifs (*e.g.* acetyl-lysine mimetics) were selected (Figure 5.6).[232,280,281] As a non-BET targeting control, benzylamine was chosen as an eleventh fragment. To minimise the effect on the fragment's binding affinity and physicochemical properties, the aliphatic diazirine **1.64** was chosen as the photoreactive group due to its low molecular weight. HATU-mediated amide coupling was used to synthesise 11 PhABits from the selected amine fragments with moderate yields (Figure 5.6).

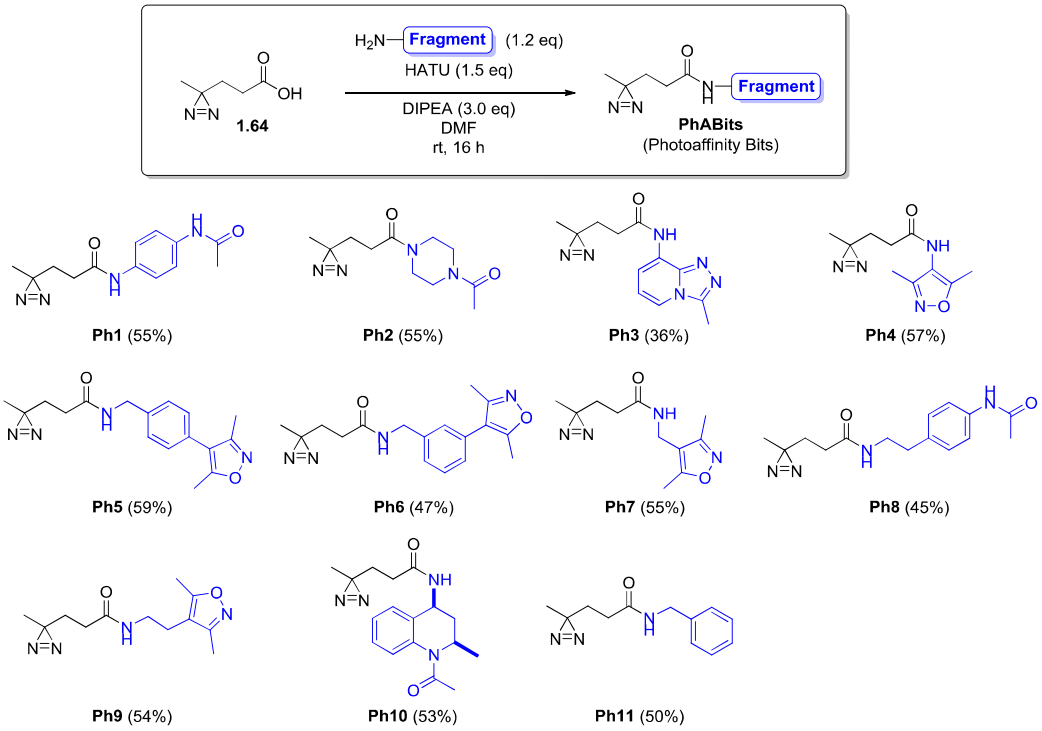


Figure 5.6 HATU-mediated amide coupling was used to synthesise ten PhABits from amine fragments containing known BET binding motifs. An additional PhABit was synthesised from benzylamine as a non-BET binding control.

5.2.2 Single high concentration PhABit screening with recombinant protein

The 11 PhABits were incubated with BRD4 BD1 at a single high concentration (1 mM). The samples were irradiated (302 nm, 10 min) and the percentage labelling was determined by intact protein LCMS (Figure 5.7). At this high concentration, single labelling by all 11 PhABits was observed (2–32%), and no double labelling was

observed. No labelling was observed in absence of PhABit (DMSO control). All PhABits except **Ph3**, **Ph5** and **Ph6** showed similar levels of labelling to the benzylamine control **Ph11** (*ca*. 5%). This suggested these PhABits gave a non-specific background level of photocrosslinking at 1 mM. Hit PhABits **Ph3**, **Ph5** and **Ph6** showed elevated levels of labelling over the 5% background observed. **Ph3** showed the highest level of single labelling (32%).

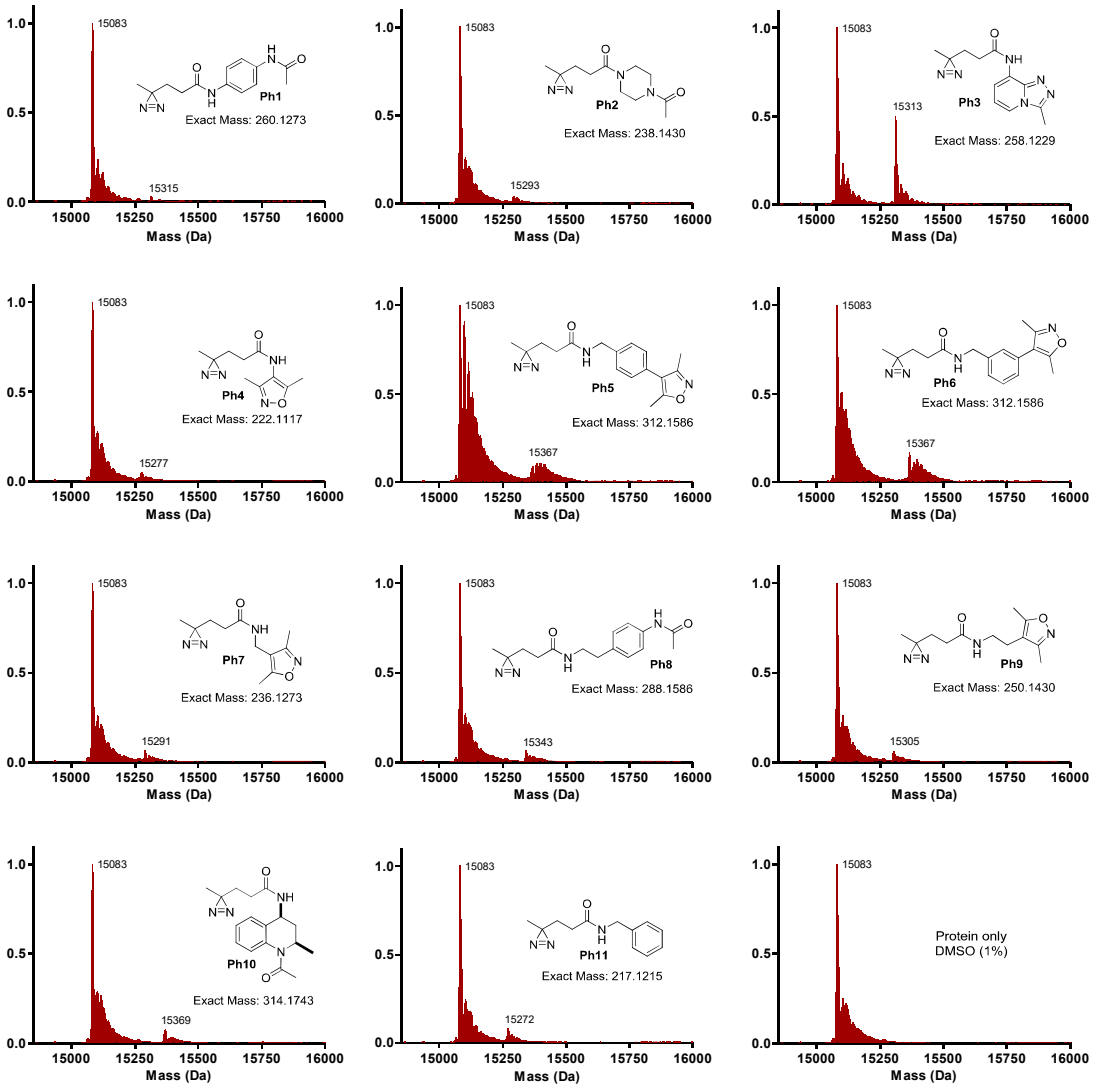


Figure 5.7 Intact protein LCMS analysis of BRD4 BD1 (3 μM) with PhABit (1 mM) after irradiation (302 nm, 10 min). **Ph3** showed the highest levels of labelling (32%). **Ph5** and **Ph6** showed high levels of protein oxidation, which could be due to fragment binding but inefficient covalent labelling.

PhABits **Ph5** and **Ph6** showed higher levels of protein oxidation (repeated +16 species, see Section 2.3.9). These two PhABits may have been bound to the active site at the

time of irradiation, but with the diazirine photoreactive group positioned in an orientation that disfavoured covalent labelling and favoured the generation reactive oxygen species in a high local concentration by reaction with the solvent or buffer components. This single shot experiment identified **Ph3**, **Ph5** and **Ph6** as hits, however further dose response experiments were required to distinguish between labelling efficiency and fragment affinity. Subsequently, the dose-response relationship between PhABit concentration and percentage labelling was examined.

5.2.3 Dose-response PhABit screening with recombinant protein

To investigate the relationship between PhABit concentration and percentage labelling, dose-response studies were conducted. BRD4 BD1 (3 μM) was incubated with a 6-point serial dilution of PhABit (1000–4.1 μM, with a DMSO control). The samples were irradiated (302 nm, 10 min) and the percentage labelling was determined by intact protein LCMS. Figure 5.7 shows an example of the deconvoluted mass spectra obtained at each concentration of **Ph3**. The average percentage crosslinking values (n=3) for each PhABit are reported in Table 5.2 (A). Based on this dose-response data, it was possible to suggest a more suitable concentration for initial single shot screening to identify hit PhABits. At 333 μM, labelling was observed for all 11 PhABits, whereas only the more potent PhABits showed appreciable levels of labelling at 111 μM (>1%). This suggests that 100–200 μM PhABit would be a more suitable single shot concentration to identify initial hits in future screening experiments.

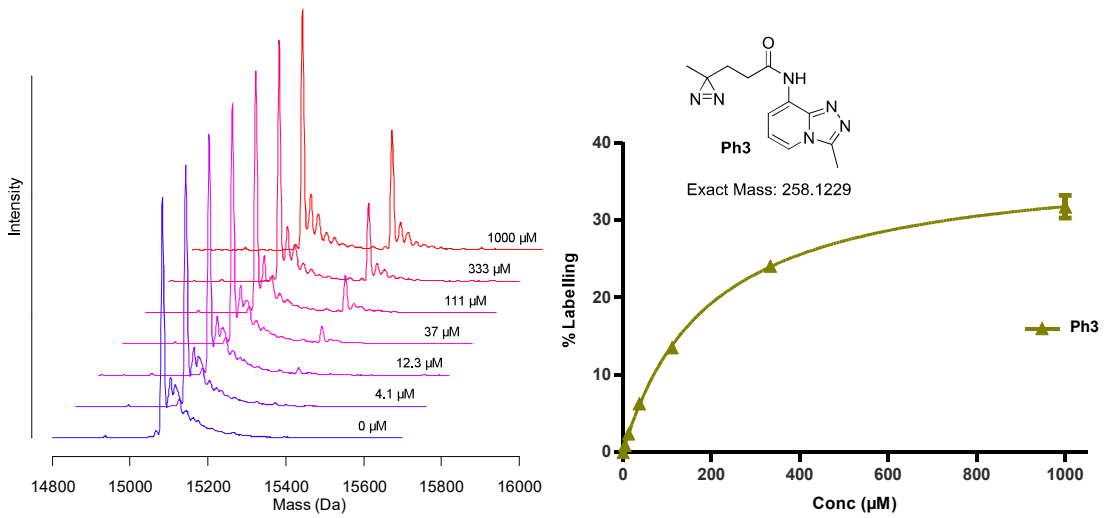


Figure 5.8 Example of dose-response relationship between PhABit **3** concentration and percentage labelling.

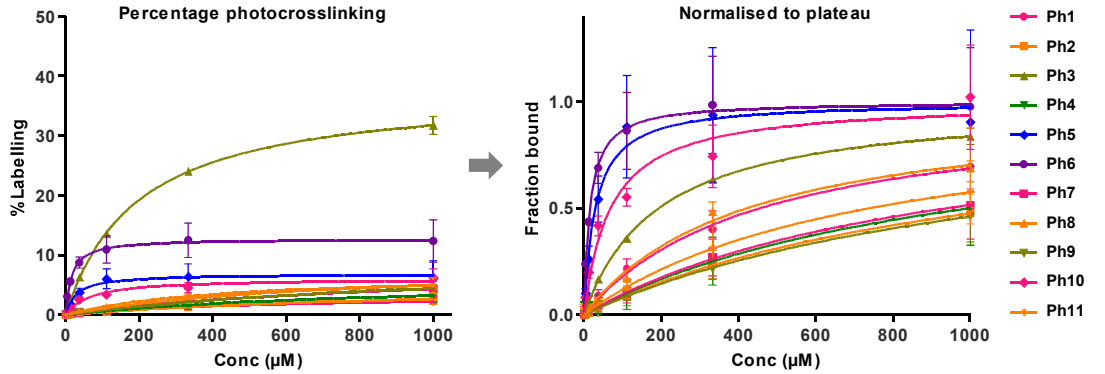


Figure 5.9 Percentage labelling for each PhABit obtained over a range of concentrations (n=3). The data was fitted using a non-linear regression (one-site specific binding, GraphPad) to obtain Y_{max} values for percentage labelling. These values were used to normalise the data to $Y_{max} = 1$ to represent the fraction of PhABit bound to the active site at the time of irradiation. K_d values for each PhABit were obtained and reported in Table 5.2.

The percentage labelling for each PhABit was plotted against concentration (Figure 5.9, left) and fitted using a non-linear regression (one-site specific binding, GraphPad) to obtain a predicted maximum in percentage labelling (Y_{max}). This value was then used to normalise the data ($Y_{max} = 1$) to represent the fraction of PhABit bound to the active site during irradiation (Figure 5.9, right). From these plots, pK_d values were calculated for each PhABit (-log[concentration] when Y = 0.5). These data are reported in Table 5.2 (B).

Table 5.2 (A) Percentage single labelling for each PhABit over a range of concentrations (n=3). (B) pK_d values were obtained from the binding curves shown in Figure 5.9 (right). (C) Affinities of the 11 PhABits for BRD4 BD1 were measured in a TR-FRET assay.

PhABit		**Ph1**	**Ph2**	**Ph3**	**Ph4**	**Ph5**	**Ph6**	**Ph7**	**Ph8**	**Ph9**	**Ph10**	**Ph11**
(A)		% Labelling in PAL experiment:										
Conc (μM) ↓	1000	2.2	2.6	31.7	3.2	6.2	12.4	4.3	4.8	4.4	6.2	4.9
	333	1.3	1.3	24.1	1.6	6.4	12.5	2.2	3.4	2.1	4.5	2.7
	111	0.7	0.5	13.6	0.6	6.0	10.9	0.8	1.2	0.9	3.4	1.0
	37	0.3	0.2	6.3	0.3	3.7	8.7	0.4	0.4	0.4	2.5	0.4
	12	0.1	0.0	2.4	0.0	1.8	5.5	0.2	0.2	0.0	1.2	0.2
	4	0.0	0.0	0.9	0.0	0.7	3.0	0.1	0.0	0.0	0.5	0.1
	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
(B)		Binding affinity from PAL experiment:										
pK_d		3.3	3.0	3.7	3.0	4.5	4.8	3.0	3.4	2.9	4.2	3.1
(C)		BRD4 BD1 TR-FRET assay:										
pIC_{50}		<4.3	<4.3	5.1	<4.3	4.8	5.2	<4.3	<4.3	<4.3	4.6	<4.3

Although **Ph3** produced the highest levels of labelling at 1 mM, the dose-response relationship indicated that **Ph3** had relatively low affinity for BRD4 BD1 (pK_d = 3.7). Despite their low levels of labelling efficiency, **Ph5**, **Ph6** and **Ph10** showed the highest affinity (pK_d = 4.5, 4.8 and 4.2 respectively). These results demonstrate the necessity for follow-up dose-response experiments to distinguish between high labelling PhABits and true fragment affinity. Once bound in the active site, the efficiency of the carbene insertion reaction with protein (*versus* quenching by water or buffer components) is dependent on the orientation of the diazirine, which is dependent on the binding mode of each PhABit. The efficiency of the insertion reaction is assumed to be fixed for each binding mode, therefore the observed percentage labelling is proportional to the occupancy of each PhABit in the active site. Preliminary single concentration screening (100–200 μM) could be used to identify PhABit hits that show labelling, and follow-up dose-response experiments could then be used to inform on the relative affinities of the hit PhABits.

To validate the relative affinities obtained in the PAL dose-response experiments, the affinities for each PhABit were measured in an orthogonal BRD4 BD1 TR-FRET assay (Table 5.2, C). **Ph5**, **Ph6** and **Ph10** showed affinity for BD1, however the affinities of remaining PhABits were below the lower limit of the assay, along with the benzylamine control **Ph11**. Pleasingly, for the three PhABits that were within the

range of the TR-FRET assay (**Ph5**, **Ph6** and **Ph10**), the affinities correlated well with the affinities derived from the photoaffinity experiments. This demonstrated that the PhABit platform could be used to identify fragment binders, and inform on their relative affinities to the target protein of interest.

5.3 Conclusions and future work

This preliminary proof-of-concept study demonstrated that the relative reversible binding affinities of PhABits to recombinant protein can be determined by MS-based PAL labelling experiments. This approach could be used to identify novel fragment binders, with a high dynamic range (≤1 mM). The relative potencies of hit PhABits could then be followed up with dose-response experiments. In future, the specificity of the PhABit labelling event should also be determined. To validate that the observed labelling is due to specific binding in the active site of BRD4 BD1, PhABit labelling should be reduced in a dose-dependent manner by the presence of a competitor (*e.g.* (+)-JQ1). It is envisaged that more extensive PhABit libraries will be synthesised and applied to proteins other than the BET family used in this work. Other coupling reactions could be used to synthesise PhABits and provide a variety of linker types (other than amides). This would also allow a wider access to fragment libraries containing functionalities other than amines.[107]

By combining the insights gained from the previous Chapters, a target-based hit identification cascade could be imagined (Figure 5.10). More extensive PhABit libraries could be screened against recombinant proteins of interest, and hits that showed labelling would be validated by competition experiments with known ligands for the active site. If little is known about the target protein or its binding site(s), and no chemical tools have been reported yet, the site of covalent modification could be determined by tryptic digestion and LC-MSMS studies as demonstrated in Section 2.3.8. Validated PhABit hits could then be used directly as PAL reporters in a photoaffinity displacement assay (Section 3.2.3), where traditional small molecule libraries could be screened to find more potent elaborated ligands for the target protein. The ability to use the initial PhABit hits immediately as PAL tools for the protein of interest has distinct advantages over other biochemical follow-up screening methods.

For example, to use fluorescent-based follow-up screening, a fluorescent tool molecule would first need to be synthesised. The relatively low amount of protein required (1 µM, 15 µL per PhABit) for this platform is one distinct advantage over NMR-based hit identification, which requires large amounts of soluble protein. The method is also more high-throughput than hit identification by x-ray crystallography. The ability to obtain structural information by tryptic digestion and LC-MSMS analysis is a distinct advantage which is not offered by SPR, TSA or biochemical screening methods (Table 5.1).

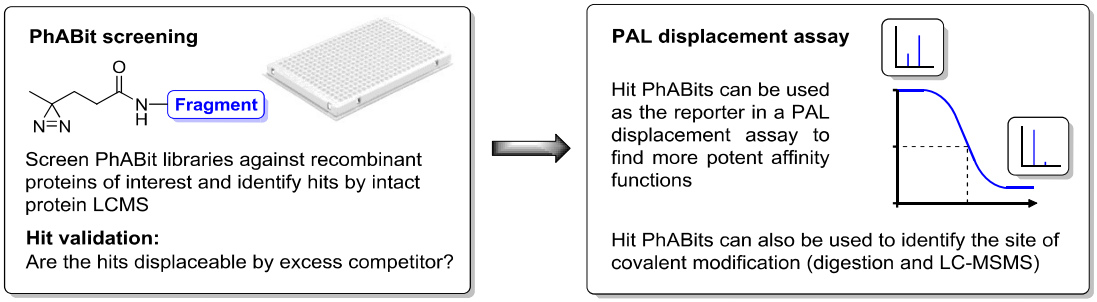


Figure 5.10 Future downstream applications of hits identified by PhABit screening. Any validated hits instantly become useful tools and chemical starting points for the target protein of interest. These can be directly used to identify the site of binding or employed in photoaffinity displacement assays to find more potent ligands.

One caveat of this approach is that PhABits that are bound in the active site but orientate the diazirine in a conformation that gives inefficient labelling (<1%) may be missed as false negatives. The low levels of labelling observed in this experiment was consistent with the performance of the alkyl diazirine in the photoreactive group comparisons conducted in Chapter 2. Other photoreactive groups (aryl azides, aryl diazirines or benzophenone) could be explored, however due to their large size and lipophilicity, the physicochemical properties of the fragment are likely to be heavily perturbed by the introduction of these photoreactive groups. This highlights the need for new or improved minimal photoreactive groups that offer increased labelling efficiency with minimal impact on the binding mode and physicochemical properties of the affinity function.

The analytical component of the PhABit workflow could also be improved. For example, PhABit libraries could be screened against multiple recombinant proteins of interest within the same sample (as demonstrated by the dual-domain PAL screening

in Chapter 3). Faster sampling technologies such as Rapidfire MS (*ca*. 8 s per sample) or RapiFlex MALDI MS (*ca.* 0.3 s per sample) may also be applicable to increase throughput.[240,241]

In conclusion, the preliminary proof-of-concept studies performed in this work demonstrate that PhABit hits can be identified by MS-based screening. The relative potencies of the PhABit hits to the target protein can be determined in follow-up dose response experiments. This new PhABit platform has an exciting potential to be optimised and expanded further to become complementary to existing hit identification technologies.

Chapter 6 Final remarks

From these investigations into the application of PAL to enhance small molecule target-based drug discovery, a number of key advancements have been made.

Chapter 2 described an Ugi multicomponent reaction protocol, which can be used to synthesise a variety of PAL probes from amine-functionalised compounds of interest. The procedure showed robust tolerance for various handles and commercially available photoreactive groups, which could all be employed in parallel synthesis to give a wide array of PAL probes in one step. Addition of the PAL scaffolds to the parent compound had only a moderate effecct on the physicochemical properties, and importantly, cellular activity was maintained. A series of investigations into the requirements for efficient labelling followed, where various photoreactive groups, linker lengths, and irradiation wavelengths were investigated. The effect of protein topology surrounding the active site was also considered, aided by X-ray crystallography and LC-MSMS analyses.

The conclusion from these studies was that efficient labelling is highly sensitive to all the above parameters, and pre-designing highly efficient probes is unfeasible due to this complexity. Instead, a variety of PAL probes should be synthesised in parallel and tested using the methods described here. This approach may be particularly advantageous when applied to new protein target families, of which less is known relative to the BET family.

Another key finding was the observed instability of isolated probes containing the aryl azide photoreactive groups and the alkyne handle. Probes containing these two groups have been reported in the literature. Thus, care should be taken with the interpretation of those results and future use should be avoided.

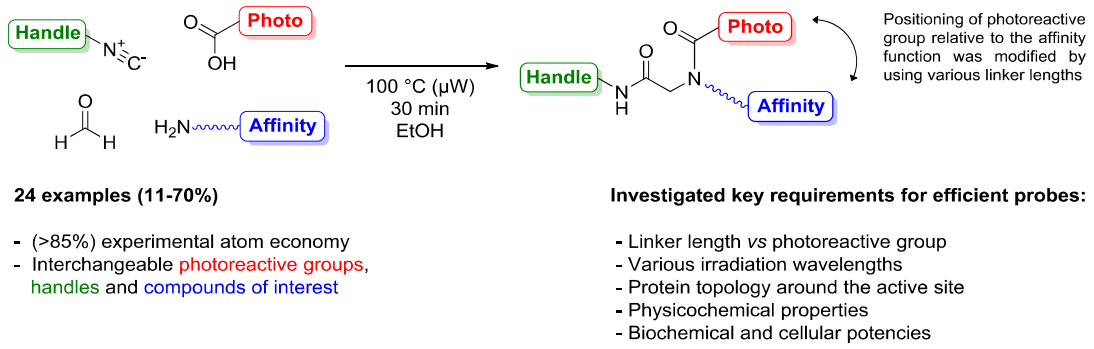


Figure 6.1 An Ugi reaction protocol for the synthesis of PAL probes was presented in Chapter 2. The procedure furnishes highly optimisable PAL probes from amine-functionalised compounds of interest in one step. An array of probes containing different photoreactive groups and linker lengths were synthesised in parallel. The requirements for efficient PAL labelling were explored in detail, and it was found that predicting the efficiency of PAL probes prior to synthesis is extremely difficult given the combination of factors that determine successful covalent labelling. Conclusions from these studies suggest synthesising and testing an array of PAL probes using the above procedure to quickly identify optimal probes.

Chapter 3 described the development of a novel biochemical assay, where a highly efficient PAL probe **2.45** was used as a reporter to inform on the binding affinities for other competitor compounds (Figure 6.2). During irradiation, a snapshot of the equilibrium between the target protein, PAL probe and competitor compound is taken, and the extent of labelling (reflective of the active site occupancy of the PAL probe) can be determined by intact protein LCMS. The ability of the competitor to displace the PAL probe from the active site prior to irradiation can be measured over a range of concentrations, forming the basis for a dose-response assay. The readout for this assay was optimised to reasonable throughput (384-well format, 5.75 min/sample) using intact protein LCMS. One key advantage of this assay *versus* other assay platforms (e.g. fluorescence-based), was that binding to multiple proteins could be analysed from the same sample. By using intact protein LCMS as the readout method, two-dimensional resolution of the target proteins was achieved, by both LC retention time and mass. BRD4 BD1 and BD2 were used as proof-of-concept target proteins for these studies. To find an appropriate test set of compounds to screen, the CHEMBL database was mined for compounds annotated with BET activity. From this list, the compounds that were available from the GSK compound collection were chosen (264 compounds) and screened at a single concentration against BD1 and BD2 in this newly developed assay. Those compounds that showed ≥ 60% displacement of the PAL probe (44 compounds) were progressed to full curve dose-response studies. The pIC_{50}

values from this dual-domain PAL assay correlated well with pIC_{50} values obtained in TR-FRET assays for BD1 and BD2. Thus, this new PAL biochemical screening assay has exciting potential for future optimisation and application. For example, compound binding to more than two domains may be analysed in tandem. Furthermore, higher-throughput mass spectrometry sampling techniques, such as MALDI or Rapidfire^{TM} sampling may reduce sampling times from minutes to seconds.[240,241]

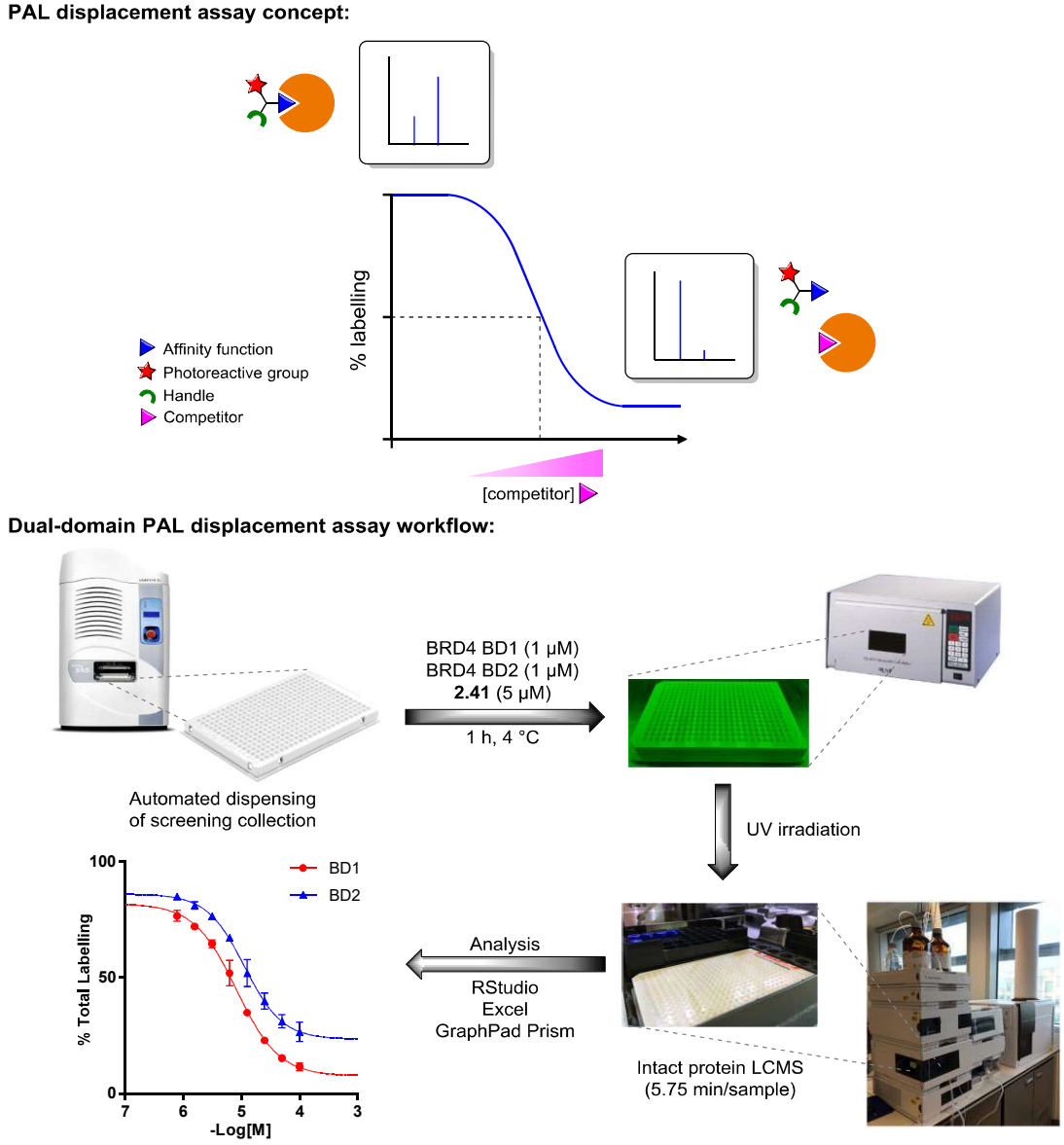


Figure 6.2 The dual-domain PAL displacement assay developed in this work. The percentage labelling by the PAL probe correlates with the fraction of PAL probe bound to the active site during irradiation. Therefore, the affinity of competitor compounds can be inferred from their dose-dependent displacement of the PAL probe from the active site. A test set of known BRD4 inhibitors obtained from the CHEMBL database were screened in this assay and their pIC_{50} values correlated well with TR-FRET data.

Chapter 4 described the development of a CuAAC-enabled proteomic pulldown workflow, which was required for the Ugi-derived PAL probes synthesised in Chapter 2 to be used in live-cells. The workflow was developed by re-designing existing TCO-based workflows and incorporating a variety of CuAAC ligation procedures reported in the literature. Key to the success of these optimisations was the development of a quick in-gel fluorescence readout using HL-60 lysate spiked with recombinant target protein (BRD4 BD1). This allowed fast readouts from early optimisation experiments (hours), *versus* full MS-based proteomic readouts (2 weeks). Once the key parameters of the CuAAC-based workflow were established, three PAL probes containing an alkyl diazirine (**4.3**), trifluoromethyl aryl diazirine (**4.4**) and benzophenone (**4.5**) photoreactive group were employed in HL-60 lysates, and then in live HL-60 cells. The alkyl diazirine probe failed to robustly capture the BET target proteins in live cells, which correlated with the poor levels of labelling observed with recombinant BRD4 BD1 and BD2. The alkyl diazirine has become a popular photoreactive group in the literature due to its small size and thus minimal modification to the parent compound, however if this probe was the only one to be synthesised and tested here, the BET family of target proteins would not have been identified. Pleasingly, the remaining two probes containing the trifluoromethyl aryl diazirine and benzophenone photoreactive groups (**4.4** and **4.5**) both robustly identified all three members of the BET family (BRD2, 3 and 4). Furthermore, these enrichments were found to be due to selective binding of the affinity function, as these enrichments could be competed by the presence of excess parent compound **4.2**. The observed variation in proteomic profiles from each probe directly demonstrated the key message learned from Chapter 2 – that due to the complex requirements needed to achieve high labelling, a range of probes should be synthesised and tested. Robust identification by two or more different photoreactive groups along with the appropriate competition experiments provides increased confidence that the observed profiles more accurately reflect the proteins engaged by the parent compound in live cells.

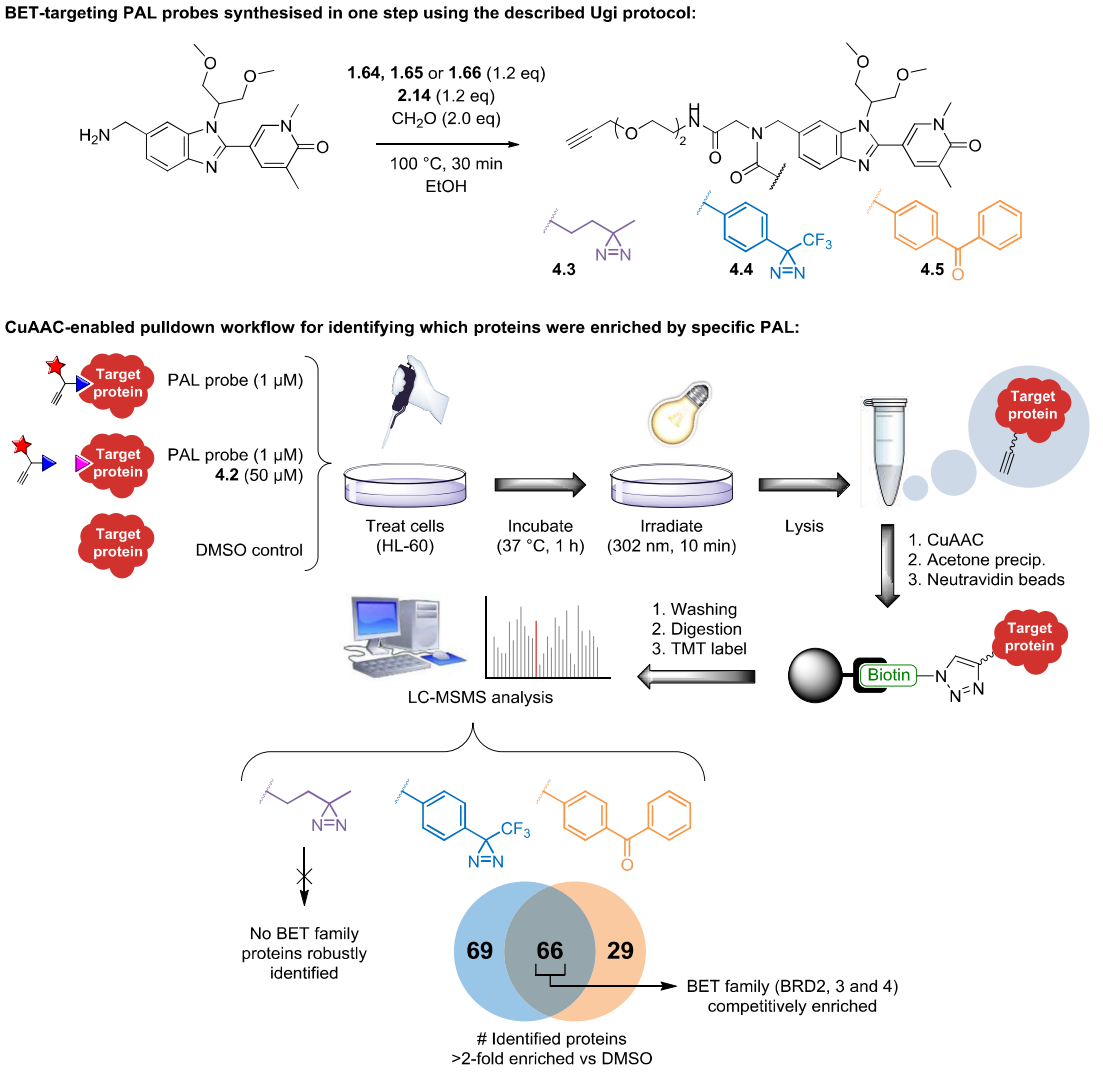


Figure 6.3 Probes synthesised using the one-step Ugi protocol could be used directly in live-cell MS-based proteomics to inform on the protein profile for the parent compound of interest.

Chapter 5 described preliminary investigations into whether or not photoaffinity fragments can be used with recombinant protein to determine the binding affinity of the fragment component. Inspired by work from the Cravatt group, 10 photoaffinity fragments were synthesised containing known BET-binding motifs and photocrosslinking experiments to recombinant BRD4 BD1 were conducted. The percentage labelling for each fragment was found to be concentration dependent and binding curves were plotted to obtain K_d values for each fragment. Gratifyingly, these values showed good correlation with the activities for each fragment obtained in a BRD4 BD1 TR-FRET assay. Expansion of this study is envisaged, where elaborated photoaffinity fragment libraries can be screened against recombinant protein and hits

identified using intact protein LCMS. A unique and attractive advantage of this PAL fragment approach is that any photoreactive fragment hits can be immediately used as photoaffinity tool compounds for the recombinant protein of interest. For example, the hit could be used in a PAL-based displacement assay, analogous to the assay presented in Chapter 3 to find more potent binding compounds for the protein of interest.

In conclusion, among the various chemical biology approaches being used to address the issues faced in small molecule drug discovery, PAL offers complementary approaches to increase our capabilities regarding fragment-based hit finding, biochemical screening and identifying which protein targets are engaged by compounds of interest. Further optimisation of the three approaches presented in this work are envisaged, along with their exciting application to new families of target proteins.

Chapter 7 Experimental methods

7.1 Synthetic chemistry

7.1.1 General synthetic methods

All reagents were used as received from commercial sources unless otherwise stated. Anhydrous solvents were used in all synthesises and used as purchased unless otherwise stated.

7.1.1.1 Nuclear Magnetic Resonance (NMR) spectroscopy

Proton Nuclear Magnetic Resonance spectra were recorded on Bruker AV-400 (^{1}H = 400 MHz, ^{13}C = 101 MHz) or AV-600 (^{1}H = 600 MHz, ^{13}C = 151 MHz) spectrometers in the stated deuterated solvent and referenced either to residual undeuterated solvent or 0.03% (*v/v*) trimethylsilane (TMS). Chemical shifts are reported in parts per million (ppm) and coupling constants (*J*) are reported in Hertz (Hz). ^{13}C experiments were recorded with total proton decoupling. HSQC, HMBC and COSY experiments were used to aid assignments when required. All NMR experiments were performed at 30 °C unless otherwise stated.

7.1.1.2 Liquid Chromatography-Mass Spectrometry (LCMS)

LCMS was used to monitor reaction progress and analyse final compound purity using one of three solvent systems stated below. For these methods, samples were prepared in methanol unless otherwise stated. The chromatography was carried out on an Aquity UPLC CSH C-18 column (internal diameter: 50 mm × 2.1 mm, packing diameter: 1.7 µm) at 40 °C with a 0.5 µL injection volume. The UV detection was a summed signal from wavelengths between 210 nm and 350 nm. For compounds with weak UV absorbance, evaporative light scattering detection (ELSD) was used (Polymer Laboratories, PL-ELS 1000). Mass detection was performed with alternate-scan positive and negative electrospray on a Waters QDa instrument, with a scan range of

100–1000 Da or 100–1200 Da (high mass range method). Scan time was 0.27 s with an inter-scan delay of 0.10 s.

7.1.1.3 LCMS using an acidic modifier (Method A)

Sample was eluted using a gradient shown in Table 7.1 with a flow rate of 1.0 mL min^{-1}.

Solvent A: 0.1% (*v/v*) solution of formic acid in water.

Solvent B: 0.1% (*v/v*) solution of formic acid in acetonitrile.

Table 7.1 Low pH gradient for LCMS analysis

Time (min)	A (%)	B (%)
0	97	3
1.5	5	95
1.9	5	95
2.0	97	3

7.1.1.4 LCMS using a basic modifier (Method B)

Sample was eluted using a gradient shown in Table 7.2 with a flow rate of 1.0 mL min^{-1}.

Solvent A: 0.1% (*v/v*) 10 mM ammonium bicarbonate in water adjusted to pH 10 with ammonia solution.

Solvent B: 0.1% (*v/v*) ammonia in acetonitrile.

Table 7.2 High pH gradient for LCMS analysis

Time (min)	A (%)	B (%)
0.00	97	3
0.05	97	3
1.5	5	95
1.9	5	95
2.00	97	3

7.1.1.5 LCMS using a trifluoroacetic acid (TFA) modifier (Method C)

Sample was eluted using a gradient shown in Table 7.3 with a flow rate of 1.0 mL min^{-1}.

Solvent A: 0.1% (*v/v*) solution of trifluoroacetic acid in water.

Solvent B: 0.1% (*v/v*) solution of trifluoroacetic acid in acetonitrile.

Table 7.3 TFA gradient for LCMS analysis

Time (min)	A (%)	B (%)
0.00	95	5
1.5	5	95
1.9	5	95
2.0	95	5

7.1.1.6 High Resolution Mass Spectrometry (HRMS)

High-resolution mass spectra were recorded on a Micromass Q-ToF Ultima hybrid quadrupole (Q) time-of-flight (ToF) mass spectrometer, with analytes separated on an Agilent 1100 Liquid Chromatography stack equipped with a Phenomenex luna C18 reverse-phase column (100 mm × 2.1 mm, 3 µm packing diameter). Samples were eluted using a gradient shown in Table 7.4 with a flow rate of 0.5 mL min^{-1}. The column was housed at 35 °C. The injection volume was 2–5 µL.

Solvent A: 0.1% (*v/v*) solution of formic acid in water.

Solvent B: 0.1% (*v/v*) solution of formic acid in acetonitrile.

Table 7.4 Liquid chromatography conditions for high resolution mass spectrometry

Time (min)	A (%)	B (%)
0.00	95	5
6.00	0	100
8.50	0	100
9.50	95	5
12.00	95	5

7.1.1.7 Normal phase preparative chromatography

Automated column chromatography was conducted on a Teledyne Isco Combiflash Rf system using RediSep Rf Silica cartridges (for normal phase) using HPLC grade solvents.

7.1.1.8 Mass Directed Automated Preparative HPLC (MDAP)

MDAP purification was conducted on a Xselect CSH C18 column (internal diameter: 150mm × 30mm, packing diameter: 5 µm) at ambient temperature, eluting with an appropriate solvent system determined by prior LCMS analysis (method A, B or C as described above). Mass detection was achieved by alternate-scan electrospray ionisation with a range of 150–1000 Da. The scan time was 0.5 s with an inter-scan delay of 0.2 s.

7.1.1.9 Infrared Spectroscopy

Infrared spectra were obtained on a Perkin Elmer Spectrum 100 FT-IR spectrometer with an ATR sampling accessory.

7.1.1.10 Melting Point

All melting points were recorded on a Stuart SMP40 melting point apparatus.

7.1.2 General synthetic procedures

7.1.2.1 General Procedure 1: Ugi reaction to synthesise PAL probes

A vial was loaded with amine affinity function (0.1 mmol), photoreactive carboxylic acid (0.12 mmol), paraformaldehyde (0.2 mmol), isonitrile handle (0.12 mmol) and ethanol (1 mL). The vial was capped and heated by microwave irradiation to 100 °C for 30 min. The reaction mixture was purified directly by mass-directed preparative reverse phase chromatography. The fractions containing desired product were combined and concentrated *in vacuo* to give the desired PAL probe.

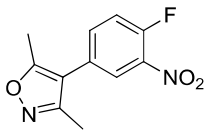
7.1.2.2 General Procedure 2: Synthesis of PhABits

A vial was loaded with amine PhABit precursor obtained from GSK compound collection (0.24 mmol) and a pre-made mixture of 3-(3-methyl-3*H*-diazirin-3-yl)propanoic acid (26 mg, 0.20 mmol), HATU (114 mg, 0.300 mmol) and DIPEA (0.1 mL, 0.6 mmol) in *N,N*-dimethylformamide (0.5 mL) and sealed. The reaction mixture was stirred at rt for 16 h. The reaction mixture was diluted with methanol (0.5 mL) and this mixture was purified by MDAP (Method B) to give the appropriate PhABit.

7.1.3 Chapter 2 – synthetic chemistry

Compound 2.5

4-(4-Fluoro-3-nitrophenyl)-3,5-dimethylisoxazole



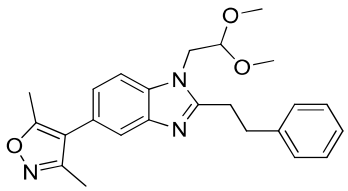
A RBF was loaded with 4-bromo-1-fluoro-2-nitrobenzene (10.7 g, 48.6 mmol), $PdCl_2$(dppf) (3.56 g, 4.86 mmol), 1,2-dimethoxyethane (100 mL), 3,5-dimethyl-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)isoxazole (14.1 g, 63.2 mmol) and the mixture was stirred while sat. aq. $NaHCO_3$ (100 mL) was added. The reaction mixture was heated to 80 °C for 4 h under nitrogen. The reaction mixture was allowed to cool to rt and partitioned between ethyl acetate (*ca.* 200 mL) and water (*ca*. 200 mL). The organic layer was collected, and the remaining aqueous layer was extracted with ethyl

acetate (3 × 100 mL). The organic layers were combined and concentrated *in vacuo*. The residue was dry-loaded on silica and this mixture was purified by flash chromatography (0–100% (*v/v*) ethyl acetate in cyclohexane over 12 CV). The fractions containing desired product were combined and concentrated *in vacuo.* The residue was triturated with diethyl ether (2 × *ca*. 10 mL) and the remaining solid was oven-dried to give 4-(4-fluoro-3-nitrophenyl)-3,5-dimethylisoxazole (9.00 g, 38.1 mmol, 78%) as a beige solid.

δ_H (400 MHz, $CDCl_3$):	7.96 (dd, J = 6.8, 2.3 Hz, 1H), 7.53 (ddd, J = 8.6, 4.3, 2.3 Hz, 1H), 7.41 (dd, J = 10.5, 8.5 Hz, 1H), 2.44 (s, 3H), 2.29 (s, 3H) ppm. Consistent with literature.[193]
LCMS (Method A):	t_R = 1.02 min, 86% by UV, $[M+H]^+$ found: 237.1.

Compound 2.8

4-(1-(2,2-Dimethoxyethyl)-2-phenethyl-1*H*-benzo[d]imidazol-5-yl)-3,5-dimethylisoxazole



4-(4-Fluoro-3-nitrophenyl)-3,5-dimethylisoxazole (8.89 g, 37.6 mmol) was dissolved in dimethyl sulfoxide (20 mL) and 2,2-dimethoxyethan-1-amine (6.0 mL, 55 mmol) was added. The mixture was heated at 80 °C for 3 h. The reaction mixture was allowed to cool to rt and methanol (80 mL) was added. 3-Phenylpropanal (13.7 mL, 113 mmol), water (100 mL) and sodium hydrosulfite (28.4 g, 188 mmol) were added and the reaction mixture was heated to 80 °C for 9 h. The reaction mixture was allowed to cool before being partitioned between ethyl acetate (150 mL) and 5% (*v/v*) aqueous ammonia solution (300 mL). The organic layer was collected, and the remaining aqueous layer was extracted with ethyl acetate (3 × 150 mL). The organics were combined and dried by passing through a hydrophobic frit before being concentrated *in vacuo*. The residue was dissolved in the minimum dichloromethane (*ca*. 10 mL) and eluted through a silica plug (*ca.* 100 g) using a gradient of 0–70% ethyl acetate in

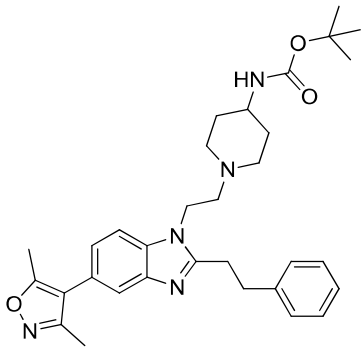
cyclohexane. The fractions containing desired product were combined and concentrated *in vacuo*. The residue was purified by flash chromatography (0–80% ethyl acetate in cyclohexane over 12 CV) to give 4-(1-(2,2-dimethoxyethyl)-2-phenethyl-1*H*-benzo[*d*]imidazol-5-yl)-3,5-dimethylisoxazole (8.11 g, 20.0 mmol, 53%) as a pale yellow solid.

δ_H (400 MHz, $CDCl_3$): 7.63 (d, J = 1.0 Hz, 1H), 7.40 (d, J = 7.6 Hz, 1H), 7.19-7.34 (m, 5H), 7.09-7.16 (m, 1H), 4.50 (s, 1H), 4.12 (d, J = 5.0 Hz, 2H), 3.35 (s, 6H), 3.19-3.32 (m, 4H), 2.42 (s, 3H), 2.30 (s, 3H) ppm. Consistent with literature.[193]

LCMS (Method A): t_R = 0.80 min, 100% by UV, $[M+H]^+$ found: 406.2.

Compound 2.10

Tert-butyl (1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)carbamate



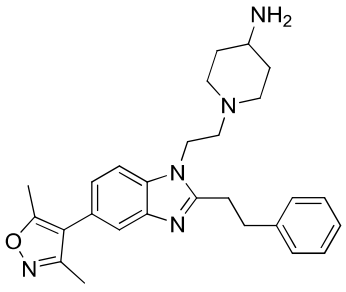
4-(1-(2,2-Dimethoxyethyl)-2-phenethyl-1*H*-benzo[d]imidazol-5-yl)-3,5-dimethylisoxazole (844 mg, 2.08 mmol) was dissolved in dichloromethane (5.0 mL) and water (2.5 mL) and trifluoroacetic acid (2.5 mL) were added. The reaction mixture was heated to 150 °C using microwave irradiation for 2 h. The reaction mixture was concentrated *in vacuo* and the residue was dissolved in tetrahydrofuran (10 mL) and acetic acid (0.5 mL). *Tert*-butyl piperidin-4-ylcarbamate (584 mg, 2.91 mmol) and sodium triacetoxyborohydride (617 mg, 2.91 mmol) were added and the reaction mixture was stirred at rt under nitrogen for 16 h. The reaction mixture was concentrated *in vacuo* and the residue was partitioned between ethyl acetate (*ca.* 50 mL) and sat. aq. sodium bicarbonate (*ca*. 100 mL). The organic layer was collected, and the remaining aqueous layer was extracted with ethyl acetate (2 × 25 mL). The

organics were combined and dried by passing through a hydrophobic frit before being concentrated *in vacuo*. The residue was purified by flash chromatography (0–20% (*v/v*) CH_3OH in CH_2Cl_2 over 12 CV). The fractions containing desired product were combined and concentrated *in vacuo*. The residue was purified by flash chromatography (0–100% (*v/v*) (1% (*v/v*) triethylamine/ethyl acetate) in (1% (*v/v*) triethylamine/cyclohexane) over 10 CV) to give tert-butyl (1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)carbamate (991 mg, 1.82 mmol, 88%) as a white gum.

δ_H (400 MHz, CD_3OD):	7.50-7.57 (m, 2H), 7.15-7.28 (m, 6H), 4.16 (t, *J* = 6.9 Hz, 2H), 3.16-3.30 (m, 5H), 2.75-2.85 (m, 2H), 2.53 (t, *J* = 6.8 Hz, 2H), 2.39 (s, 3H), 2.24 (s, 3H), 2.13 (td, *J* = 11.7, 2.3 Hz, 2H), 1.74-1.85 (m, 2H), 1.35-1.48 (m, 11H) ppm. Exchangeable proton not observed.
δ_C (101 MHz, CD_3OD):	166.8, 160.3, 157.9, 157.5, 143.6, 142.2, 135.6, 129.8, 129.7, 127.7, 125.6, 125.1, 120.0, 118.5, 111.7, 80.1, 58.0, 54.1, 48.9, 42.8, 35.3, 33.2, 30.5, 28.9, 11.6, 10.9 ppm.
LCMS (Method B):	t_R = 1.28 min, 96% by UV, $[M+H]^+$ found: 544.5.
HRMS:	($C_{32}H_{42}N_5O_3$) $[M+H]^+$ requires: 544.3288, $[M+H]^+$ found: 544.3284.
ν_{max} (neat, cm^{-1}):	2930, 2810, 2464, 1686, 1452, 1407, 1364, 1292, 1247, 1166, 1143, 1043, 980, 882, 807, 754, 701.

Compound 2.2

1-(2-(5-(3,5-Dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-amine

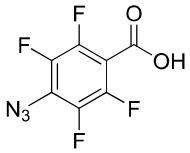


Tert-butyl (1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)carbamate (991 mg, 1.82 mmol) was dissolved in dichloromethane (9.1 mL) and trifluoroacetic acid (2.81 mL, 36.5 mmol) was added. The reaction mixture was stirred at rt for 3 h before being concentrated *in vacuo*. The residue was dissolved in methanol (*ca.* 5 mL) and loaded onto a SCX-2 10 g cartridge which was pre-equilibrated with methanol. The cartridge was then washed with methanol (*ca*. 200 mL) and then 2 M ammonia in methanol (*ca.* 200 mL). The basic wash was concentrated *in vacuo* to give 1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-amine (800 mg, 1.80 mmol, 99%) as a yellow oil.

δ_H (400 MHz, CD_3OD):	7.52-7.57 (m, 2H), 7.15-7.29 (m, 6H), 4.18 (t, *J* = 6.9 Hz, 2H), 3.16-3.30 (m, 4H), 2.78-2.87 (m, 2H), 2.56-2.66 (m, 1H), 2.53 (t, *J* = 6.9 Hz, 2H), 2.40 (s, 3H), 2.26 (s, 3H), 2.10 (td, *J* = 11.7, 2.4 Hz, 2H), 1.73-1.83 (m, 2H), 1.35 (qd, *J* = 11.9, 3.8 Hz, 2H) ppm. Exchangeable protons not observed.
δ_C (101 MHz, CD_3OD):	166.8, 160.3, 157.5, 143.6, 142.2, 135.6, 129.8, 129.7, 127.7, 125.6, 125.1, 120.0, 118.5, 111.7, 58.0, 54.1, 49.5, 42.8, 35.8, 35.3, 30.5, 11.6, 10.9 ppm.
LCMS (Method B):	t_R = 0.95 min, 98% by UV, $[M+H]^+$ found: 444.5.
HRMS:	($C_{27}H_{34}N_5O$) $[M+H]^+$ requires: 444.2763, $[M+H]^+$ found: 444.2762.
ν_{max} (neat, cm^{-1}):	3279, 2931, 2816, 1608, 1508, 1482, 1381, 1322, 1452, 1408, 1381, 1322, 1247, 1135, 1077, 1030, 886, 810, 720, 754, 701, 637, 526, 497.

Compound 1.16

4-Azido-2,3,5,6-tetrafluorobenzoic acid

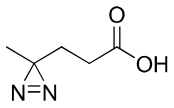


2,3,4,5,6-Pentafluorobenzoic acid (9.79 g, 46.2 mmol) was dissolved in dry methanol (150 mL) and sulfuric acid (3.00 mL, 55.4 mmol) was added dropwise with magnetic stirring under nitrogen at rt. The reaction mixture was concentrated under a flow of nitrogen to give a residue which was partitioned between ethyl acetate (200 mL) and saturated aqueous sodium bicarbonate solution (200 mL). The organic layer was collected, and the remaining aqueous layer was extracted with ethyl acetate (2 × 150 mL). The organics were combined and concentrated *in vacuo*. The residue was suspended in acetone (24 mL) and water (8.0 mL) and sodium azide (0.863 g, 13.3 mmol) was added. The reaction mixture was heated to reflux (80 °C) with magnetic stirring for 2 h. The reaction mixture was partitioned between water (100 mL) and diethyl ether (100 mL). The organic layer was collected, and the remaining aqueous layer was extracted with diethyl ether (3 × 50 mL). The extractions were combined and concentrated *in vacuo*. The residue was suspended in tetrahydrofuran (10 mL), water (10 mL) and 2.0 M sodium hydroxide (4.2 mL, 8.4 mmol). The reaction mixture was stirred at rt for 6 h. The reaction mixture was partitioned between dichloromethane (50 mL) and water (50 mL). The aqueous layer was collected and washed with dichloromethane (50 mL) before being acidified to pH *ca.* 1 using 2 M aqueous HCl. The aqueous layer was then extracted with dichloromethane (3 × 50 mL) and 5:1 (*v/v*) chloroform:2-propanol (3 × 50 mL). The extractions were combined and dried by passing through a hydrophobic frit before being concentrated *in vacuo* to give 4-azido-2,3,5,6-tetrafluorobenzoic acid (2.0 g, 8.5 mmol, 60%) as a white powder.

δ_F (376 MHz, CD_3OD):	-141.45 to -141.19 (m, 2F), -151.72 to -151.42 (m, 2F) ppm. Consistent with literature.[69]
LCMS (Method B):	t_R = 0.97 min, 93% by UV, $[M-H]^-$ found: 234.2.

Compound 1.64

3-(3-Methyl-3*H*-diazirin-3-yl)propanoic acid



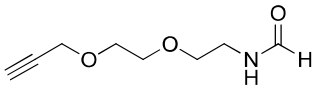
A flask was loaded with levulinic acid (3.65 g, 31.4 mmol) and 7 M ammonia in methanol (40 mL, 280 mmol) and the reaction mixture was stirred at 0 °C for 3 h.

(Aminooxy)sulfonic acid (5.33 g, 47.2 mmol) was added portion-wise and the reaction mixture was stirred under nitrogen for 40 h while allowed to warm to rt. Nitrogen was bubbled through the reaction mixture for 1 h and the mixture was filtered. The retentate was washed with methanol (2 × 20 mL) and the filtrate was concentrated *in vacuo*. The residue was dissolved in methanol (*ca*. 20 mL) and triethylamine (13 mL, 94 mmol) was added. The mixture was stirred for 5 mins before cooling to 0 °C. Iodine granules (7.98 g, 31.4 mmol) were added portion-wise at 0 °C under nitrogen in the dark with vigorous stirring until the dark colour remained. The mixture was then acidified with 2 M aqueous HCl (*ca*. 30 mL) and 10% (*w/v*) aqueous sodium thiosulfate (100 mL) was added. The mixture was then extracted with ethyl acetate (3 × 50 mL). The extractions were combined and dried by passing through a hydrophobic frit before being concentrated *in vacuo*. The residue was purified by flash column chromatography (0–20% (*v/v*) methanol in dichloromethane over 15 CV). The fractions containing desired product were combined and concentrated *in vacuo*. The residue was purified by flash column chromatography (0–50% (*v/v*) (3:1 (*v/v*) ethyl acetate:ethanol) in cyclohexane over 15 CV). The fractions containing desired product were combined and concentrated *in vacuo* to give 3-(3-methyl-3*H*-diazirin-3-yl)propanoic acid (675 mg, 5.27 mmol, 34%) as a colourless oil.

δ_H (400 MHz, $CDCl_3$):	9.08 (br s, 1H), 2.24 (t, J = 7.8 Hz, 2H), 1.73 (t, J = 7.8 Hz, 2H), 1.05 (s, 3H) ppm. Consistent with literature.[282]
δ_C (101 MHz, $CDCl_3$):	178.2, 29.4, 28.5, 25.0, 19.7 ppm. Consistent with literature.[282]

Compound 2.16

N-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)formamide



Sodium hydride (1.07 g, 26.8 mmol) was added to a dried 100 mL RBF and the flask was purged with nitrogen three times. Tetrahydrofuran (30 mL) was added and the mixture was cooled to 0 °C under nitrogen. *Tert*-butyl (2-(2-hydroxyethoxy)ethyl)carbamate (5.24 g, 25.5 mmol) was added dropwise over 10 min

and the mixture was stirred for 1 h at 0 °C. 3-Bromoprop-1-yne (3.41 mL, 30.6 mmol) was added dropwise over 10 min and the mixture was allowed warm to rt and stirred under nitrogen for 16 h. The solvent was removed in vacuo and the mixture was partitioned between ethyl acetate (100 mL) and water (100 mL). The organic layer was collected and washed with water (2 × 50 mL) before being dried by passing through a hydrophobic frit and concentrated *in vacuo*. The residue was purified by flash chromatography (0–50% (*v/v*) ethyl acetate in cyclohexane over 15 CV). The fractions containing desired product were combined and concentrated *in vacuo.* The residue was dissolved in dichloromethane (41 mL) and trifluoroacetic acid (31.7 mL, 412 mmol) was added. The reaction mixture was stirred at rt for 30 min before being concentrated *in vacuo*. The residue was dissolved in methanol (10 mL) and divided into aliquots (2 × 5 mL). Each aliquot was loaded onto a Biotage 50 g SCX-2 cartridge which was prewashed with methanol. The cartridges were washed with methanol (500 mL) followed by 2 M ammonia in methanol (250 mL). The basic washes were combined and concentrated *in vacuo*. The residue was suspended in ethyl formate (8.38 mL, 103 mmol) and the reaction mixture was heated to reflux (80 °C) for 16 h. The reaction mixture was concentrated *in vacuo* and the residue was purified by flash chromatography (0–100% (*v/v*) (3:1 (*v/v*) ethyl acetate:ethanol) in cyclohexane over 12 CV); 40 mL fractions. The fractions containing desired product (R_f 0.59 with 75% (3:1 (*v/v*) ethyl acetate:ethanol) in cyclohexane on silica TLC with $KMnO_4$ staining) were combined and concentrated *in vacuo* to give *N*-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)formamide (3.51 g, 20.5 mmol, 81% (three steps)) as a yellow oil.

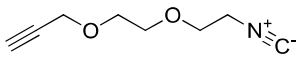
δ_H (400 MHz, CD_3OD):	7.97-8.09 (m, 1H), 4.74 (s, 1H), 4.18-4.21 (m, 2H), 3.66-3.71 (m, 2H), 3.61-3.66 (m, 2H), 3.51-3.59 (m, 2H), 3.34-3.44 (m, 1H), 2.83-2.87 ppm (m, 1H). Exchangeable proton not observed. Additional signals observed due to rotamers.
δ_C (101 MHz, CD_3OD):	167.9, 164.0, 80.7, 76.1, 71.8, 71.4, 71.2, 70.6, 70.3, 59.2, 43.0, 39.0 ppm. Additional signals observed due to rotamers.
LCMS (method A):	t_R = 0.42 min, 92% by UV, $[M+H]^+$ found: 172.1.

HRMS: ($C_8H_{14}NO_3$) $[M+H]^+$ requires: 172.0974, $[M+H]^+$ found: 172.0970.

ν_{max} (neat, cm^{-1}): 3293, 3058, 2869, 2113, 1660, 1526, 1444, 1385, 1350, 1243, 1127, 1088, 1081, 920, 880, 835, 666, 542.

Compound 2.14

3-(2-(2-Isocyanoethoxy)ethoxy)prop-1-yne



A vial was loaded with *N*-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)formamide (514 mg, 3.0 mmol), dichloromethane (3.0 mL), triethylamine (1.25 mL, 9.00 mmol) and tosyl chloride (1.26 g, 6.60 mmol) and capped. The reaction mixture was heated to 60 °C with magnetic stirring for 1 h. Analysis by TLC (50% ethyl acetate in cyclohexane on silica TLC with iodine staining) showed desired product (Rf. 0.5) and no starting material remaining. The reaction mixture was concentrated *in vacuo* and the residue was purified by flash chromatography (0–70% (*v/v*) (20% (*v/v*) methanol in dichloromethane) in dichloromethane over 20 CV) to give 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne (345 mg, 2.25 mmol, 75%) as a clear yellow oil.

δ_H (400 MHz, CD_3OD): 4.20 (d, J = 2.4 Hz, 2H), 3.62-3.73 (m, 8H), 2.83 (t, J = 2.2 Hz, 1H) ppm.

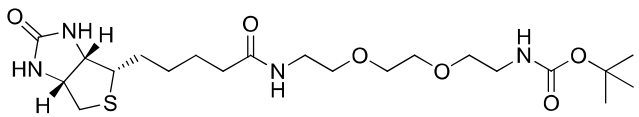
δ_C (101 MHz, CD_3OD): 156.7, 80.7, 76.0, 71.6, 70.3, 69.9, 59.2, 42.9 ppm.

HRMS: ($C_8H_{12}NO_2$) $[M+H]^+$ requires: 154.0868, $[M+H]^+$ found: 154.0864.

ν_{max} (neat, cm^{-1}): 3286, 2874, 2153, 1615, 1442, 1350, 1248, 1134, 1098, 1032, 954, 920, 880, 840.

Compound 2.19

Tert-butyl (2-(2-(2-(5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamido)ethoxy)ethoxy)ethyl)carbamate



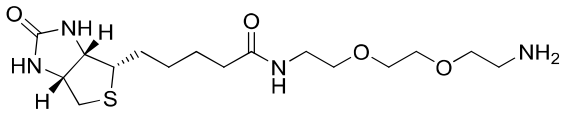
A RBF was loaded with 5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanoic acid (2.04 g, 8.36 mmol), *N*,*N*-dimethylformamide (40 mL), *tert*-butyl (2-(2-(2-aminoethoxy)ethoxy)ethyl)carbamate (2.38 mL, 10.0 mmol), *N*-ethyl-*N*-isopropylpropan-2-amine (4.37 mL, 25.1 mmol) and 1-(bis(dimethylamino)methylene)-1*H*-[1,2,3]triazolo[4,5-*b*]pyridine-1-ium 3-oxide hexafluorophosphate(V) (HATU) (3.82 g, 10.0 mmol). The reaction mixture was stirred at rt for 6 h. The reaction mixture was then concentrated under a flow of nitrogen. The residue was purified by flash chromatography (0–20% (*v*/*v*) methanol in dichloromethane over 15 CV). The fractions containing desired product (R_f 0.5 with 10% (*v*/*v*) methanol in dichloromethane on silica TLC with $KMnO_4$ staining) were combined and concentrated *in vacuo* to give *tert*-butyl (2-(2-(2-(5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamido) ethoxy)ethoxy)ethyl)carbamate (3.97 g, 8.36 mmol, 100%) as a white crystalline solid.

δ_H (400 MHz, CD_3OD): 4.44-4.54 (m, 1H), 4.30 (dd, *J* = 7.8, 4.9 Hz, 1H), 3.62 (s, 4H), 3.48-3.58 (m, 4H), 3.35-3.40 (m, 2H), 3.16-3.26 (m, 3H), 2.93 (dd, *J* = 12.7, 4.9 Hz, 1H), 2.71 (d, *J* = 12.7 Hz, 1H), 2.22 (t, *J* = 7.5 Hz, 2H), 1.54-1.82 (m, 4H), 1.39-1.53 (m, 11H) ppm. Exchangeable protons not observed. Consistent with literature.[69]

LCMS (Method B): t_R = 0.75 min, 97% by UV, $[M+H]^+$ found: 475.4.

Compound 2.20

N-(2-(2-(2-aminoethoxy)ethoxy)ethyl)-5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide



Tert-butyl (2-(2-(2-(5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl) pentanamido)ethoxy)ethoxy)ethyl)carbamate (2.07 g, 4.35 mmol) was dissolved in dichloromethane (22 mL) and trifluoroacetic acid (16.8 mL, 218 mmol) was added. The reaction mixture was stirred at rt for 30 min. The reaction mixture was concentrated *in vacuo* and the residue was dissolved in methanol (*ca*. 3 mL) and loaded

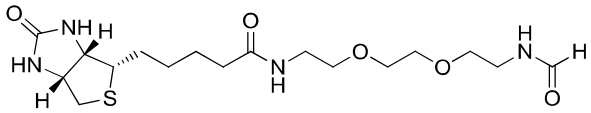
onto a Biotage 20 g SCX-2 cartridge which was prewashed with methanol. The cartridge was then eluted with methanol (*ca.* 100 mL) followed by 2 M ammonia in methanol (*ca.* 100 mL). The basic wash was concentrated *in vacuo* to give *N*-(2-(2-(2-aminoethoxy)ethoxy)ethyl)-5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide (1.51 g, 4.03 mmol, 93%) as a white gum.

δ_H (400 MHz, CD_3OD): 4.45-4.53 (m, 1H), 4.30 (dd, *J* = 7.8, 4.4 Hz, 1H), 3.63 (s, 4H), 3.54 (s, 4H), 3.36 (t, *J* = 5.4 Hz, 2H), 3.17-3.24 (m, 1H), 2.93 (dd, *J* = 12.7, 4.9 Hz, 1H), 2.80 (t, *J* = 5.1 Hz, 2H), 2.71 (d, *J* = 12.7 Hz, 1H), 2.22 (t, *J* = 7.5 Hz, 2H), 1.54-1.80 (m, 4H), 1.38-1.51 (m, 2H) ppm. Consistent with literature.[69]

LCMS (Method B): t_R = 0.49 min, 98% by UV, $[M+H]^+$ found: 375.3.

Compound 2.21

N-(2-(2-(2-formamidoethoxy)ethoxy)ethyl)-5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide



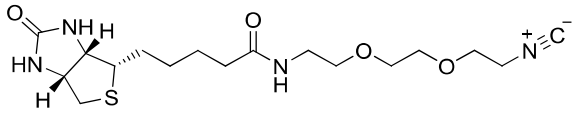
A RBF was charged with *N*-(2-(2-(2-aminoethoxy)ethoxy)ethyl)-5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide (370 mg, 0.988 mmol), ethyl formate (6.0 mL, 74 mmol), *N,N*-dimethylformamide (7.0 mL) and dichloromethane (1.0 mL). The reaction mixture was heated to reflux for 24 h. The reaction mixture was concentrated *in vacuo* and the residue was purified by flash chromatography (0–100% (20% (*v/v*) methanol/dichloromethane) in dichloromethane over 12 CV) to give *N*-(2-(2-(2-formamidoethoxy)ethoxy)ethyl)-5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide (265 mg, 0.658 mmol, 67%) as a white gum.

δ_H (400 MHz, CD_3OD): 8.07 (s, 1H), 4.44-4.56 (m, 1H), 4.30 (dd, *J* = 7.8, 4.4 Hz, 1H), 3.63 (s, 4H), 3.51-3.59 (m, 4H), 3.41 (t, *J* = 5.8 Hz, 2H), 3.36 (t, *J* = 5.8 Hz, 2H), 3.21 (ddd, *J* = 8.8, 5.9,

	4.4 Hz, 1H), 2.93 (dd, *J* = 13.0, 5.1 Hz, 1H), 2.71 (d, *J* = 12.7 Hz, 1H), 2.22 (t, *J* = 7.5 Hz, 2H), 1.54-1.81 (m, 4H), 1.45 (quin, *J* = 8.0 Hz, 2H) ppm. Exchangeable protons not observed.
δ_C (101 MHz, CD_3OD):	176.3, 166.2, 164.0, 71.50, 71.45, 70.8, 70.6, 63.5, 61.8, 57.1, 41.2, 40.5, 39.1, 36.9, 29.9, 29.6, 27.0 ppm.
LCMS (Method A):	t_R = 0.47 min, 100% by UV, $[M+H]^+$ found: 403.2.
HRMS:	($C_{17}H_{31}N_4O_5S$) $[M+H]^+$ requires: 403.2015, $[M+H]^+$ found: 403.2017.
ν_{max} (neat, cm^{-1}):	3276, 3072, 2936, 2867, 1686, 1635, 1549, 1463, 1416, 1386, 1262, 1238, 1102, 1079, 691, 592, 555, 527, 497.

Compound 2.17

N-(2-(2-(2-isocyanoethoxy)ethoxy)ethyl)-5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide



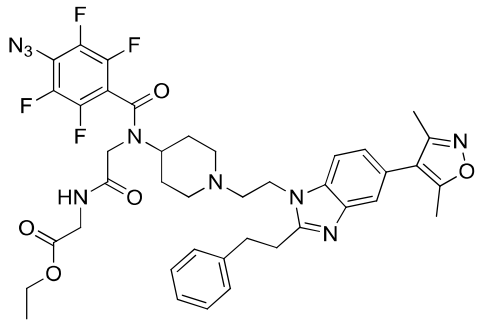
A 25 mL RBF was loaded with *N*-(2-(2-(2-formamidoethoxy)ethoxy)ethyl)-5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide (0.257 g, 0.638 mmol), triethylamine (3.0 mL), tosyl chloride (0.365 g, 1.92 mmol) and dichloromethane (3.0 mL). The reaction mixture was heated to reflux for 16 h. The reaction mixture was concentrated *in vacuo* and the residue was purified by passing through a silica plug (eluting with 0–100% (20% (*v/v*) methanol/dichloromethane) in dichloromethane over 12 CV). The fractions containing desired product were combined and concentrated *in vacuo*. The residue was partitioned between 4:1 (*v/v*) chloroform:isopropanol (25 mL) and 5% (*w/v*) aqueous citric acid solution (25 mL). The organic layer was collected, and the remaining aqueous layer was extracted with 4:1 (*v/v*) chloroform:isopropanol (2 × 25 mL). The organics were combined and dried by passing through a hydrophobic frit before being concentrated *in vacuo*. The residue was purified by flash chromatography (10% (*v/v*) methanol/dichloromethane for 10 CV) to give *N*-(2-(2-(2-isocyanoethoxy)ethoxy)ethyl)-5-((3a*S*,4*S*,6a*R*)-2-

oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide (53 mg, 0.14 mmol, 22%) as a white gum.

δ_H (400 MHz, CD_3OD):	4.45-4.53 (m, 1H), 4.31 (dd, *J* = 7.8, 4.5 Hz, 1H), 3.61-3.72 (m, 8H), 3.52-3.60 (m, 2H), 3.34-3.41 (m, 2H), 3.21 (ddd, *J* = 8.8, 5.7, 4.8 Hz, 1H), 2.93 (dd, *J* = 12.7, 4.9 Hz, 1H), 2.71 (d, *J* = 12.8 Hz, 1H), 2.18-2.27 (m, 2H), 1.66 (d, *J* = 7.5 Hz, 4H), 1.46 (d, *J* = 7.8 Hz, 2H) ppm. Exchangeable protons not observed.
δ_C (101 MHz, CD_3OD):	176.3, 166.2, 157.5, 71.8, 71.4, 70.8, 69.9, 63.5, 61.8, 57.1, 42.9, 41.2, 40.5, 36.9, 29.9, 29.6, 27.0 ppm.
LCMS (Method A):	t_R = 0.56 min, 100% by UV, $[M+H]^+$ found: 385.2.
HRMS:	($C_{17}H_{29}N_4O_4S$) $[M+H]^+$ requires: 385.1910, found $[M+H]^+$ found: 385.1912.
ν_{max} (neat, cm^{-1}):	3287, 2923, 2862, 2154, 1691, 1642, 1548, 1462, 1423, 1324, 1265, 1241, 1103, 759, 688, 656, 595, 549.

Compound 2.27

Ethyl *N*-(4-azido-2,3,5,6-tetrafluorobenzoyl)-*N*-(1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)glycylglycinate

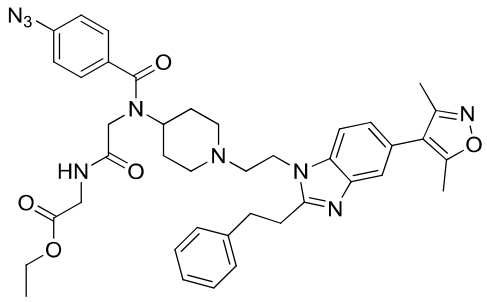


Following General Procedure 1 (Section 7.1.2.1), 1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-amine, 4-azido-2,3,5,6-tetrafluorobenzoic acid and ethyl 2-isocyanoacetate were used to give ethyl *N*-(4-azido-2,3,5,6-tetrafluorobenzoyl)-*N*-(1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)glycylglycinate (54 mg, 0.067 mmol, 67%) as a white powder.

δ_H (400 MHz, $CDCl_3$):	7.63 (d, *J* = 1.0 Hz, 1H), 7.21-7.37 (m, 3H), 7.09-7.16 (m, 1H), 6.75 (t, *J* = 5.4 Hz, 1H), 6.18 (t, *J* = 5.1 Hz, 1H), 4.49-4.65 (m, 1H), 4.19-4.27 (m, 2H), 4.13-4.17 (m, 2H), 4.05 (d, *J* = 5.4 Hz, 4H), 3.93 (d, *J* = 5.4 Hz, 1H), 3.82 (s, 1H), 3.72 (q, *J* = 7.0 Hz, 1H), 3.32-3.41 (m, 1H), 3.25-3.32 (m, 2H), 3.13-3.25 (m, 2H), 2.81-2.99 (m, 2H), 2.62 (t, *J* = 6.8 Hz, 1H), 2.54 (t, *J* = 7.1 Hz, 2H), 2.41-2.45 (m, 3H), 2.28-2.32 (m, 3H), 2.09-2.09 (m, 1H), 1.94-2.07 (m, 2H), 1.79-1.92 (m, 4H), 1.66 (br d, *J* = 3.9 Hz, 2H), 1.22-1.33 ppm (m, 4H) ppm. Additional signals observed due to rotamers.
δ_C (101 MHz, $CDCl_3$):	169.4, 169.2, 168.1, 167.5, 165.0, 159.9, 158.9, 155.3, 155.1, 143.0, 140.8, 134.2, 128.7, 128.3, 126.5, 124.3, 123.4, 120.0, 117.0, 109.3, 61.8, 61.5, 58.4, 58.2, 56.8, 56.6, 53.3, 53.1, 53.0, 47.3, 45.9, 41.9, 41.4, 41.2, 33.9, 30.2, 29.5, 14.1, 11.5, 10.8 ppm. Carbon atoms in the tetrafluoroaryl azide ring were not identified due to splitting by the adjacent fluorine atoms. Additional signals observed due to rotamers.
LCMS (Method B):	t_R = 1.29 min, 97% by UV, $[M+H]^+$ found: 804.5.
HRMS:	($C_{40}H_{42}F_4N_9O_5$) $[M+H]^+$ requires: 804.3240, $[M+H]^+$ found: 804.3239.
ν_{max} (neat, cm^{-1}):	3322, 2938, 2129, 1742, 1696, 1650, 1485, 1454, 1437, 1376, 1321, 1291, 1252, 1196, 1027, 996, 823, 808, 732, 701, 638, 522, 485.
M.pt. (MeCN/water):	82–85 °C

Compound 2.28

Ethyl *N*-(4-azidobenzoyl)-*N*-(1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)glycylglycinate



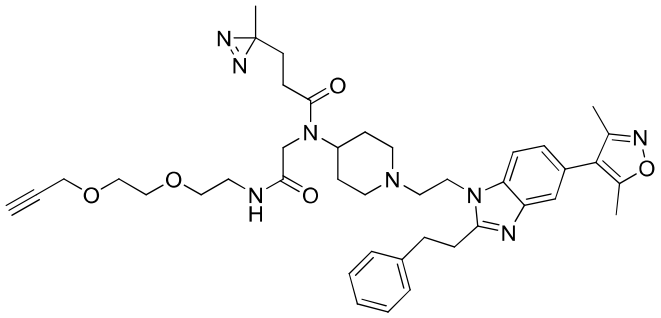
Following General Procedure 1 (Section 7.1.2.1), 1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-amine, 4-azidobenzoic acid and ethyl 2-isocyanoacetate were used to give ethyl *N*-(4-azidobenzoyl)-*N*-(1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)glycylglycinate (49 mg, 0.066 mmol, 66%) as an orange powder.

δ_H (400 MHz, $CDCl_3$):	7.63 (d, *J* = 1.0 Hz, 1H), 7.43 (d, *J* = 8.8 Hz, 2H), 7.20-7.35 (m, 6H), 7.12 (dd, *J* = 8.3, 1.5 Hz, 1H), 7.08 (d, *J* = 8.8 Hz, 2H), 4.22 (q, *J* = 7.2 Hz, 2H), 4.00-4.14 (m, 5H), 3.68-3.79 (m, 1H), 3.23-3.33 (m, 2H), 3.11-3.23 (m, 2H), 2.86 (br d, *J* = 9.3 Hz, 2H), 2.53 (br t, *J* = 6.8 Hz, 2H), 2.43 (s, 3H), 2.29 (s, 3H), 1.63-2.08 (m, 7H), 1.30 (t, *J* = 7.1 Hz, 3H), 1.25 (t, *J* = 7.1 Hz, 1H) ppm.
δ_C (101 MHz, $CDCl_3$):	172.1, 169.7, 165.0, 158.9, 155.2, 143.0, 141.9, 140.8, 134.2, 132.3, 128.7, 128.4, 128.1, 126.5, 124.2, 123.4, 119.9, 119.2, 117.1, 109.3, 61.5, 58.4, 56.7, 53.3, 46.3, 41.9, 41.2, 33.9, 30.0, 29.5, 18.4, 14.1, 11.5, 10.8 ppm.
LCMS (Method B):	t_R = 1.20 min, 98% by UV, $[M+H^+]$ found: 732.6.
HRMS:	($C_{40}H_{46}N_9O_5$) $[M+H]^+$ requires: 732.3622, $[M+H]^+$ found: 732.3614.
ν_{max} (neat, cm^{-1}):	2932, 2126, 2101, 1747, 1687, 1621, 1603, 1506, 1407, 1375, 1281, 1195, 1152, 1131, 1027, 844, 809, 759, 727, 701, 638, 532, 496.

M.pt. (MeCN/water): 66–68 °C.

Compound 2.22

N-(1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)-3-(3-methyl-3*H*-diazirin-3-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)propenamide



Following General Procedure 1 (Section 7.1.2.1), 1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[d]imidazol-1-yl)ethyl)piperidin-4-amine, 3-(3-methyl-3*H*-diazirin-3-yl)propanoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to give *N*-(1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)-3-(3-methyl-3H-diazirin-3-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)propenamide as a clear gum.

δ_H (400 MHz, CD_3OD): 7.49-7.60 (m, 2H), 7.14-7.32 (m, 6H), 4.34 (tt, *J* = 11.8, 3.9 Hz, 1H), 4.13-4.23 (m, 4H), 3.95 (s, 1H), 3.85 (s, 1H), 3.69-3.77 (m, 1H), 3.58-3.69 (m, 4H), 3.47-3.56 (m, 2H), 3.32-3.42 (m, 2H), 3.17-3.30 (m, 4H), 2.87-2.97 (m, 2H), 2.79-2.86 (m, 1H), 2.50-2.64 (m, 2H), 2.41 (s, 3H), 2.35 (t, *J* = 7.6 Hz, 1H), 2.26 (s, 3H), 2.01-2.22 (m, 3H), 1.37-1.87 (m, 6H), 1.02 (s, 1H), 0.99 ppm (s, 2H) ppm. Additional signals observed due to rotamers. Exchangeable proton not observed.

δ_C (101 MHz, CD_3OD): 175.0, 174.3, 171.8, 171.4, 166.8, 160.3, 157.5, 143.6, 142.2, 135.7, 135.6, 129.9, 129.7, 127.7, 125.6, 125.1, 120.0, 118.5, 111.7, 80.8, 80.7, 76.3, 76.2, 71.2, 70.7, 70.6, 70.3, 59.3, 57.9, 56.5, 54.7, 54.5, 53.5, 47.1, 46.0,

42.9, 42.8, 40.6, 40.5, 35.3, 35.2, 32.8, 31.7, 31.5, 31.3, 31.1, 30.9, 30.4, 30.2, 29.1, 28.4, 26.5, 26.5, 20.1, 20.0, 11.6, 10.9 ppm. Additional signals observed due to rotamers.

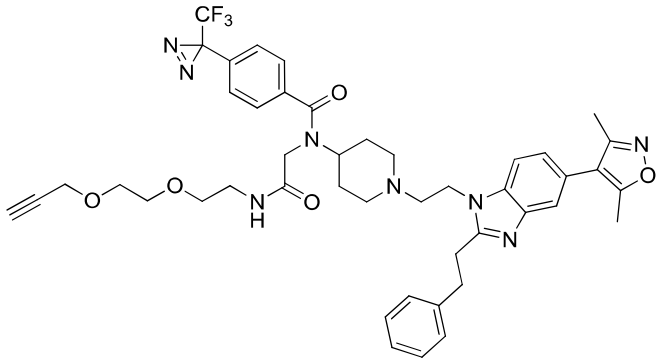
LCMS (Method B): t_R = 1.15 min, 86% by UV, $[M+H]^+$ found: 737.7.

HRMS: ($C_{41}H_{53}N_8O_5$) $[M+H]^+$ requires: 737.4139, $[M+H]^+$ found: 737.4135.

ν_{max} (neat, cm^{-1}): 3288, 2929, 1645, 1511, 1453, 1354, 1322, 1293, 1244, 1133, 1101, 1032, 888, 810, 756, 702.

Compound 2.23

N-(1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)-4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzamide



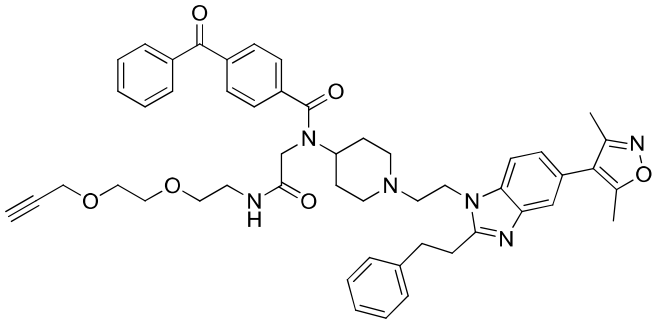
Following General Procedure 1 (Section 7.1.2.1), 1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1H-benzo[d]imidazol-1-yl)ethyl)piperidin-4-amine, 4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to give *N*-(1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)-4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzamide (41 mg, 0.049 mmol, 49%) as an orange powder.

δ_H (400 MHz, CD_3OD): δ = 7.45-7.70 (m, 4H), 7.13-7.42 (m, 8H), 4.09-4.48 (m, 5H), 4.03 (br s, 1H), 3.85 (br s, 1H), 3.36-3.76 (m, 8H),

	3.18-3.30 (m, 4H), 3.03-3.17 (m, 1H), 2.78-2.97 (m, 2H), 2.63-2.77 (m, 1H), 2.49-2.60 (m, 1H), 2.34-2.47 (m, 4H), 2.26 (s, 3H), 1.65-2.02 (m, 5H) ppm. Exchangeable protons not observed. Additional signals due to the presence of rotamers.
δ_C (101 MHz, CD_3OD):	173.1, 171.1, 166.9, 160.3, 157.4, 143.1, 142.1, 135.5, 129.9, 129.7, 127.8, 126.0, 125.3, 119.9, 118.4, 111.9, 80.7, 76.2, 71.2, 70.7, 70.4, 59.3, 58.4, 57.4, 54.5, 54.1, 45.7, 42.5, 42.2, 40.6, 35.2, 31.0, 30.3, 29.8, 29.4, 11.6, 10.9 ppm. Additional signals due to the presence of rotamers and splitting by fluorine.
LCMS (Method B):	t_R = 1.30 min, 93% by UV, $[M+H]^+$ found: 839.6.
HRMS:	($C_{45}H_{50}F_3N_8O_5$) $[M+H]^+$ requires: 839.3856, $[M+H]^+$ found: 839.3849.
ν_{max} (neat, cm^{-1}):	3285, 2931, 1678, 1629, 1513, 1454, 1344, 1231, 1183, 1151, 1095, 1030, 977, 878, 740, 637, 527.
M.pt. (MeCN/water):	62–64 °C.

Compound 2.24

4-Benzoyl-*N*-(1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)benzamide



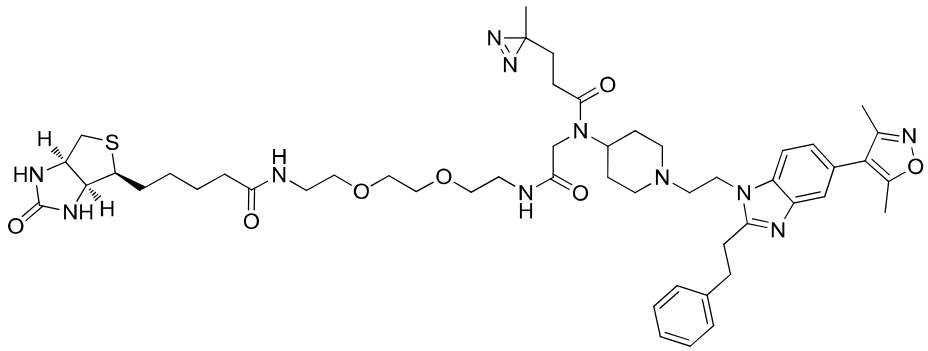
Following General Procedure 1 (Section 7.1.2.1), 1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-amine, 4-benzoylbenzoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to give 4-benzoyl-*N*-(1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-

yl)ethyl)piperidin-4-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)benzamide (54 mg, 0.064 mmol, 64%) as a clear gum.

δ_H (400 MHz, CD_3OD):	7.72-7.93 (m, 4H), 7.44-7.70 (m, 7H), 7.11-7.32 (m, 6H), 4.30-4.49 (m, 1H), 3.85-4.27 (m, 6H), 3.47-3.74 (m, 6H), 3.42 (br s, 2H), 3.17-3.28 (m, 4H), 2.91-3.04 (m, 1H), 2.83 (br s, 2H), 2.54-2.64 (m, 1H), 2.45-2.53 (m, 1H), 2.41 (s, 3H), 2.26 (s, 4H), 1.80-1.94 (m, 2H), 1.60-1.80 (m, 3H) ppm. Exchangeable protons not observed. Additional signals observed due to the presence of rotamers.
δ_C (101 MHz, CD_3OD):	143.6, 142.2, 138.6, 135.7, 134.2, 131.5, 131.24, 131.16, 129.9, 129.8, 129.7, 128.0, 127.7, 127.6, 125.6, 125.1, 120.0, 118.5, 111.7, 80.7, 76.2, 71.3, 71.2, 70.7, 70.4, 59.3, 57.8, 54.7, 54.3, 45.7, 42.8, 40.6, 35.2, 31.3, 30.4, 30.2, 11.6, 10.9 ppm. Additional signals observed due to the presence of rotamers.
LCMS (Method B):	t_R = 1.22 min, 98% by UV, $[M+H^+]$ found: 835.6.
HRMS:	($C_{50}H_{55}N_6O_6$) $[M+H]^+$ requires: 835.4183, $[M+H]^+$ found: 835.4174.
ν_{max} (neat, cm^{-1}):	2930, 1633, 1506, 1447, 1317, 1275, 1131, 1096, 1031, 976, 924, 862, 805, 754, 701, 655, 638, 496.
M.pt. (MeCN/water):	62–63 °C.

Compound 2.29

N-(12-(1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)-15-(3-methyl-3*H*-diazirin-3-yl)-10,13-dioxo-3,6-dioxa-9,12-diazapentadecyl)-5-((3*aS*,4*S*,6*aR*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide



Following General Procedure 1 (Section 7.1.2.1), 1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-amine, 3-(3-methyl-3*H*-diazirin-3-yl)propanoic acid and *N*-(2-(2-(2-isocyanoethoxy)ethoxy)ethyl)-5-((3*aS*,4*S*,6*aR*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide were used to give *N*-(12-(1-(2-(5-(3,5-dimethylisoxazol-4-yl)-2-phenethyl-1*H*-benzo[*d*]imidazol-1-yl)ethyl)piperidin-4-yl)-15-(3-methyl-3*H*-diazirin-3-yl)-10,13-dioxo-3,6-dioxa-9,12-diazapentadecyl)-5-((3*aS*,4*S*,6*aR*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide (27 mg, 0.028 mmol, 28%) as a clear gum.

δ_H (400 MHz, CD_3OD): 7.50-7.60 (m, 2H), 7.14-7.34 (m, 6H), 4.42-4.51 (m, 1H), 4.35 (br t, *J* = 12.0 Hz, 1H), 4.28 (dd, *J* = 7.8, 4.4 Hz, 1H), 4.18 (q, *J* = 6.7 Hz, 2H), 3.79-4.01 (m, 2H), 3.66-3.78 (m, 1H), 3.60 (d, *J* = 4.4 Hz, 4H), 3.49-3.58 (m, 4H), 3.33-3.43 (m, 4H), 3.13-3.30 (m, 5H), 2.85-2.97 (m, 3H), 2.69 (d, *J* = 12.7 Hz, 1H), 2.52-2.63 (m, 2H), 2.42 (s, 3H), 2.35 (t, *J* = 7.6 Hz, 1H), 2.27 (s, 3H), 2.09-2.24 (m, 5H), 1.50-1.79 (m, 10H), 1.36-1.48 (m, 2H), 0.96-1.04 (m, 3H) ppm. Exchangeable protons not observed. Additional signals observed due to the presence of rotamers.

δ_C (101 MHz, CD_3OD): 176.2, 175.0, 174.3, 171.8, 171.5, 166.9, 166.2, 160.4, 157.5, 143.6, 142.2, 135.70, 135.65, 129.9, 129.7, 127.7, 125.6, 125.1, 120.0, 118.5, 111.8, 71.5, 70.82, 70.77, 70.7, 63.5, 61.8, 57.9, 57.1, 56.5, 54.7, 54.5, 53.6, 47.1, 46.0, 42.9, 42.8, 41.2, 40.6, 40.49, 40.45, 36.9, 35.3, 35.2, 31.3, 31.0, 30.9, 30.5, 30.2, 29.9, 29.7, 29.1, 28.4,

27.0, 26.5, 20.12, 20.07, 11.6, 10.9 ppm. Additional signals observed due to the presence of rotamers.

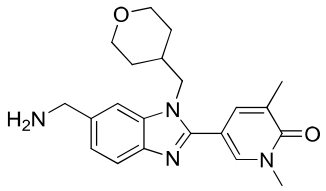
LCMS (Method B): t_R = 1.02 min, 98% by UV, $[M+H]^+$ found: 968.8.

HRMS: ($C_{50}H_{70}N_{11}O_7S$) $[M+H]^+$ requires: 968.5175, $[M+H]^+$ found: 968.5168.

ν_{max} (neat, cm^{-1}): 3293, 2929, 2864, 1693, 1647, 1453, 1244, 1116, 887, 702.

Compound 2.31

5-(6-(Aminomethyl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one



A RBF was loaded with 2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazole-6-carbaldehyde (700 mg, 1.92 mmol), hydroxylamine hydrochloride (160 mg, 2.30 mmol), sodium acetate (189 mg, 2.30 mmol) and methanol (6.0 mL) and the mixture was stirred at rt for 2 h. The reaction mixture was concentrated *in vacuo* to give a residue which was partitioned between 10% (*v/v*) 2-propanol/chloroform (30 mL) and water (30 mL). The organic layer was collected, and the remaining aqueous layer was extracted with 10% (*v/v*) 2-propanol/chloroform (3 × 20 mL). The extractions were combined and dried by passing through a hydrophobic frit before being concentrated *in vacuo*. The residue was dissolved in methanol (30 mL) and 37% (*w/w*) hydrogen chloride in water (0.25 mL, 2.54 mmol) was added. This mixture was then hydrogenated (1 atm) over 10% (*w/w*) palladium on carbon (204 mg, 0.192 mmol) at rt for 16 h. The reaction mixture was filtered through celite and the filtrate was concentrated *in vacuo* to give 5-(6-(aminomethyl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one (675 mg, 1.84 mmol, 96% yield) as a white foam.

δ_H (400 MHz, $CDCl_3$): 7.65-7.75 (m, 2H), 7.47 (dd, *J* = 2.5, 1.0 Hz, 1H), 7.37 (s, 1H), 7.22 (s, 1H), 4.12 (d, *J* = 7.1 Hz, 2H), 4.01-4.07

(m, 2H), 3.88 (br dd, *J* = 11.6, 2.5 Hz, 2H), 3.63 (s, 3H), 3.23 (td, *J* = 11.6, 2.5 Hz, 2H), 2.23 (s, 3H), 2.01-2.16 (m, 1H), 1.78 (br s, 2H), 1.18-1.39 (m, 4H) ppm.

δ_C (101 MHz, $CDCl_3$): 162.6, 150.5, 141.8, 138.2, 137.4, 136.0, 135.9, 129.9, 122.4, 119.7, 109.3, 108.5, 67.1, 50.3, 46.7, 38.1, 35.5, 30.5, 17.4 ppm.

LCMS (Method A): t_R = 0.32 min, 96% by UV, $[M+H]^+$ found: 367.2

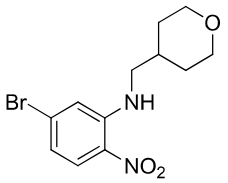
HRMS: ($C_{21}H_{27}N_4O_2$) $[M+H]^+$ requires 367.2129, found $[M+H]^+$ 367.2135.

ν_{max} (neat, cm^{-1}): 3365, 2922, 2842, 1652, 1602, 1510, 1455, 1428, 1363, 1238, 1205, 1141, 1089, 1015, 985, 899, 850, 809, 769, 723, 505, 479.

M.pt. (CH_2Cl_2): 127–129 °C.

Compound 2.39

5-bromo-2-nitro-*N*-((tetrahydro-2*H*-pyran-4-yl)methyl)aniline



(Tetrahydro-2*H*-pyran-4-yl)methanamine (3.89 mL, 31.8 mmol) was dissolved in 2-methyltetrahydrofuran (60 mL) and 4-bromo-2-fluoro-1-nitrobenzene (5.0 g, 23 mmol) and DIPEA (6.0 mL, 34 mmol) were added under nitrogen. The reaction mixture was heated to 90 °C for 2 h. The reaction mixture was partitioned between ethyl acetate (100 mL) and 5% (*w/v*) aqueous citric acid solution (200 mL). The organic layer was collected, and the remaining aqueous layer was extracted with ethyl acetate (3 × 50 mL). The organics were combined and dried by passing through a hydrophobic frit before being concentrated *in vacuo* to give 5-bromo-2-nitro-*N*-((tetrahydro-2*H*-pyran-4-yl)methyl)aniline (7.02 g, 22.3 mmol, 98%) as an orange powder.

δ_H (400 MHz, $CDCl_3$): 8.07-8.21 (m, 1H), 8.04 (d, *J* = 9.1 Hz, 1H), 7.02 (d, *J* = 2.0 Hz, 1H), 6.77 (dd, *J* = 9.1, 2.0 Hz, 1H), 4.03 (ddd, *J*

= 11.1, 4.0, 1.0 Hz, 2H), 3.44 (td, *J* = 11.8, 2.0 Hz, 2H), 3.19 (dd, *J* = 6.5, 5.5 Hz, 2H), 1.96 (s, 1H), 1.70-1.81 (m, 2H), 1.45 (qd, *J* = 13.0, 4.0 Hz, 2H) ppm.

δ_C (101 MHz, $CDCl_3$): 145.8, 131.8, 131.0, 128.2, 118.8, 116.4, 67.5, 49.1, 34.7, 30.9 ppm.

LCMS (Method A): t_R = 1.22 min, 100% by UV, $[M+H]^+$ found: 313.0/315.2 (Br signature).

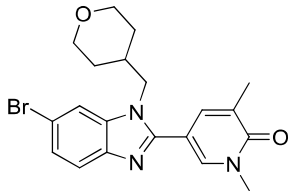
HRMS: ($C_{12}H_{16}{}^{79}BrN_2O_3$) $[M+H]^+$ requires 315.0344, $[M+H]^+$ found: 315.0337.

ν_{max} (neat, cm^{-1}): 3371, 2920, 2852, 1605, 1492, 1308, 1233, 1184, 1156, 1137, 1091, 1061, 1041, 875, 832, 751, 623, 602.

M.pt. (MeCN/water): 100–102 °C.

Compound 2.36

5-(6-Bromo-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one



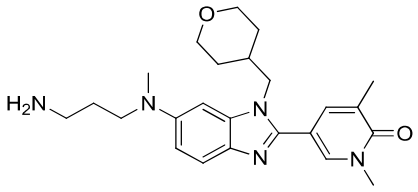
5-Bromo-2-nitro-*N*-((tetrahydro-2*H*-pyran-4-yl)methyl)aniline (7.02 g, 22.3 mmol) and 1,5-dimethyl-6-oxo-1,6-dihydropyridine-3-carbaldehyde (4.04 g, 26.7 mmol) were suspended in ethanol (66 mL) and water (33 mL) and the mixture was heated to 60 °C. Sodium dithionite (11.6 g, 66.8 mmol) was added and the reaction mixture was heated at 60 °C for 2.5 d. The reaction mixture was allowed to cool to rt and then concentrated to half the orginal volume *in vacuo* before being partitioned between ethyl acetate (100 mL) and water (120 mL). The organic layer was collected, and the remaining aqueous layer was extracted with ethyl acetate (3 × 50 mL). The organics were combined and dried by passing through a hydrophobic frit before being concentrated *in vacuo*. The residue was purified by flash chromatography (0–30% (*v/v*) ethanol/ethyl acetate over 10 CV) to give 5-(6-bromo-1-((tetrahydro-2*H*-pyran-4-

yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one (5.64 g, 13.6 mmol, 61%) as a white powder.

δ_H (400 MHz, $CDCl_3$):	7.73 (d, *J* = 2.5 Hz, 1H), 7.61 (d, *J* = 8.6 Hz, 1H), 7.53 (d, *J* = 2.0 Hz, 1H), 7.44-7.50 (m, 1H), 7.40 (dd, *J* = 8.6, 1.5 Hz, 1H), 4.10 (d, *J* = 7.6 Hz, 2H), 3.83-3.99 (m, 2H), 3.65 (s, 3H), 3.26 (td, *J* = 11.7, 2.8 Hz, 2H), 2.24 (s, 3H), 1.99-2.15 (m, 1H), 1.16-1.41 (m, 4H) ppm.
δ_C (101 MHz, $CDCl_3$):	162.6, 151.2, 141.7, 137.7, 136.8, 135.6, 130.1, 126.1, 121.1, 116.1, 113.2, 108.8, 67.1, 50.5, 38.2, 35.5, 30.5, 17.4 ppm.
LCMS (Method B):	t_R = 0.94 min, 100% by UV, $[M+H]^+$ found: 416.0/417.9 (Br signature).
HRMS:	($C_{20}H_{23}{}^{79}BrN_3O_2$) $[M+H]^+$ requires 416.0974, found $[M+H]^+$ 416.0972
ν_{max} (neat, cm^{-1}):	2927, 1655, 1616, 1492, 1456, 1424, 1384, 1363, 1278, 1242, 1139, 1090, 1055, 1018, 984, 972, 886, 851, 819, 769, 503.
M.pt. (ethyl acetate):	204 °C (decomp.).

Compound 2.33

Tert-butyl (3-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)propyl)carbamate



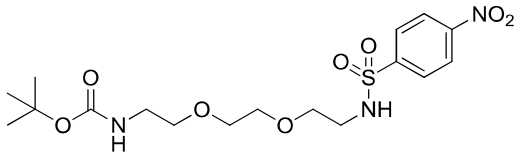
A vial was charged with 5-(6-bromo-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one (0.83 g, 2.00 mmol), $Pd_2(dba)_3$ (0.183 g, 0.200 mmol), Cs_2CO_3 (1.96 g, 6.00 mmol), RuPhos (0.113 g, 0.242 mmol), tetrahydrofuran (13.3 mL) and *tert*-butyl (3-(methylamino)propyl)carbamate (0.45 g, 2.40 mmol). The vial was sealed and heated 120 °C for 2 h. The reaction

mixture was filtered through celite and the retentate was washed with ethyl acetate (100 mL). The filtrate was concentrated *in vacuo* and the residue was purified by flash chromatography (0–20% (*v/v*) methanol/dichloromethane over 15 CV). The fractions containing desired product were combined and concentrated *in vacuo*. The residue was dissolved in dichloromethane (10 mL) and trifluoroacetic acid (1.47 mL, 19.1 mmol) was added. The reaction mixture was stirred at rt for 0.5 h before being concentrated *in vacuo*. The residue was dissolved in methanol (10 mL) and loaded onto a Biotage SCX-2 cartridge which was prewashed with methanol. The cartridge was washed with methanol (100 mL), followed by 2 M ammonia in methanol (100 mL). The basic wash was collected and concentrated *in vacuo* to give 5-(6-((3-aminopropyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one (410 mg, 0.968 mmol, 48%) as a yellow oil.

δ_H (400 MHz, CD_3OD):	7.93 (d, *J* = 2.4 Hz, 1H), 7.62-7.73 (m, 1H), 7.48 (d, *J* = 8.8 Hz, 1H), 6.90 (dd, *J* = 9.0, 2.2 Hz, 1H), 6.78 (d, *J* = 2.0 Hz, 1H), 4.19 (d, *J* = 7.3 Hz, 2H), 3.83 (br dd, *J* = 11.7, 2.4 Hz, 2H), 3.67 (s, 3H), 3.47 (t, *J* = 7.0 Hz, 2H), 3.24 (td, *J* = 11.7, 2.4 Hz, 2H), 3.00 (s, 3H), 2.73 (t, *J* = 7.0 Hz, 2H), 2.21 (s, 3H), 1.99-2.13 (m, 1H), 1.79 (quin, *J* = 7.2 Hz, 2H), 1.29-1.38 (m, 2H), 1.24 (qd, *J* = 12.5, 5.5 Hz, 2H) ppm. Exchangeable protons not observed.
δ_C (101 MHz, CD_3OD):	164.7, 149.9, 149.0, 139.4, 139.2, 138.5, 135.7, 130.5, 120.1, 112.9, 111.7, 95.1, 68.5, 52.7, 50.8, 40.7, 39.7, 38.7, 36.9, 31.8, 31.0, 17.3 ppm.
LCMS (Method B):	t_R = 0.69 min, 97% by UV, $[M+H]^+$ found: 424.4.
HRMS:	($C_{24}H_{34}N_5O_2$) $[M+H]^+$ requires: 424.2713, $[M+H]^+$ found: 424.2709.
ν_{max} (neat, cm^{-1}):	3366, 2930, 2852, 1654, 1603, 1494, 1456, 1425, 1402, 1364, 1279, 1238, 1090, 798, 768, 634, 605, 510.

Compound 2.44

Tert-butyl (2-(2-(2-((4-nitrophenyl)sulfonamido)ethoxy)ethoxy)ethyl)carbamate

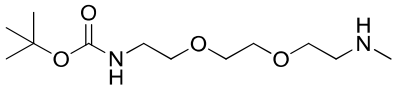


A vial was charged with *tert*-butyl (2-(2-(2-aminoethoxy)ethoxy)ethyl)carbamate (1.95 g, 7.85 mmol), dichloromethane (15 mL), potassium carbonate (1.25 g, 9.04 mmol) and 4-nitrobenzenesulfonyl chloride (1.9 g, 8.6 mmol) and the vial was capped. The reaction mixture was stirred at rt for 16 h. The reaction mixture was partitioned between dichloromethane (30 mL) and water (50 mL). The organic layer was collected, and the remaining aqueous layer was extracted with dichloromethane (2 × 25 mL). The extractions were combined and dried by passing through a hydrophobic frit before being concentrated *in* vacuo. The residue was purified by flash chromatography (0–80% (*v/v*) ethyl acetate in cyclohexane over 12 CV) to give *tert*-butyl (2-(2-(2-((4-nitrophenyl)sulfonamido)ethoxy)ethoxy)ethyl)carbamate (3.08 g, 7.11 mmol, 90%) as a yellow oil.

δ_H (400 MHz, CD_3OD):	8.40 (d, J = 9.3 Hz, 2H), 8.10 (d, J = 8.8 Hz, 2H), 3.45-3.56 (m, 8H), 3.21 (t, J = 5.6 Hz, 2H), 3.14 (t, J = 5.5 Hz, 2H), 1.43 (s, 9H) ppm.
δ_C (101 MHz, CD_3OD):	158.5, 151.5, 148.3, 129.6, 125.5, 80.3, 71.4, 71.3, 71.2, 70.9, 44.2, 41.3, 28.9 ppm.
LCMS (Method B):	t_R = 1.02 min, 100% by UV, $[M+H]^+$ found: 432.4.
HRMS:	($C_{17}H_{27}N_3NaO_8S$) $[M+Na]^+$ requires 456.1411, $[M+Na]^+$ found: 456.1413.
ν_{max} (neat, cm^{-1}):	2977, 2867, 1688, 1528, 1366, 1348, 1308, 1249, 1162, 1090, 854, 746, 686, 735, 686, 609, 564, 462.

Compound 2.42

Tert-butyl (2-(2-(2-(methylamino)ethoxy)ethoxy)ethyl)carbamate

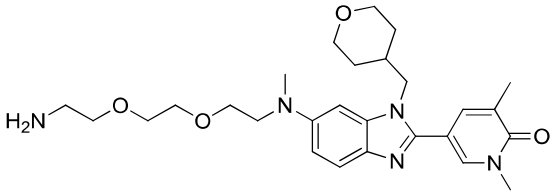


Tert-butyl (2-(2-(2-((4-nitrophenyl)sulfonamido)ethoxy)ethoxy)ethyl)carbamate (3.06 g, 7.06 mmol) was dissolved in *N,N*-dimethylformamide (20 mL) and potassium carbonate (1.46 g, 10.6 mmol) was added. The reaction mixture was cooled to -15 °C and methyl iodide (0.66 mL, 10.6 mmol) was added portion-wise over 5 min. The mixture was then allowed to warm to rt under nitrogen and stirred for 2 h. The reaction mixture was concentrated *in vacuo*. The residue was dissolved in *N,N*-dimethylformamide (20 mL) and cooled to 0 °C before the addition of thiophenol (1.82 mL, 17.7 mmol) portion-wise over 5 min. The reaction mixture was allowed warm to rt and stirred for 40 h under nitrogen. The reaction mixture was filtered through celite and the retentate was washed with methanol (150 mL). The filtrate was concentrated *in vacuo* and the residue was partitioned between dichloromethane (100 mL) and 10% (*v/v*) formic acid in water (100 mL). The aqueous layer was retained and washed with dichloromethane (2 × 50 mL). The aqueous layer was then basified by the addition of 2 M NaOH in water (100 mL). The aqueous layer was extracted with dichloromethane (3 × 100 mL) and these extractions were combined and dried by passing through a hydrophobic frit before being concentrated *in vacuo* to give *tert*-butyl (2-(2-(2-(methylamino)ethoxy)ethoxy)ethyl)carbamate (1.45 g, 5.53 mmol, 78%) as a clear oil.

δ_H (400 MHz, CD_3OD):	4.78 (s, 2H), 3.57-3.63 (m, 6H), 3.51 (t, J = 5.9 Hz, 2H), 3.22 (t, J = 5.6 Hz, 2H), 2.73 (t, J = 5.1 Hz, 2H), 2.40 (s, 3H), 1.44 (s, 9H) ppm.
δ_C (101 MHz, CD_3OD):	158.5, 80.2, 71.5, 71.4, 71.2, 70.7, 51.9, 41.4, 36.1, 28.9 ppm.
LCMS (Method C):	t_R = 0.48 min, 100% by ELSD, $[M+H]^+$ found: 263.2.
HRMS:	($C_{12}H_{27}N_2O_4$) $[M+H]^+$ requires: 263.1971, $[M+H]^+$ found: 263.1970.
ν_{max} (neat, cm^{-1}):	3329, 2974, 2868, 1699, 1517, 1477, 1453, 1391, 1365, 1274, 1250, 1171, 1099, 969, 864, 780, 528, 462.

Compound 2.34

5-(6-((2-(2-(2-Aminoethoxy)ethoxy)ethyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one



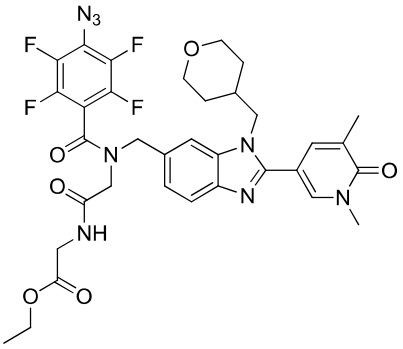
A vial was charged with *tert*-butyl (2-(2-(2-(methylamino)ethoxy)ethoxy)ethyl)carbamate (0.69 g, 2.6 mmol), tetrahydrofuran (13.3 mL), 5-(6-bromo-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one (0.83 g, 2.0 mmol), RuPhos G2 (0.16 g, 0.20 mmol) and Cs_2CO_3 (1.96 g, 6.00 mmol). The vial was sealed and heated to 120 °C for 4 h. The reaction mixture was filtered through celite and the retentate was washed with ethyl acetate (120 mL). The filtrate was concentrated *in vacuo* and the residue was purified by flash chromatography (0–20% (*v/v*) methanol/dichloromethane over 12 CV). The fractions containing desired product were combined and concentrated *in vacuo*. The residue was dissolved in dichloromethane (10 mL) and trifluoroacetic acid (2.03 mL, 26.3 mmol) was added. The reaction mixture was stirred for 1 h at rt before being concentrated *in vacuo*. The residue was dissolved in methanol and loaded onto a Biotage SCX-2 cartridge which was prewashed with methanol (100 mL). The cartridge was then washed with 2 M ammonia in methanol (100 mL). The basic wash was concentrated *in vacuo* to give 5-(6-((2-(2-(2-aminoethoxy)ethoxy)ethyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one (618 mg, 1.24 mmol, 62% yield) as a yellow oil.

δ_H (400 MHz, CD_3OD): 7.93 (d, *J* = 2.4 Hz, 1H), 7.65-7.70 (m, 1H), 7.47 (d, *J* = 8.8 Hz, 1H), 6.90 (dd, *J* = 8.8, 2.4 Hz, 1H), 6.79 (d, *J* = 2.4 Hz, 1H), 4.19 (d, *J* = 7.8 Hz, 2H), 3.82 (br dd, *J* = 11.5, 2.7 Hz, 2H), 3.72 (t, *J* = 5.5 Hz, 2H), 3.67 (s, 3H), 3.58-3.65 (m, 6H), 3.49 (t, *J* = 5.5 Hz, 2H), 3.24 (td, *J* =

	11.7, 2.4 Hz, 2H), 3.05 (s, 3H), 2.75 (t, *J* = 5.4 Hz, 2H), 2.21 (s, 3H), 1.99-2.13 (m, 1H), 1.16-1.38 (m, 4H) ppm. Exchangeable protons not observed.
δ_C (101 MHz, CD_3OD):	164.7, 149.9, 148.9, 139.4, 139.2, 138.5, 135.6, 130.4, 120.1, 112.5, 111.7, 94.8, 73.5, 71.8, 71.6, 70.1, 68.5, 54.6, 50.8, 42.3, 40.2, 38.7, 36.9, 31.8, 17.3 ppm.
LCMS (Method B):	t_R = 0.73 min, 97% by UV, $[M+H]^+$ found: 498.5.
HRMS:	($C_{27}H_{40}N_5O_4$) $[M+2H]^{2+}/2$ requires: 249.6580, $[M+2H]^{2+}/2$ found: 249.6585.
ν_{max} (neat, cm^{-1}):	3371, 2920, 2857, 1654, 1603, 1495, 1456, 1426, 1364, 1238, 1091, 1014, 985, 798, 769, 634, 605, 510.

Compound 2.45

Ethyl *N*-(4-azido-2,3,5,6-tetrafluorobenzoyl)-*N*-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)glycylglycinate



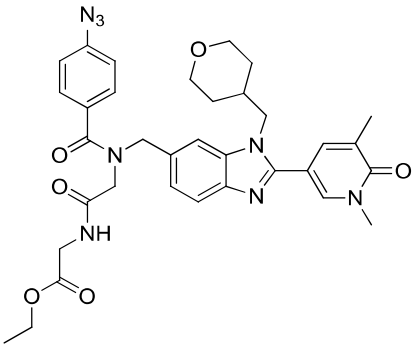
Following General Procedure 1 (Section 7.1.2.1), 5-(6-(aminomethyl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-azido-2,3,5,6-tetrafluorobenzoic acid and ethyl 2-isocyanoacetate were used to give ethyl *N*-(4-azido-2,3,5,6-tetrafluorobenzoyl)-*N*-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)glycylglycinate (44 mg, 0.061 mmol, 61%) as a white gum.

δ_H (400 MHz, $CDCl_3$):	7.67-7.78 (m, 2H), 7.42-7.53 (m, 1H), 7.28 (d, *J* = 0.8 Hz, 1H), 7.21-7.25 (m, 1H), 7.06 (dd, *J* = 8.3, 1.5 Hz, 1H), 6.52 (t, *J* = 5.3 Hz, 1H), 6.19 (t, *J* = 5.3 Hz, 1H),

	5.04 (s, 1H), 4.73 (s, 1H), 4.18-4.28 (m, 2H), 4.10-4.18 (m, 3H), 4.05 (d, *J* = 5.3 Hz, 1H), 3.95 (d, *J* = 5.5 Hz, 1H), 3.90 (br dd, *J* = 10.9, 3.4 Hz, 2H), 3.85 (s, 1H), 3.65 (s, 3H), 3.19-3.31 (m, 2H), 2.25 (s, 3H), 2.00-2.17 (m, 1H), 1.27-1.34 (m, 7H) ppm. Additional signals observed due to rotamers.
δ_C (101 MHz, $CDCl_3$):	169.4, 169.2, 167.2, 166.7, 162.6, 160.1, 151.6, 151.4, 142.9, 137.72, 137.69, 136.3, 136.1, 135.8, 135.7, 130.12, 130.07, 130.0, 128.7, 123.4, 122.4, 120.4, 120.0, 110.6, 110.4, 109.6, 109.0, 108.9, 67.1, 61.8, 61.7, 53.9, 50.7, 50.5, 50.0, 47.9, 41.4, 41.3, 38.2, 35.52, 35.47, 32.9, 31.9, 30.5, 30.1, 29.7, 22.7, 17.4, 14.12, 14.09 ppm. Additional signals observed due to rotamers.
LCMS (Method A):	t_R = 0.85 min, 99% by UV, $[M+H]^+$ found: 727.3.
HRMS:	($C_{34}H_{35}F_4N_8O_6$) $[M+H]^+$ requires: 727.2616, $[M+H]^+$ found: 727.2609.
ν_{max} (neat, cm^{-1}):	2919, 2857, 2127, 1744, 1652, 1611, 1487, 1460, 1376, 1241, 1197, 1092, 1000, 980, 820.

Compound 2.46

N-(4-azidobenzoyl)-*N*-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)glycylglycinate



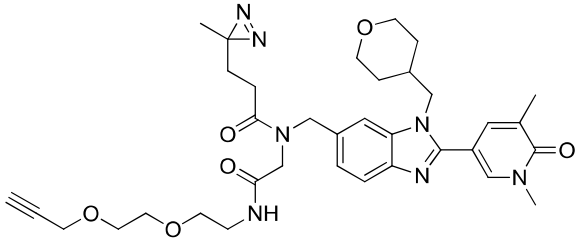
Following General Procedure 1 (Section 7.1.2.1), 5-(6-(aminomethyl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-azidobenzoic acid and ethyl 2-isocyanoacetate were used to give ethyl *N*-(4-

azidobenzoyl)-*N*-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)glycylglycinate (20 mg, 0.031 mmol, 31%) as a brown gum.

δ_H (400 MHz, $CDCl_3$):	7.70-7.77 (m, 2H), 7.60 (br d, *J* = 8.3 Hz, 2H), 7.49 (dd, *J* = 2.4, 1.1 Hz, 1H), 7.34 (br s, 1H), 7.15 (br s, 1H), 7.01-7.10 (m, 2H), 4.86 (br s, 2H), 4.24 (q, *J* = 7.3 Hz, 2H), 4.03 (d, *J* = 5.3 Hz, 6H), 3.89 (br dd, *J* = 11.7, 2.6 Hz, 2H), 3.66 (s, 3H), 3.24 (td, *J* = 11.5, 2.5 Hz, 2H), 2.25 (s, 3H), 2.01-2.14 (m, 1H), 1.28-1.38 (m, 7H) ppm. Exchangeable proton not observed.
δ_C (101 MHz, $CDCl_3$):	171.9, 169.6, 162.6, 151.2, 142.6, 142.4, 137.6, 136.2, 135.8, 131.4, 130.6, 130.1, 129.0, 120.3, 119.1, 109.1, 67.2, 61.7, 50.4, 41.2, 38.2, 35.7, 30.6, 29.7, 17.4, 14.2 ppm.
LCMS (Method A):	t_R = 0.73 min, 100% by UV, $[M+H]^+$ found: 655.3.
HRMS:	($C_{34}H_{39}N_8O_6$) $[M+H]^+$ requires: 655.2993, $[M+H]^+$ found: 655.2994.
ν_{max} (neat, cm^{-1}):	2925, 2850, 2131, 2094, 1744, 1655, 1603, 1456, 1428, 1280, 1239, 1196, 1143, 1092, 1001, 847, 768.

Compound 2.47

N-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)-3-(3-methyl-3*H*-diazirin-3-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)propenamide



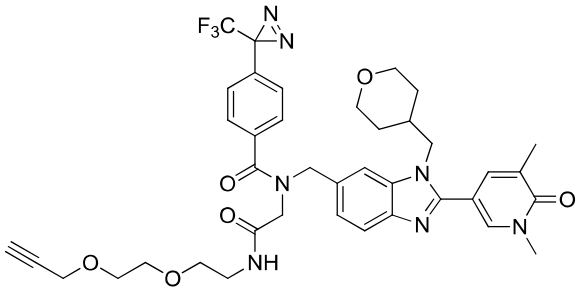
Following General Procedure 1 (Section 7.1.2.1), 5-(6-(aminomethyl)-1-((tetrahydro-2H-pyran-4-yl)methyl)-1H-benzo[d]imidazol-2-yl)-1,3-dimethylpyridin-2(1H)-one, 3-(3-methyl-3H-diazirin-3-yl)propanoic acid and 3-(2-(2-

isocyanoethoxy)ethoxy)prop-1-yne were used to give *N*-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)-3-(3-methyl-3*H*-diazirin-3-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)propanamide (29 mg, 0.043 mmol, 43%) as a clear gum.

δ_H (400 MHz, CD_3OD):	7.97-8.05 (m, 1H), 7.71 (br s, 1H), 7.59-7.70 (m, 1H), 7.60 (s, 1H), 7.55 (d, *J* = 3.5 Hz, 1H), 7.19-7.26 (m, 1H), 4.81 (s, 1H), 4.77 (s, 1H), 4.24-4.34 (m, 2H), 4.18 (d, *J* = 2.5 Hz, 2H), 4.01-4.08 (m, 2H), 3.77-3.87 (m, 2H), 3.65-3.70 (m, 5H), 3.56-3.63 (m, 2H), 3.47-3.54 (m, 1H), 3.42-3.47 (m, 1H), 3.32-3.39 (m, 2H), 3.21-3.30 (m, 2H), 2.82-2.86 (m, 1H), 2.40 (t, *J* = 7.4 Hz, 1H), 2.31 (t, *J* = 7.4 Hz, 1H), 2.22 (s, 3H), 2.04-2.17 (m, 1H), 1.73-1.84 (m, 2H), 1.14-1.37 (m, 4H), 0.94-1.08 (m, 3H) ppm.
δ_C (101 MHz, CD_3OD):	175.1, 175.0, 171.2, 170.6, 164.7, 152.7, 152.5, 143.0, 142.9, 140.0, 139.9, 138.89, 138.87, 137.5, 137.3, 133.8, 133.4, 130.7, 130.6, 124.7, 123.4, 120.5, 119.9, 112.2, 111.1, 111.0, 80.72, 80.69, 76.3, 76.2, 71.21, 71.16, 70.6, 70.5, 70.3, 68.4, 59.3, 59.2, 56.6, 54.0, 51.8, 51.2, 50.5, 40.5, 38.8, 36.9, 36.8, 31.7, 30.9, 30.6, 28.5, 28.4, 26.7, 20.3, 20.2, 17.3 ppm. Additional signals observed due to rotamers.
LCMS (Method A):	t_R = 0.68 min, 100% by UV, $[M+H]^+$ found: 660.4.
HRMS:	($C_{35}H_{46}N_7O_6$) $[M+H]^+$ requires: 660.3510, $[M+H]^+$ found: 660.3502.
ν_{max} (neat, cm^{-1}):	3290, 2924, 2858, 1654, 1613, 1560, 1514, 1363, 1281, 1239, 1205, 1140, 1092, 1015, 987, 920, 851, 811, 770, 728, 508.

Compound 2.48

N-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)-4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzamide



Following General Procedure 1 (Section 7.1.2.1), 5-(6-(aminomethyl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to give *N*-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)-4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzamide (14 mg, 0.019 mmol, 19%) as a clear gum.

δ_H (400 MHz, CD_3OD): 8.01 (d, *J* = 2.2 Hz, 1H), 7.50-7.75 (m, 5H), 7.28-7.40 (m, 2H), 7.14 (br d, *J* = 7.7 Hz, 1H), 4.94 (s, 1H), 4.69-4.77 (m, 1H), 4.25-4.33 (m, 2H), 4.15 (br d, *J* = 5.1 Hz, 3H), 3.79-3.88 (m, 3H), 3.56-3.70 (m, 8H), 3.50-3.55 (m, 1H), 3.37-3.48 (m, 2H), 3.21-3.29 (m, 2H), 2.81 (t, *J* = 2.4 Hz, 1H), 2.22 (s, 3H), 1.99-2.17 (m, 1H), 1.19-1.38 (m, 4H) ppm. Exchangeable proton not observed. Additional signals observed due to rotamers.

δ_C (101 MHz, CD_3OD): 173.7, 173.6, 170.5, 170.2, 164.7, 152.7, 143.1, 140.0, 139.0, 138.9, 138.7, 137.5, 137.4, 133.2, 131.8, 130.7, 129.0, 128.9, 128.1, 126.3, 124.8, 124.5, 123.4, 122.7, 120.5, 120.2, 112.9, 111.6, 111.5, 111.0, 80.6, 76.2, 71.2, 71.1, 70.7, 70.5, 70.4, 68.4, 59.2, 57.8, 57.6, 57.5, 57.3, 55.8, 52.4, 51.4, 51.3, 40.54, 40.48, 38.8, 37.0,

31.8, 30.0, 29.7, 29.4, 29.2, 17.6, 17.4, 17.3 ppm. Additional signals observed due to rotamers.

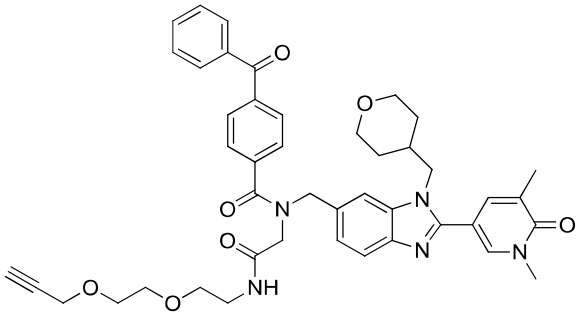
LCMS (Method B): t_R = 1.02 min, 90% by UV, $[M+H]^+$ found: 762.5.

HRMS: ($C_{39}H_{43}F_3N_7O_6$) $[M+H]^+$ requires: 762.3227, $[M+H]^+$ found: 762.3225.

ν_{max} (neat, cm^{-1}): 3283, 2927, 2856, 1655, 1612, 1428, 1346, 1232, 1183, 1142, 1091, 1002, 938, 767, 730.

Compound 2.49

4-Benzoyl-*N*-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)benzamide



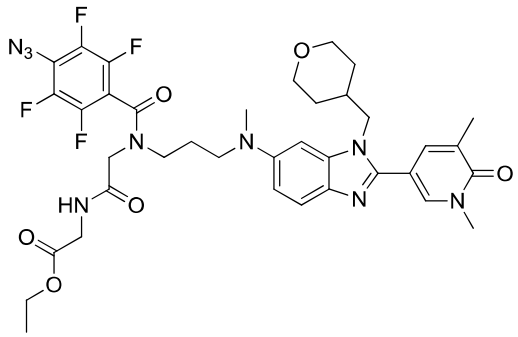
Following General Procedure 1 (Section 7.1.2.1), 5-(6-(aminomethyl)-1-((tetrahydro-2H-pyran-4-yl)methyl)-1H-benzo[d]imidazol-2-yl)-1,3-dimethylpyridin-2(1H)-one, 4-benzoylbenzoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to give 4-benzoyl-*N*-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)benzamide (28 mg, 0.037 mmol, 37%) as a clear oil.

δ_H (600 MHz, CD_3OD): 8.01 (br s, 1H), 7.47-7.88 (m, 12H), 7.36 (br d, *J* = 8.4 Hz, 1H), 7.18 (br d, *J* = 8.1 Hz, 1H), 4.97 (br s, 1H), 4.24-4.35 (m, 2H), 4.06-4.24 (m, 3H), 3.92 (s, 1H), 3.81 (br t, *J* = 10.6 Hz, 2H), 3.51-3.72 (m, 9H), 3.39-3.48 (m, 2H), 3.32-3.35 (m, 1H), 3.18-3.29 (m, 2H), 2.81 (br d, *J* = 11.4 Hz, 1H), 2.22 (br s, 3H), 1.95-2.17 (m, 1H), 1.15-

	1.39 (m, 4H) ppm. Exchangeable proton not observed. Additional signals observed due to rotamers.
δ_C (151 MHz, CD_3OD):	197.6, 174.0, 173.9, 170.5, 170.3, 164.7, 152.8, 152.7, 143.1, 141.1, 140.9, 140.3, 140.2, 140.0, 138.9, 138.8, 138.50, 138.47, 137.5, 137.4, 134.2, 133.3, 132.9, 131.4, 131.3, 131.2, 130.6, 129.8, 128.6, 128.3, 128.1, 124.8, 123.5, 120.4, 120.2, 112.9, 111.7, 111.0, 110.9, 80.7, 76.3, 71.2, 71.1, 70.7, 70.5, 70.3, 68.4, 59.2, 55.8, 52.4, 51.3, 51.3, 51.2, 40.6, 40.5, 38.8, 37.0, 31.8, 31.7, 17.4 ppm. Additional signals observed due to rotamers.
LCMS (Method B):	t_R = 0.95 min, 99% by UV, $[M+H]^+$ found: 758.5.
HRMS:	($C_{44}H_{48}N_5O_7$) $[M+H]^+$ requires: 758.3554, $[M+H]^+$ found: 758.3550.
ν_{max} (neat, cm^{-1}):	3293, 2927, 2856, 1654, 1609, 1447, 1276, 1239, 1138, 1090, 1001, 924, 769, 702, 656.

Compound 2.50

N-(4-azido-2,3,5,6-tetrafluorobenzoyl)-*N*-(3-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)propyl)glycylglycinate



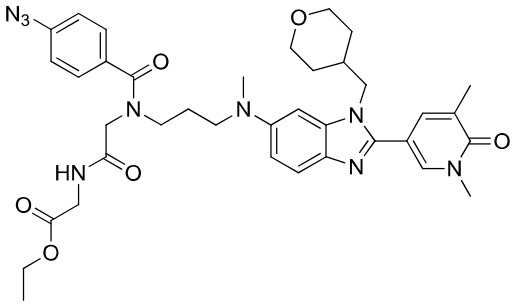
Following General Procedure 1 (Section 7.1.2.1), 5-(6-((3-aminopropyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-azido-2,3,5,6-tetrafluorobenzoic acid and ethyl 2-isocyanoacetate were used to give ethyl *N*-(4-azido-2,3,5,6-tetrafluorobenzoyl)-*N*-(3-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-

3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)propyl)glycylglycinate (47 mg, 0.060 mmol, 60%) as a brown gum.

δ_H (400 MHz, $CDCl_3$):	7.71 (d, *J* = 2.5 Hz, 1H), 7.51-7.62 (m, 1H), 7.42-7.51 (m, 1H), 6.72-6.92 (m, 1H), 6.56-6.67 (m, 1H), 6.36-6.50 (m, 1H), 4.16-4.30 (m, 3H), 4.01-4.12 (m, 3H), 3.84-4.00 (m, 3H), 3.71 (br t, *J* = 7.1 Hz, 1H), 3.62-3.67 (m, 3H), 3.36-3.52 (m, 2H), 3.17-3.35 (m, 3H), 2.84-3.04 (m, 3H), 2.20-2.28 (m, 3H), 1.84-2.15 (m, 5H), 1.33-1.42 (m, 2H), 1.26-1.31 (m, 4H) ppm. Additional signals observed due to rotamers.
δ_C (101 MHz, $CDCl_3$):	169.23, 169.17, 167.8, 166.9, 162.6, 160.0, 149.1, 148.9, 148.6, 146.8, 146.2, 141.3, 137.12, 137.06, 136.2, 136.1, 135.6, 129.82, 129.76, 120.0, 111.2, 110.8, 110.3, 109.8, 109.3, 93.7, 93.3, 91.2, 67.2, 61.8, 61.6, 53.4, 51.9, 51.3, 50.6, 50.0, 49.9, 48.5, 45.4, 41.4, 41.2, 40.2, 39.4, 38.13, 38.11, 35.6, 35.5, 30.7, 30.6, 29.7, 26.5, 24.5, 17.4, 14.10, 14.06 ppm. Additional signals observed due to rotamers.
LCMS (Method B):	t_R = 1.00 min, 95% by UV, $[M+H]^+$ found: 784.5.
HRMS:	($C_{37}H_{42}F_4N_9O_6$) $[M+H]^+$ requires: 784.3194, $[M+H]^+$ found: 784.3191.
ν_{max} (neat, cm^{-1}):	2919, 2844, 2127, 1744, 1652, 1611, 1489, 1427, 1427, 1375, 1241, 1197, 1093, 995, 798, 769.

Compound 2.51

Ethyl *N*-(4-azidobenzoyl)-*N*-(3-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)propyl)glycylglycinate



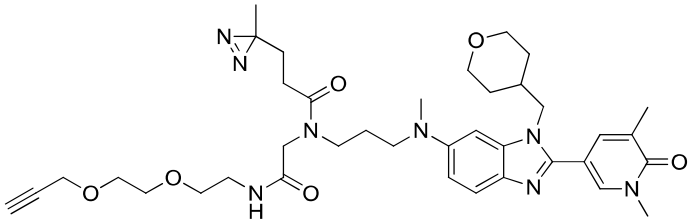
Following General Procedure 1 (Section 7.1.2.1), 5-(6-((3-aminopropyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-azidobenzoic acid and ethyl 2-isocyanoacetate were used to give ethyl *N*-(4-azidobenzoyl)-*N*-(3-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)propyl)glycylglycinate (31 mg, 0.044 mmol, 44%) as a yellow gum.

δ_H (400 MHz, $CDCl_3$):	7.70 (d, *J* = 2.5 Hz, 1H), 7.57 (d, *J* = 8.9 Hz, 1H), 7.47-7.52 (m, 1H), 7.42 (br d, *J* = 5.9 Hz, 1H), 6.90-7.24 (m, 2H), 6.36-6.87 (m, 2H), 3.98-4.27 (m, 8H), 3.89 (br dd, *J* = 11.6, 2.7 Hz, 2H), 3.64 (s, 3H), 3.40-3.58 (m, 2H), 2.75-3.39 (m, 6H), 2.24 (s, 3H), 1.74-2.12 (m, 5H), 1.32-1.39 (m, 2H), 1.27-1.31 (m, 4H) ppm. Additional signals observed due to rotamers.
δ_C (101 MHz, $CDCl_3$):	171.93, 171.88, 169.5, 162.6, 148.6, 141.9, 137.2, 137.1, 136.1, 135.3, 131.7, 129.8, 128.7, 120.0, 119.0, 111.1, 109.8, 93.5, 67.2, 61.6, 58.6, 57.3, 53.4, 50.9, 50.0, 41.2, 39.5, 38.1, 35.6, 30.6, 29.7, 17.4, 14.1 ppm.
LCMS (Method A):	t_R = 0.69 min, 100% by UV, $[M+H]^+$ found: 712.3.
HRMS:	($C_{37}H_{46}N_9O_6$) $[M+H]^+$ requires: 712.3571, $[M+H]^+$ found: 712.3571.

ν_{max} (neat, cm^{-1}): 2919, 2850, 2125, 2100, 1655, 1603, 1497, 1463, 1426, 1280, 1238, 1199, 1092, 847, 798, 768.

Compound 2.52

N-(3-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)propyl)-3-(3-methyl-3*H*-diazirin-3-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)propanamide



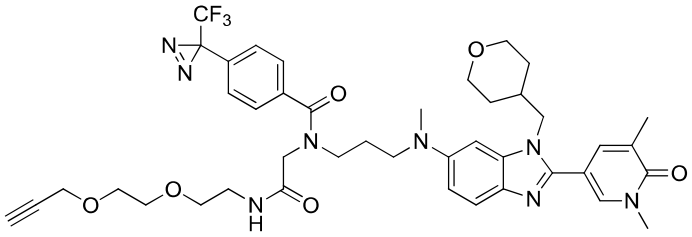
Following General Procedure 1 (Section 7.1.2.1), 5-(6-((3-aminopropyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 3-(3-methyl-3*H*-diazirin-3-yl)propanoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to give *N*-(3-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)propyl)-3-(3-methyl-3*H*-diazirin-3-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)propanamide (24 mg, 0.033 mmol, 33% as a clear oil.

δ_H (600 MHz, CD_3OD): 7.93 (s, 1H), 7.68 (s, 1H), 7.44-7.54 (m, 1H), 6.87-6.96 (m, 1H), 6.79-6.87 (m, 1H), 4.22 (d, *J* = 7.3 Hz, 2H), 4.17 (t, *J* = 2.2 Hz, 2H), 4.07 (s, 1H), 4.00 (s, 1H), 3.83 (br dd, *J* = 11.4, 2.9 Hz, 2H), 3.68 (s, 3H), 3.63-3.66 (m, 2H), 3.58-3.61 (m, 2H), 3.50-3.55 (m, 2H), 3.41-3.49 (m, 4H), 3.34-3.41 (m, 2H), 3.25 (td, *J* = 11.7, 1.8 Hz, 2H), 2.97-3.04 (m, 3H), 2.86 (d, *J* = 2.2 Hz, 1H), 2.17-2.24 (m, 5H), 2.01-2.13 (m, 1H), 1.90-1.99 (m, 1H), 1.85 (quin, *J* = 7.3 Hz, 1H), 1.65-1.72 (m, 1H), 1.56-1.62 (m, 1H), 1.30-1.38 (m, 2H), 1.18-1.28 (m, 2H), 0.87-1.02

	(m, 3H) ppm. Additional signals observed due to the presence of rotamers. Exchangeable proton not observed.
δ_C (151 MHz, CD_3OD):	174.9, 174.5, 171.3, 170.9, 164.7, 150.1, 149.9, 148.88, 148.86, 139.5, 139.4, 139.2, 139.1, 138.5, 136.0, 135.6, 130.5, 130.4, 120.3, 120.1, 113.2, 112.9, 111.74, 111.68, 95.8, 95.3, 80.73, 80.71, 76.3, 76.2, 71.21, 71.18, 70.7, 70.6, 70.3, 68.5, 68.4, 59.26, 59.25, 52.6, 52.2, 51.7, 50.85, 50.80, 50.4, 48.4, 47.3, 40.6, 40.5, 40.1, 39.7, 38.7, 36.9, 36.8, 31.8, 31.0, 30.8, 28.5, 28.1, 27.3, 26.55, 26.52, 26.1, 20.1, 20.0, 17.4 ppm. Additional signals observed due to rotamers.
LCMS (Method B):	t_R = 0.89 min, 97% by UV, $[M+H]^+$ found: 717.6.
HRMS:	($C_{38}H_{53}N_8O_6$) $[M+H]^+$ requires: 717.4088, $[M+H]^+$ found: 717.4091.
ν_{max} (neat, cm^{-1}):	3280, 2923, 2863, 1684, 1610, 1496, 1426, 1364, 1238, 1092, 986, 799, 769.

Compound 2.53

N-(3-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)propyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)-4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzamide



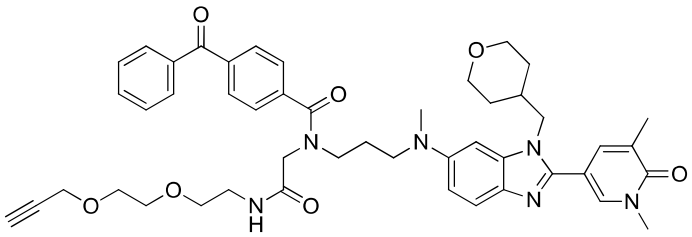
Following General Procedure 1 (Section 7.1.2.1), 5-(6-((3-aminopropyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used

to give *N*-(3-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)propyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)-4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzamide (9.3 mg, 0.011 mmol, 11%) as a yellow gum.

δ_H (600 MHz, CD_3OD):	7.82-7.97 (m, 1H), 7.69 (s, 1H), 7.42-7.58 (m, 3H), 7.15-7.36 (m, 2H), 6.81-6.96 (m, 1H), 6.63-6.80 (m, 1H), 4.11-4.25 (m, 5H), 3.92 (s, 1H), 3.76-3.88 (m, 2H), 3.54-3.70 (m, 10H), 3.48-3.54 (m, 1H), 3.40-3.47 (m, 2H), 3.34-3.40 (m, 1H), 3.17-3.30 (m, 3H), 3.01-3.06 (m, 1H), 2.81-2.86 (m, 1H), 2.79 (s, 1H), 2.22 (s, 3H), 1.96-2.13 (m, 2H), 1.86-1.95 (m, 1H), 1.28-1.35 (m, 2H), 1.09-1.28 (m, 3H) ppm. Additional signals observed due to the presence of rotamers. Exchangeable proton not observed.
δ_C (151 MHz, CD_3OD):	173.6, 173.2, 170.8, 170.6, 164.7, 150.0, 148.9, 148.6, 144.1, 139.4, 139.2, 139.1, 139.0, 138.5, 135.7, 130.5, 128.8, 128.7, 128.5, 128.01, 127.95, 124.5, 122.7, 120.21, 120.17, 113.0, 112.7, 111.74, 111.72, 111.5, 95.5, 95.1, 76.2, 71.24, 71.16, 70.7, 70.6, 70.4, 70.3, 68.4, 59.2, 53.6, 52.7, 51.7, 50.8, 50.2, 49.8, 49.7, 46.9, 40.60, 40.56, 39.8, 38.7, 36.9, 31.8, 29.7, 29.4, 27.6, 25.8, 17.4, 15.6 ppm. Additional signals observed due to the presence of rotamers.
LCMS (Method B):	t_R = 1.04 min, 98% by UV, $[M+H]^+$ found: 819.5.
HRMS:	($C_{42}H_{50}F_3N_8O_6$) $[M+H]^+$ requires: 819.3805, found $[M+H]^+$ 819.3800.
ν_{max} (neat, cm^{-1}):	2926, 2863, 1655, 1610, 1497, 1426, 1345, 1231, 1153, 1092, 979, 938, 832, 799, 768.

Compound 2.54

4-Benzoyl-*N*-(3-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)propyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)benzamide



Following General Procedure 1 (Section 7.1.2.1), 5-(6-((3-aminopropyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-benzoylbenzoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to give 4-benzoyl-*N*-(3-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)propyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)benzamide (38 mg, 0.047 mmol, 47% as a yellow gum.

δ_H (600 MHz, CD_3OD): 7.86-7.98 (m, 1H), 7.74-7.82 (m, 2H), 7.59-7.73 (m, 4H), 7.38-7.58 (m, 5H), 6.93 (br d, J = 8.8 Hz, 1H), 6.86 (s, 1H), 6.70 (br d, J = 7.7 Hz, 1H), 6.57 (s, 1H), 4.16-4.26 (m, 3H), 3.94-4.12 (m, 3H), 3.73-3.84 (m, 2H), 3.61-3.70 (m, 6H), 3.55-3.61 (m, 2H), 3.49-3.55 (m, 2H), 3.37-3.48 (m, 3H), 3.32-3.36 (m, 2H), 3.15-3.26 (m, 2H), 3.04 (s, 1H), 2.82 (s, 2H), 2.20 (s, 3H), 1.90-2.12 (m, 3H), 1.10-1.38 (m, 4H) ppm. Additional signals observed due to rotamers. Exchangeable proton not observed.

δ_C (151 MHz, CD_3OD): 197.7, 197.4, 173.8, 173.5, 170.8, 170.6, 164.6, 150.0, 149.9, 148.9, 148.4, 141.4, 141.1, 140.0, 139.6, 139.4, 139.2, 138.5, 138.4, 135.7, 135.6, 134.2, 131.3, 131.22, 131.17, 130.43, 130.36, 129.82, 129.77, 128.1, 127.8, 120.2, 112.9, 112.4, 111.7, 111.6, 95.5, 94.8, 80.74,

80.68, 80.3, 76.3, 76.0, 75.8, 71.2, 71.1, 70.7, 70.6, 70.4, 70.3, 68.42, 68.39, 59.24, 59.21, 53.6, 52.7, 51.5, 50.8, 50.3, 50.0, 49.8, 46.9, 40.61, 40.57, 40.0, 39.8, 38.8, 36.84, 36.78, 31.8, 31.7, 27.9, 25.8, 17.4 ppm. Additional signals observed due to rotamers.

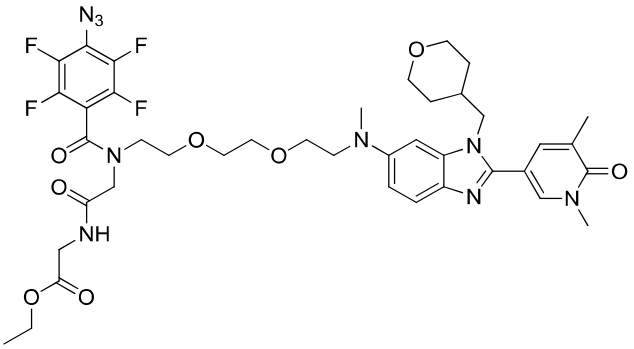
LCMS (Method B): t_R = 0.93 min, 99% by UV, $[M+H]^+$ found: 815.5.

HRMS: ($C_{47}H_{55}N_6O_7$) $[M+H]^+$ requires: 815.4132, $[M+H]^+$ found: 815.4132.

ν_{max} (neat, cm^{-1}): 3283, 2927, 2856, 1655, 1608, 1497, 1447, 1425, 1275, 1239, 1091, 924, 796, 768, 702.

Compound 2.55

Ethyl (11-(4-azido-2,3,5,6-tetrafluorobenzoyl)-2-(2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)-5,8-dioxa-2,11-diazatridecan-13-oyl)glycinate



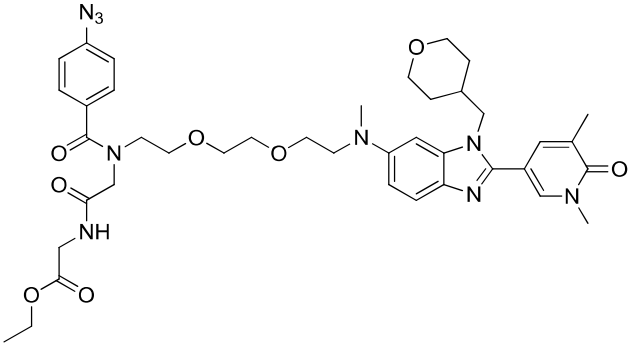
Following General Procedure 1 (Section 7.1.2.1), 5-(6-((2-(2-(2-aminoethoxy)ethoxy)ethyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-azido-2,3,5,6-tetrafluorobenzoic acid and ethyl 2-isocyanoacetate were used to give ethyl (11-(4-azido-2,3,5,6-tetrafluorobenzoyl)-2-(2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)-5,8-dioxa-2,11-diazatridecan-13-oyl)glycinate (46 mg, 0.053 mmol, 53%) as a brown gum.

δ_H (400 MHz, $CDCl_3$): 7.75-7.85 (m, 1H), 7.56-7.63 (m, 1H), 7.45-7.51 (m, 1H), 6.91-7.23 (m, 1H), 6.87 (dd, *J* = 9.0, 2.2 Hz, 1H), 6.59 (d, *J* = 2.0 Hz, 1H), 4.24 (s, 1H), 4.15-4.23 (m, 2H),

	4.05-4.12 (m, 2H), 4.00 (d, *J* = 10.3 Hz, 2H), 3.85-3.96 (m, 3H), 3.79 (s, 1H), 3.57-3.73 (m, 9H), 3.45-3.57 (m, 6H), 3.24 (td, *J* = 11.6, 2.2 Hz, 2H), 3.04 (s, 3H), 2.24 (s, 3H), 1.97-2.10 (m, 1H), 1.32-1.40 (m, 2H), 1.26-1.31 (m, 4H) ppm. Additional signals observed due to rotamers.
δ_C (101 MHz, $CDCl_3$):	169.5, 169.3, 168.0, 162.6, 148.2, 148.19, 148.16, 147.10, 147.06, 137.6, 136.9, 135.9, 135.8, 129.9, 119.5, 111.5, 111.1, 92.92, 92.91, 70.61, 70.55, 70.3, 68.83, 68.75, 67.6, 67.2, 61.5, 61.4, 53.43, 53.38, 50.5, 50.1, 50.0, 47.6, 41.5, 41.3, 39.7, 39.6, 38.2, 35.58, 35.56, 30.6, 29.7, 17.4, 14.2 ppm. Additional signals observed due to rotamers.
LCMS (Method A):	t_R = 0.80 min, 96% by UV, $[M+H]^+$ found: 858.3.
HRMS:	($C_{40}H_{48}F_4N_9O_8$) $[M+H]^+$ requires: 858.3562, $[M+H]^+$ found: 858.3570.
ν_{max} (neat, cm^{-1}):	2925, 2856, 2126, 1744, 1653, 1612, 1490, 1460, 1376, 1243, 1198, 1093, 997, 798, 769.

Compound 2.56

Ethyl (11-(4-azidobenzoyl)-2-(2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)-5,8-dioxa-2,11-diazatridecan-13-oyl)glycinate



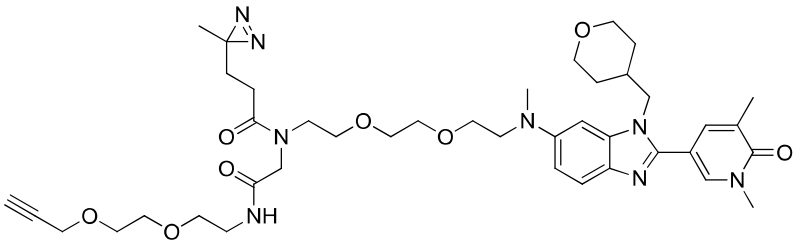
Following General Procedure 1 (Section 7.1.2.1), 5-(6-((2-(2-(2-aminoethoxy)ethoxy)ethyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-azidobenzoic acid and

ethyl 2-isocyanoacetate were used to give ethyl (11-(4-azidobenzoyl)-2-(2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)-5,8-dioxa-2,11-diazatridecan-13-oyl)glycinate (32 mg, 0.041 mmol, 41%) as a yellow gum.

δ_H (400 MHz, $CDCl_3$):	7.78 (br s, 1H), 7.50-7.64 (m, 3H), 7.44-7.49 (m, 1H), 6.98-7.04 (m, 2H), 6.85 (dd, *J* = 8.8, 2.4 Hz, 1H), 6.57 (d, *J* = 2.0 Hz, 1H), 3.83-4.28 (m, 11H), 3.47-3.83 (m, 16H), 3.23 (td, *J* = 11.6, 2.2 Hz, 2H), 3.02 (s, 3H), 2.23 (s, 3H), 1.97-2.12 (m, 1H), 1.31-1.39 (m, 2H), 1.27-1.30 (m, 4H) ppm. Additional signals observed due to rotamers.
δ_C (101 MHz, $CDCl_3$):	172.3, 169.7, 169.3, 162.6, 148.17, 148.15, 147.0, 142.4, 142.0, 137.64, 137.62, 136.9, 135.9, 131.8, 129.9, 129.2, 119.5, 119.0, 118.9, 111.2, 93.1, 70.6, 70.4, 68.9, 67.2, 61.4, 53.6, 50.1, 41.2, 39.7, 38.1, 35.5, 30.6, 29.7, 17.4, 14.2 ppm. Additional signals observed due to rotamers.
LCMS (Method A):	t_R = 0.72 min, 100% by UV, $[M+H]^+$ found: 786.4.
HRMS:	($C_{40}H_{52}N_9O_8$) $[M+H]^+$ requires: 786.3939, $[M+H]^+$ found: 786.3940.
ν_{max} (neat, cm^{-1}):	2919, 2850, 2131, 2088, 1655, 1603, 1497, 1456, 1280, 1198, 1093, 1015, 984, 847, 797, 768.

Compound 2.57

N-(2-(2-(2-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)ethoxy)ethoxy)ethyl)-3-(3-methyl-3*H*-diazirin-3-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)propanamide



Following General Procedure 1 (Section 7.1.2.1), 5-(6-((2-(2-(2-aminoethoxy)ethoxy)ethyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 3-(3-methyl-3*H*-diazirin-3-yl)propanoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to give *N*-(2-(2-(2-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)ethoxy)ethoxy)ethyl)-3-(3-methyl-3*H*-diazirin-3-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)propanamide (31 mg, 0.039 mmol, 39%) as a yellow oil.

δ_H (600 MHz, CD_3OD): 7.94 (d, *J* = 2.6 Hz, 1H), 7.69 (d, *J* = 1.1 Hz, 1H), 7.48 (d, *J* = 8.8 Hz, 1H), 6.88-6.93 (m, 1H), 6.80 (s, 1H), 4.20 (d, *J* = 7.3 Hz, 2H), 4.17 (s, 2H), 3.98-4.11 (m, 2H), 3.83 (br dd, *J* = 11.4, 2.9 Hz, 2H), 3.70-3.74 (m, 2H), 3.68 (s, 3H), 3.56-3.67 (m, 12H), 3.48-3.55 (m, 4H), 3.33-3.41 (m, 2H), 3.25 (td, *J* = 11.7, 1.8 Hz, 2H), 3.06 (s, 3H), 2.86 (t, *J* = 2.4 Hz, 1H), 2.35-2.39 (m, 1H), 2.21 (s, 3H), 2.14 (t, *J* = 7.7 Hz, 1H), 2.02-2.11 (m, 1H), 1.62-1.67 (m, 2H), 1.30-1.36 (m, 2H), 1.24 (qd, *J* = 12.3, 4.4 Hz, 2H), 0.98 (s, 3H) ppm. Additional signals observed due to rotamers. Exchangeable proton not observed.

δ_C (151 MHz, CD_3OD): 175.1, 175.0, 171.6, 171.2, 164.7, 149.9, 148.89, 148.87, 139.4, 139.2, 138.5, 135.6, 135.5, 130.4, 120.1, 112.6, 111.7, 111.5, 94.9, 94.8, 80.77, 80.76, 80.4, 80.32, 80.27, 76.29, 76.26, 76.2, 76.0, 75.7, 72.0, 71.9, 71.8, 71.6, 71.22, 71.20, 70.64, 70.61, 70.4, 70.3, 70.2, 70.14, 70.08, 68.5, 59.24, 59.23, 59.22, 54.62, 54.57, 53.4, 51.0, 50.8, 50.4, 48.8, 40.6, 40.5, 40.27, 40.26, 38.7, 36.9, 36.8, 31.8, 30.94, 30.89, 28.5, 28.3, 26.6, 26.5, 20.11, 20.05, 17.4 ppm. Additional signals observed due to rotamers.

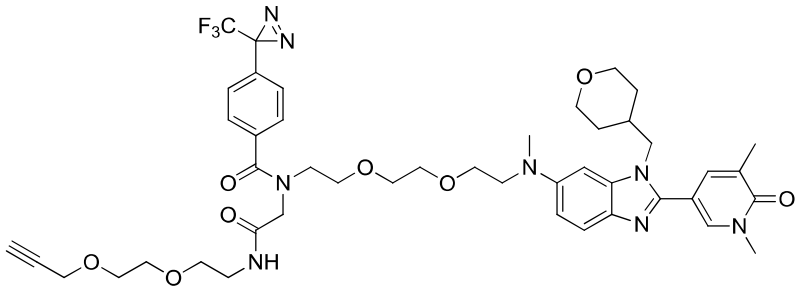
LCMS (Method B): t_R = 0.93 min, 99% by UV, $[M+H]^+$ found: 791.6.

HRMS: ($C_{41}H_{59}N_8O_8$) $[M+H]^+$ requires: 791.4456, $[M+H]^+$ found: 791.4453.

ν_{max} (neat, cm^{-1}): 2916, 2864, 1654, 1611, 1559, 1496, 1447, 1425, 1364, 1262, 1239, 1206, 1015, 986, 915, 850, 800, 769, 636, 511.

Compound 2.58

N-(2-(2-(2-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)ethoxy)ethoxy)ethyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)-4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzamide



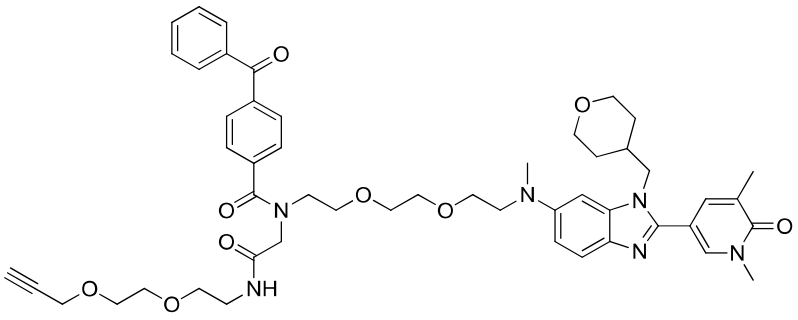
Following General Procedure 1 (Section 7.1.2.1), 5-(6-((2-(2-(2-aminoethoxy)ethoxy)ethyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to give *N*-(2-(2-(2-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)ethoxy)ethoxy)ethyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)-4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzamide (17.4 mg, 0.019 mmol, 19.49 % yield) as a yellow gum.

δ_H (600 MHz, CD_3OD): 7.84-7.97 (m, 1H), 7.65-7.71 (m, 1H), 7.43-7.64 (m, 3H), 7.23-7.31 (m, 2H), 6.90 (dd, *J* = 9.0, 2.4 Hz, 1H), 6.74-6.82 (m, 1H), 4.11-4.24 (m, 5H), 3.98 (br s, 1H), 3.77-3.85 (m, 2H), 3.70-3.76 (m, 3H), 3.49-3.69 (m, 17H), 3.38-3.49 (m, 3H), 3.17-3.27 (m, 2H), 2.98-3.07 (m, 3H), 2.81-2.87 (m, 1H), 2.21 (s, 3H), 1.99-2.10 (m,

	1H), 1.30 (br d, J = 6.6 Hz, 2H), 1.16-1.27 (m, 2H) ppm. Additional signals observed due to rotamers. Exhangeable proton not observed.
δ_C (151 MHz, CD_3OD):	172.1, 169.5, 163.1, 148.3, 147.4, 147.3, 137.9, 137.6, 137.5, 136.9, 133.9, 130.1, 129.9, 129.7, 128.9, 127.9, 127.6, 127.2, 126.5, 126.2, 122.9, 121.1, 119.3, 118.5, 111.0, 110.1, 110.0, 93.4, 79.2, 74.7, 70.3, 70.2, 70.1, 69.70, 69.65, 69.2, 69.0, 68.8, 68.7, 68.5, 68.0, 66.9, 57.7, 53.1, 53.0, 50.1, 49.3, 48.5, 46.6, 39.1, 38.7, 37.2, 35.3, 30.2, 28.4, 28.1, 27.9, 27.6, 15.8, 14.0 ppm. Additional signals observed due to rotamers.
LCMS (Method B):	t_R = 1.08 min, 97% by UV, $[M+H]^+$ found: 893.7.
HRMS:	($C_{45}H_{56}F_3N_8O_8$) $[M+H]^+$ requires: 893.4173, found $[M+H]^+$ found: 893.4166.
ν_{max} (neat, cm^{-1}):	2922, 2866, 1656, 1613, 1560, 1497, 1460, 1426, 1345, 1262, 1232, 1184, 1140, 1094, 986, 938, 833, 801, 769, 718, 511.

Compound 2.59

4-Benzoyl-*N*-(2-(2-(2-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)ethoxy)ethoxy)ethyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)benzamide



Following General Procedure 1 (Section 7.1.2.1), 5-(6-((2-(2-(2-aminoethoxy)ethoxy)ethyl)(methyl)amino)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-benzoylbenzoic acid

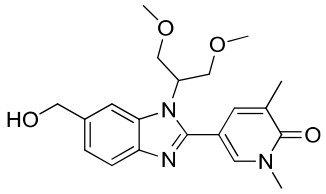
and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to give 4-benzoyl-*N*-(2-(2-(2-((2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1-((tetrahydro-2*H*-pyran-4-yl)methyl)-1*H*-benzo[*d*]imidazol-6-yl)(methyl)amino)ethoxy)ethoxy)ethyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)benzamide (47 mg, 0.052 mmol, 52%) as a yellow gum.

δ_H (600 MHz, CD_3OD):	7.88-7.94 (m, 1H), 7.47-7.77 (m, 10H), 7.39-7.47 (m, 1H), 6.82-6.93 (m, 1H), 6.71-6.82 (m, 1H), 4.24 (s, 1H), 4.08-4.20 (m, 4H), 4.05 (br s, 1H), 3.69-3.83 (m, 6H), 3.64-3.68 (m, 6H), 3.53-3.63 (m, 10H), 3.49-3.52 (m, 1H), 3.45-3.49 (m, 1H), 3.43 (br t, J = 4.8 Hz, 1H), 3.32-3.37 (m, 2H), 3.15-3.25 (m, 2H), 2.97-3.07 (m, 3H), 2.19 (s, 3H), 1.96-2.08 (m, 1H), 1.13-1.36 (m, 4H) ppm. Additional signals observed due to rotamers. Exchangeable proton not observed.
δ_C (151 MHz, CD_3OD):	197.7, 197.6, 174.0, 173.9, 171.0, 170.8, 169.3, 164.6, 149.83, 149.81, 148.93, 148.90, 148.86, 148.8, 141.4, 141.3, 141.2, 139.9, 139.7, 139.4, 139.2, 138.5, 138.4, 135.6, 135.5, 134.3, 134.20, 134.16, 131.3, 131.13, 131.08, 130.4, 129.8, 128.7, 128.5, 128.0, 120.1, 112.6, 112.5, 111.7, 94.90, 94.87, 80.33, 80.28, 76.3, 76.0, 75.8, 71.9, 71.8, 71.7, 71.6, 71.3, 71.2, 70.7, 70.6, 70.34, 70.27, 70.2, 70.1, 69.5, 68.4, 59.2, 54.7, 54.6, 54.6, 51.7, 50.8, 50.04, 50.00, 48.1, 41.2, 40.65, 40.58, 40.3, 40.32, 40.25, 38.7, 36.83, 36.79, 31.8, 17.4 ppm. Additional signals observed due to rotamers.
LCMS (Method B):	t_R = 1.00 min, 95% by UV, $[M+H]^+$ found: 889.7.
HRMS:	($C_{50}H_{61}N_6O_9$) $[M+H]^+$ requires: 889.4500, $[M+H]^+$ found: 889.4494.
ν_{max} (neat, cm^{-1}):	3304, 2927, 2861, 1653, 1609, 1497, 1447, 1425, 1275, 1239, 1091, 924, 798, 768, 702.

7.1.4 Chapter 4 – synthetic chemistry

Compound 4.8

5-(1-(1,3-Dimethoxypropan-2-yl)-6-(hydroxymethyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one

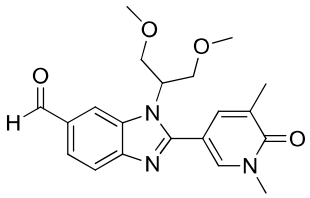


A flask was loaded with (3-fluoro-4-nitrophenyl)methanol (10 g, 58 mmol), 2-methyltetrahydrofuran (200 mL), DIPEA (20.4 mL, 117 mmol) and 1,3-dimethoxypropan-2-amine (11 mL, 88 mmol) and heated to reflux (110 °C) for 5 d under nitrogen. The reaction mixture was concentrated *in vacuo* and the residue was partitioned between ethyl acetate (300 mL) and 50% saturated ammonium chloride in water (400 mL). The organic layer was collected, and the remaining aqueous layer was extracted with ethyl acetate (2 × 200 mL). The organics were combined and dried by passing through a hydrophobic frit before being concentrated *in vacuo*. The residue was suspended in 2:1 (*v/v*) ethanol:water (225 mL) and 1,5-dimethyl-6-oxo-1,6-dihydropyridine-3-carbaldehyde (10.6 g, 70.1 mmol) was added. The reaction mixture was heated to 80 °C and sodium hydrosulfite (30.5 g, 175 mmol) was added portion-wise with magnetic stirring. The reaction mixture was heated under reflux (100 °C) for 1.5 h. The reaction mixture was filtered through cotton wool and the retentate was washed with ethyl acetate (*ca.* 200 mL). The filtrate was concentrated *in vacuo* and partitioned between ethyl acetate (400 mL) and 10% (*v/v*) brine in water (600 mL). The organic layer was collected, and the remaining aqueous layer was extracted with ethyl acetate (2 × 300 mL), followed by 4:1 (*v/v*) chloroform:2-propanol (2 × 200 mL) followed by 10% (*v/v*) 2-propanol in dichloromethane (2 × 200 mL). The organic extractions were dried by passing through a hydrophobic frit and combined before being concentrated *in vacuo*. The residue was purified by flash chromatography (0–16% (*v/v*) methanol in dichloromethane over 15 CV) to give 5-(1-(1,3-dimethoxypropan-2-yl)-6-(hydroxymethyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one (6.96 g, 18.7 mmol, 32%) as a white powder.

δ_H (400 MHz, CD_3OD):	8.02 (d, *J* = 2.4 Hz, 1H), 7.72-7.82 (m, 2H), 7.64 (d, *J* = 8.3 Hz, 1H), 7.30 (dd, *J* = 8.3, 1.5 Hz, 1H), 4.89 (tt, *J* = 9.0, 4.4 Hz, 1H), 4.76 (s, 2H), 4.11 (dd, *J* = 10.3, 9.3 Hz, 2H), 3.83 (dd, *J* = 10.3, 4.4 Hz, 2H), 3.66 (s, 3H), 3.26 (s, 6H), 2.20 (s, 3H) ppm. Exchangeable proton not observed.
δ_C (101 MHz, CD_3OD):	164.8, 153.6, 143.3, 140.7, 139.9, 138.2, 134.9, 130.0, 123.5, 120.0, 112.1, 111.0, 71.4, 65.7, 59.52, 59.49, 38.9, 17.4 ppm.
LCMS (Method B):	t_R = 0.66 min, 100% by UV, $[M+H]^+$ found: 372.3.
HRMS:	($C_{20}H_{26}N_3O_4$) $[M+H]^+$ requires: 372.1923, $[M+H]^+$ found: 372.1917.
ν_{max} (neat, cm^{-1}):	3349, 2810, 1652, 1600, 1581, 1557, 1516, 1446, 1431, 1383, 1224, 1178, 1113, 1095, 1068, 1054, 980, 970, 809, 769, 762, 517.
M.pt. (CH_2Cl_2):	219–220 °C.

Compound 4.9

1-(1,3-Dimethoxypropan-2-yl)-2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1*H*-benzo[*d*]imidazole-6-carbaldehyde

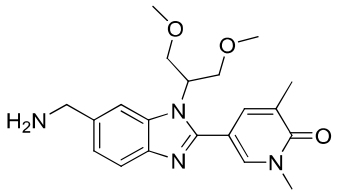


5-(1-(1,3-Dimethoxypropan-2-yl)-6-(hydroxymethyl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one (8.82 g, 23.8 mmol) was suspended in acetone (250 mL) and manganese dioxide (83.0 g, 950 mmol) was added. The reaction mixture was stirred at rt under nitrogen for 16 h. The reaction mixture was filtered through celite and the retentate was washed with acetone (*ca*. 500 mL). The filtrate was concentrated *in vacuo* 1-(1,3-dimethoxypropan-2-yl)-2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1*H*-benzo[*d*]imidazole-6-carbaldehyde (6.84 g, 18.5 mmol, 78%) as a white powder.

δ_H (400 MHz, CD_3OD):	10.07 (s, 1H), 8.40 (s, 1H), 8.07-8.15 (m, 1H), 7.85-7.92 (m, 1H), 7.80 (d, *J* = 8.3 Hz, 2H), 5.01 (tt, *J* = 8.8, 4.4 Hz, 1H), 4.15 (dd, *J* = 10.3, 9.3 Hz, 2H), 3.87 (dd, *J* = 10.3, 4.4 Hz, 2H), 3.67 (s, 3H), 3.27 (s, 6H), 2.21 (s, 3H) ppm.
δ_C (101 MHz, CD_3OD):	193.9, 164.8, 157.2, 148.6, 141.1, 139.5, 135.2, 133.4, 130.2, 125.1, 120.7, 117.1, 110.3, 71.5, 60.0, 59.5, 38.9, 17.4 ppm.
LCMS (Method B):	t_R = 0.79 min, 100% by UV, $[M+H]^+$ found: 370.3.
HRMS:	($C_{20}H_{24}N_3O_4$) $[M+H]^+$ requires: 370.1767, $[M+H]^+$ found: 370.1761.
ν_{max} (neat, cm^{-1}):	3229, 3077, 2920, 2815, 2406, 2055, 1688, 1659, 1622, 1516, 1379, 1327, 1187, 1112, 1092, 1073, 978, 965, 821, 720, 516.
M.pt. (acetone):	166–167 °C.

Compound 4.2

5-(6-(Aminomethyl)-1-(1,3-dimethoxypropan-2-yl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one



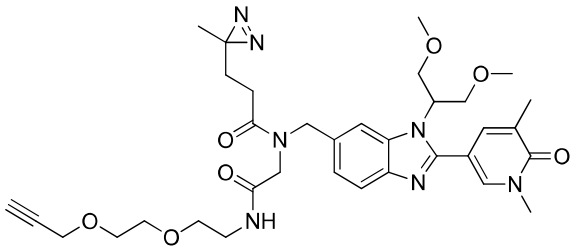
A flask was loaded with 1-(1,3-dimethoxypropan-2-yl)-2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1*H*-benzo[*d*]imidazole-6-carbaldehyde (3.0 g, 8.1 mmol), hydroxylamine hydrochloride (0.68 g, 9.74 mmol), sodium acetate (0.80 g, 9.74 mmol) and methanol (80 mL). The reaction mixture was stirred under nitrogen for 2 h at rt. The reaction mixture was concentrated *in vacuo* and the residue was partitioned between 4:1 (*v/v*) chloroform:2-isopropanol (100 mL) and water (100 mL). The organic layer was collected, and the remaining aqueous layer was extracted with 4:1 (*v/v*) chloroform:2-isopropanol (2 × 50 mL). The organics were combined and dried

by passing through a hydrophobic frit before being concentrated *in vacuo*. The residue was dissolved in 0.1 M HCl in methanol (*ca.* 100 mL) and hydrogenated (1 atm) over 10% (*w/w*) palladium on carbon (0.864 g, 0.812 mmol) at rt for 16 h. The reaction mixture was then filtered through celite and the retentate was washed with methanol (*ca*. 200 mL). The filtrate was concentrated *in vacuo* and the residue was dissolved in methanol (*ca.* 50 mL) and loaded onto a Biotage SCX-2 cartridge which was pre-washed with methanol. The cartridge was then washed with methanol (300 mL) and then 2M ammonia in methanol (300 mL). The basic wash was concentrated *in vacuo* and the residue was purified by flash chromatography 0–100% (*v/v*) (10:1 (*v/v*) methanol:ammonium hydroxide) in ethyl acetate over 18 CV) to give 5-(6-(aminomethyl)-1-(1,3-dimethoxypropan-2-yl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one (2.27 g, 6.14 mmol, 76%) as a white powder.

δ_H (400 MHz, CD_3OD):	8.02 (d, *J* = 2.4 Hz, 1H), 7.75-7.79 (m, 1H), 7.74 (s, 1H), 7.63 (d, *J* = 8.3 Hz, 1H), 7.29 (dd, *J* = 8.3, 1.5 Hz, 1H), 4.89 (tt, *J* = 9.2, 4.5 Hz, 1H), 4.13 (dd, *J* = 10.3, 8.8 Hz, 2H), 3.95 (s, 2H), 3.84 (dd, *J* = 10.3, 4.4 Hz, 2H), 3.66 (s, 3H), 3.26 (s, 6H), 2.20 (s, 3H) ppm. Exchangeable protons not observed.
δ_C (101 MHz, CD_3OD):	164.8, 153.6, 143.0, 140.7, 139.9, 139.0, 135.0, 130.0, 124.0, 120.2, 112.4, 111.0, 71.4, 59.52, 59.49, 47.2, 38.9, 17.4 ppm.
LCMS (Method B):	t_R = 0.64 min, 98% by UV, $[M+H]^+$ found: 371.3.
HRMS:	($C_{20}H_{27}N_4O_3$) $[M+H]^+$ requires: 371.2083, $[M+H]^+$ found: 371.2080.
ν_{max} (neat, cm^{-1}):	3318, 2920, 1654, 1608, 1514, 1447, 1428, 1360, 1110, 1090, 1067, 965, 822, 769, 658, 554, 538.
M.pt. (CH_2Cl_2):	175–177 °C.

Compound 4.3

N-((1-(1,3-dimethoxypropan-2-yl)-2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)-3-(3-methyl-3*H*-diazirin-3-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)propanamide



Following General Procedure 1 (Section 7.1.2.1), 5-(6-(aminomethyl)-1-(1,3-dimethoxypropan-2-yl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 3-(3-methyl-3*H*-diazirin-3-yl)propanoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to give *N*-((1-(1,3-dimethoxypropan-2-yl)-2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)-3-(3-methyl-3*H*-diazirin-3-yl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)propanamide (35 mg, 0.053 mmol, 53%) as a clear yellow gum.

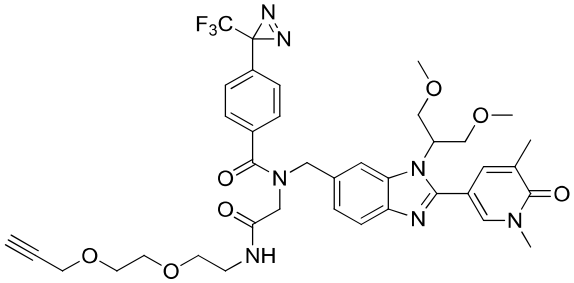
δ_H (400 MHz, CD_3OD): 8.03 (d, J = 2.4 Hz, 1H), 7.59-7.79 (m, 3H), 7.22 (ddd, J = 8.3, 4.4, 1.5 Hz, 1H), 4.85-4.95 (m, 1H), 4.80-4.82 (m, 1H), 4.76 (s, 1H), 4.15-4.20 (m, 2H), 4.11 (t, J = 9.8 Hz, 2H), 4.02-4.08 (m, 2H), 3.82 (td, J = 10.0, 4.4 Hz, 2H), 3.63-3.68 (m, 5H), 3.58-3.62 (m, 1H), 3.54-3.57 (m, 1H), 3.46-3.51 (m, 1H), 3.34-3.41 (m, 2H), 3.32 (br d, J = 1.5 Hz, 1H), 3.25-3.28 (m, 6H), 2.81-2.88 (m, 1H), 2.41 (t, J = 7.3 Hz, 1H), 2.32 (t, J = 7.3 Hz, 1H), 2.20 (s, 3H), 1.71-1.84 (m, 2H), 0.96-1.08 (m, 3H) ppm. Additional signals observed due to rotamers. Exchangeable proton not observed.

δ_C (101 MHz, CD_3OD): 175.1, 175.0, 171.1, 170.5, 164.7, 154.1, 153.9, 143.6, 143.5, 140.72, 140.69, 139.82, 139.79, 135.3, 135.1, 133.4, 133.0, 130.05, 130.00, 124.6, 123.3, 120.7, 120.2,

	113.6, 112.5, 110.84, 110.79, 80.7, 76.24, 76.20, 71.6, 71.4, 71.2, 71.1, 70.6, 70.5, 70.33, 70.30, 59.6, 59.5, 59.2, 54.1, 51.9, 51.3, 50.5, 40.5, 38.9, 30.9, 30.7, 28.51, 28.45, 26.75, 26.70, 20.22, 20.17, 17.4 ppm. Additional signals observed due to rotamers.
LCMS (Method B):	t_R = 0.87 min, 100% by UV, $[M+H]^+$ found: 664.5.
HRMS:	($C_{34}H_{46}N_7O_7$) $[M+H]^+$ requires: 664.3459, $[M+H]^+$ found: 664.3465.
ν_{max} (neat, cm^{-1}):	3287, 2924, 1655, 1613, 1562, 1517, 1361, 1285, 1101, 1033, 962, 919, 809, 770, 660, 511.

Compound 4.4

N-((1-(1,3-dimethoxypropan-2-yl)-2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)-4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzamide



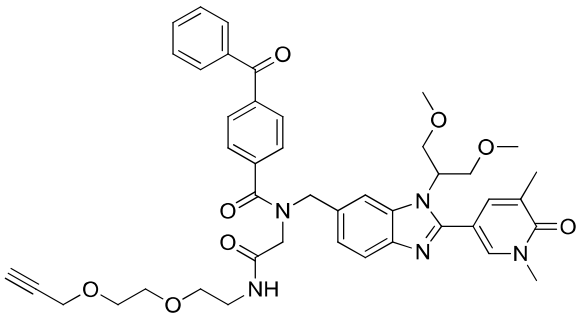
Following General Procedure 1 (Section 7.1.2.1), 5-(6-(aminomethyl)-1-(1,3-dimethoxypropan-2-yl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to give *N*-((1-(1,3-dimethoxypropan-2-yl)-2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)-4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzamide (40 mg, 0.053 mmol, 53%) as a clear glass.

δ_H (400 MHz, CD_3OD):	8.02 (br s, 1H), 7.57-7.89 (m, 5H), 7.05-7.42 (m, 3H), 4.85-4.97 (m, 2H), 4.73 (br s, 1H), 4.02-4.22 (m, 5H), 3.76-3.91 (m, 3H), 3.37-3.70 (m, 11H), 3.26 (s, 6H),

	2.81 (t, *J* = 2.2 Hz, 1H), 2.20 (s, 3H) ppm. Additional signals observed due to rotamers. Exchangeable proton not observed.
δ_C (101 MHz, CD_3OD):	173.7, 173.6, 170.5, 170.3, 164.8, 154.2, 154.1, 143.6, 140.7, 139.8, 139.0, 138.7, 135.30, 135.27, 135.20, 135.17, 132.8, 132.5, 131.9, 130.1, 129.0, 128.9, 128.1, 124.65, 124.62, 123.35, 123.30, 122.2, 120.6, 120.4, 114.3, 113.0, 110.8, 80.6, 76.2, 71.61, 71.56, 71.2, 71.1, 70.6, 70.5, 70.3, 59.6, 59.5, 59.2, 55.8, 52.4, 51.5, 40.5, 38.9, 29.8, 29.4, 17.4 ppm. Additional signals observed due to rotamers.
LCMS (Method B):	t_R = 1.08 min, 100% by UV, $[M+H]^+$ found: 766.5.
HRMS:	($C_{38}H_{43}F_3N_7O_7$) $[M+H]^+$ requires: 766.3176, $[M+H]^+$ found: 766.3179.
ν_{max} (neat, cm^{-1}):	3279, 2927, 1654, 1610, 1515, 1451, 1429, 1345, 1286, 1230, 1185, 1151, 1099, 1033, 1001, 962, 938, 881, 769, 730, 546, 513.
M.pt. (CH_3OH):	60–62 °C

Compound 4.5

4-Benzoyl-*N*-((1-(1,3-dimethoxypropan-2-yl)-2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)benzamide



Following General Procedure 1 (Section 7.1.2.1), 5-(6-(aminomethyl)-1-(1,3-dimethoxypropan-2-yl)-1*H*-benzo[*d*]imidazol-2-yl)-1,3-dimethylpyridin-2(1*H*)-one, 4-benzoylbenzoic acid and 3-(2-(2-isocyanoethoxy)ethoxy)prop-1-yne were used to

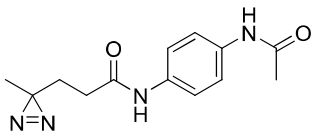
give 4-benzoyl-*N*-((1-(1,3-dimethoxypropan-2-yl)-2-(1,5-dimethyl-6-oxo-1,6-dihydropyridin-3-yl)-1*H*-benzo[*d*]imidazol-6-yl)methyl)-*N*-(2-oxo-2-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)ethyl)benzamide (53 mg, 0.070 mmol, 70%) as a clear glass.

δ_H (400 MHz, CD_3OD):	8.02 (br d, *J* = 8.8 Hz, 1H), 7.58-7.92 (m, 10H), 7.46-7.57 (m, 2H), 7.12-7.39 (m, 1H), 4.84-5.02 (m, 2H), 4.02-4.28 (m, 5H), 3.75-3.99 (m, 3H), 3.58-3.70 (m, 6H), 3.49-3.57 (m, 2H), 3.38-3.47 (m, 2H), 3.32-3.36 (m, 2H), 3.22-3.29 (m, 6H), 2.76-2.84 (m, 1H), 2.19 (s, 3H) ppm. Additional signals observed due to rotamers. Exchangeable proton not observed.
δ_C (101 MHz, CD_3OD):	197.6, 174.0, 173.9, 170.5, 170.3, 164.7, 154.2, 154.1, 143.6, 141.1, 140.9, 140.7, 140.3, 140.2, 139.8, 138.5, 135.3, 135.2, 134.2, 132.8, 132.6, 131.4, 131.3, 131.2, 130.1, 129.8, 128.2, 128.1, 124.6, 123.4, 120.6, 120.4, 114.3, 113.1, 110.80, 110.76, 80.7, 76.2, 71.63, 71.56, 71.2, 71.1, 70.7, 70.5, 70.3, 59.6, 59.5, 59.2, 55.9, 52.5, 51.4, 40.5, 38.9, 17.4 ppm. Additional signals observed due to rotamers.
LCMS (Method B):	t_R = 0.99 min, 100% by UV, $[M+H]^+$ found: 762.6.
HRMS:	($C_{43}H_{48}N_5O_8$) $[M+H]^+$ requires: 762.3503, $[M+H]^+$ found: 762.3506.
ν_{max} (neat, cm^{-1}):	3285, 2922, 1652, 1607, 1557, 1506, 1360, 1315, 1276, 1098, 1033, 1000, 962, 923, 859, 803, 769, 702, 655, 540, 511.
M.pt. (CH_3OH):	61–63 °C

7.1.5 Chapter 5 – synthetic chemistry

Ph1

N-(4-acetamidophenyl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide

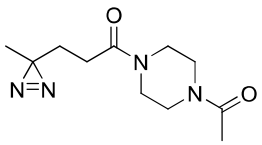


Following General Procedure 2 (Section 7.1.2.2), *N*-(4-aminophenyl)acetamide was used to afford *N*-(4-acetamidophenyl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide (29 mg, 0.11 mmol, 55%) as a white powder.

δ_H (400 MHz, CD_3OD):	7.47 (m, 4H), 2.24 (m, 2H), 2.10 (s, 3H), 1.74 (m, 2H), 1.03 (s, 3H) ppm. Exchangeable protons not observed.
δ_C (101 MHz, CD_3OD):	172.8, 171.6, 136.2, 136.1, 121.8, 121.7, 32.2, 31.4, 26.4, 23.8, 19.9 ppm.
LCMS (Method B):	t_R = 0.72 min, 100% by UV, $[M+H]^+$ found: 261.2.
HRMS:	($C_{13}H_{17}N_4O_2$) $[M+H]^+$ requires: 261.1352, $[M+H]^+$ found: 261.1351.
ν_{max} (neat, cm^{-1}):	3299, 1646, 1599, 1550, 1535, 1517, 1448, 1392, 1315, 1300, 1265, 1248, 968, 845, 814, 795, 712, 602, 511, 487.
M.pt. (MeCN/water):	189–191 °C

Ph2

1-(4-Acetylpiperazin-1-yl)-3-(3-methyl-3*H*-diazirin-3-yl)propan-1-one

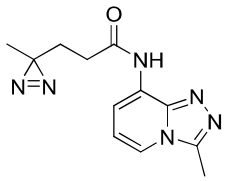


Following General Procedure 2 (Section 7.1.2.2), 1-(piperazin-1-yl)ethan-1-one was used to afford 1-(4-acetylpiperazin-1-yl)-3-(3-methyl-3*H*-diazirin-3-yl)propan-1-one (26 mg, 0.11 mmol, 55%) as a yellow oil.

δ_H (400 MHz, CD_3OD):	3.43-3.77 (m, 8H), 2.30 (t, J = 7.5 Hz, 2H), 2.13 (s, 3H), 1.69 (t, J = 7.5 Hz, 2H), 1.04 (s, 3H) ppm.
δ_C (101 MHz, CD_3OD):	172.9, 172.2, 46.8, 42.7, 31.0, 28.2, 26.4, 21.3, 20.0 ppm.
LCMS (Method B):	t_R = 0.60 min, 100% by UV, $[M+H^+]$ found: 239.2.
HRMS:	($C_{11}H_{18}N_4O_2Na$) $[M+Na]^+$ requires: 261.1327, $[M+Na]^+$ found: 261.1326.
ν_{max} (neat, cm^{-1}):	3463, 2923, 2864, 1627, 1427, 1387, 1365, 1284, 1240, 1225, 1173, 1115, 1050, 1021, 592, 545.

Ph3

3-(3-Methyl-3*H*-diazirin-3-yl)-*N*-(3-methyl-[1,2,4]triazolo[4,3-a]pyridin-8-yl)propenamide



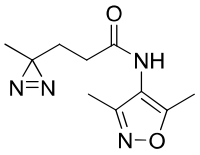
Following General Procedure 2 (Section 7.1.2.2), 3-methyl-[1,2,4]triazolo[4,3-*a*]pyridin-8-amine was used to afford 3-(3-methyl-3*H*-diazirin-3-yl)-*N*-(3-methyl-[1,2,4]triazolo[4,3-*a*]pyridin-8-yl)propanamide (19 mg, 0.072 mmol, 36%) as a brown powder.

δ_H (400 MHz, CD_3OD):	8.09 (dd, J = 7.5, 1.0 Hz, 1H), 7.98 (dd, J = 7.0, 1.0 Hz, 1H), 6.97 (dd, J = 7.5, 7.0 Hz, 1H), 2.74 (s, 3H), 2.44 (t, J = 7.5 Hz, 2H), 1.80 (t, J = 7.5 Hz, 2H), 1.06 (s, 3H) ppm. Exchangeable proton not observed.
δ_C (101 MHz, CD_3OD):	173.7, 146.8, 146.3, 127.2, 119.5, 115.9, 115.7, 32.2, 31.1, 26.4, 19.9, 10.0 ppm.
LCMS (Method B):	t_R = 0.72 min, 100% by UV, $[M+H]^+$ found: 259.2.
HRMS:	($C_{12}H_{15}N_6O$) $[M+H]^+$ requires: 259.1307, found $[M+H]^+$ found: 259.1302.

ν_{max} (neat, cm^{-1}):	1687, 1551, 1493, 1451, 1434, 1394, 1384, 1347, 1321, 1282, 1219, 1176, 1126, 1052, 766, 738, 722, 671, 624, 539.
M.pt. (MeCN/water):	144 °C (decomp.)

Ph4

N-(3,5-dimethylisoxazol-4-yl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide

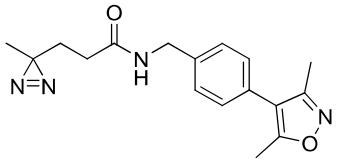


Following General Procedure 2 (Section 7.1.2.2), 3,5-dimethylisoxazol-4-amine was used to afford *N*-(3,5-dimethylisoxazol-4-yl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide (25 mg, 0.11 mmol, 57%) as a brown powder.

δ_H (400 MHz, CD_3OD):	2.30 (s, 3H), 2.26 (t, J = 7.5 Hz, 2H), 2.16 (s, 3H), 1.78 (t, J = 7.5 Hz, 2H), 1.05 (s, 3H) ppm. Exchangeable proton not observed.
δ_C (101 MHz, CD_3OD):	174.3, 165.3, 159.4, 115.2, 31.14, 31.06, 26.5, 20.0, 11.0, 9.8 ppm.
LCMS (Method B):	t_R = 0.71 min, 98% by UV, $[M+H]^+$ found: 223.2.
HRMS:	($C_{10}H_{15}N_4O_2$) $[M+H]^+$ requires: 223.1195, $[M+H]^+$ found: 223.1194.
ν_{max} (neat, cm^{-1}):	3263, 1668, 1651, 1524, 1449, 1419, 1384, 1314, 1280, 1238, 1194, 1051, 968, 879, 753, 696, 674, 612, 584, 496.
M.pt. (MeCN/water):	79–81 °C

Ph5

N-(4-(3,5-dimethylisoxazol-4-yl)benzyl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide

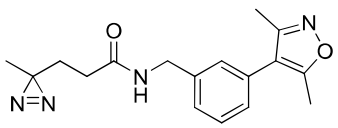


Following General Procedure 2 (Section 7.1.2.2), (4-(3,5-dimethylisoxazol-4-yl)phenyl)methanamine was used to afford *N*-(4-(3,5-dimethylisoxazol-4-yl)benzyl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide (37 mg, 0.12 mmol, 59%) as a brown solid.

δ_H (400 MHz, CD_3OD):	7.41 (d, *J* = 8.5 Hz, 2H), 7.29 (d, *J* = 8.5 Hz, 2H), 4.41 (s, 2H), 2.38 (s, 3H), 2.22 (s, 3H), 2.15 (m, 2H), 1.70 (m, 2H), 1.00 (s, 3H) ppm. Exchangeable proton not observed.
δ_C (101 MHz, CD_3OD):	174.6, 167.0, 160.2, 139.8, 130.6, 130.4, 129.3, 117.8, 44.0, 31.6, 31.4, 26.4, 19.9, 11.5, 10.8 ppm.
LCMS (Method B):	t_R = 0.97 min, 93% by UV, $[M+H]^+$ found: 313.2.
HRMS:	($C_{17}H_{21}N_4O_2$) $[M+H]^+$ requires: 313.1665, $[M+H]^+$ found: 313.1658.
ν_{max} (neat, cm^{-1}):	3297, 2925, 1644, 1538, 1422, 1440, 1406, 1385, 1289, 1033, 1000, 843, 828, 808, 739, 722, 680, 653, 586, 539, 557.
M.pt. (MeCN/water):	96–98 °C

Ph6

N-(3-(3,5-dimethylisoxazol-4-yl)benzyl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide



Following General Procedure 2 (Section 7.1.2.2), (3-(3,5-dimethylisoxazol-4-yl)phenyl)methanamine was used to afford *N*-(3-(3,5-dimethylisoxazol-4-yl)benzyl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide (30 mg, 0.10 mmol, 47%) as a yellow oil.

δ_H (400 MHz, CD_3OD): 7.39-7.46 (m, 1H), 7.32 (d, J = 8.0 Hz, 1H), 7.27 (s, 1H), 7.22 (d, J = 7.5 Hz, 1H), 4.41 (s, 2H), 2.39 (s, 3H), 2.23 (s, 3H), 2.13 (t, J = 7.5 Hz, 2H), 1.69 (t, J = 7.5 Hz, 2H), 0.98 (s, 3H) ppm. Exchangeable proton not observed.

δ_C (101 MHz, CD_3OD): 174.6, 167.1, 160.1, 141.0, 131.8, 130.3, 129.6, 129.2, 128.2, 118.0, 44.1, 31.5, 31.4, 26.4, 19.9, 11.5, 10.8 ppm.

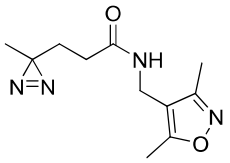
LCMS (Method B): t_R = 0.98 min, 100% by UV, $[M+H]^+$ found: 313.2.

HRMS: ($C_{17}H_{21}N_4O_2$) $[M+H]^+$ requires 313.1665, found $[M+H]^+$ 313.1663.

ν_{max} (neat, cm^{-1}): 3297, 2926, 1648, 1606, 1585, 1545, 1485, 1422, 1385, 1347, 1323, 1295, 1244, 1188, 1033, 895, 796, 707, 681, 617.

Ph7

N-((3,5-dimethylisoxazol-4-yl)methyl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide



Following General Procedure 2 (Section 7.1.2.2), (3,5-dimethylisoxazol-4-yl)methanamine was used to afford *N*-((3,5-dimethylisoxazol-4-yl)methyl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide (26 mg, 0.11 mmol, 55%) as a white powder.

δ_H (400 MHz, CD_3OD): 4.13 (s, 2H), 2.39 (s, 3H), 2.24 (s, 3H), 2.06 (t, J = 7.5 Hz, 2H), 1.66 (t, J = 7.5 Hz, 2H), 0.97 (s, 3H) ppm. Exchangeable proton not observed.

δ_C (101 MHz, CD_3OD): 174.4, 168.8, 161.2, 112.8, 32.4, 31.4, 31.2, 26.4, 19.9, 11.0, 10.1 ppm.

LCMS (Method B): t_R = 0.73 min, 100% by UV, $[M+H]^+$ found: 237.2.

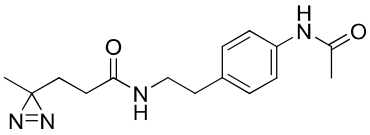
HRMS: ($C_{11}H_{17}N_4O_2$) $[M+H]^+$ requires: 237.1352, $[M+H]^+$ found: 237.1350.

ν_{max} (neat, cm^{-1}): 3299, 1631, 1585, 1545, 1448, 1387, 1287, 1264, 1251, 1238, 1198, 1086, 884, 767, 735, 669, 605, 573, 526, 478.

M.pt. (MeCN/water): 116–117 °C

Ph8

N-(4-acetamidophenethyl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide



Following General Procedure 2 (Section 7.1.2.2), *N*-(4-(2-aminoethyl)phenyl)acetamide was used to afford *N*-(4-acetamidophenethyl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide (26 mg, 0.091 mmol, 45%) as a white powder.

δ_H (400 MHz, CD_3OD): 7.45 (d, J = 8.5 Hz, 2H), 7.16 (d, J = 8.5 Hz, 2H), 3.38 (t, J = 7.0 Hz, 2H), 2.76 (t, J = 7.0 Hz, 2H), 2.10 (s, 3H), 1.99-2.07 (m, 2H), 1.57-1.66 (m, 2H), 0.97 (s, 3H) ppm. Exchangeable protons not observed.

δ_C (101 MHz, CD_3OD): 174.6, 171.7, 138.3, 136.5, 130.3, 121.6, 42.1, 36.0, 31.7, 31.4, 26.4, 23.9, 19.8 ppm.

LCMS (Method B): t_R = 0.73 min, 100% by UV, $[M+H]^+$ found: 289.2.

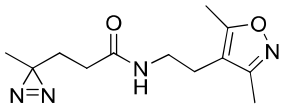
HRMS: ($C_{15}H_{21}N_4O_2$) $[M+H]^+$ requires: 289.1665, $[M+H]^+$ found: 289.1661.

ν_{max} (neat, cm^{-1}): 3302, 1634, 1596, 1553, 1529, 1518, 1446, 1398, 1372, 1312, 1265, 1192, 1115, 839, 793, 714, 602, 550, 513, 464.

M.pt. (MeCN/water): 144–145 °C

Ph9

N-(2-(3,5-dimethylisoxazol-4-yl)ethyl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide

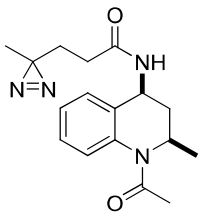


Following General Procedure 2 (Section 7.1.2.2), 2-(3,5-dimethylisoxazol-4-yl)ethan-1-amine was used to afford *N*-(2-(3,5-dimethylisoxazol-4-yl)ethyl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide (27 mg, 0.11 mmol, 54%) as a yellow oil.

δ_H (400 MHz, CD_3OD):	3.28 (t, *J* = 7.0 Hz, 2H), 2.55 (t, *J* = 7.0 Hz, 2H), 2.32 (s, 3H), 2.22 (s, 3H), 1.99-2.08 (m, 2H), 1.59-1.66 (m, 2H), 0.99 (s, 3H) ppm. Exchangeable proton not observed.
δ_C (101 MHz, CD_3OD):	174.6, 167.6, 161.3, 112.5, 40.0, 31.5, 31.3, 26.4, 23.2, 19.9, 10.9, 10.2 ppm.
LCMS (Method B):	t_R = 0.76 min, 98% by UV, $[M+H]^+$ found: 251.2.
HRMS:	($C_{12}H_{19}N_4O_2$) $[M+H]^+$ requires: 251.1508, $[M+H]^+$ found: 251.1507.
ν_{max} (neat, cm^{-1}):	3289, 2930, 1643, 1551, 1447, 1426, 1385, 1329, 1293, 1327, 1260, 1223, 1194, 1105, 1032, 893, 749, 680, 657, 567.

Ph10

N-((2*R*,4*S*)-1-acetyl-2-methyl-1,2,3,4-tetrahydroquinolin-4-yl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide

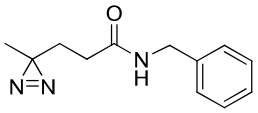


Following General Procedure 2 (Section 7.1.2.2), 1-((2*R*,4*S*)-4-amino-2-methyl-3,4-dihydroquinolin-1(2*H*)-yl)ethan-1-one was used to afford *N*-((2*R*,4*S*)-1-acetyl-2-methyl-1,2,3,4-tetrahydroquinolin-4-yl)-3-(3-methyl-3*H*-diazirin-3-yl)propanamide (33 mg, 0.11 mmol, 53%) as a white powder.

δ_H (400 MHz, CD_3OD):	7.22-7.38 (m, 4H), 4.80-4.89 (m, 1H), 4.71-4.78 (m, 1H), 2.47-2.58 (m, 1H), 2.19-2.33 (m, 2H), 2.12 (s, 3H), 1.71-1.86 (m, 2H), 1.24-1.37 (m, 1H), 1.12 (d, J = 6.5 Hz, 3H), 1.06 (s, 3H) ppm. Exchangeable proton not observed.
δ_C (101 MHz, CD_3OD):	174.6, 172.2, 138.0, 137.4, 128.5, 127.64, 127.62, 124.1, 48.9, 46.9, 41.1, 31.4, 31.3, 26.5, 22.9, 21.5, 20.0 ppm.
LCMS (Method B):	t_R = 0.86 min, 100% by UV, $[M+H]^+$ found: 315.2.
HRMS:	($C_{17}H_{23}N_4O_2$) $[M+H]^+$ requires: 315.1821, $[M+H]^+$ found: 315.1816.
ν_{max} (neat, cm^{-1}):	3281, 2925, 1629, 1605, 1583, 1547, 1488, 1455, 1386, 1372, 1341, 1288, 1251, 1145, 1032, 762, 748, 609, 556, 533, 483, 457.
M.pt. (MeCN/water):	85 °C

Ph11

N-benzyl-3-(3-methyl-3*H*-diazirin-3-yl)propanamide



Following General Procedure 2 (Section 7.1.2.2), benzylamine was used to afford *N*-benzyl-3-(3-methyl-3*H*-diazirin-3-yl)propanamide (22 mg, 0.10 mmol, 50%) as a white powder.

δ_H (400 MHz, CD_3OD):	7.19-7.35 (m, 5H), 4.35 (s, 2H), 2.08-2.17 (m, 2H), 1.64-1.73 (m, 2H), 0.99 (s, 3H) ppm. Exchangeable proton not observed.
δ_C (101 MHz, CD_3OD):	174.5, 140.1, 129.7, 128.8, 128.3, 44.3, 31.6, 31.4, 26.4, 19.9 ppm.
LCMS (Method B):	t_R = 0.89 min, 100% by UV, $[M+H]^+$ found: 218.3.
HRMS:	($C_{12}H_{16}N_3O$) $[M+H-N_2]^+$ requires: 190.1232, $[M+H-N_2]^+$ found: 190.1236.

ν_{max} (neat, cm^{-1}):	3286, 2441, 1633, 1586, 1546, 1498, 1448, 1422, 1384, 1361, 1316, 1223, 1211, 1075, 1031, 742, 694, 576, 517, 480.
M.pt. (MeCN/water):	65–67 °C

7.2 Biochemistry

7.2.1 General biochemical methods

7.2.1.1 Materials list

Material	Supplier	Catalogue number	Lot number	Other comments
1 M $MgCl_2$	Sigma Aldrich	M1028-100ML	NA	NA
1,4-Dithiothreitol (DTT)	Biomol	04010.100	3063	CAS: 3483-12-3
1.0 M HEPES buffer pH 8.5	Alfa Aesar	J61360	P30D503	CAS: 7365-45-9
10 cm culture plates	Corning BV	353003	NA	NA
12-well plates	Corning (Sigma Aldrich)	CLS3512-100EA	NA	NA
15 cm culture plates	Fisher Scientific	10314601	NA	NA
15 mL falcon tubes	VWR	TPPA91015	NA	NA
20% SDS	Bio-Rad	1610418	NA	NA
24-well plates	Corning (Sigma Aldrich)	CLS3527-100EA	NA	NA
2-chloroacetamide	Sigma Aldrich	22790-250G-F	BCBN8771V	NA
302 nm UV lightbox	Fisher Scientific	12873948	NA	CL1000-M
365 nm UV lightbox	Fisher Scientific	11798211	NA	CL1000-L
5 M Sodium chloride solution	Sigma Aldrich	S5150-1L	SLBW5952	NA
5 mL eppendorf vials	Eppendorf	0030108310	F170142P	NA
50 mL falcon tubes	Falcon (Fisher Scientific)	10788561	NA	NA
500 mM NaF	Sigma Aldrich	S7920	089K0029V	CAS: 7681-49-4
6-well plates	Corning (Sigma Aldrich)	CLS3516-50EA	NA	NA
85% glycerol	Merck	1.04094.2500	Z242694 129	NA
96-well clear bottom plates for BCA assay	BRANDplates	781602	NA	NA
96-well plates	Greiner Bio-one	651201	E17083C4	Microplate, 96 well, PP, V-bottom, Chimney well.
96-well top square sealing mat	ABgene, ThermoScientific	AB-0932	NA	NA
Acetone	Fisher Scientific	A949-1	NA	NA
Adhesive foil plate seals	GE healthcare	28-9758-16	NA	NA
Aprotinin (for PIM5)	Sigma Aldrich	A1153	SLBP2656V	NA
Benzonase nuclease	Sigma Aldrich	E1014-25KU	SLBV2674	CAS: 9025-65-4, PCode 1002538308
Bestatin (for PIM5)	Sigma Aldrich	B8386	SLBQ6027V	NA
Biotin	Apollo Scientific	BIB0603	NA	NA
Biotin azide	TCI	A2523	EW8XE-IF	CAS: 875770-34-6
Brilliant Blue G – Colloidal Concentrate for Coomassie staining	Sigma Aldrich	B2025-1EA	SLBQ0154V	NA
Bromophenol blue	Sigma Aldrich	B0126	NA	NA
Copper(II) sulfate pentahydrate	Sigma Aldrich	C8027-500G	SLBV2520	NA
Corning 8-channel reservoir liner	Fisher Scientific	10407611	NA	NA
Cy5 azide	Sigma Aldrich	777323	NA	NA
Cy5.5 azide	Jena Bioscience	CLK-1059-5	NA	NA
Dimethyl sulfoxide (DMSO)	Sigma Aldrich	276855-250ML	STBH1891	NA
Drain cap mat	Fisherbrand	219005	057982	

Echo Qualified 384-Well Low Dead Volume Microplate (384LDV-Black)	Labcyte	LPL-0200	NA	NA
Echo555 dispenser	Labcyte	NA	NA	NA
Eppendforf centrifuge 5415 D	Eppendorf	NA	NA	For ultracentrifugation (clearing lysate)
Eppenforf Thermomixer Comfort	Eppendorf	5355 000.011	NA	With plate block
Gibco IMDM medium (without phenol red)	Thermofisher Scientific	21056023	1929922	Without phenol red
Greiner 384 low volume plates white	Greiner BioOne	784075	NA	Round well, 20 µL working volume, high-base.
Greiner 384 square well PP plates with V-bottom	Greiner BioOne	781280	B170439B	Square well, polypropylene, 100 µL working volume, V-bottom.
Heraeus Multifuge 3S	ThermoScientific	NA	NA	For centrifuging 15 mL and 50 mL tubes and 500 mL centrifugation vessels
High capacity neutravidin agarose resin	ThermoScientific	29204	SK262607	10 mL
IGEPAL CA-630	Sigma Aldrich	I3021-500ML	MKBV1477V	NA
Invitrogen SeeBlue Plus2 molecular weight marker	Invitrogen (ThermoScientific)	LC5925	NA	NA
Leupeptin (for PIM5)	Sigma Aldrich	L2884	SLBN2885V	NA
Licor Odyssey gel imager	Li-Cor	Version 2.1.15	NA	NA
Licor Odyssey One-colour protein molecular weight marker	Li-cor	928-40000	NA	NA
LysC	Wako Chemicals	125-05061	NA	NA
Microlute 96-deep well filter plate	Porvair Sciences	600033	058018	
MultiScreen HTS vacuum manifold	Millipore	MSVMHTS00	NA	NA
NuPAGE 12% bis-tris protein gels (1.0 mm, 17-well)	Invitrogen (ThermoScientific)	NP0343BOX	NA	NA
NuPAGE 12% bis-tris protein gels (1.0 mm, 17-well)	Invitrogen (ThermoScientific)	NP0349BOX	NA	NA
NuPAGE 4–12% bis-tris protein gels (1.0 mm, 10-well)	Invitrogen (ThermoScientific)	NP0321BOX	NA	NA
NuPAGE 4–12% bis-tris protein gels (1.0 mm, 12-well)	Invitrogen (ThermoScientific)	NP0322BOX	NA	NA
NuPAGE 4–12% bis-tris protein gels (1.0 mm, 15-well)	Invitrogen (ThermoScientific)	NP0323BOX	NA	NA
NuPAGE LDS sample buffer (4x)	Invitrogen (ThermoScientific)	NP0007	1879570	NA
NuPage MES SDS running buffer 20x	ThermoScientific	NP0002	NA	NA
Open pipette reservoir (Corning)	Fisher Scientific	10320551	NA	NA
Pepstatin (for PIM5)	Sigma Aldrich	P5318	087M4005	NA
Phosphoramidon (for PIM5)	Sigma Aldrich	R7386	SLBT6465	NA
Pierce BCA protein assay kit	ThermoScientific	23225	SG251390	NA
Quick Start Bovine Serum Albumin (BSA) Standard Set	Bio-Rad	500-0207	64053247	NA

(Protein standards for BCA assay)				
Sequencing grade trypsin	Promega	V5111	NA	NA
Sodium orthovanadate	Sigma Aldrich	S6508	NA	NA
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)	Sigma Aldrich	C4706-10G	NA	NA
Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA)	Sigma Aldrich	678937-50MG	MKCD9254	CAS: 510578-28-8
Urea	Sigma Aldrich	U4884-500G	SLBF4607V	NA
Vacuum pump	Vacuubrand	PC2001 VARIO	NA	NA
Vortex Genie 2	Scientific Industries	NA	NA	NA

7.2.1.2 BRD4 BD1 and BRD4 BD2 recombinant protein for photocrosslinking experiments

BRD4 BD1

BRD4 BD1 (44–168) was obtained from GenScript:

6H-tev-BRD4 (44–168) was expressed in *E. coli* and purified by Ni-affinity chromatography (HisTRAP HP Affinity Column, GE Healthcare 17-5248-02). TEV-cleavage and size-exclusion chromatography (Superdex 75pg SEC column) was performed to obtain BRD4 BD1 (44–168).

BRD4 BD2

6H-tev-Brd4 (333-460) was produced by Emma Jones:

6H-tev-Brd4 (333-460) was expressed in *E. coli* and purified by Ni-affinity chromatography (HisTRAP HP column, GE Healthcare 17-5248-02) followed by size-exclusion chromatography (Superdex 75pg SEC column).

7.2.1.3 Intact protein LCMS analysis

Samples were injected using an Agilent 1200 series AutoSampler (Model No. G1367B) with sample temperature maintained at 10 °C. Chromatography was performed on an Agilent PLRP-S reverse phase column (1000 Å, 5 µm × 50 mm × 1.0 mm, PL1312-1502) at 70 °C and using an Agilent 1200 series binary pump system (Model No. G1312B) with 0.2% formic acid in water (Solvent A) and 0.2% formic acid in acetonitrile (Solvent B) with a flow rate of 0.5 mL min^{-1}. Detection was performed using an Agilent ToF mass spectrometer (Model No. G6224A) with dual ESI source with a scan rate of 1.03 s in positive mode. Analysis was performed using Agilent MassHunter Qualitative Analysis Software (Version B.06.00).

Intact protein LCMS method A (Open access method)

Injection volume = 10 µL, Scan range = 900–3200 Da.

Table 7.5 Solvent gradient for intact protein LCMS method A.

Time (min)	Solvent A (%)	Solvent B (%)
0.00	90	10
0.50	70	30
5.0	42	58
5.7	35	65
5.71	0	100
6.2	0	100
6.21	90	10
8.00	90	10
8.10	90	10

Intact protein LCMS method B (7.7 min method)

Injection volume = 10 µL, Scan range = 600–3200 Da.

Table 7.6 Solvent gradient for intact protein LCMS method B.

Time (min)	Solvent A (%)	Solvent B (%)
0.00	80	20
0.50	80	20
0.51	60	40
2.50	20	80
2.51	0	100
3.50	0	100
3.51	80	20
4.00	80	20

Intact protein LCMS method C (high injection volume method)

Injection volume = 40 µL, Scan range = 600–3200 Da.

Table 7.7 Solvent gradient for intact protein LCMS method C.

Time (min)	Solvent A (%)	Solvent B (%)
0.00	80	20
0.50	80	20
0.51	60	40
2.50	20	80
2.51	0	100
3.50	0	100
3.51	80	20
4.00	80	20

Intact protein LCMS method D (5.75 min method)

Injection volume = 10 µL, Scan range = 600–3200 Da.

Table 7.8 Solvent gradient for intact protein LCMS method D.

Time (min)	Flow rate (mL min^{-1})	Solvent A (%)	Solvent B (%)
0.00	0.5	80	20
0.50	0.5	80	20
0.51	0.5	60	40
2.50	0.5	20	80
2.51	0.5	0	100
3.60	0.5	0	100
3.61	0.5	80	20
3.80	1.2	80	20
4.45	1.2	80	20
4.46	0.5	80	20

7.2.1.4 Software details

Software	Use
GraphPad Prism for Windows (Version 5.04, 2010).	Plotting binding curves, dose-response curves, calculating K_d values, plotting proteomic experimental data
ACD Spectrus Processor 2016.1	Analysing 1D and 2D NMR experiments
Image Studio Version 2.1.10	Analysing in-gel fluorescence
Agilent MassHunter Qualitative Analysis Software (Version B.06.00).	Analysing intact protein LCMS data
Rstudio (Version 0.98.978)	Interpreting intact protein LCMS deconvoluted spectra

7.2.2 Compound characterisation assays

Potency assays (BD1 and BD2 TR-FRET, MCP-1 whole blood) were performed by Alex Phillipou. Physicochemical measurements (ChromLogD, HSA binding, Solubility and artificial membrane permeability) were performed by Iain Reid, Ferdausi Mazumder, Terrence Johnson and Robert Armstrong.

7.2.2.1 BD1 and BD2 TR-FRET assays

BET proteins were produced using published protocols.[283] The TR-FRET assays for both BD1 and BD2 domains were performed following the published protocols.[284]

7.2.2.2 MCP-1 assay in human whole blood

Compounds to be tested (3 mM in DMSO) were serial diluted (1:3) in a 96 well plate (col 1–10) with a final volume of 0.5 µL. 130 µL of whole blood (containing 1 mL sodium heparin per 100 mL) was added to each well and the plate was incubated (37 °C, 5% CO_2, 30 min). LPS (from Salmonella Typhosa) (2.8 µg/mL in PBS, 10 µL) was added and the plate was incubated overnight (37 °C, 5% CO_2). PBS (140 µL) was added and the plate was centrifuged (1800 rpm, 10 min). The supernatant (20 µL) was transferred to a new plate containing 1% BSA in PBS (20 µL) and shaken (700 rpm, 2 min). 10 µL was transferred to a new plate that was then centrifuged (1300 rpm, 1 min). BD CBA Human MCP1 Capture Beads (BD Bioscience, #558297) were diluted 1:200 in PBS (1% BSA). 10 µL of this mixture was added to the plate that was then shaken (700 rpm, 2 h, rt, dark). BD CBA MCP1 PE Detection Reagent (BD Bioscience, #588287) was diluted 1:400 in PBS (1% BSA). 10 µL was added to the plate that was then shaken (700 rpm, 5 min, rt, dark) before incubation (3 h, rt). The bead populations were analysed by flow cytometry (iQUE Screener, Intellicyt, Ex 396 nm, Em 578 nm) to determine levels of MCP-1 production. The data was fitted (4 parameter fit) between a negative control (DMSO) and an in-house positive control (representing 100% MCP-1 inhibition) to determine an IC_{50} value for each compound screened.

7.2.2.3 ChromLogD

ChromLogD = 0.0857CHI − 2.

ChromLogD at pH 7.4 was determined by fast-gradient HPLC, according to literature procedures,[285] using a Waters Aquity UPLC System, Phenomenex Gemini NX 50 × 2 mm, 3 µm HPLC column, 0−100% pH 7.40 ammonium acetate buffer/acetonitrile gradient. Retention time was compared to standards of known pH to derive the Chromatographic Hydrophobicity Index.[286,287]

7.2.2.4 Human Serum Albumin Binding (%)

Percentage binding to human serum albumin was measured using a published protocol.[286]

7.2.2.5 Solubility

Compounds to be tested (10 mM, 5 µL in DMSO) were diluted to 100 µL in PBS (pH 7.4), equilibrated for 1 hour at room temperature and filtered through Millipore Multiscreen$_{HTS}$-PCF filter plates (MSSL BPC). The filtrate was quantified using a suitably calibrated Charged Aerosol Detector.[288]

7.2.2.6 Artificial Membrane Permeability

The artificial permeability rate was measured across a phospholipid bilayer system. The lipid was egg phosphatidyl choline (1.8%) and cholesterol (1%) dissolved in *n*-decane. This was applied to the bottom of the microfiltration filter inserts in a Transwell plate. Phosphate buffer (50 mM Na_2HPO_4 with 0.5% 2-hydroxypropyl-*b*-cyclodextrin), pH 7.4 was added to the top and bottom of the plate. The lipids were allowed to form bilayers across the small holes in the filter. The permeation experiment was initiated by adding the compound to the bottom well and stopped at a pre-determined elapsed time. The compound permeated through the membrane to enter the acceptor well. The compound concentration in both the donor and acceptor compartments was determined by liquid chromatography after 3 hours incubation at

room temperature.[265,289] The permeability ($logP_{app}$) measuring how fast molecules pass through the lipid membrane was expressed in nm/s.

7.2.3 Chapter 2 – biochemistry

7.2.3.1 Photocrosslinking timecourses for probes 2.27, 2.28, 2.22, 2.23 and 2.24 at 302 and 365 nm (Figure 2.4) and for probes 2.45–2.59 at 365 nm (Figure 2.22)

For each PAL probe, 6 µL of a 2 mM solution in DMSO (FAC = 20 µM) was added to BRD4 (BD1/N-terminal domain, 44–168, GenScript) (3 µM, 594 µL in PBS) on ice. 15 µL was transferred to row A, with a separate column per compound (Table 7.9) in an irradiation plate (Greiner 384-well low volume, 784075). The plate was irradiated (302 or 365 nm) on ice for the time given in Table 7.9. This was repeated sequentially for the following rows. The summed irradiation time for each well is shown in Table 7.9. The experiment was performed in duplicate.

Table 7.9 Irradiation times for the photocrosslinking timecourses shown in Figure 2.4. and Figure 2.22.

Row:	Irradiation time for each row (min):	Total irradiation time for each row (min):
A	10	20
B	5	10
C	2.5	5
D	1.3	2.5
E	0.6	1.2
F	0.3	0.3
G	0.3	0.3
H	0	0

The plate was sealed and sampled directly for intact protein LCMS analysis (method B, Section 0). The total ion chromatograms (TIC) were extracted (region containing protein) and the summed scans were deconvoluted using a maximum entropy algorithm between 800–2600 Da with an expected mass range of 14000–20000 Da. The peak areas for unmodified protein and single labelled protein were recorded and % labelling was determined using Equation 2.1. This data was plotted against irradiation time (GraphPad Prism) and fit using one-phase decay non-linear regression to create a photocrosslinking timecourse. Error bars show ±1 standard deviation.

7.2.3.2 Photocrosslinking timecourses for probes 2.45–2.59 at 302 nm (Figure Figure 2.11) and competition experiment (Figure 2.13)

For each PAL probe, a 400 µL stock solution containing BRD4 (BD1/N-terminal domain, 44–168, GenScript) (3 µM), PAL probe (6 µM), 1% DMSO in PBS was prepared on ice. 50 µL was transferred to row A of an irradiation plate (96-well, V-bottom, ThermoScientific, AB1300), with a separate column per compound (Table 7.10) The plate was irradiated (302 nm) on ice for the time given in Table 7.10. This was repeated for the following rows, except for row C, which contained a stock solution of BRD4 BD1 (3 µM), PAL probe (6 µM), displacement compound **2.60** (100 µM), 1% DMSO in PBS (50 µL). The summed irradiation time for each well is shown in Table 7.10. The experiment was performed in triplicate.

Table 7.10 Irradiation times for the photocrosslinking timecourses shown in Figure 2.13.

Row:	Irradiation time for each row (min):	Total irradiation time for each row (min):
A	10	20
B + C	5	10
D	3.0	5
E	1.0	2
F	1.0	1
G	0	0

The plate was sealed and sampled directly for intact protein LCMS analysis (method A, Section 0). The total ion chromatograms (TIC) were extracted (region containing protein) and the summed scans were deconvoluted using a maximum entropy algorithm with an expected mass range of 14000–20000 Da. The peak areas for unmodified protein and single labelled protein were recorded and % labelling was determined using Equation 2.1. This data was plotted against irradiation time (GraphPad Prism) and fit using one-phase decay non-linear regression to create a photocrosslinking timecourse. Error bars show ±1 standard deviation.

7.2.3.3 Identification of the site of covalent modification by probe 2.59 to BRD4 BD1 by LC-MSMS

Purification by SDS-PAGE and LC-MSMS analysis was performed by Ken Fantom.

A stock solution of BRD4 BD1 (10 µM), PAL probe **2.59** (10 µM) in PBS (1% DMSO) was prepared on ice (480 µL). 113 µL of this mixture was irradiated on ice (302 nm, 6 min) in a 96-well plate. Irradiated and non-irradiated samples were analysed by intact protein LCMS (method B, Section 0). The total ion chromatograms (TIC) were extracted (region containing protein) and the summed scans were deconvoluted (between 750–2200 Da) using a maximum entropy algorithm with an expected mass range of 14000–21000 Da. The peak heights for unmodified and single labelled protein were used to calculate percentage labelling using Equation 7.4.

Each sample (30 µL, 4.5 µg protein) was loaded onto a Novex 4–20% Tris-Glycine gel (Thermo Fisher, XP04200BOX) which was run with 50 V (10 min), then 200 V (40 min) constant voltage. The gel was stained with Coomassie (Instant Blue) and the stained protein bands were excised. Each band was washed with 100 mM ammonium bicarbonate for 1 h with gentle shaking. The supernatant was removed, and the samples were reduced using 10 mM TCEP in 100 mM ammonium bicarbonate (100 µL). The samples were incubated (65 °C, 30 min). The supernatant was removed and 2-iodoacetamide (10 mM in 100 mM ammonium bicarbonate, 100 µL) was added. The samples were incubated at rt in the dark for 30 min. The supernatant was removed, and the samples were incubated with 50% (*v/v*) acetonitrile in 100 mM ammonium bicarbonate (200 µL) for 15 min at rt. The supernatant was removed and acetonitrile (200 µL) was added for 5 min and then removed. Trypsin (20 µg/mL, 20 µL, 0.4 µg added) in 100 mM ammonium bicarbonate was added, followed by 25 mM ammonium bicarbonate (30 µL) and the samples were incubated at 37 °C overnight with gentle shaking. 25 mM ammonium bicarbonate (50 µL) was added and the samples were mixed for 15 min at rt. LysC (20 µg/mL, 20 µL, 0.4 µg added) in 100 mM ammonium bicarbonate was added and the samples were incubated (37 °C, 3 h). The supernatant was transferred to a new eppendorf tube. The remaining gel pieces were extracted with acetonitrile (50 µL × 1, 20 µL × 1). Supernatants were added to the previously transferred supernatant. The samples were concentrated to *ca*. 40 µL and 1% formic acid (4 µL) was added.

LC-MSMS analysis was carried out on a Q Exactive MS (Thermo Fisher) coupled to an Easy-nLC chromatographic system (Thermo Fisher) with a nano ESI source

(Thermo Fisher). 2 µL of each sample was injected and trapped on a C18 Acclaim pepMap 100 enrichment column (75 µm internal diameter × 2 cm, 3 µm nanoviper Thermo Fisher, #164946). The sample was transferred to a C18 Acclaim PepMap RSLC analytical column (75um internal diameter × 15cm, 2um Thermo Fisher, #164940) at 300 nL min^{-1}. Mobile phase A consisted of 0.1% (*v/v*) formic acid in water, and mobile phase B consisted of 0.1% (*v/v*) formic acid in acetonitrile. A gradient of (2–60% B over 90mins), (60–95% B over 10 min), and isocratic wash of 95% B for 9 min was used. The nanospray ionisation source was operated in positive mode with spray voltage set at 2 kV, and heated capillary at 250 °C. MS-based peptide sequencing data were acquired using data dependent LC-MSMS and the top 10 most intense ions were subjected to HCD fragmentation with normalised collision energy set at 30. The MS1 scans were performed at a resolution of 70,000 from 400-2000 *m/z* (automatic gain control (AGC) value 1e6, maximum injection time 200 ms). MS2 scans were performed at a resolution of 17,500 with AGC set at 5e4 and max injection time 200 ms and data acquired using a 15 sec exclusion window. Data were acquired and elaborated using Xcalibur 4.0.27.19. Alternatively, samples were analysed with a Parallel Reaction Monitoring (PRM) method using a 1.0 *m/z* isolation window on the precursor ion value and a collision energy of 30.

Uninterpreted spectra were searched for peptide matches against the BRD4 sequence using the Mascot software (Version 2.6.0) (Matrix Science) using a 5-ppm mass tolerance for peptide precursors and 20 mDa mass tolerance for fragment ions. Oxidation (M) and PAL probe **2.59** modification on any residue were allowed as variable modifications. Carbamidomethylation (C) was used as a fixed modification. To confirm site assignment, samples were analysed using an LC-MSMS PRM method targeting **2.55**-modified and unmodified peptide(s). MS2 spectra were manually validated.

7.2.4 Chapter 3 – biochemistry

7.2.4.1 Percentage labelling of 15 probes (2.45–2.59) to BD1 and BD2 after irradiation (302 nm, 10 min)

Each PAL probe (0.5 mM in DMSO, 150 nL, FAC = 5 µM) was transferred to four wells of a Greiner low volume 384-well plate using a Labcyte 555 Echo acoustic dispenser. The plate was placed on ice and a solution of BRD4 BD1 (1 µM in PBS, 15 µL) was added to the first two wells and a solution of BRD4 BD1 (1 µM in PBS, 15 µL) was added to the second two wells. The plate was sealed and allowed to equilibrate on ice for 1 h. The plate was centrifuged (1000 rpm, 1 min, rt). The seal was removed, and the plate was irradiated on ice (302 nm, 10 min). The UV lamp was warmed for 2 min prior to sample irradiation. The plate was sealed and centrifuged (1000 rpm, 1 min, rt) before being sampled directly for intact protein LCMS analysis (method B, Section 0). The total ion chromatograms (TIC) were extracted (region containing protein) and the summed scans were deconvoluted using a maximum entropy algorithm between 700–2200 Da with an expected mass range of 14000–20000 Da for both BD1 and BD2. The peak areas for unmodified protein and for single, double and triple labelled protein were recorded for each duplicate. Percentage single labelling and excess labelling were calculated using Equation 7.1 and Equation 7.2 respectively and then averaged over the two replicates. These data were plotted on a stacked and grouped column plot (GraphPad Prism). Error bars show ±1 standard deviation.

Equation 7.1

$$\%\ single\ labelling = \frac{Peak\ area\ for\ single\ labelled\ protein}{(Sum\ of\ Peak\ areas\ for\ all\ recorded\ peaks)} \times 100$$

Equation 7.2

$$\%\ excess\ labelling = \frac{Sum\ of\ peak\ areas\ for\ excess\ labelling}{(Sum\ of\ Peak\ areas\ for\ all\ recorded\ peaks)} \times 100$$

7.2.4.2 Experiments to reduce non-specific binding

Three 400 µL protein solutions were prepared on ice containing;
(1) BRD4 BD1 (0.2 µM in PBS (2% DMSO))
(2) BRD4 BD1 (0.2 µM in PBS (2% DMSO with 10 µM CHAPS))
(3) BRD4 BD1 (0.2 µM in PBS (2% DMSO with 50 µM CHAPS))
2.45 (0.2 mM in DMSO, 4 µL, FAC = 2 µM) was added to each solution (FAC of DMSO = 3%) and the samples were equilibrated on ice for 1h. Each solution (100 µL) was added in triplicate to a 384-square well plate (Greiner, 781280) on ice and irradiated (302 nm, 2 min). The plate was sealed and centrifuged (1000 rpm, 1 min, rt) before being sampled directly for intact protein LCMS analysis (method C, Section 0). The total ion chromatograms (TIC) were extracted (region containing protein) and the summed scans were deconvoluted using a maximum entropy algorithm between 1000–2600 Da with an expected mass range of 14000–20000 Da. The peak areas for unmodified protein and for single, double and triple labelled protein were recorded. Percentage labelling for each species (single double or triple) were calculated using Equation 7.3 and then averaged over the three replicates. These data were plotted on a grouped column plot (GraphPad Prism). Error bars show ±1 standard deviation.

Equation 7.3

$$\%\ species\ labelling = \frac{Peak\ area\ for\ species}{(Sum\ of\ Peak\ areas\ for\ all\ recorded\ peaks)} \times 100$$

7.2.4.3 Photocrosslinking timecourse for probe 2.45 with BD1 and BD2

2.45 (0.5 mM in DMSO, 60 µL) was added to a stock solution of BRD4 BD1 and BD2 (1 µM each, 6 mL in PBS) on ice. The mixture was equilibrated on ice for 30 min. 15 µL was transferred to row A, columns 1 and 2 (duplicates) of an irradiation plate (Greiner 384-well low volume, 784075). The plate was irradiated (302 nm) on ice for the time given in Table 7.11. This was repeated sequentially for the following rows. The summed irradiation time for each well is shown in Table 7.11.

Table 7.11 Irradiation times for the photocrosslinking timecourses shown in Figure 3.7.

Row:	Irradiation time for each row (min):	Total irradiation time for each row (min):
A	1.0	2.0
B	0.5	1.0
C	0.2	0.5
D	0.2	0.3
E	0.1	0.1
F	0.0	0.0

The plate was sealed and sampled directly for intact protein LCMS analysis (method B, Section 0). The total ion chromatograms (TIC) were extracted (region containing protein) and the summed scans were deconvoluted using a maximum entropy algorithm between 750–2200 Da with an expected mass range of 14000–21000 Da. A csv file of the deconvoluted spectra (Counts *vs* Mass (Da)) was exported and interpreted using RStudio (Version 0.98.978). The peak heights for unmodified, single labelled and double labelled protein were used to calculate % labelling using Equation 7.4 for BD1 and BD2 for each replicate.

Equation 7.4

$$\%\ species\ labelling = \frac{Peak\ height\ for\ species}{(Sum\ of\ Peak\ heights\ for\ all\ recorded\ peaks)} \times 100$$

These data were exported as a csv file and interpreted in excel to obtain average percentage labelling and standard deviation for the duplicates. These data were transferred to GraphPad Prism and plotted on an XY scatter plot *vs* irradiation time and fitted with a non-linear regression (one-phase decay). Error bars show ±1 standard deviation.

7.2.4.4 Experiments to determine the active concentration of BD1 and BD2

A 19-point serial dilution (0.75x) of **2.45** was prepared in a Greiner 384 square well plate from a stock of 1 mM in DMSO (1 mM to 0.0056 mM over 19 points, with DMSO control as point 20). 150 nL of each well was transferred to two daughter Greiner 384 low volume plates using an Labcyte Echo 555 acoustic dispenser in triplicate (Rows A–C, columns 1–20). The daughter plates were placed on ice and a solution of BRD4 BD1 (estimated from stock to be 1 µM, 15 µL) was added to each

well of the first plate, and BRD4 BD2 (estimated from stock to be 2 µM, 15 µL) was added to each well of the second plate. The plates were sealed and centrifuged (1000 rpm, 1 min, rt) and equilibrated on ice for 20 min. The seals were removed, and the plates were irradiated (302 nm, 0.6 min) on ice. The UV lamp was warmed for 2 min prior to sample irradiation. The plate was sealed and sampled directly for intact protein LCMS analysis (method B, Section 0). The total ion chromatograms (TIC) were extracted (regions containing protein) and the summed scans were deconvoluted using a maximum entropy algorithm between 800–2200 Da for BD1 and 750–2200 Da for BD1 with an expected mass range of 14000–21000 Da for both proteins. A csv file of the deconvoluted spectra (Counts *vs* Mass (Da)) was exported and interpreted using RStudio (Version 0.98.978). The peak heights for unmodified, single labelled, double labelled and triple labelled protein were used to calculate % labelling using Equation 7.4 for BD1 and BD2 for each replicate. % total labelling was calculated as the sum of % single, double and triple labelling. The average % total labelling and standard deviation for the triplicates for both proteins were transferred to GraphPad Prism and plotted on an XY scatter plot *vs* concentration of **2.45**. The linear portions of each plot were re-plotted and fitted with linear regression to obtain the active protein concentration. Error bars show ±1 standard deviation.

7.2.4.5 Single concentration PAL displacement assay with the CHEMBL set with BRD4 BD1 and BD2

CHEMBL set compounds (10 mM in DMSO, 150 nL, FAC = 100 µM) were transferred to two duplicate daughter Greiner low-volume 384 well plates (one compound per well). A mixture of BRD4 BD1 (1 µM), BRD4 BD2 (1 µM) and **2.45** (5 µM) in PBS (0.1% DMSO) was prepared and 15 µL was added to each sample well on ice (final concentration of DMSO = 1.1%). The plates were sealed and centrifuged (1000 rpm, 1 min, rt) and equilibrated on ice for 1 h. The seals were removed, and the plates were irradiated on ice (302 nm, 0.6 min). The UV lamp was warmed for 2 min prior to sample irradiation. The plates were sealed and sampled directly for intact protein LCMS analysis (method B, Section 0). The total ion chromatograms (TIC) were extracted (regions containing protein) and the summed scans were deconvoluted

using a maximum entropy algorithm between 750–2200 Da with an expected mass range of 14000–21000 Da for both proteins. A csv file of the deconvoluted spectra (Counts *vs* Mass (Da)) was exported and interpreted using RStudio (Version 0.98.978). The peak heights for unmodified and single labelled protein were used to calculate percentage labelling using Equation 3.1.

Equation 3.1

$$\%\ Labelling = \frac{Peak\ height\ for\ single\ labelled\ protein}{(Peak\ height\ for\ unlabelled\ protein + single\ labelled\ protein)} \times 100$$

Equation 3.2

$$\%\ Normalised\ labelling = \frac{\%\ labelling - A}{B - A} \times 100$$

where:

A = lowest % labelling observed (most potent competitor)

B = highest % labelling observed (DMSO control)

Equation 7.5

$$\%\ Displacement = 100 - \%\ normalised\ labelling$$

The percentage normalised labelling and percentage displacement were calculated for each compound following Equation 3.2 and Equation 3.3 respectively. The experiment was performed in duplicate. The percentage displacement from both replicates were plotted against each other for both BD1 and BD2 in XY scatter plots using Graphpad Prism (Figure 3.13). The average percentage displacement for BD1 and BD2 over the two replicates given in Table 7.12 were plotted against each other in an XY scatter plot using GraphPad Prism (Figure 3.14). Compounds that showed and average percentage displacement of ≥60% to either BD1 or BD2 were chosen for follow-up dose response experiments.

Table 7.12 SMILES for 264 compounds from the CHEMBL set screened at a single concentration in the PAL displacement assay with BRD4 BD1 and BD2.

SMILES	% Displacement for BRD4 BD1[a]	% Displacement for BRD4 BD2[a]
COC[C@H](C)n1c(nc2ccc(cc12)N1CCOC[C@@H]1C)-c1cc(C)c(=O)n(C)c1	91	96
CCS(=O)(=O)Nc1ccc(Oc2ccc(F)cc2F)c(c1)-c1cn(C)c(=O)c2[nH]ccc12	89	95
CC(C)OC(=O)N[C@@H]1C[C@H](C)N(C(C)=O)c2ccc(cc12)-c1ccc(NC(=O)CCCCCCC(=O)NO)cc1	87	97
C[C@@H](c1ccccc1)n1c(nc2cnc(cc12)C1CCOCC1)-c1cc(C)c(=O)n(C)n1	87	83
C[C@H]1C[C@@H](Nc2ccc(Cl)cc2)c2cc(ccc2N1C(C)=O)-c1ccc(cc1)C(O)=O	90	80
Cc1cc(cn(C)c1=O)-c1nc2cc(cnc2n1Cc1ccccc1)N1CCOCC1	100	69
CC[C@H]1N(C2CCCC2)c2nc(Nc3ccc(cc3OC)C(=O)NC3CCN(C)CC3)ncc2N(C)C1=O	73	82
Cc1cc(cn(C)c1=O)-c1nc2ccccc2n1Cc1ccccc1	84	70
CNCc1ccc(cc1)-c1ccc2N([C@@H](C)C[C@@H](NC(=O)OC(C)C)c2c1)C(C)=O	54	100
COc1cc2c3n(Cc4ccccn4)c(=O)[nH]c3cnc2cc1-c1c(C)noc1C	79	74
Cc1nnc2C(NC(=O)OCc3ccccc3)N=C(c3ccc(C)cc3)c3ccccc3-n12	75	78
COc1cc2c3n(Cc4ccccc4)c(=O)[nH]c3cnc2cc1-c1c(C)noc1C	78	73
CCOC(=O)Nc1cc(nn2c(C)nnc12)-c1ccc(C)c(NS(C)(=O)=O)c1	60	87
COC(=O)C[C@@H]1N=C(c2ccc(Cl)cc2)c2cc(OC)ccc2-n2c(C)nnc12	72	74
C[C@H]1C[C@@H](Nc2ccc(Cl)cc2)c2cc(ccc2N1C(C)=O)-c1ccc(cc1)C(N)=O	69	77
COc1cccc(c1)C(=O)Nc1cc2n(C)c(=O)n(C)c2cc1N1CCCC1	58	88
COc1cc2c3n([C@H](C)c4ccccc4)c(=O)[nH]c3cnc2cc1-c1c(C)noc1C	77	67
CN1CCN(CC1)c1ccc(Nc2ncc(C)c(Nc3cccc(c3)S(=O)(=O)NC(C)(C)C)n2)cc1	58	80
COc1cc2c3n(-c4ccccc4C(C)(C)C)c(=O)[nH]c3cnc2cc1-c1c(C)noc1C	78	60
CCOC(=O)NC1N=C(c2ccccc2)c2ccccc2-n2c(C)nnc12	69	68
Cc1nnc2C(NC(=O)OCc3ccccc3)N=C(c3ccccc3)c3ccccc3-n12	69	67
COc1ccc-2c(c1)C(=N[C@@H](CC(=O)Nc1ccccnc1)c1nnc(C)n-21)c1ccc(Cl)cc1	62	74
Cn1c2cc(NC(=O)Cc3ccccc3)c(cc2n(C)c1=O)N1CCCC1	55	81
COc1cc2c3n([C@H](C)c4ccccn4)c(=O)[nH]c3cnc2cc1-c1c(C)noc1C	79	57
CCOC(=O)NC1N=C(c2ccc(Cl)cc2)c2ccccc2-n2c(C)nnc12	73	61
COc1ccc-2c(c1)C(=N[C@@H](CC(=O)Nc1ccccn1)c1nnc(C)n-21)c1ccc(Cl)cc1	61	71
COc1ccc-2c(c1)C(=N[C@@H](CC(=O)NCC1CC1)c1nnc(C)n-21)c1ccc(Cl)cc1	58	74
COC[C@H](C)n1c(nc2ccc(cc12)C1CCOCC1)-c1cc(C)c(=O)n(C)c1	86	43
Cc1nnc2C(NC(=O)OCc3ccc(F)cc3)N=C(c3ccccc3)c3ccccc3-n12	55	73
COc1ccccc1C(=O)Nc1cc2n(C)c(=O)n(C)c2cc1N1CCCC1	52	75
Cc1cnc(Nc2ccc(OCCN3CCCC3)cc2)nc1Nc1cccc(c1)S(=O)(=O)NC(C)(C)C	48	78
COc1ccc(cc1)C1=NC(NC(=O)OCc2ccccc2)c2nnc(C)n2-c2ccccc12	61	61
CCCCOC(=O)C[C@@H]1N=C(c2ccc(Cl)cc2)c2cc(OC)ccc2-n2c(C)nnc12	63	58
CN(C)c1cccc(c1)-c1cn(C)c(=O)c2cc(sc12)C(=N)NC1CCS(=O)(=O)CC1	57	63
COc1cc2c3n([C@@H](C)c4ccccn4)c(=O)[nH]c3cnc2cc1-c1c(C)noc1C	70	48
COc1cc2c(Nc3ccccc3F)c(cnc2cc1-c1c(C)noc1C)C(N)=O	57	59
COc1cccc(c1)C1=NC(NC(=O)OCc2ccccc2)c2nnc(C)n2-c2ccccc12	48	66
CNC(=N)c1cc2c(s1)c(cn(C)c2=O)-c1ccc(OC)c(OC)c1	43	71
COc1cc2c3n([C@@H](C)c4ccccc4)c(=O)[nH]c3cnc2cc1-c1c(C)noc1C	60	53
CC(=O)n1cc(C2C(=O)NN=C2C)c2ccccc12	45	66
COc1ccc-2c(c1)C(=N[C@@H](CC(=O)Nc1ccc(F)cc1)c1nnc(C)n-21)c1ccc(Cl)cc1	42	68
CC(C)n1nc(-c2cc3cc(O)ccc3[nH]2)c2c(N)ncnc12	37	67
Cc1nnc2C(NC(=O)NCc3ccccc3)N=C(c3ccccc3)c3ccccc3-n12	23	78
C[C@H]1C[C@@H](NC(=O)CCCCCCC(=O)NO)c2cc(ccc2N1C(C)=O)-c1ccccc1	27	73
Cl.Cc1noc(C)c1-c1ccc2c(Nc3ccccc3C(C)(C)C)c(cnc2c1)C(O)=O	50	49
CC(=O)n1cc(C2C(=O)NN=C2C)c2ccccc12	41	55
CC(C)OC(=O)N[C@@H]1C[C@H](C)N(C(C)=O)c2ccc(NC(=O)CCCCCCC(O)=O)cc12	36	59
Cc1cc2c(cnc(N[C@@H]3CCNC[C@H]3OCC3CCCCC3)c2[nH]c1=O)-c1cncc(N)c1	47	48
CCNC(=O)C[C@@H]1N=C(c2cccc(F)c2)c2cc(OC)ccc2-n2c(C)nnc12	44	50
Cc1noc(C)c1-c1ccc2c(Nc3ccccc3F)c(cnc2c1)C(O)=O	51	41

COc1cc2c(NC3CCCCC3)c(cnc2cc1-c1c(C)noc1C)C(N)=O	50	42
CCNC(=O)C[C@@H]1N=C(c2ccc(Cl)cc2)c2ccc(OC)cc2-n2c(C)nnc12	41	49
Cc1noc(C)c1-c1ccc(O)c(c1)S(=O)(=O)NC1CCCC1	27	62
COc1cccc(c1)-c1cn(C)c(=O)c2cc(sc12)C(=N)NC1CCS(=O)(=O)CC1	41	47
COc1ccc-2c(c1)C(=N[C@@H](CC(O)=O)c1nnc(C)n-21)c1ccc(Cl)cc1	18	69
Cc1nnc2c3ccccc3c(nn12)-c1ccc(C)c(NS(C)(=O)=O)c1	42	44
COc1ccccc1S(=O)(=O)Nc1ccc2NC(=O)N(C)Cc2c1	58	28
COc1cccc(c1)-c1cc2CN(CCOc2c(OCC2CCN(C)C2)c1)C(=O)C1CC1	31	55
CCN1C(=O)c2cccc3c(NS(=O)(=O)c4ccc(cc4)N(=O)=O)ccc1c23	23	61
Cc1nnc2C(NS(=O)(=O)Cc3ccccc3)N=C(c3ccccc3)c3ccccc3-n12	9	73
Cc1nnc2C(NC(=O)OCc3ccccc3)N=C(c3ccccc3)c3cc(Cl)ccc3-n12	29	51
Cn1c2ccccc2n(CCSc2nc3ccccc3[nH]2)c1=S	33	46
Cc1cc2c(cnc(NC3CCNCC3)c2[nH]c1=O)-c1cncc(N)c1	28	51
CN(C)CCCNc1cccc2cc(C)c(=O)[nH]c12	24	53
Cc1cc2cccc(NCCO)c2[nH]c1=O	17	59
CCN1C(=O)c2cccc3c(ccc1c23)S(=O)(=O)NC1CCCC1	49	26
CCN1C(=O)c2cccc3c(NS(=O)(=O)c4ccc(C)cc4)ccc1c23	37	39
Cl.O=c1cc(oc2c(cccc12)-c1ccccc1)N1CCOCC1	23	52
CN(C)CCNc1cccc2cc(C)c(=O)[nH]c12	21	54
Cc1cc2cccc(NCc3ncc[nH]3)c2[nH]c1=O	18	56
COCCNc1cccc2cc(C)c(=O)[nH]c12	18	55
COc1cc2c3n(-c4ccccc4F)c(=O)[nH]c3cnc2cc1-c1c(C)noc1C	31	41
Cc1nnc2C(NC(=O)OCc3ccccc3)N=C(c3ccccc3)c3ccc(C)cc3-n12	29	42
Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)NC1CC1	19	53
CN1N=C(c2ccccc2)c2cc(Cl)ccc2-n2c(C)nnc12	16	55
Cc1nc(-c2c[nH]c(=O)c(Cl)c2)n(n1)-c1ccc(F)cc1	37	34
Cc1cc2cccc(N[C@@H]3CCNC3)c2[nH]c1=O	20	50
Cc1cc2cccc(NCCCO)c2[nH]c1=O	16	54
Oc1ccc(cc1)-c1nc(c([nH]1)-c1ccncc1)-c1ccc(F)cc1	26	43
Cc1cc2cccc(N[C@H]3CCNC3)c2[nH]c1=O	16	52
Nc1ccccc1NC(=O)c1ccc(CNC(=O)OCc2ccccc2)cc1	15	53
CCC(=O)N1CCOc2c(C1)cc(cc2OC[C@H]1CCCN(C)C1)-c1ccc(OC)c(OC)c1	20	47
Cl.Cc1noc(C)c1-c1ccc(OCCCN2CCOCC2)c(c1)S(=O)(=O)NC1CCCC1	10	57
CS(=O)c1ccc(cc1)-c1nc(c([nH]1)-c1ccncc1)-c1ccc(F)cc1	40	26
CCc1cnn2c(NCc3cccn(=O)c3)cc(nc12)N1CCCC[C@H]1CCO	22	44
CN1Cc2cc(ccc2N(C)C1=O)S(=O)(=O)N1CCOCC1	20	45
CCNS(=O)(=O)c1cc(ccc1C)-c1c(C)noc1C	12	54
Cc1ccc(cn1)-c1cnc(NC2CCNCC2)c2[nH]c(=O)c(C)cc12	35	30
CCN1C(=O)c2cccc3c(ccc1c23)S(=O)(=O)Nc1ccc(F)cc1	20	44
Cc1cc2c(cnc(N[C@@H]3CCNC[C@H]3OCC3CCS(=O)(=O)CC3)c2[nH]c1=O)-c1ccccnc1	23	41
Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)NC1CCCC1	18	46
Cl.Cc1cc2c(cnc(NC3CCNCC3)c2[nH]c1=O)-c1cn[nH]c1	38	25
CCN1C(=O)c2cccc3c(ccc1c23)S(=O)(=O)NC1CCCCC1	51	11
CC(=O)N1CCc2[nH]nc(Nc3ccccc3)c2C1	19	43
Cc1noc(C)c1-c1ccc2cc[nH]c2c1	24	37
Cc1cc2ccnc(NCCC3CCCNC3)c2[nH]c1=O	12	48
CCc1cc2c(CCCC2=O)n1Cc1ccccc1	12	47
COc1ccc(cc1S(=O)(=O)NC1CC1)-c1c(C)noc1C	13	45
Cc1cncc(c1C)-c1cnc(NC2CCNCC2)c2[nH]c(=O)c(C)cc12	26	32
Cc1cncc(c1)-c1cnc(N[C@@H]2CCNC[C@H]2OCC2CCCCC2)c2[nH]c(=O)c(C)cc12	29	29
Cc1cc(-c2ccccc2)n2nc(N)nc2n1	22	35
COc1ccccc1C1=NC(NC(=O)OCc2ccccc2)c2nnc(C)n2-c2ccccc12	16	40
CSc1ccc(CC(=O)N2CC(=O)Nc3ccccc23)cc1	12	44
Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)NCCCO	10	46
CCN(C1CCCC1)S(=O)(=O)c1cc(ccc1OC)-c1c(C)noc1C	9	46
COCCCN1CCC(CC1)Nc1nccc2cc(C)c(=O)[nH]c12	17	37
Cn1c2cc(NC(=O)Cc3ccccc3)c(cc2n(C)c1=O)N1CCCCC1	14	40
Cc1cc2c(cnc(N[C@@H]3CCNC[C@H]3OCC3CCOCC3)c2[nH]c1=O)-c1ccncc1	25	28
Cc1cc2ccnc(NCC3CCCNC3)c2[nH]c1=O	18	34

COc1cccc(NS(=O)(=O)c2cc(ccc2OC)-c2c(C)noc2C)c1	12	40
COc1ccccc1C(=O)Nc1cc2n(C)c(=O)[nH]c2cc1N1CCCCC1	14	37
COc1ccc(cc1OC)-c1cn(C)c(=O)c2cc(sc12)C(=O)N1CCCCC1	18	34
Cn1cc(-c2ccccc2)c2sc(cc2c1=O)C(=N)NC1CCS(=O)(=O)CC1	24	28
COc1ccc(cn1)-c1cnc(NC2CCNCC2)c2[nH]c(=O)c(C)cc12	20	31
Cc1cncc(c1)-c1cnc(N[C@H]2CCNC[C@@H]2OCC2CCCCC2)c2[nH]c(=O)c(C)cc12	19	32
COc1cccc(NS(=O)(=O)c2cc(ccc2C)-c2c(C)noc2C)c1	11	39
COc1cncc(c1)-c1cnc(N[C@@H]2CCNC[C@H]2OCC2CCOCC2)c2[nH]c(=O)c(C)cc12	27	22
CCc1cc2c(cnc(N[C@@H]3CCNC[C@H]3OCC3CCS(=O)(=O)CC3)c2[nH]c1=O)-c1cncc(C)c1	19	30
Cc1nnc2C(CC(=O)OC(C)(C)C)N=C(c3c(C)c(C)sc3-n12)c1ccc(Cl)cc1	15	34
Cc1cc2ccnc(N[C@@H]3CCNC3)c2[nH]c1=O	12	38
Cc1cc2ccnc(N[C@H]3CCNC[C@@H]3OCC3CCS(=O)(=O)CC3)c2[nH]c1=O	20	28
Cc1cc2c(cnc(N[C@@H]3CCNC[C@H]3OCC3CCOCC3)c2[nH]c1=O)-c1cncnc1	17	31
COc1cc(c(OC)cc1CN(C)C)-c1cn(C)c(=O)c2cnccc12	9	39
Cc1cc2c(cnc(NC3CCNCC3)c2[nH]c1=O)-c1ccccnc1C	21	26
Clc1c(sc2ccccc12)C(=O)N1CC(=O)Nc2ccccc12	12	35
COc1cc2c(Nc3ccccc3F)ccnc2cc1-c1c(C)noc1C	13	34
Cc1cc2c(cnc(NC3CCNCC3)c2[nH]c1=O)-c1ccncc1	27	20
CC(C)CO[C@@H]1CNCC[C@H]1Nc1nccc2cc(C)c(=O)[nH]c12	23	23
COc1ccc(cc1OC)-c1cn(C)c(=O)c2cc(sc12)C(=O)N(C)C	18	28
FC(F)(F)c1cccc(c1)C(=O)N1CC(=O)Nc2ccccc12	14	31
CN(Cc1ccc(C)cc1)c1cc(C)nc2ncnn12	17	28
Cc1ccccc1C(=O)Nc1cc2n(C)c(=O)n(C)c2cc1N1CCCCC1	21	23
COc1cccc(c1)C(=O)Nc1ccc2n(C)c(=O)n(C)c2c1	12	32
COc1ccc(cc1S(=O)(=O)NC1CCCC1)-c1c(C)noc1C	12	32
Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)N1CCCC1	12	31
COc1ccc(cc1OC)-c1cc2CN(CCOc2c(OC)c1)C(=O)C1CC1	22	21
COc1cccc(c1)C(=O)Nc1cc2n(C)c(=O)n(C)c2cc1C	12	31
COc1ccccc1C(=O)Nc1cc2[nH]c(=O)n(C)c2cc1N1CCCCC1	11	32
COc1ccc(cc1)C(=O)Nc1ncnc2[nH]cnc12	13	28
COc1ccc(NS(=O)(=O)c2cc(ccc2C)-c2c(C)noc2C)cc1	11	30
Cc1cc2c(cnc(NC3CCNCC3)c2[nH]c1=O)-c1cncc(Cl)c1	24	17
Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)Nc1ccc(Cl)cc1	10	32
Cc1cc2ccnc(N[C@@H]3CCNC[C@H]3OCC3CCOCC3)c2[nH]c1=O	23	18
CCC(=O)N1CCc2n[nH]c(c2C1)-c1ccccc1	10	31
ONC(=O)CCCCCCC(=O)Nc1ccccc1	10	30
CCn1ccc(n1)C(=O)N1CC(=O)Nc2ccccc12	8	32
CCN1C(=O)c2cccc3c(ccc1c23)S(=O)(=O)NC1CCCCCC1	20	20
C[C@@H]1CC(=O)Nc2cccc(-c3ccc4n(C)nc(-c5cnn(C)c5)c4c3)c2N1	13	27
Cc1cc2cccc(NCCC(O)=O)c2[nH]c1=O	7	33
CN1CC[C@@H](Nc2ncc(-c3cncc(C)c3)c3cc(C)c(=O)[nH]c23)[C@@H](C1)OCC1CCS(=O)(=O)CC1	17	23
Cl.Cc1noc(C)c1-c1ccc2c(Nc3ccccc3F)c(cnc2c1)C(N)=O	8	32
O=C(Nc1cccc2ncccc12)c1ccccc1	18	22
Cc1noc(C)c1-c1ccc2[nH]cnc2c1	11	29
Cl.CN1c2ccccc2NC(CC1=O)c1ccccc1	8	31
CC(=O)c1cc(-c2ccccc2S(C)(=O)=O)c2cc(Oc3ccccc3)ccn12	9	30
O=C(Cc1ccccs1)N1CC(=O)Nc2ccccc12	7	31
COCCN1CCC(CC1)Nc1nccc2cc(C)c(=O)[nH]c12	11	27
Cn1cc(-c2cccc(c2)C#N)c2sc(cc2c1=O)C(=N)NC1CCS(=O)(=O)CC1	17	21
O=C(Nc1ncnc2nc[nH]c12)c1ccccc1	7	31
Cc1cc2ccnc(N[C@H]3CCNC[C@@H]3OCC3CCOCC3)c2[nH]c1=O	9	28
COc1cccc(n1)-c1cc2CN(CCOc2c(OC)c1)C(=O)C1CC1	10	27
COc1cc(OC)c2c(c1)nc([nH]c2=O)-c1cc(C)c(OCCO)c(C)c1	19	18
Cl.Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)N1CCNCC1	9	27
COc1ccccc1C(=O)Nc1cc2n(C)c(=O)n(C)c2cc1C	7	30
Cc1cc2c(cnc(NC3CCNCC3)c2[nH]c1=O)-c1cncnc1	16	21
Cc1cc2ccnc(N[C@@H]3CCNC[C@H]3OCC3CCS(=O)(=O)CC3)c2[nH]c1=O	12	24
COc1ccccc1NS(=O)(=O)c1cc(ccc1C)-c1c(C)noc1C	6	30
CSc1ccc(Cl)c(c1)C(=O)N1CC(=O)Nc2ccccc12	10	25

Cc1noc(C)c1NC(=O)c1cccc(c1)-c1ccccc1	4	30
Cc1nnc2C(N)N=C(c3ccccc3)c3ccccc3-n12	8	26
COc1ccc(cc1OC)-c1cn(C)c(=O)c2cc(sc12)C(=O)N1CCN(CC1)S(C)(=O)=O	9	25
Cc1cc2ccnc(NC[C@H]3CCNC3)c2[nH]c1=O	10	24
Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)NC(C)(C)C	11	23
Cl.Cc1noc(C)c1-c1ccc(OCCN2CCNCC2)c(c1)S(=O)(=O)NC1CCCC1	8	26
Cc1nnc2C(NC(=O)OCc3ccccc3)N=C(c3ccccc3)c3cc(ccc3-n12)N(=O)=O	12	21
CC(C)(C)NS(=O)(=O)c1cncc(c1)-c1ccn2nc(N)nc2c1	11	22
COc1cccc(c1)-c1cc2CN(CCOc2c(OC)c1)C(C)=O	11	21
COc1ccccc1C(=O)Nc1cc2n(C)c(=O)n(C)c2cc1N1CCCCC1	14	18
Cl.C(Nc1ncnc2nc[nH]c12)c1ccccc1	7	25
COc1ccccc1C(=O)Nc1ccc2n(C)c(=O)n(C)c2c1	20	12
Cc1noc(C)c1-c1ccc(C)c(c1)S(N)(=O)=O	5	27
Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)NCCCc1ccccc1	4	27
Cc1ccc(cc1)-c1cc([nH]n1)C(=O)N1CC(=O)Nc2ccccc12	6	26
CC(=O)c1cc(-c2ccccc2CO)c2ccccn12	16	15
O=C1CCNc2ccccc2N1	9	22
Cc1cc2ccnc(NC3CCNCC3)c2[nH]c1=O	17	14
Cc1noc(C)c1NC(=O)c1cccnc1	9	22
Cc1cc2ccnc(NCC3CCNCC3)c2[nH]c1=O	14	17
COc1ccccc1NS(=O)(=O)c1cc(ccc1OC)-c1c(C)noc1C	11	19
CN(C)CC(C)(C)CNc1nccc2cc(C)c(=O)[nH]c12	6	25
O=C1Nc2ccc(c3cccc1c23)S(=O)(=O)NC1CCCCC1	5	25
Cc1cc2ccnc(NC3CCN(CC3)C3CC3)c2[nH]c1=O	16	14
COc1cc(cc(OC)c1CN(C)C)-c1cn(C)c(=O)c2cnccc12	9	21
Cn1c2cc(NC(=O)C(C)(C)C)c(cc2n(C)c1=O)N1CCCCC1	8	22
CCCNc1cccc2cc(C)c(=O)[nH]c12	8	22
COc1ccc(NS(=O)(=O)c2cc(ccc2OC)-c2c(C)noc2C)cc1	11	18
COc1ccccc1C(=O)Nc1cc2n(C)c(=O)n(C)c2cc1OC	9	20
Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)Nc1ccccc1Cl	6	23
CC1CC(=O)Nc2ccccc2S1	3	25
CC(C)Sc1ccccc1C(=O)N1CC(=O)Nc2ccccc12	9	20
COc1cccc(c1)C(=O)Nc1cc2n(C)c(=O)n(C)c2cc1N1CCCCC1	9	20
CCC(=O)Nc1cc(ccc1C)-c1nn2c(C)nnc2c2ccccc12	10	18
CN(Cc1ccc(Cl)s1)c1cc(C)nc2ncnn12	15	13
Cl.Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)NC1CCCNC1	5	23
CC(C)n1cc(-c2cccc(c2)C(F)(F)F)c2sc(cc2c1=O)C(=N)NC1CCS(=O)(=O)CC1	6	22
CC1CC(=O)Nc2ccccc2S1	7	21
CC(=O)c1ccc(OCc2c(C)noc2C)cc1	18	9
COc1ccn2c(cc(-c3ccccc3S(C)(=O)=O)c2c1)C(C)=O	15	12
Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)Nc1cccc(Cl)c1	12	15
CCCCc1c(C)nc2nc(SCc3nc4n(CCCC)c(=O)[nH]c(=O)c4n3CC)nn2c1C	11	16
COc1cccc(c1)-c1cc2CN(CCOc2c(OC2CCOC2)c1)C(=O)C1CC1	12	15
Cl.Cc1noc(C)c1-c1cccc(c1)S(=O)(=O)NC1CC1	10	17
Cl.COc1ccc(cc1S(=O)(=O)N1CCNCC1)-c1c(C)noc1C	8	19
COc1ccc(CC(=O)N2CC(=O)Nc3ccccc23)cc1OC	11	16
Cc1cc2ccnc(N[C@@H]3CCCNC3)c2[nH]c1=O	11	15
CCC1=NCC(=O)Nc2ccc(Cl)cc12	4	22
Cc1nc(no1)-c1ccc(C)c(c1)S(=O)(=O)NC1CCCC1	2	23
Cc1ccc(CC(=O)N2CC(=O)Nc3ccccc23)cc1	6	19
CSc1ccccc1C(=O)N1CC(=O)Nc2ccccc12	7	18
Cc1cc(-c2ccccc2)n2ncnc2n1	8	17
CN1Cc2ccccc2NC1=O	6	19
COc1ccc(cc1S(=O)(=O)Nc1ccccc1Cl)-c1c(C)noc1C	4	22
COc1cccc(OC)c1-c1cc2CN(CCOc2c(OC)c1)C(=O)C(C)C	5	21
Cc1cccc(OCCC(=O)N2CC(=O)Nc3ccccc23)c1	11	15
COc1ccc(cc1S(=O)(=O)Nc1ccccc1)-c1c(C)noc1C	11	14
Cc1nnc2c3ccccc3c(nn12)-c1ccccc1	9	16
Cc1noc(N)c1-c1ccccc1	8	17

CN1N=C(c2ccccc2)c2cc(Cl)ccc2-n2c(N)nnc12	11	13
COCCNS(=O)(=O)c1cc(ccc1C)-c1c(C)noc1C	5	19
CS(=O)(=O)N(CC(=O)N1CC(=O)Nc2ccccc12)c1ccccc1F	4	20
Cl.Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)NC1CCNCC1	4	20
Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)N1CCCCC1	9	15
Cl.COc1ccc(cc1S(=O)(=O)NC1CCNCC1)-c1c(C)noc1C	2	22
Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)NCc1ccccc1	3	20
OC(=O)C(F)(F)F.OC(=O)C(F)(F)F.OC(=O)C(F)(F)F.Cc1cccc(Oc2nccc(n2)-c2c(ncn2C2CCNCC2)-c2ccc(F)cc2)c1C	7	16
CN1CCCC1=O	2	21
CCCOc1ccc(CCc2nc3cc(ccc3n2CCN2CCOCC2)-c2c(C)noc2C)cc1	6	16
CC1CCc2ccccc2N1C(C)=O	3	18
Cn1c2ccc(cc2n(C)c1=O)C(N)=O	6	15
Nc1ccc2nnc(-c3ccccc3)n2n1	7	14
Cn1c2cc(NC(=O)c3ccccc3)c(cc2n(C)c1=O)N1CCCCC1	6	13
COc1cccc(c1)C(=O)Nc1cc2n(C)c(=O)n(C)c2cc1N1CCOCC1	8	11
Clc1ccc(CC(=O)N2CC(=O)Nc3ccccc23)cc1	3	15
CN(C)CCNc1nccc2cc(C)c(=O)[nH]c12	8	11
Cc1cncc(c1)-c1cnc(N[C@@H]2CCNC[C@H]2OCC2CCS(=O)(=O)CC2)c2[nH]c(=O)c(C)cc12	10	8
Cl.COc1ccc(cc1S(=O)(=O)NC1CCCNC1)-c1c(C)noc1C	7	12
Cl.Cc1cc2ccnc(N[C@H]3CCNC3)c2[nH]c1=O	13	5
CC1CC(=O)Nc2ccccc2N1	6	11
Cc1noc(C)c1-c1ccccc1	5	12
C\C(CO)=C/CNc1ncnc2[nH]cnc12	3	13
Cc1nnc2c3ccccc3c(nn12)-c1ccc(C)c(c1)S(=O)(=O)N1CCCCCC1	4	12
CN1CCCC(COc2cc(cc3CN(CCOc23)C(=O)C2CC2)-c2cccc(Cl)c2)C1	9	6
COCC(=O)N1CCOc2c(C1)cc(cc2OC)-c1cccc(OC)c1	7	8
CN(C)S(=O)(=O)c1cc(ccc1C)-c1nn2c(C)nnc2c2ccccc12	4	11
O=c1cc(oc2c(cccc12)-c1cccc2c3ccccc3sc12)N1CCOCC1	8	6
COc1ccc(cc1S(=O)(=O)NCc1ccccc1)-c1c(C)noc1C	4	10
O=C(N1CCCCC1c1ccccnc1)c1ccccc1	4	10
Cc1noc(C)c1NC(=O)c1ccccc1	4	10
CN(C1CCCCC1)S(=O)(=O)c1cc(ccc1C)-c1nn2c(C)nnc2c2ccccc12	5	9
Cc1nnc2C(CC(=O)NCc3ccccc3)N=C(c3ccccc3)c3ccccc3-n12	4	9
Cc1noc(C)c1-c1ccc(C)c(c1)S(=O)(=O)NCCc1ccccc1	6	5
Cc1noc(C)c1-c1ccccc1	3	5
CC(=O)Nc1cc(ccc1C)-c1nn2c(C)nnc2c2ccccc12	5	3
COc1ccc(cc1S(=O)(=O)N1CCOCC1)-c1c(C)noc1C	2	1

[a] Average value over two replicates.

7.2.4.6 Full-curve dose-response PAL displacement assay with the compounds selected from single shot screening

An 8-point, 1-in-2 serial dilution of the 44 competitor compounds chosen from the preliminary single shot PAL screen (+ DMSO only treated as a 45th compound) was prepared in an Echo Qualified 384-Well Low Dead Volume Microplate (384LDV-Black). This plate was then transferred to two duplicate daughter Greiner low-volume 384 well plates (150 nL per well) using a Labcyte Echo555 acoustic dispenser. A mixture of BRD4 BD1 (1 µM), BRD4 BD2 (1 µM) and **2.45** (5 µM) in PBS (0.1% DMSO) was prepared and 15 µL was added to each sample well of both daughter

plates on ice (final concentration of DMSO = 1.1%). The plates were sealed and centrifuged (1000 rpm, 1 min, rt) and equilibrated on ice for 1 h. The seals were removed, and the plates were irradiated on ice (302 nm, 1.0 min). The UV lamp was warmed for 2 min prior to sample irradiation. The plates were sealed and sampled directly for intact protein LCMS analysis (method D, Section 0). The total ion chromatograms (TIC) were extracted (region containing protein) and the summed scans were deconvoluted using a maximum entropy algorithm between 750–2200 Da with an expected mass range of 14000–21000 Da for both proteins. A csv file of the deconvoluted spectra (Counts *vs* Mass (Da)) was exported and interpreted using RStudio (Version 0.98.978). The peak heights for unmodified, single labelled and double labelled protein were used to calculate percentage total labelling using Equation 3.4.

Equation 3.4

$$\%\ Total\ labelling = \frac{B + 2C}{A + B + 2C} \times 100$$

where:

A = peak height for unlabelled protein

B = peak height for single labelled protein

C = peak height for double labelled protein

The values for total percentage labelling were averaged over the two replicates and plotted against the $-\log_{10}$[competitor]. These data were fit with a "log(agonist) vs. response – Variable slope (four parameters)" non-linear regression with no top constraint, and a bottom constraint of greater than zero using Graphpad Prism. The average percentage total labelling for the DMSO controls was included as the ninth data point for each compound. These plots for the 44 compounds (+DMSO) are shown in Figure 7.1 to Figure 7.4. The IC_{50} values obtained from each curve were transformed to pIC_{50} values using $IC_{50} = -\log_{10}(IC_{50})$. These values were plotted against the pIC_{50} values obtained for the 44 compounds in BD1 and BD2 TR-FRET assays (conducted following the methods described in Section 7.2.2.1) to produce Figure 3.18. The IC_{50} values obtained from the PAL displacement assay were transformed to K_i values following the Munson-Rodbard correction to the Cheng-Prusoff equation (Equation 3.5).

Equation 3.5

$$K_i = \frac{IC_{50}}{1 + \frac{p^*(y_0 + 2)}{2K_d(y_0 + 1)} + y_0} + K_d\left(\frac{y_0}{y_0 + 2}\right)$$

where:

IC_{50} = value obtained in the PAL displacement assay.

p* = Initial concentration of PAL probe **2.45** (5×10^{-6} M).

K_d = IC_{50} value for **2.45** in BD1 or BD2 TR-FRET assay ($10^{-7.8}$ or $10^{-7.3}$ M respectively).

y_0 = initial ratio of bound/free PAL probe (2/3).

These K_i values were then transformed to pK_i values by $pK_i = -\log_{10}(K_i)$. These values were plotted against the pIC_{50} values obtained in BD1 and BD2 TR-FRET assays to produce Figure 3.19. The pK_i values for BD1 and BD2 were plotted against each other on an XY scatter plot to produce Figure 3.20. The TR-FRET data, PAL displacement single-shot percentage displacement data, PAL displacement pIC_{50} data and the transformed pK_i data for both BD1 and BD2 domains for the 44 compounds screened is given in Table 3.2.

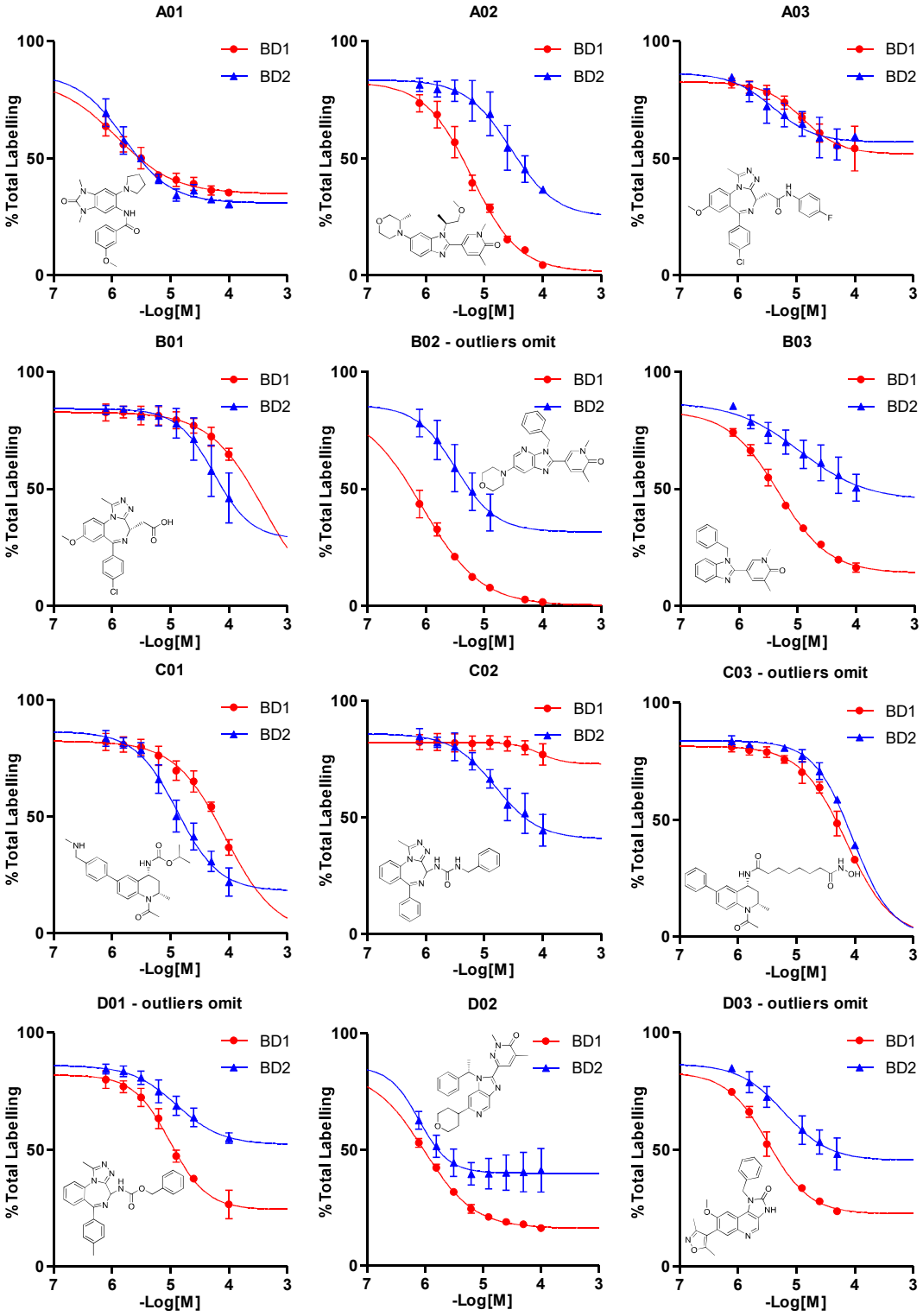


Figure 7.1 Dose-response curves for compounds **A01–D03** obtained from the dual-domain PAL displacement assay performed in Section 3.3.3.

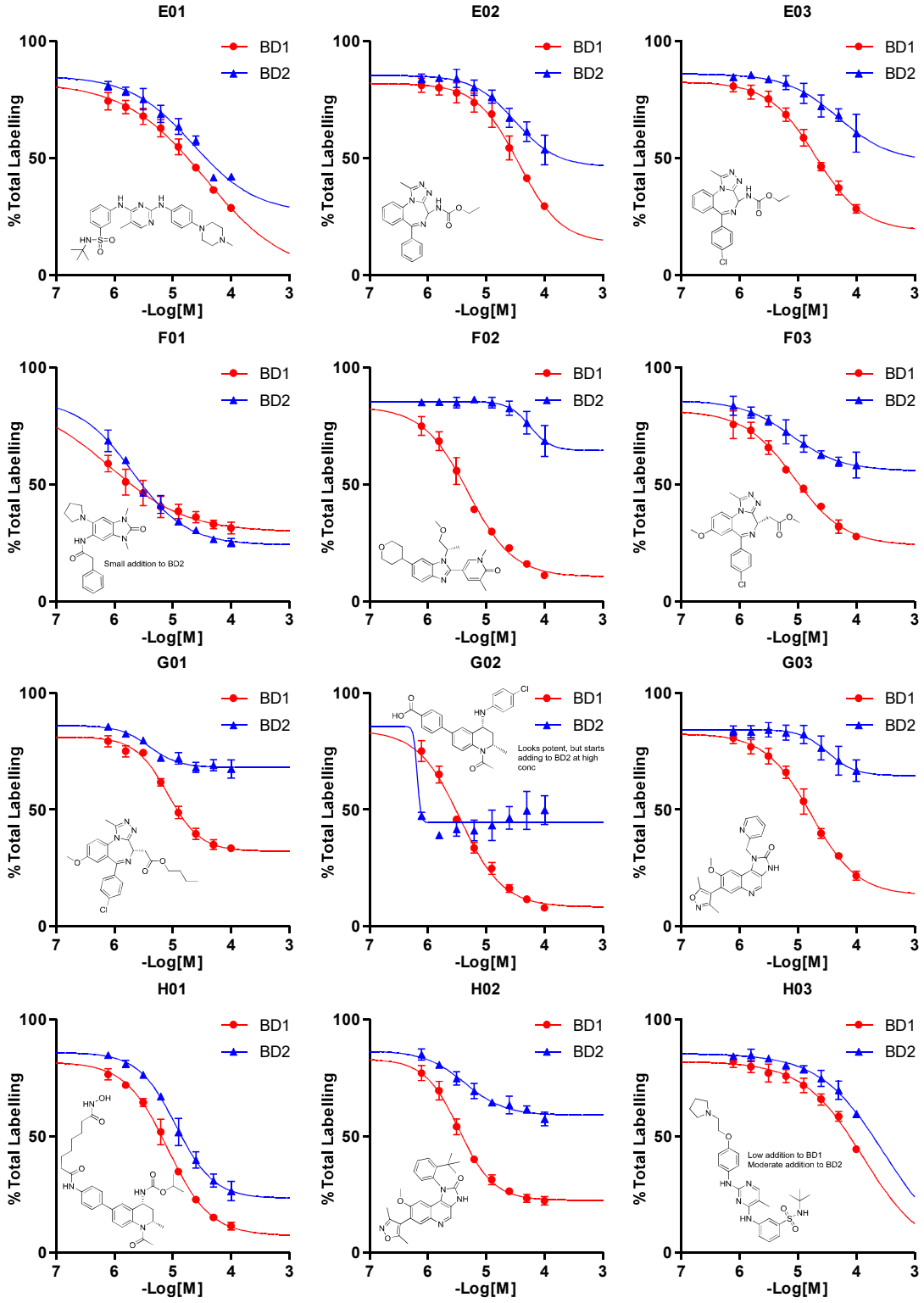


Figure 7.2 Dose-response curves for compounds **E01–H03** obtained from the dual-domain PAL displacement assay performed in Section 3.3.3.

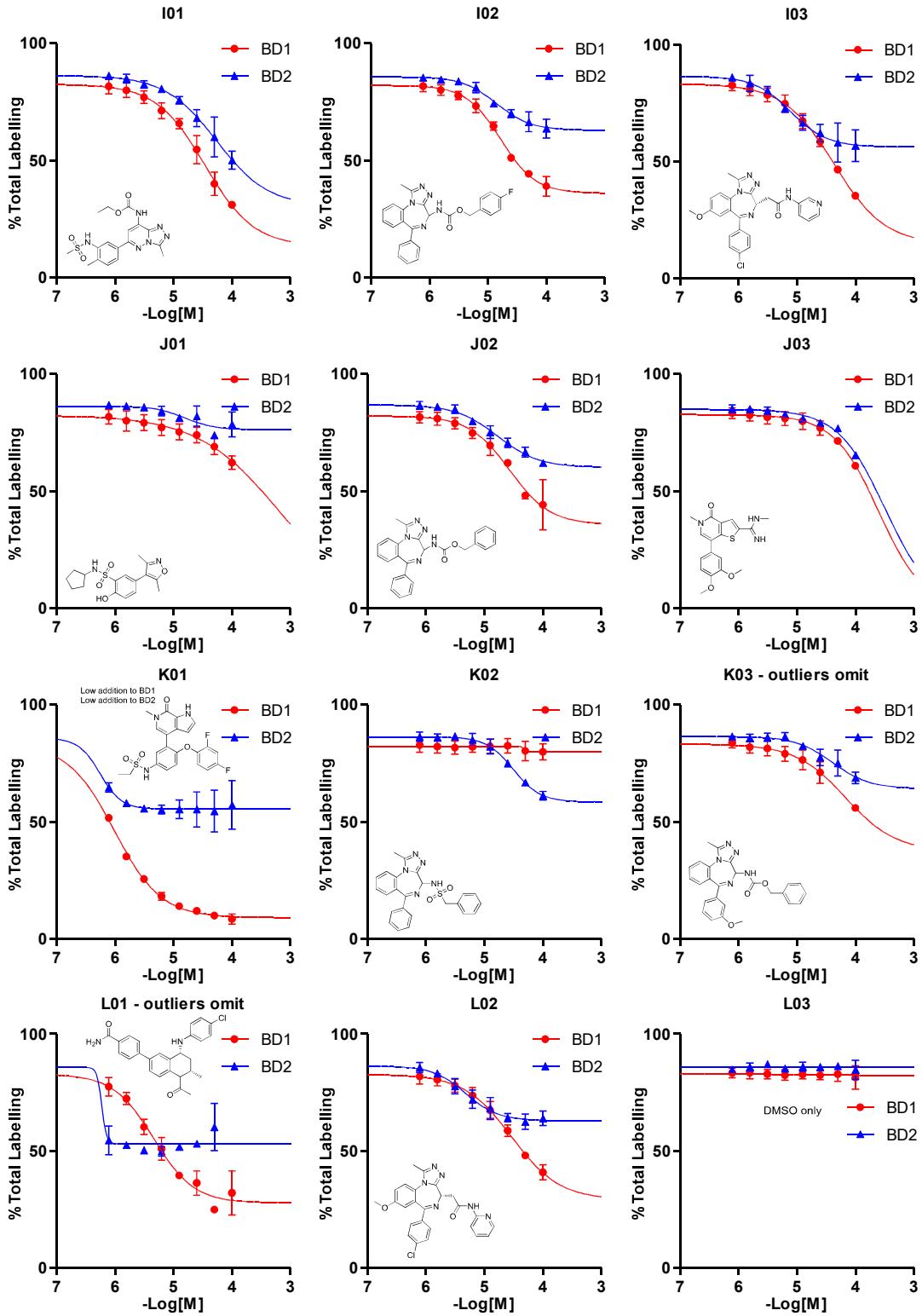


Figure 7.3 Dose-response curves for compounds **I01–L03** obtained from the dual-domain PAL displacement assay performed in Section 3.3.3.

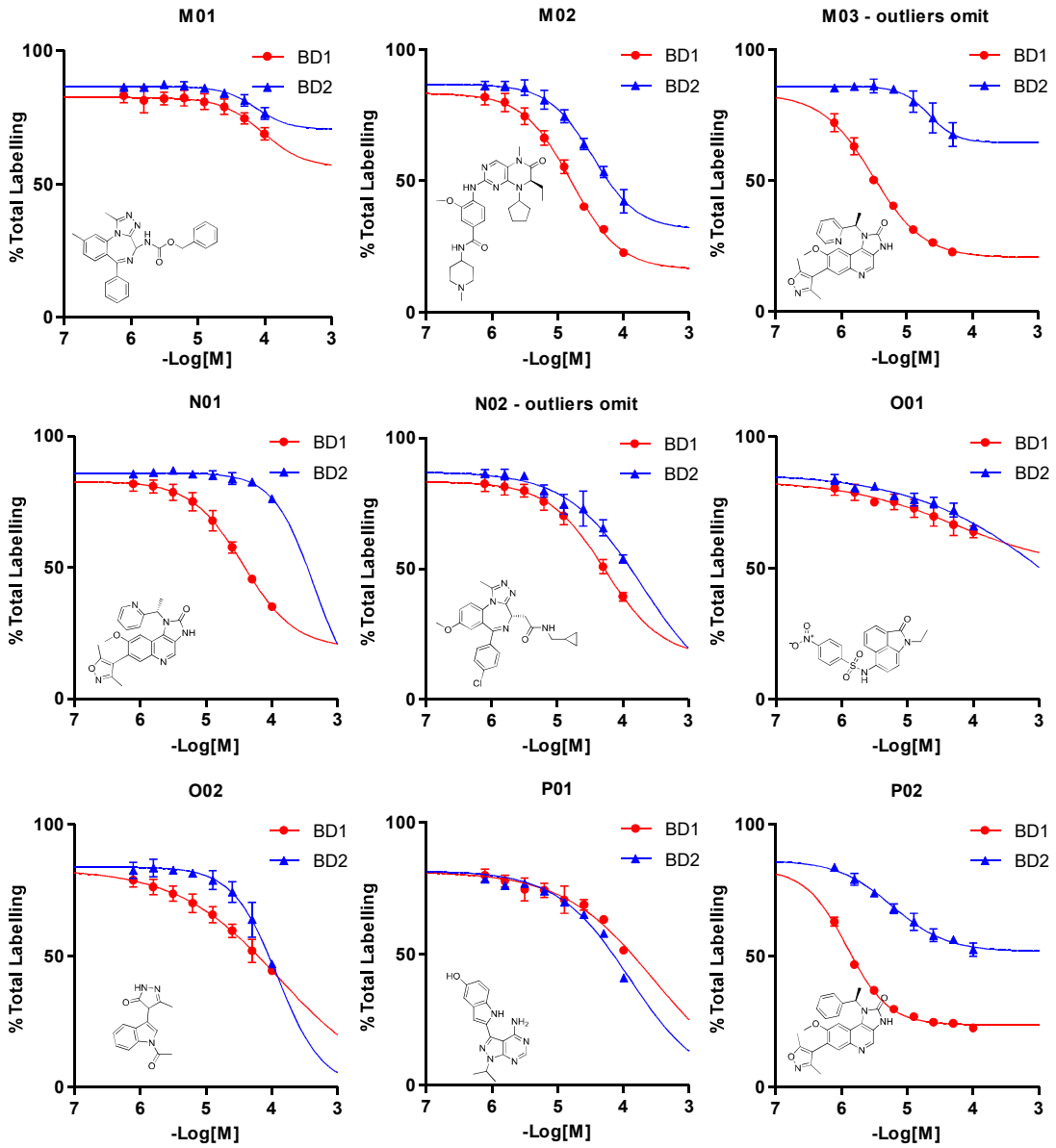


Figure 7.4 Dose-response curves for compounds **M01–P02** obtained from the dual-domain PAL displacement assay performed in Section 3.3.3.

Table 3.2 The TR-FRET, single-shot PAL displacement, and full-curve PAL displacement data for the 44 competitor compounds screened against BRD4 BD1 and BD2 in this proof-of-concept work.

Competitor compound	Compound details	Competitor labelling	BD1 TR-FRET (pIC_{50})	BD2 TR-FRET (pIC_{50})	% Disp. BD1 (single-shot)	% Disp. BD2 (single-shot)	BD1 PAL (full-curve) (pIC_{50})	BD2 PAL (full-curve) (pIC_{50})	BD1 PAL (full-curve) (pK_i)	BD2 PAL (full-curve) (pK_i)
A01	BRPF1 probe analogue[233,234]		<4.3	<4.3	58	88	5.9	5.8	8.1	7.5
A02	Analogue of PAL probe affinity function[196,197]		7.2	6.2	91	96	5.2	4.5	7.6	6.4
A03	Analogue of iBET-762[236]		6.8	7	42	68	4.9	5.4	7.3	7.2
B01	Analogue of iBET-762[236]		5.5	5.9	18	69	<3.8	4.3	<6.2	6.2
B02	Analogue of PAL probe affinity function[196,197]		7.9	7.2	100	69	6.1	5.5	8.1	7.3
B03	Analogue of PAL probe affinity function[196,197]		7.6	7.2	84	70	5.3	5	7.7	6.8
C01	iBET-295[227]		6.9	7.2	54	100	4.1	4.9	6.5	6.8
C02			6.4	7.5	23	78	4.1	4.8	6.5	6.7
C03	Dual HDAC/BET probe[230]		5.9	5.7	27	73	4.1	4	6.5	5.9
D01			7.2	7	75	78	5	4.9	7.4	6.8
D02	Analogue of PAL probe affinity function[196,197]		7.8	7.6	87	83	6	6.1	8.1	7.7
D03	Analogue of iBET-151[228]		7.5	7	78	73	5.5	5.2	7.8	7.1
E01	TG-101209 (Dual JAK2/BRD4 inhibitor)[183,231]	Y	5.9	5.6	58	80	4.4	4.6	6.8	6.5
E02			6.9	6.7	69	68	4.4	4.5	6.8	6.4
E03			7.4	7	73	61	4.7	4.3	7.1	6.2
F01	BRPF1 probe analogue[233,234]		<4.3	<4.3	55	81	6.1	5.7	8.1	7.4
F02	Analogue of PAL probe affinity function[196,197]		7.4	6.2	86	43	5.3	4.3	7.6	6.2
F03	Analogue of iBET-762[236]		7.4	7	72	74	5.1	5.1	7.4	7.0
G01	Analogue of iBET-762[236]		7.5	7.3	63	58	5.1	5.4	7.4	7.2
G02	iBET-726[227]		7.8	8.2	90	80	5.4	>6.3	7.7	>7.7
G03	Analogue of iBET-151[228]		6.8	6.2	79	74	4.8	4.5	7.2	6.4
H01	Dual HDAC/BET probe[230]		8	7.4	87	97	5.1	4.9	7.4	6.8
H02	Analogue of iBET-151[228]		7.5	7.3	78	60	5.5	5.4	7.8	7.2
H03	Dual kinase/BRD4 inhibitor[183,231]	Y	6.4	6	48	78	3.9	<3.8	6.3	<5.7
I01	Bromosporine[237]		7	6.5	60	87	4.5	4.3	6.9	6.2
I02			7.1	7.1	55	73	4.8	4.9	7.2	6.7
I03	Analogue of iBET-762[236]		6.9	7.1	62	74	4.4	5.1	6.8	7.0
J01		Y	5.8	5.7	27	62	<3.8	4.8	<6.2	6.7
J02			6.9	6.8	69	67	4.6	4.8	7	6.7
J03			5.9	5.9	43	71	<3.8	<3.8	<6.2	<5.7
K01	ABBV-075/Mivebresib[229]	Y	8.1	8.2	89	95	6	6.3	8.1	7.7
K02			5.3	6.4	9	73	NA	4.5	–	6.4
K03			6.7	6.6	48	66	4.2	4.4	6.6	6.3
L01	Analogue of iBET-726[227]		8	8.5	69	77	5.3	>6.3	7.7	7.7
L02	Analogue of iBET-762[236]		7	7.5	61	71	4.5	5.3	6.9	7.1
L03	DMSO control		NA	NA	0	0	NA	NA	–	–
M01			6.4	6.6	29	42	4	4.1	6.4	6.0
M02	BI-2356 (Dual PLK1/BRD4 inhibitor)[231,238]		7	6.3	73	82	4.8	4.5	7.2	6.4
M03	iBET-151[228]		7.6	6.5	79	57	5.5	4.7	7.8	6.6
N01	Analogue of iBET-151[228]		6.4	5.8	70	48	4.4	<3.8	6.8	<5.7
N02	Analogue of iBET-762[236]		6.9	6.9	58	74	4.3	<3.8	6.7	<5.7
O01	Benzo[cd]indol-2(1H)-one BET inhibitor[239]	Y	5.6	4.5	23	61	4.3	<3.8	6.7	<5.7
O02		Y	NA	NA	45	66	3.9	3.9	6.3	5.8
P01		Y	4.8	4.8	37	67	<3.8	3.9	<6.2	5.8
P02	Analogue of iBET-151[228]		7.9	7.2	77	67	5.9	5.2	8.1	7.1

7.2.5 Chapter 5 – biochemistry

7.2.5.1 Dose-response photocrosslinking studies with BRD4 BD1 and PhABit compounds (Ph1–Ph11)

A serial dilution (1 in 3, 6-point, 100–0.41 mM) of each PhABit in DMSO was prepared in a 96-well plate with each compound occupying a separate column and the concentrations desending by row with a final volume of 50 µL. DMSO (50 µL) was added to row G. 1 µL from each well was transferred to a 96-well daughter plate containing PBS (79 µL) in each well on ice. BRD4 (BD1/N-terminal domain, 44–168, GenScript) (15 µM, 20 µL in PBS, FAC = 3 µM) was added to each well. The plate was incubated on ice for 30 min before irradiation (302 nm, 10 min). The plate was sealed and sampled directly for intact protein LCMS analysis (method A, Section 0). The total ion chromatograms (TIC) were extracted (region containing protein) and the summed scans were deconvoluted using a maximum entropy algorithm between *m/z* 800–2600 with an expected mass range of 14000–20000 Da. The peak areas for unmodified protein and single labelled protein were recorded and % labelling was determined using Equation 2.1 (pg 62). The experiment was performed in triplicate and the average percentage crosslinking for each PhABit was plotted against concentration (GraphPad Prism) using a one site specific binding regression to create a dose-response curve. Error bars show ±1 standard deviation.

7.3 X-ray crystallography

X-ray crystallography was performed by Dr Chun-wa Chung.

All the statistics for the data collection and refined co-ordinates are given in Table 7.13. Ligand and protein geometries were validated using MOGUL and MolProbity respectively.

7.3.1.1 Cocrystallisation and structure determination of BRD4 BD1 with 2.31 (Figure 2.9)

1 µL of compound **2.31** at 200 mM DMSO stock concentration (FAC = 4 mM) was added to 50 µL of BRD4 BD1 (44-168) and incubated on ice for 15 min. The complex was centrifuged (13,000 rpm, 4 °C, 15 min) on a bench top centrifuge before setting up in 100 nL + 100 nL sitting drops using a 96-well MRC plate. Crystals grown with a well solution of 20% (*w/v*) PEG 3350, 0.2 M sodium sulfate, 0.1 M bis-tris propane pH 8.5 at 20 °C were cryoprotected using well solution with 10% ethylene glycol prior to flash freezing in liquid nitrogen. Data from a single crystal was collected at 100K on I02 at the Diamond Light Source (Harwell) and processed to 1.80 Å using XIA2, DIALS and AIMLESS.[290,291] A molecular replacement solution was determined using Phaser and a previously determined in-house structure as a starting model. The $P2_12_12_1$ cell ($\alpha=\beta=\gamma=90^{\circ}$, a=38.694 Å, b=43.252 Å, c=79.181 Å) had one molecule in the asymmetric unit (ASU). Manual model building was performed using COOT and refined using REFMAC with twin refinement within the CCP4 software suite.[292-294]

7.3.1.2 Cocrystallisation and structure determination of BRD4 BD1 with 2.46

1 µL of compound **2.46** at 100 mM DMSO stock concentration (FAC = 2 mM) was added to 50 µL of TEV-cleaved BRD4 BD1 (44-145) and incubated on ice for 15 min. The complex was centrifuged (13,000 rpm, 4 °C) on a bench top centrifuge before setting up in 100 nL + 100 nL sitting drops using a 96-well MRC plate. Crystals grown with a well solution of 30% (*w/v*) PEG 4000, 0.2 M sodium acetate, 0.1 M Tris (pH 8.5) at 20 °C were cryoprotected using well solution with 10% ethylene glycol prior

to flash freezing in liquid nitrogen. Data from a single crystal was collected at 100K on I04-1 at the Diamond Light Source (Harwell) and processed to 1.89 Å using XIA2, DIALS and AIMLESS.[290,291] A molecular replacement solution was determined using Phaser and a previously determined in-house structure as a starting model. The $P2_12_12_1$ cell ($\alpha=\beta=\gamma=90^\circ$, a=42.351 Å, b=52.103 Å, c=57.176 Å) had one molecule in the asymmetric unit (ASU). Manual model building was performed using COOT and refined using REFMAC with twin refinement within the CCP4 software suite.[292-294] There was clear difference density for the ligand in the acetylated lysine binding site, allowing the ligand to be well modelled within the conserved site. There was good density for most of the ligand including the aryl azide photoreactive group, whilst the density for the ester chain was sparse.

7.3.1.3 Cocrystallisation and structure determination of BRD4 BD1 with 2.45

1 µL of compound **2.45** at 250 mM DMSO stock concentration (FAC = 5 mM) was added to 50 µL of TEV-cleaved BRD4 BD1 (44-145) and incubated on ice for 15 min. The complex was centrifuged (13,000 rpm, 4 °C) on a bench top centrifuge before setting up in 100 nL + 100 nL sitting drops using a 96-well MRC plate. Crystals grown with a well solution of 25% (*w/v*) PEG 1500, 0.1 M SPG (pH 5.0) at 20 °C were cryoprotected using well solution with 10% ethylene glycol prior to flash freezing in liquid nitrogen. Data from a single crystal was collected at 100 K on I04-1 at the Diamond Light Source (Harwell) and processed to 1.89 Å using XIA2, DIALS and AIMLESS.[290,291] A molecular replacement solution was determined using Phaser and a previously determined in-house structure as a starting model. The $P2_12_12_1$ cell ($\alpha=\beta=\gamma=90^\circ$, a=42.611 Å, b=52.134 Å, c=57.186 Å) had one molecule in the ASU. Manual model building was performed using COOT and refined using REFMAC with twin refinement within the CCP4 software suite.[292-294] There was clear difference density for the ligand in the acetylated lysine binding site, allowing the ligand to be well modelled within the conserved site. There was good density for most of the ligand including the tetrafluoroaryl azide group, whilst the density for the ester chain was sparse.

7.3.1.4 Cocrystallisation and structure determination of BRD4 BD1 with 2.48

1 µL of compound **2.48** at 200 mM DMSO stock concentration (FAC = 4 mM) was added to 50 µL of TEV-cleaved BRD4 BD1 (44-145) and incubated on ice for 15 min. The complex was centrifuged (13,000 rpm, 4 °C) on a bench top centrifuge before setting up in 100 nL + 100 nL sitting drops using a 96-well MRC plate. Crystals grown with a well solution of 10% (*w/v*) PEG 4000, 0.2 M $MgCl_2$, 0.1 M MES (pH 6.5) at 20 °C were cryoprotected using well solution with 10% ethylene glycol prior to flash freezing in liquid nitrogen. Data from a single crystal was collected at 100 K on I04-1 at the Diamond Light Source (Harwell) and processed to 1.75 Å using XDS, AIMLESS and STARANISO within autoPROC.[291,295,296] A molecular replacement solution was determined using Phaser and a previously determined in house structure as a starting model. The $P2_1$ cell (α=γ=90˚, β=93.435˚, a=59.640 Å, b=41.798 Å, c=109.529 Å) had four molecules in the ASU. Manual model building was performed using COOT and refined using REFMAC with twin refinement within the CCP4 software suite.[292-294] There was clear difference density for the ligand in each of the acetylated lysine binding site, allowing the ligand to be well modelled within the conserved site. There was good density for all the ligand.

7.3.1.5 Cocrystallisation and structure determination of BRD4 BD1 with 2.59

1 µL of compound **2.59** at 500 mM DMSO stock concentration (FAC = 10 mM) was added to 50 µL of TEV-cleaved BRD4 BD1 (44-145) and incubated on ice for 15 min. The complex was centrifuged (13,000 rpm, 4 °C) on a bench top centrifuge before setting up in 100 nL + 100 nL sitting drops using a 96-well MRC plate. Crystals grown with a well solution of 0.1 M sodium citrate (pH 5.6), 10% (*w/v*) PEG 4000, 10% (*w/v*) 2-propanol at 20 °C were cryoprotected using well solution with 10% ethylene glycol prior to flash freezing in liquid nitrogen. Data from a single crystal was collected at 100K on I04-1 at the Diamond Light Source (Harwell) and processed to 1.89 Å using AutoPROC, XDS and AIMLESS.[291,295,296] A molecular replacement solution was determined using Phaser and a previously determined in-house structure as a starting model. The $P2_1$ cell (α=γ=90˚, β=90.14˚, a=42.035 Å, b=52.166 Å, c=59.957 Å) had two molecules in the ASU. Manual model building was performed using COOT and

refined using REFMAC with twin refinement within the CCP4 software suite.[292-294] There was clear difference density for the ligand in the acetylated lysine binding site of Chains A and B, allowing the ligand to be well modelled within the conserved site. Away from this site the density was sparse, but the pendent chain could still be positioned.

7.3.1.6 Data collection and refinement statistics

Table 7.13 Data collection and refinement statistics for BRD4 BD1 X-ray structures

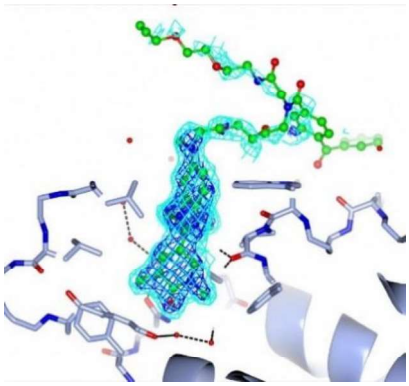
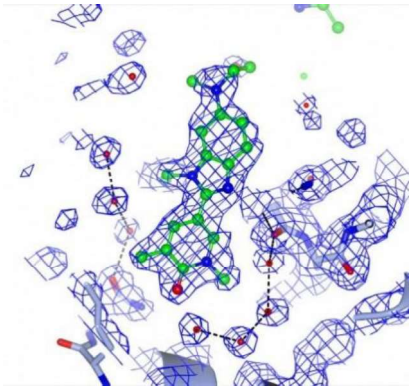
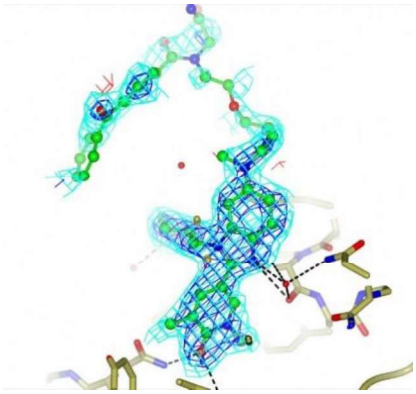
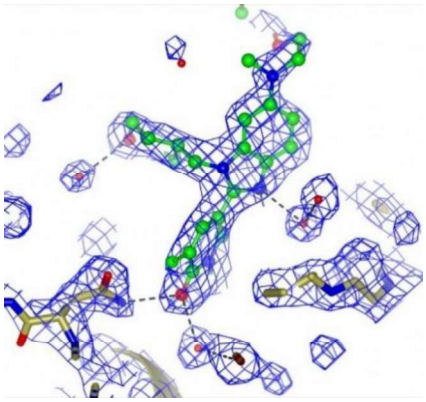
Collection on a single crystal	**2.59**	**2.45**	**2.46**	**2.48**	**2.31**
Crystallisation conditions at 20 °C	0.1 M sodium citrate (pH 5.6), 10% (*w/v*) PEG 4000, 10% (*w/v*) 2-propanol	25% (*w/v*) PEG 1500, 0.1 M SPG (pH 5.0)	30% (*w/v*) PEG 4000, 0.2 M sodium acetate, 0.1 M Tris (pH 8.5)	10% (*w/v*) PEG 4000, 0.2 M $MgCl_2$, 0.1 M MES (pH 6.5)	0.1 M Bis-tris propane pH 8.5, 20% PEG3350, 0.2 M potassium sodium tartrate
Data collection					
Data collection	DLS I04-1 λ 0.97951Å	DLS I04-1 λ 0.91587Å	DLS I04-1 λ 0.91587Å	DLS I04-1 λ 0.97951Å	DLS I02
Space group	$P2_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$	$P2_12_12_1$
Cell dimensions					
a, *b*, *c* (Å)	42.035, 52.166, 59.957	42.611, 52.134, 57.186	42.351, 52.103, 57.176	59.640, 41.798, 109.529	38.694, 43.252, 79.181
α, β, γ (°)	90.000, 90.140, 90.000	90.000, 90.000, 90.000	90.000, 90.000, 90.000	90.000, 93.435, 90.000	90.000, 90.000, 90.000
Resolution (Å)[a]	59.96-1.89 (1.92-1.89)	42.59-1.17 (1.20-1.17)	34.02-1.31 (1.34-1.31)	109.33-1.75 (1.78-1.75)	39.59-1.80 (1.85-1.80)
R_{merge}	0.043 (0.478)	0.078 (0.766)	0.035 (0.996)	0.101 (0.674)	0.109 (0.064)
ccI (1/2)		0.989 (0.650)		0.997 (0.620)	
$I/\sigma I$	16.8 (2.4)	10.3 (1.1)	17.9 (1.3)	8.8 (1.8)	10.5 (2.7)
Completeness (%)	96.3 (96.3)	96.2 (87.1)	98.6 (97.2)	86.2 (87.9)[b]	99.9 (99.8)
Redundancy	3.4 (3.3)	5.7 (4.1)	6.1 (5.2)	3.3 (2.7)	6.1 (5.4)
Refinement					
Resolution (Å)	59.96-1.89	42.59-1.17	34.02-1.31	109.332-1.75	39.59-1.80
No. reflections	68933 (3451)	240330 (11370)	188117 (11418)	118283 (485)	78000 (4248)
No. uniq reflections	20355 (1034)	41918 (2743)	30602 (2182)	36111 (177)	12890 (789)
Twin fraction (H, K, L)	0.695	–	–	–	–
Twin fraction (H, K, L)	0.305	–	–	–	–
R_{work}/ R_{free}	0.177/0.220	0.218/0.242	0.304/0.230	0.179/0.239	0.203/0.254
No. atoms	2442	1331	1359	5316	1301

Protein	2104	1087	1117	4267	–
Ligand/ion	130	52	48	330	–
Water	208	192	194	719	–
B-factors					
Protein	34.60	17.70	23.76	24.76	–
Ligand/ion	55.32	22.31	30.29	25.60	–
Water	38.98	30.70	36.68	33.70	–
RMS deviations					
Bond lengths (Å)	0.059	0.055	0.056	0.061	0.007
Bond angles (°)	1.354	1.130	1.174	1.280	1.158

[a] Highest resolution shell is shown in parenthesis. [b] Ellipsoidal using STARANISO.

7.3.1.7 Electron density maps for BRD4 BD1 X-ray structures.

Table 7.14 Electron density maps for compounds **2.59**, **2.45**, and **2.46**.

Compound/ Chain	OMIT (Fo-Fc) map in the active site contoured at:	OMIT (2Fo-Fc) map in the active site contoured at:
2.59/ Chain A	**±3.0sigma (blue/red), +1.5sigma (cyan)** 	**+1.0 sigma (blue)** 
2.59/ Chain B	**±3.0sigma (blue/red), +1.5sigma (cyan)** 	**+1.0 sigma (blue)** 

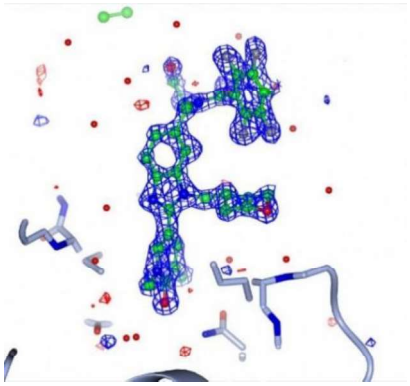
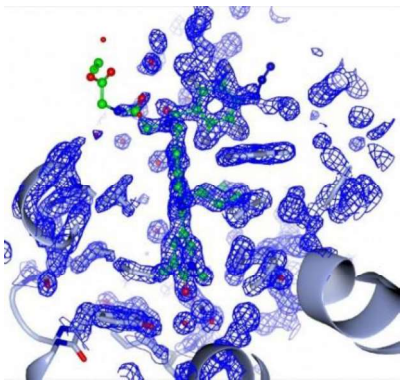
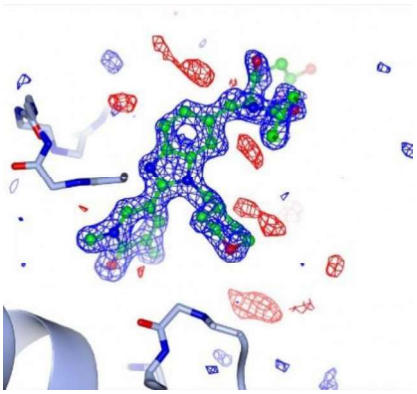
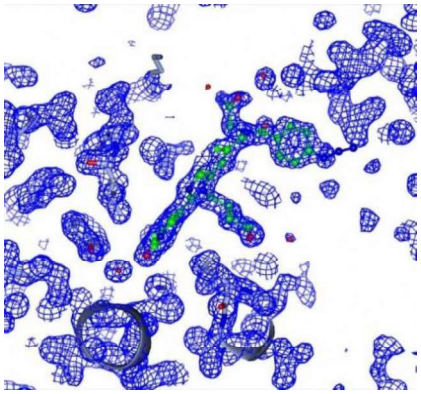
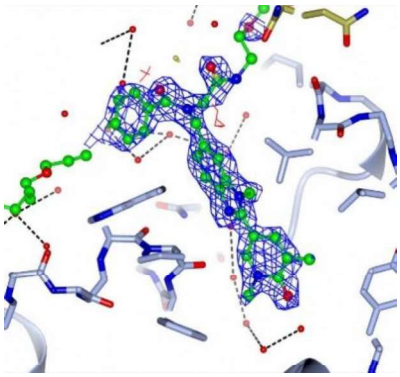
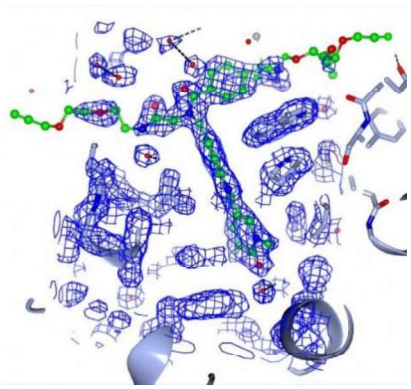
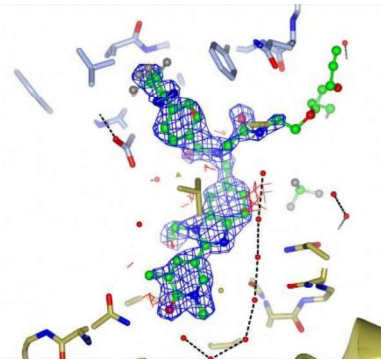
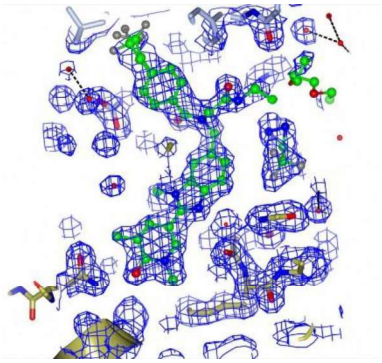
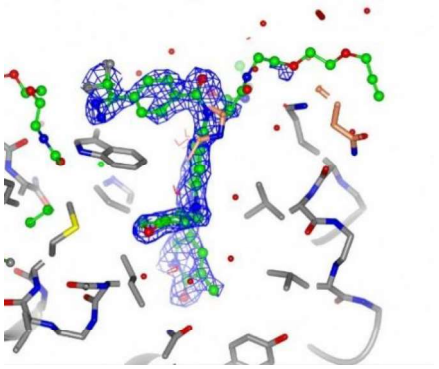
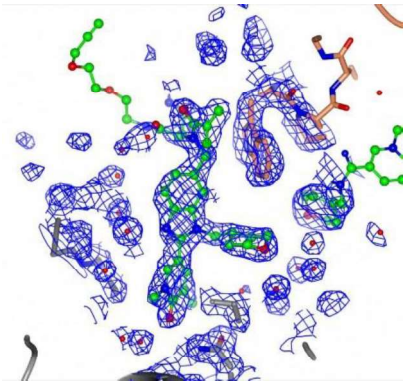
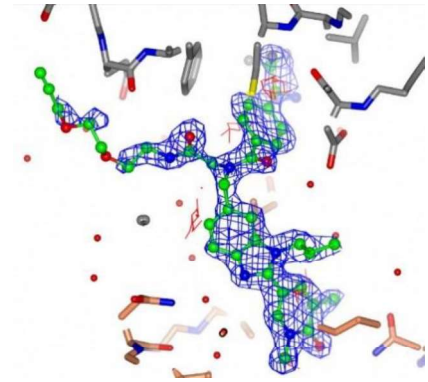
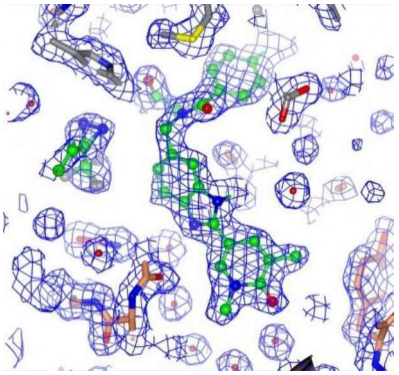
2.45/ Chain A	**±3.0sigma (blue/red)**	**+1.5 sigma (blue)**
		
2.46/ Chain A	**±3.0sigma (blue/red)**	**+1.5 sigma (blue)**
		

Table 7.15 Electron density maps for compound **2.48**.

Compound /Chain	OMIT (Fo-Fc)map in the active site contoured at ±3.0sigma (blue/red)	OMIT (2Fo-Fc) map in the active site contoured at +1.0 sigma (blue)
2.48/ Chain A		

2.48/ Chain B		
2.48/ Chain C		
2.48/ Chain D		

7.4 Chemoproteomics

7.4.1 General chemoproteomic methods

7.4.1.1 General Procedure 1: Preparation of HL-60 lysate (method A)

HL-60 cell culturing was performed by Jürgen Stuhlfauth and Nilma Garcia-Altrieth. HL-60 lysate (method A) was prepared by Doris Uhlenbruch as follows:

3 volumes of chilled lysis buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 1 mM Na_3VO_4, 0.8% IGEPAL CA-630, 25 mM NaF, 1x PIM5 (in-house protease inhibitor mix containing 500 nM Aprotinin (Sigma, A1153), 40 µM Bestatin (Sigma, B8386), 100 µM Leupeptin (Sigma, L2884), 1.46 µM Pepstatin (Sigma, P5318), 10 µM Phosphoramidon (Sigma, R7386)) was added to one volume of HL-60 cell pellet and the mixture was rotated overhead (15 min , 4 °C). The cell suspension was homogenised using a Wheaton Dounce tissue grinder (Sartorius) before mixing by overhead rotation (30 min, 4 °C). The mixture was centrifuged (20,000 × g, 10 min, 4 °C) and the supernatant was transferred and centrifuged (140,000 × g, 1 h, 4 °C). The supernatant was aliquoted and flash-frozen. The protein concentration was determined by BCA assay (Pierce BCA protein assay kit, Thermo Scientific, 23225, using Quick Start BSA Standard Set, Bio-Rad, 500-0207).

7.4.1.2 General Procedure 2: TMT labelling and LC-MSMS analysis for chemoproteomic experiments

Sample preparation (TMT labelling and desalting), and LC-MSMS instrument operation was performed by Kerstin Kammerer as described below.

Sample preparation

Lyophilized peptides were resuspended in HPLC grade water then labelled with 6-plex or 8-plex TMT (Thermo Fisher Scientific) reagents and quenched with hydroxylamine. Labelled peptide extracts were combined to a single sample per experiment. Combined samples were then desalted using C18-SCX stage tips: for each stage tip, 3 plugs SCX material (Cation 47 mm Extraction Disks 2252, Empore/3M) were overlaid with 3

plugs C18 material (Octadecyl C18 47 nm Extraction Disks, Empore/3M). After equilibration with 0.5% trifluoroacetic acid (TFA) and 2% acetonitrile, combined samples were loaded on the stage tips. Stage tips were washed with 0.5% TFA in 2% acetonitrile, and twice with 0.5% TFA in 60% acetonitrile, followed by elution of bound peptides with 5% ammonia in 80% acetonitrile. Desalted peptides were dried *in vacuo* and resuspended in 0.05% TFA in water and injected into an Ultimate3000 nanoRLSC (Dionex, Sunnyvale, CA) coupled to a Q-Exactive HF (Thermo Fisher Scientific).

LC-MSMS acquisition

Peptides were separated on custom 50 cm × 100 µM (ID) reverse-phase columns (Reprosil) at 40 °C. Gradient elution was performed from 2–40% acetonitrile in 0.1% formic acid over 200 minutes. Samples were online injected into a Q-Exactive HF mass spectrometer operating with a data-dependent top 10 method. MS spectra were acquired by using 60,000 resolution and an ion target of 3E6. Higher energy collisional dissociation (HCD) scans were performed with 35% (normalised collision energy) NCE at 30,000 resolution (at *m/z* 200), and the ion target setting was set to 2E5 to avoid coalescence. The instrument was operated with Tune 2.8 and Xcalibur 4.0.27.19. Mascot 2.5 (Matrix Science) was used for protein identification by using a 10-ppm mass tolerance for peptide precursors and 20 mDa (HCD) mass tolerance for fragment ions. Carbamidomethylation of cysteine residues and TMT modification of lysine residues were set as fixed modifications. N-terminal acetylation of proteins and TMT modification of peptide N-termini were set as variable modifications.

Data treatment

Swissprot human (downloaded on January 2018) was used as search database, combined with a decoy version of this database created by using a script supplied by MatrixScience. We accepted protein identifications as follows: (i) For single-spectrum to sequence assignments, we required this assignment to be the best match and a minimum Mascot score of 31 and a 10× difference of this assignment over the next best assignment. Based on these criteria, the decoy search results indicated <1% false discovery rate (FDR). (ii) For multiple spectrum to sequence assignments and using

the same parameters, the decoy search results indicate <0.1% FDR. All identified proteins were quantified; FDR for quantified proteins was <1%. For the analysis an in house version of isobarQuant 1 was used.[297] Reporter ion intensities were read from raw data and multiplied with ion accumulation times (the unit is milliseconds, so as to yield a measure proportional to the number of ions; this measure is referred to as ion area).[298] Spectra matching to peptides were filtered according to the following criteria: mascot ion score >15, signal; signal-to-background of the precursor ion >4, and signal-to-interference >0.5. Fold-changes were corrected for isotope purity and adjusted for interference caused by co-eluting nearly isobaric peaks.[299] Protein quantification was derived from individual spectra matching to distinct peptides by using a sum-based bootstrap algorithm; 95% confidence intervals were calculated for all protein fold-changes that were quantified with more than three spectra.[297] Relative protein abundances were generated on the basis of MS1 abundances.[300] Briefly, XIC peaks were matched to identified peptides. The apex of the XIC peak was acquired to be within 30 seconds from the time of the MS2 event performed on the peptide precursor. The raw abundances of the XIC peaks of the peptides with identical sequences were summed (*i.e.* same sequence, but different charge states and/or different modifications), and the resulting single entity was referred to as a sequence. For each protein the 3 sequences with the highest raw XIC intensity from a given sample were selected and log 10 transformed. These values were then summed, and the mean was calculated. In cases where fewer than 3 sequences were associated, the mean was calculated using XIC intensities from either one or two sequences. The data was filtered to a minimum of two quantified unique spectra matches (qusm).

7.4.1.3 General Procedure 3: CuAAC-enabled PAL pulldown workflow

Click ligation

To 1 mL of irradiated sample in a 15 mL tube, 110 µL of a CuAAC click mix containing $CuSO_4$ (50 mM in water, 20 µL), TBTA ligand (1.7 mM in 1:4 (*v/v*) DMSO:*t*-butanol, 60 µL), biotin-azide (10 mM in DMSO, 10 µL) and freshly prepared TCEP (50 mM in water, 20 µL) was added. The samples were mixed by rotary shaking (550 rpm, rt, 1 h).

Acetone precipitation

Acetone (4 mL, pre-chilled to -25 °C) was added to each sample. The samples were vortexed briefly and stored at -25 °C overnight. The samples were centrifuged (4700 g, 10 min, 4 °C). The supernatant was aspirated, and the pellets were washed with 4:1 (*v/v*) acetone:water (5 mL). The samples were centrifuged (4700 g, 10 min, 4 °C) and the supernatant was aspirated and 5 M urea in 50 mM HEPES (150 µL) was added. The samples were sonicated (*ca*. 2 min per sample) and buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 0.4% IGEPAL CA-630, 1x PIM5 and 2% SDS) (150 µL) was added. The samples were sonicated (*ca*. 2 min) before diluting with buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 0.4% IGEPAL CA-630 and 1x PIM5) (1.2 mL).

Bead pulldown and on-bead digestion

Each sample (1.4 mL) was incubated (overhead rotation, 4 °C, 2 h) with Pierce high capacity neutravidin beads (35 µL dry bead volume) which were pre-washed with PBS. The samples were filtered (96-well filter plates) and the samples were washed with buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 0.4% IGEPAL CA-630 (10 × 1 mL) and 50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF and 0.2% IGEPAL CA-630) (5 × 1 mL). The samples were then incubated (50 °C, 30 min, 750 rpm rotary shaking) with 0.25x NuPAGE LDS sample buffer containing 50 mM DTT (50 µL). The samples were filtered and further washed with 50 mM HEPES, 400 mM NaCl, 0.5% SDS (5 × 1 mL), 50 mM HEPES, 400 mM NaCl (10 × 1 mL), 50 mM HEPES, 2 M urea (10 × 1 mL) and 50 mM HEPES (5 × 1 mL). 60 µL of an on-bead digestion mix containing 40 mM HEPES, 15 mM 2-chloroacetamide, 5 mM TCEP, 4 ng/µL LysC, 4 ng/µL Trypsin was added, and the samples were incubated overnight (rt, 750 rpm rotary shaking). The digests were filtered and the retentate was washed with 50 mM HEPES (60 µL). The filtrate was frozen (-80 °C), lyophilised and stored at -20 °C.

7.4.2 Chapter 4 – optimisation experiments and chemoproteomics

7.4.2.1 Investigation into CuAAC ligation rates with increasing concentrations of click mix A

Sample preparation and irradiation

4.5 (1 mM in DMSO, 30 µL, FAC = 10 µM) was added to BRD4 BD1 (1 µM in PBS, 3 mL) and incubated on ice for 30 min. DMSO (30 µL) was added to BRD4 BD1 (1 µM in PBS, 3 mL) and incubated on ice for 30 min. The samples (2.5 mL) were irradiated (302 nm, 10 min) on ice. The UV lamp was warmed for 2 min prior to sample irradiation. The sample containing **4.5** was analysed by intact protein LCMS (method B, Section 0) to determine percentage labelling (55%). 2 mL of each irradiated mixture was added to 1% DMSO in PBS (18 mL, 1 in 10 dilution, BRD4 BD1 FAC = 0.1 µM). These two solutions were labelled "probe sample" and "DMSO sample" respectively.

Click ligation

- 4x Loading buffer = 0.1% bromophenol blue, 100 mM DTT, 5% SDS, 300 mM Tris-HCl, 50% glycerol.
- Click mix A = Prepared by combining 50 mM $CuSO_4$ in water (15 µL, FAC = 8.33 mM), 1.7 mM TBTA in 1:4 DMSO:*t*-BuOH (45 µL, FAC = 0.85 mM), 1.25 mM Cy5 azide in DMSO (15 µL, FAC = 0.21 mM) and 50 mM TCEP in water (15 µL, FAC = 8.33 mM).

For the 0 min timepoint, "probe sample" (4 × 200 µL) was added to 4 × 1.5 mL eppendorf vials labelled A1, B1, C1 and D1 (Table 7.16), followed by 4x loading buffer (66.7 µL). Click mix A (1.0 µL, 2.0 µL, 4.0 µL and 8.0 µL) was added to vials A1, B1, C1 and D1 to give a final ratio of 1:1.05, 1:2.1, 1:4.2 and 1:8.4 of alkyne:azide, respectively. Eppendorf vials A1-D1 (40 µL) were transferred to a 96-well plate following the plate map described in Table 7.17. To 4 new eppendorf vials (A2–D2) containing "probe sample" (4 × 200 µL), click mix A (1.0, 2.0, 4.0 and 8.0 µL) was added respectively, initiating t=0 for the click reaction. At the indicated timepoint shown in Table 7.17, 30 µL of sample was transferred to the 96-well plate, followed by quenching with 4x loading buffer (10 µL). Four eppendorf vials (A3–D3)

containing "DMSO sample" (4 × 200 µL) were also prepared. Click mix A (1.0, 2.0, 4.0 and 8.0 µL) was added to A3–D3 respectively, and at the 140 min timepoint, 30 µL of each sample was transferred to the 96-well plate and quenched by the addition of 4x loading buffer (10 µL) following the plate map in Table 7.17.

Table 7.16 1.5 mL eppendorf vial layout and additions.

Eppendorf position	**1**	**2**	**3**
	"Probe sample" For 0 min timepoint	**"Probe sample" For remaining timepoints**	**"DMSO control"**
A (1:1 alkyne to azide)	Probe sample (200 µL) 4x Loading buffer (66.7 µL) Click mix A (1 µL)	Probe sample (200 µL) Click mix A (1 µL)	DMSO sample (200 µL) Click mix A (1 µL)
B (1:2 alkyne to azide)	Probe sample (200 µL) 4x Loading buffer (66.7 µL) Click mix A (2 µL)	Probe sample (200 µL) Click mix A (2 µL)	DMSO sample (200 µL) Click mix A (2 µL)
C (1:4 alkyne to azide)	Probe sample (200 µL) 4x Loading buffer (66.7 µL) Click mix A (4 µL)	Probe sample (200 µL) Click mix A (4 µL)	DMSO sample (200 µL) Click mix A (4 µL)
D (1:8 alkyne to azide)	Probe sample (200 µL) 4x Loading buffer (66.7 µL) Click mix A (8 µL)	Probe sample (200 µL) Click mix A (8 µL)	DMSO sample (200 µL) Click mix A (8 µL)

Table 7.17 96-well plate additions to quench the CuAAC reaction at the specified timepoints.

Plate position	**1**	**2**	**3**	**4**	**5**	**6**	**7** (DMSO control)
Timepoint for loading buffer	**0 min**	**10 min**	**20 min**	**40 min**	**80 min**	**140 min**	**140 min**
A (1:1 alkyne to azide)	Epi. A1 (40 µL)	Epi. A2 (30 µL) After 10 min: Loading buffer (10 µL)	Epi. A2 (30 µL) After 20 min: Loading buffer (10 µL)	Epi. A2 (30 µL) After 40 min: Loading buffer (10 µL)	Epi. A2 (30 µL) After 80 min: Loading buffer (10 µL)	Epi. A2 (30 µL) After 140 min: Loading buffer (10 µL)	Epi. A3 (30 µL) After 140 min: Loading buffer (10 µL)
B (1:2 alkyne to azide)	Epi. B1 (40 µL)	Epi. B2 (30 µL) After 10 min: Loading buffer (10 µL)	Epi. B2 (30 µL) After 20 min: Loading buffer (10 µL)	Epi. B2 (30 µL) After 40 min: Loading buffer (10 µL)	Epi. B2 (30 µL) After 80 min: Loading buffer (10 µL)	Epi. B2 (30 µL) After 140 min: Loading buffer (10 µL)	Epi. B3 (30 µL) After 140 min: Loading buffer (10 µL)
C (1:4 alkyne to azide)	Epi. C1 (40 µL)	Epi. C2 (30 µL) After 10 min: Loading buffer (10 µL)	Epi. C2 (30 µL) After 20 min: Loading buffer (10 µL)	Epi. C2 (30 µL) After 40 min: Loading buffer (10 µL)	Epi. C2 (30 µL) After 80 min: Loading buffer (10 µL)	Epi. C2 (30 µL) After 140 min: Loading buffer (10 µL)	Epi. C3 (30 µL) After 140 min: Loading buffer (10 µL)
D (1:8 alkyne to azide)	Epi. D1 (40 µL)	Epi. D2 (30 µL) After 10 min: Loading buffer (10 µL)	Epi. D2 (30 µL) After 20 min: Loading buffer (10 µL)	Epi. D2 (30 µL) After 40 min: Loading buffer (10 µL)	Epi. D2 (30 µL) After 80 min: Loading buffer (10 µL)	Epi. D2 (30 µL) After 140 min: Loading buffer (10 µL)	Epi. D3 (30 µL) After 140 min: Loading buffer (10 µL)

From the 96-well plate, samples (10 µL) were loaded onto NuPAGE 4–12% bis-tris protein gels (1.0 mm, 12-well) with an Invitrogen SeeBlue Plus2 ladder. Using 1x NuPAGE MES SDS running buffer, the gels were run for 35 min with constant voltage (200 V). The gels were imaged on a Li-cor Odyssey gel reader (version 2.1.15) and analysed using Image Studio (version 2.1.10). The background fluorescence (0 min timepoint) was subtracted and the values were averaged across the two duplicate gels. These values were plotted against click reaction time in GraphPad and fitted with one-phase decay non-linear regression to obtain a y_{max} value. This value was used

normalise the data to 100% reaction completion and these values were plotted against reaction time and fitted using a one-phase decay non-linear regression (Figure 4.3 right).

7.4.2.2 Investigating the compatibility of CuAAC ligation with various lysis buffer components

Sample preparation and irradiation

4.5 (1 mM in DMSO, 13.5 µL, FAC = 10 µM) was added to solutions of:

A: BRD4 BD1 (1 µM in PBS, 1.35 mL),

B: BRD4 BD1 (1 µM in 50 mM Tris-HCl, 1.35 mL)

C: BRD4 BD1 (1 µM in 50 mM HEPES, 1.35 mL)

These mixtures were incubated on ice for 30 min. 1 mL of each mixture was irradiated (302 nm, 10 min) on ice. The irradiated mixtures were labelled “Sample A”, Sample B” and “Sample C” respectively.

Preparation of buffer solutions, click ligation and analysis by in-gel fluorescence

A solution of each buffer component listed in Table 4.2 was made up in 50 mM Tris-HCl and in 50 mM HEPES. These buffer solutions (75 µL) were added to a 96-well plate following the plate map described in Table 7.18. The irradiated samples (A–C) in the respective buffers were added (7.5 µL), followed by click mix A (1.42 µL, alkyne:azide = 1:4, see Section 7.4.2.1 for click mix A details). This initiated t=0 for the click reaction, and after 1 h at rt, 4x loading buffer (25 µL, see Section 7.4.2.1 for 4x loading buffer details) was added to quench the click reaction.

Table 7.18 96-well plate map for assessing if the click reaction could proceed in various lysis buffer components.

96-well plate position XY		1	2	3	4	5	6	7	8	9
		Parent buffer only (Tris or HEPES)	**5% glycerol**	**1.5 mM $MgCl_2$**	**150 mM NaCl**	**1 mM Na_3VO_4**	**25 mM NaF**	**0.4% IGEPAL**	**0.2% SDS**	**No UV controls**
A	**Tris-HCl based**	Buffer (75 µL) Sample A (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample A (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample A (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample A (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample A (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample A (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample A (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample A (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Non-irradiated sample A (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)
B	**PBS based (controls)**	Buffer (75 µL) Sample B (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	NA	NA	NA	NA	NA	NA	NA	Buffer (75 µL) Non-irradiated sample B (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)
C	**HEPES based**	Buffer (75 µL) Sample C (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample C (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample C (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample C (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample C (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample C (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample C (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Sample C (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)	Buffer (75 µL) Non-irradiated sample C (7.5 µL) Click mix A (1.42 µL) After rt, 1 h: Loading buffer (25 µL)

The samples (10 µL) were loaded onto NuPAGE 4–12% bis-tris protein gels (1.0 mm, 15-well) with an Invitrogen SeeBlue Plus2 ladder. Using 1x NuPAGE MES SDS running buffer, the gels were run for 35 min with constant voltage (200 V). The gels were imaged on a Li-cor Odyssey gel reader (version 2.1.15) and analysed using Image Studio (version 2.1.10).

7.4.2.3 Development of an in-gel fluorescent protocol with HL-60 lysate spiked with BRD4 BD1

Preparation of lysate

A 2 mL aliquot of HL-60 lysate prepared by method A (Section 7.4.1.1) was thawed and 2-fold diluted with 2 mL of buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 1 mM Na_3VO_4, 25 mM NaF and 1x PIM5) to reduce the IGEPAL CA-630 concentration from 0.8% to 0.4%. The lysate was then further diluted to 1.0 mg/mL with buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 1 mM Na_3VO_4, 0.4% IGEPAL CA-630, 25 mM NaF and 1x PIM5) and aliquoted (7 × 200 µL) into a 96-well plate on ice.

Compound treatment and irradiation

A 6-point, 1 in 2 serial dilution of BRD4 BD1 was prepared (80, 40, 20, 10, 5, 2.5 µM in buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 1 mM Na_3VO_4, 0.4% IGEPAL CA-630, 25 mM NaF and 1x PIM5) and 2 µL was added to the respective well of lysate to give final BRD4 BD1 concentrations of (0.8, 0.4, 0.2, 0.1, 0.05, 0.025 µM) respectively. 2 µL of buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 1 mM Na_3VO_4, 0.4% IGEPAL CA-630, 25 mM NaF and 1x PIM5) was added to well 7 (no BRD4 BD1 added). For a DMSO control, buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 1 mM Na_3VO_4, 0.4% IGEPAL CA-630, 25 mM NaF and 1x PIM5) (200 µL) was added to well 8, followed by BRD4 BD1 (80 µM, 2 µL, FAC = 0.8 µM). **4.5** (1.0 mM in DMSO, 2 µL, FAC = 10 µM) was added to each well and the plate was incubated on ice for 30 min before irradiation (302 nm, 10 min). The UV lamp was warmed for 2 min prior to sample irradiation.

Click ligation and analysis by in-gel fluorescence

50 µL of each irradiated sample was transferred to a new 96-well plate at rt. Click mix B was prepared by combining 50 mM CuSO4 in water (15 µL), 1.7 mM TBTA in 1:4 (*v/v*) DMSO:*t*-BuOH (45 µL), 1.25 mM Cy5.5 azide (15 µL) and freshly made 50 mM TCEP in water (15 µL). Click mix B (5 µL, giving a final ratio of alkyne:azide = 1:2) was added to each sample and the plate was incubated at rt for 1 h. 2x loading buffer with DTT was prepared by combining 4x NuPAGE LDS sample buffer (500 µL), 1 M DTT (50 µL) and water (450 µL). After 1 h, 2x loading buffer (55 µL) was added to the samples and the plate was sealed with adhesive foil and heated (70 °C, 10 min). The samples (10 µL) were loaded onto NuPAGE 4–12% bis-tris protein gels (1.0 mm, 10-well) with a Li-cor Odyssey One-colour protein molecular weight marker. Using 1x NuPAGE MES SDS running buffer, the gels were run for 50 min with constant voltage (200 V). The gels were imaged on a Li-cor Odyssey gel reader (version 2.1.15) and analysed using Image Studio (version 2.1.10).

7.4.2.4 In-gel fluorescence for PAL probes 4.3–4.5 in BRD4 BD1 spiked HL-60 lysate with and without competition by (+)-JQ1

Preparation of lysate

HL-60 lysate prepared by method A (Section 7.4.1.1) (17.4 mg/mL, 2.0 mL) was 2-fold diluted with 2 mL of buffer A (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF and 1x PIM5) to reduce the IGEPAL CA-630 concentration from 0.8% to 0.4%. The lysate was further diluted with buffer B (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 0.4% IGEPAL CA-630 and 1x PIM5) to a final protein concentration of 1.0 mg/mL. BRD4 BD1 (676 µM, 1.5 µL, FAC = 0.1 µM) was added to 10 mL of 1.0 mg/mL lysate on ice and this mixture was aliquoted into a 96-well plate (8 × 220 µL).

Compound treatment and irradiation

The appropriate PAL probe **4.3**, **4.4** or **4.5** (200 µM in DMSO, 1.1 µL, FAC = 1 µM) was added followed by DMSO (1.1 µL) or (+)-JQ1 (40 mM in DMSO, 1.1 µL, FAC = 200 µM), or DMSO (2.2 µL) for DMSO control sample. The plate was sealed with adhesive film and incubated (4 °C, 30 min). Each sample (30 µL) was added in duplicate to two 96-well plates on ice and these were irradiated at 302 nm and 365 nm respectively (10 min) on ice. The UV lamp was warmed for 2 min prior to sample irradiation.

Click ligation and analysis by in-gel fluorescence

Click mix B was prepared by combining Cy5.5 azide (1.25 mM in DMSO, 50 µL), TBTA (1.7 mM in 1:4 (*v*/*v*) DMSO:*t*-BuOH, 150 µL), $CuSO_4$ (50 mM in water) and freshly prepared TCEP (50 mM in water, 50 µL). Click mix B (1 µL, giving a final ratio of alkyne:azide = 1:6.8) was added to each irradiated sample and the plates were incubated at rt for 1 h. 2x loading buffer was prepared by combining 4x NuPAGE sample buffer (500 µL), DTT (1 M in water, 50 µL) and water (450 µL). 2x loading buffer (30 µL) was added to each sample and the plates were sealed and heated (70 °C, 10 min). The samples (10 µL) were loaded onto NuPAGE 12% bis-tris protein gels (1.0 mm, 17-well) with an Li-cor Odyssey One-colour protein molecular weight marker. Using 1x NuPAGE MES SDS running buffer, the gels were run for 70 min

with constant voltage (180 V). The gels were imaged on a Li-cor Odyssey gel reader (version 2.1.15) and analysed using Image Studio (version 2.1.10).

7.4.2.5 Investigating CuAAC ligation efficiency with biotin-azide immobilised on neutravidin beads (Section 4.3.5)

Preparation of lysate

HL-60 lysate prepared by method A (Section 7.4.1.1) (17.4 mg/mL, 2.0 mL) was diluted with 50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 1x PIM5 (2.0 mL) to reduce the IGEPAL CA-630 concentration from 0.8% to 0.4%. The lysate was further diluted with 50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 0.4% IGEPAL CA-630, 1x PIM5 to a final protein concentration of 1.0 mg/mL. BRD4 BD1 (676 µM, 1.5 µL, FAC = 0.1 µM) was added to 10 mL of 1.0 mg/mL lysate on ice.

Compound treatment and irradiation

4.3 (0.2 mM in DMSO, 50 µL, FAC = 1 µM) was added and the mixture was incubated on ice for 30 min. 8.0 mL of this mixture was irradiated (302 nm, 10 min) in a reservoir on ice. The UV lamp was warmed for 2 min prior to sample irradiation. The irradiated mixture was transferred to a 15 mL tube on ice and labelled “irradiated lysate”.

Bead preparation

Pierce high capacity neutravidin beads (50% slurry, 2 × 280 µL (dry bead volume = 140 µL)) were transferred to 2 × 1.5 mL eppendorf vials and washed with PBS (3 × 0.5 mL). After the last aspiration, PBS (140 µL) was added to make a 50% slurry of beads in PBS (280 µL). Biotin-azide or biotin only (50 mM, 0.56 µL, FAC by dry bead volume = 0.2 mM) was added to each sample and briefly vortexed. The samples were incubated (4 °C, 30 min, 1400 rpm orbital mixing). To block the remaining neutravidin sites, samples were centrifuged (1200 rpm, 1 min) and biotin (100 mM in DMSO, 1.4 µL, FAC by dry bead volume = 1 mM) was added to both samples and briefly vortexed. Each sample was transferred to a 15 mL tube and centrifuged (1200 rpm, 1 min). The supernatant was removed, and the beads were washed with buffer (50 mM HEPES,

5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF and 0.4% IGEPAL CA-630) (3 × 2 mL). After the last aspiration, the beads were resuspended with buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF and 0.4% IGEPAL CA-630) (2.66 mL) to give a 5% slurry (2.8 mL) on ice. The 5% bead slurries (one labelled "biotin-azide" and the other "biotin only") were transferred in duplicate (2 × 700 µL each) to 1.5 mL eppendorf vials on ice to give 4 samples (1–4). The vials were centrifuged (1200 rpm, 1 min) and the supernatant was removed.

"Irradiated lysate" (700 µL) was added and the samples were briefly vortexed. For the zero min timepoint, a fifth vial was loaded with buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF and 0.4% IGEPAL CA-630) (35 µL) and "irradiated lysate" (700 µL) (no beads present).

Click ligation and analysis by in-gel fluorescence

Click mix was prepared by combing combining $CuSO_4$ (50 mM in water, 100 µL), TBTA (1.7 mM in 1:4 DMSO:*t*-BuOH, 300 µL) and freshly prepared TCEP (50 mM in water, 100 µL). Click mix (100 µL) was added to each sample (1–4), initiating t_1=0. For the zero timepoint, 20 µL samples from the fifth vial were transferred to a 96-well plate on ice. At each timepoint (7.5, 15, 30, 60 and 120 min), the first four vials (2 containing "biotin-azide" and 2 containing "biotin only" beads) were centrifuged (1200 rpm, 1 min) and 20 µL per sample was transferred to the 96-well plate. After each sampling, the vials were briefly vortexed and returned to mixing. Click mix B (90 µL) was prepared by combining $CuSO_4$ (50 mM in water, 15 µL), TBTA (1.7 mM in 1:4 DMSO:*t*-BuOH, 45 µL), Cy5.5 azide (1.25 mM in DMSO, 15 µL) and freshly prepared TCEP (50 mM in water, 15 µL). The 96-well plate was taken off ice to rt and click mix B (1 µL, giving a final ratio of alkyne:azide = 1:10) was added to each sample, initiating t=0 for the click reaction with fluorescent dye. The plate was sealed and incubated (rt, 1 h, 450 rpm). 2x loading buffer was prepared by combining 4x NuPAGE sample buffer (500 µL), DTT (1 M in water, 50 µL) and water (450 µL). 2x loading buffer (21 µL) was added to each sample. The plate was sealed and heated (70 °C, 10 min, 750 rpm). The samples (10 µL) were loaded onto NuPAGE 12% bis-tris protein gels (1.0 mm, 15-well) with an Li-cor Odyssey One-colour protein molecular weight marker. Using 1x NuPAGE MES SDS running buffer, the gels were run for 70

min with constant voltage (180 V). The gels were imaged on a Li-cor Odyssey gel reader (version 2.1.15) and analysed using Image Studio (version 2.1.10). The average fluorescent intensities at each timepoint were plotted using Graphpad Prism and fitted using a non-linear regression (one-phase decay). Error bars represent ±1 standard deviation.

7.4.2.6 Solution-based CuAAC ligation with biotin-azide (Section 4.3.5)

Protein preparation, compound treatment and irradiation

PAL probe **4.3** (0.2 mM in DMSO, 50 µL, FAC = 1 µM) was added to recombinant BRD4 BD1 (0.1 µM) in buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF and 0.4% IGEPAL CA-630) (10 mL) on ice and incubated for 30 min. 8 mL of this mixture was irradiated (302 nm, 10 min) on ice to give "irradiated mixture". The UV lamp was warmed for 2 min prior to sample irradiation.

Click ligation and analysis by in-gel fluorescence

"Irradiated mixture" was added to duplicate eppendorf vials containing biotin azide (0.2 mM in 50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 0.4% IGEPAL CA-630 and 1x PIM5, 35 µL) and biotin (0.2 mM in 50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 0.4% IGEPAL CA-630 and 1x PIM5, 35 µL). Click mix was prepared by combining $CuSO_4$ (50 mM in water, 200 µL), TBTA (1.7 mM in 1:4 DMSO:t-BuOH, 600 µL) and freshly prepared TCEP (50 mM in water (200 µL). Click mix (100 µL, giving final ratio of alkyne:biotin-azide = 1:7) was added to each vial initiating t_1=0 (time for first click reaction). Click mix B was prepared by combining $CuSO_4$ (50 mM in water, 50 µL), TBTA (1.7 mM in 1:4 DMSO:*t*-BuOH, 150 µL), Cy5.5 azide (1.25 mM in DMSO, 50 µL) and freshly prepared TCEP (50 mM in water, 50 µL). At t_1 = 7.5, 15, 30 and 60 min, a 20 µL sample from each vial was transferred to a 96-well plate at rt and click mix B (3 µL, giving Cy5.5 azide FAC = 27.4 µM, 5x biotin-azide concentration) was added, which initiated t_2=0 for each sample. 2x loading buffer was prepared by combining 4x NuPAGE sample buffer (500 µL), DTT (1 M in water, 50 µL) and water (450 µL). When t_2=1 h for each sample, 2x loading buffer (23 µL) was added. The plate was

sealed and heated (70 °C, 10 min, 750 rpm). The samples (10 µL) were loaded onto NuPAGE 4–12% bis-tris protein gels (1.0 mm, 12-well) with a Li-cor Odyssey One-colour protein molecular weight marker. Using 1x NuPAGE MES SDS running buffer, the gels were run for 70 min with constant voltage (180 V). The gels were imaged on a Li-cor Odyssey gel reader (version 2.1.15) and analysed using Image Studio (version 2.1.10). The average fluorescent intensities at each timepoint were plotted using Graphpad Prism and fitted using a non-linear regression (one-phase decay). Error bars represent ±1 standard deviation.

7.4.2.7 Acetone precipitation trial (Section 4.3.6)

Preparation of lysate

HL-60 lysate prepared by method A (Section 7.4.1.1) (17.4 mg/mL, 2.0 mL) was diluted 2-fold with 2 mL of buffer A (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF and 1x PIM5) to reduce the IGEPAL CA-630 concentration from 0.8% to 0.4%. The lysate was further diluted with buffer B (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 0.4% IGEPAL CA-630 and 1x PIM5) to a final protein concentration of 1.0 mg/mL. BRD4 BD1 (676 µM, 1.5 µL, FAC = 0.1 µM) was added to 10 mL of 1.0 mg/mL of "irradiated lysate" on ice.

Compound treatment, irradiation and click ligation

"Irradiated lysate" (4.5 mL) was added to wells A1 and A2 of a 6-well plate. **4.3** (0.2 mM in DMSO, 22.5 µL, FAC = 1 µM) was added to A1 and DMSO (22.5 µL, FAC = 0.5%) was added to A2. The plate was incubated (30 min, 4 °C) and then irradiated (302 nm, 10 min) on ice. Click mix was prepared by combining $CuSO_4$ (50 mM in water, 200 µL), TBTA (1.7 mM in 1:4 DMSO:*t*-BuOH, 600 µL), biotin-azide (5.0 mM in DMSO, 100 µL) and freshly prepared TCEP (50 mM in water, 200 µL). To 4.0 mL of each irradiated sample, click mix (440 µL, giving biotin-azide FAC = 45 µM) was added and the samples were mixed by overhead rotation (rt, 1 h) to give two samples "PAL probe" and "DMSO control" respectively.

No acetone precipitation workflow

1 mL of each sample ("PAL probe" and "DMSO control") were incubated in duplicate (overhead rotation, 4 °C, 2 h) with neutravidin beads (35 µL dry bead volume) which were pre-washed with PBS.

Acetone precipitation workflow

To 1.0 mL of each sample ("PAL probe" and "DMSO control") in duplicate, acetone (4.0 mL, pre-chilled to -25 °C) was added and the samples were briefly vortexed before being stored at -25 °C for 1.5 h. The samples were centrifuged (16000 × g, 10 min, 4 °C) and the supernatant was aspirated. 4:1 (*v/v*) acetone:water (4 mL, -25 °C) was added and the samples were centrifuged (16000 × g, 10 min, 4 °C). The supernatant was aspirated and 2.5 M urea in buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 0.4% IGEPAL CA-630 and 1x PIM5) (200 µL) was added. The samples were sonicated (*ca*. 5 min) and 20% SDS (8 µL) was added. The samples were further sonicated (*ca*. 1 min) before being diluted with buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 0.4% IGEPAL CA-630 and 1x PIM5) (800 µL) to give a final concentration of 0.5 M urea and 0.16% SDS. Each 1 mL sample was incubated (overhead rotation, 4 °C, 2 h) with neutravidin beads (35 µL dry bead volume) which were pre-washed with PBS.

Bead washing, digestion and analysis (both workflows)

For both workflows, each replicate was filtered and washed with buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF and 0.4% IGEPAL CA-630) (10 × 1 mL) and buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF and 0.2% IGEPAL CA-630 (5 × 1 mL). The samples were then incubated (50 °C, 30 min, 750 rpm rotary shaking) with 0.5x NuPAGE LDS sample buffer (Thermo Scientific, NP0007) containing 50 mM DTT (50 µL). The samples were filtered and further washed with 50 mM HEPES, 400 mM NaCl, 0.5% SDS (5 × 1 mL), 50 mM HEPES, 400 mM NaCl (10 × 1 mL), 50 mM HEPES, 2 M urea (10 × 1 mL) and 50 mM HEPES (5 × 1 mL). 60 µL of an on-bead digestion mix containing 40 mM HEPES, 15 mM 2-chloroacetamide, 5 mM TCEP, 4 ng/µL LysC and 4 ng/µL Trypsin was added and the samples were incubated overnight (rt, 750 rpm

rotary shaking). The digests were filtered and the retentate was washed with 50 mM HEPES (60 µL). The combined filtrate was frozen (-80 °C), lyophilised and stored at -20 °C. Further sample preparation and LC-MSMS analysis was performed following General Procedure 4 (Section 7.4.1.2). The average relative TMT values for channels 1–4 (no acetone precipitation) and 5–8 (with acetone precipitation) were calculated and plotted in a bar chart to show the difference in overall signal for when acetone precipitation was used. For TMT channels 5–8, log_2(PAL probe *vs* DMSO) was calculated for both replicates. The average log_2(PAL probe *vs* DMSO) values for each replicate were calculated and subtracted to centre the XY scatter plot on (0,0). The corrected log_2(PAL probe *vs* DMSO) for both replicates were plotted (replicate 1 on x-axis, replicate 2 on y axis) in Graphpad Prism (Version 5.04) (Figure 4.14).

7.4.2.8 PAL MS-based proteomics experiment in HL-60 lysates with 4.3, 4.4 and 4.5

HL-60 cells were cultured by Jürgen Stuhlfauth and Nilma Garcia-Altrieth.

Preparation of cell lysate

1 L of HL-60 cells (3.6×10^6 cells/mL, 80% viable) were harvested by centrifugation. The cells were washed with Gibco IMDM medium (without Phenol Red) (Thermo Scientific, 21056023), re-centrifuged and the pellet was diluted with IMDM medium (without Phenol Red) to give a final cell concentration of 5×10^6 cells/mL. Cells (5×10^6 cells/mL, 4 × 40 mL) were transferred to 50 mL tubes and centrifuged (1500 rpm, 5 min, rt) and the supernatant was aspirated. Lysis buffer containing 50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 1% IGEPAL CA-630, 1x PIM5 (in-house protease inhibitor mix containing 500 nM Aprotinin (Sigma, A1153), 40 µM Bestatin (Sigma, B8386), 100 µM Leupeptin (Sigma, L2884), 1.46 µM Pepstatin (Sigma, P5318), 10 µM Phosphoramidon (Sigma, R7386)) and Benzonase nuclease (Sigma, E1014) (1 µL/mL) was prepared. Lysis buffer (5x cell pellet volume) was added and the samples were mixed by rotary shaking (550 rpm, 30 min, rt). The lysate was centrifuged (20000 g, 20 min, 4 °C) and the supernatant was flash frozen and stored at -80 °C until required. The mixture was thawed on ice and cleared by

centrifugation (20000 × g, 20 min, 4 °C). The supernatant was transferred to a 15 mL tube and diluted (1 in 2.5) with buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF and 1x PIM5) to reduce the IGEPAL-CA630 concentration from 1% to 0.4%. The protein concentration was determined by BCA assay (Pierce BCA protein assay kit, Thermo Scientific, 23225, using Quick Start BSA Standard Set, Bio-Rad, 500-0207). The protein concentration was adjusted to 2.2 mg/mL by diluting with buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 0.4% IGEPAL CA-630 and 1x PIM5) to give "HL-60 lysate" (2.2 mg/mL, 20.8 mL).

Compound treatment, irradiation and pulldown

PAL probe (**4.3**, **4.4** or **4.5**), (+)-JQ1 and DMSO were added to 2 × 12-well plates on ice following Table 7.19.

Table 7.19 Irradiation plate compound additions.

XY	1	2	3	4	5	6
A	DMSO (3.75 µL) **4.3** (0.4 mM, 2.75 µL, FAC = 1 µM)	DMSO (3.75 µL) **4.3** (0.4 mM, 2.75 µL, FAC = 1 µM)	JQ1 (80 mM, 2.75 µL, FAC = 200 µM) **4.3** (0.4 mM, 2.75 µL, FAC = 1 µM)	JQ1 (80 mM, 2.75 µL, FAC = 200 µM) **4.3** (0.4 mM, 2.75 µL, FAC = 1 µM)	DMSO (5.5 µL)	DMSO (5.5 µL)
B	DMSO (3.75 µL) **4.4** (0.4 mM, 2.75 µL, FAC = 1 µM)	DMSO (3.75 µL) **4.4** (0.4 mM, 2.75 µL, FAC = 1 µM)	JQ1 (80 mM, 2.75 µL, FAC = 200 µM) **4.4** (0.4 mM, 2.75 µL, FAC = 1 µM)	JQ1 (80 mM, 2.75 µL, FAC = 200 µM) **4.4** (0.4 mM, 2.75 µL, FAC = 1 µM)	DMSO (5.5 µL)	DMSO (5.5 µL)
C	DMSO (3.75 µL) **4.5** (0.4 mM, 2.75 µL, FAC = 1 µM)	DMSO (3.75 µL) **4.5** (0.4 mM, 2.75 µL, FAC = 1 µM)	JQ1 (80 mM, 2.75 µL, FAC = 200 µM) **4.5** (0.4 mM, 2.75 µL, FAC = 1 µM)	JQ1 (80 mM, 2.75 µL, FAC = 200 µM) **4.5** (0.4 mM, 2.75 µL, FAC = 1 µM)	DMSO (5.5 µL)	DMSO (5.5 µL)

"HL-60 lysate" (2.2 mg/mL, 1.1 mL) was added to each well and the plate was incubated (4 °C, 1 h, 200 rpm) before being irradiated (302 nm, 10 min). The UV lamp was warmed for 2 min prior to sample irradiation. 1 mL of each irradiated sample was transferred to a 15 mL tube. CuAAC-enabled pulldown was performed following General Procedure 5 described in Section 7.4.1.3. Further sample preparation and LC-MSMS analysis was performed following General Procedure 4 (Section 7.4.1.2).

Data treatment

After filtering for ≥2 qusm, the relative TMT values for each replicate were averaged and the fold-changes were calculated for PAL probe *vs* DMSO control and PAL probe *vs* competition. The log_2 values of these fold-changes were calculated and an average log_2 value was subtracted to centralise the plots on (0,0). An XY scatter plot of these corrected log_2 values was created in GraphPad Prism (version 5.04 for Windows) (Figure 4.16). These data for PAL probes **4.3**, **4.4** and **4.5** are shown in Table 7.20, Table 7.21 and Table 7.22, respectively.

Table 7.20 Proteins significantly enriched (Log_2(fold change for probe vs DMSO) ≥ 1) for alkyl diazirine probe **4.3** from HL-60 cell lysate.

Gene name	Uniprot ID	qusm	MS1 intensity (log_{10})	TMT 126[a]	TMT 127L[a]	TMT 127H[b]	TMT 128L[b]	TMT 128H[c]	TMT 129L[c]	Log_2(fold change for probe vs competition corrected)	Log_2(fold change for probe vs DMSO corrected)
BRD3	Q15059	4	7.2305	1.00	1.24	0.14	0.12	0.10	0.09	3.19	3.64
GSTP1	P09211	5	7.6412	1.00	1.16	1.28	1.32	0.34	0.33	-0.15	1.78
HMGCS1	Q01581	2	6.5295	1.00	1.14	1.55	1.28	0.34	0.34	-0.29	1.74
P4HA1	P13674	4	6.8928	1.00	1.11	1.13	1.24	0.42	0.35	-0.05	1.53
GART	P22102	3	6.5061	1.00	1.70	1.61	1.38	0.52	0.55	-0.03	1.42
HADH	Q16836	2	6.7234	1.00	1.42	1.42	1.68	0.55	0.49	-0.24	1.31
HIST1H2A	P0C0S8	4	7.2060	1.00	5.62	1.02	1.73	1.87	1.16	1.38	1.21
PRCP	P42785	2	6.2289	1.00	1.29	1.15	1.56	0.57	0.56	-0.13	1.10
DCTPP1	Q9H773	2	6.9763	1.00	1.18	1.22	1.26	0.54	0.55	-0.07	1.09

[a] Duplicate samples treated with PAL probe **4.3** + DMSO. [b] Duplicate samples treated with PAL probe **4.3** + (+)-JQ1. [c] Duplicate samples treated with DMSO only.

Table 7.21 Proteins significantly enriched (Log_2(fold change for probe vs DMSO) ≥ 1) for trifluoromethyl aryl diazirine probe **4.4** from HL-60 cell lysate.

Gene name	Uniprot ID	qusm	MS1 intensity (log_{10})	TMT 126[a]	TMT 127L[a]	TMT 127H[b]	TMT 128L[b]	TMT 128H[c]	TMT 129L[c]	Log_2(fold change for probe vs competition corrected)	Log_2(fold change for probe vs DMSO corrected)
SLC16A3	O15427	2	6.9486	1.00	1.05	0.79	0.88	0.07	0.04	0.18	3.97
CALM1	P0DP23	4	7.6322	1.00	0.80	0.72	0.85	0.04	0.07	0.09	3.90
BRD3	Q15059	4	7.2284	1.00	0.95	0.09	0.10	0.06	0.08	3.24	3.64
FDFT1	P37268	11	7.7000	1.00	1.17	0.63	0.68	0.09	0.15	0.62	2.96
SLC25A3	Q00325	3	7.0677	1.00	1.03	0.75	0.88	0.10	0.13	0.21	2.91
BRD4	O60885	3	7.1949	1.00	0.89	0.12	0.11	0.08	0.15	2.91	2.81
NOC3L	Q8WTT2	2	6.5343	1.00	1.02	0.78	1.00	0.14	0.15	0.07	2.55
PRTN3	P24158	3	7.6838	1.00	1.01	0.67	0.93	0.17	0.14	0.22	2.48
SURF4	O15260	2	6.4377	1.00	1.47	0.98	1.20	0.15	0.24	0.07	2.41
DAD1	P61803	2	6.8803	1.00	1.21	0.95	1.11	0.19	0.21	-0.01	2.23
LMAN2	Q12907	2	7.3701	1.00	1.00	0.93	1.02	0.13	0.23	-0.07	2.22
MDN1	Q9NU22	2	6.6817	1.00	1.01	0.76	0.78	0.23	0.17	0.27	2.11
EIF3A	Q14152	3	6.6309	1.00	1.14	0.88	1.00	0.22	0.26	0.07	1.92
PWP2	Q15269	2	6.2681	1.00	0.99	0.71	0.87	0.24	0.21	0.22	1.90
PUM3	Q15397	2	6.3335	1.00	1.03	0.72	0.82	0.36	0.11	0.28	1.87
RAP1B	P61224	2	7.3401	1.00	0.99	0.77	0.83	0.26	0.26	0.20	1.67
PRKDC	P78527	31	7.3649	1.00	1.04	0.78	0.90	0.26	0.29	0.17	1.64
BRIX1	Q8TDN6	2	6.3374	1.00	1.07	0.90	1.07	0.31	0.26	-0.04	1.60
IPO7	O95373	4	6.6874	1.00	1.15	0.88	1.04	0.21	0.39	0.06	1.59

ARL6IP1	Q15041	2	7.2457	1.00	1.10	0.90	0.98	0.29	0.31	0.05	1.56
ATP5J2	P56134	2	6.7250	1.00	1.23	0.89	1.07	0.33	0.31	0.07	1.55
SEC63	Q9UGP8	2	6.3364	1.00	1.03	0.89	0.91	0.31	0.31	0.06	1.47
PDS5B	Q9NTI5	2	6.8483	1.00	0.99	0.80	0.89	0.16	0.46	0.12	1.44
STT3A	P46977	3	6.5908	1.00	1.14	0.95	1.07	0.30	0.38	-0.03	1.41
SAR1A	Q9NR31	2	6.4740	1.00	1.12	0.87	0.89	0.33	0.36	0.16	1.38
CANX	P27824	8	7.5737	1.00	1.00	0.76	0.72	0.29	0.36	0.32	1.36
M6PR	P20645	3	6.7389	1.00	1.02	0.89	0.95	0.34	0.33	0.02	1.35
YARS	P54577	3	7.0623	1.00	1.01	0.79	0.88	0.33	0.34	0.16	1.34
DHRS7	Q9Y394	3	6.5063	1.00	1.23	0.84	1.00	0.41	0.36	0.16	1.30
IPO5	O00410	7	7.0959	1.00	1.02	0.77	0.89	0.34	0.36	0.17	1.28
NCBP1	Q09161	2	5.7592	1.00	1.29	0.66	1.11	0.46	0.35	0.26	1.25
UBR4	Q5T4S7	3	5.7968	1.00	0.87	0.67	0.83	0.31	0.37	0.21	1.22
ATP1A1	P05023	4	6.6721	1.00	1.14	0.97	0.99	0.39	0.40	0.01	1.20
SLC25A6	P12236	2	6.8056	1.00	1.08	0.87	1.02	0.33	0.45	0.03	1.18
XPO4	Q9C0E2	2	6.2254	1.00	1.10	0.79	0.97	0.41	0.39	0.14	1.15
LAMP1	P11279	3	6.4708	1.00	0.90	0.83	0.94	0.33	0.44	-0.01	1.06
EIF4G1	Q04637	4	6.8955	1.00	1.05	0.82	0.90	0.37	0.47	0.14	1.03
HLA-B	P10319	6	7.2054	1.00	1.21	0.79	0.77	0.46	0.47	0.40	1.01
TMEM33	P57088	3	6.5245	1.00	1.30	1.06	1.20	0.47	0.50	-0.08	1.00

[a] Duplicate samples treated with PAL probe **4.4** + DMSO. [b] Duplicate samples treated with PAL probe **4.4** + (+)-JQ1. [c] Duplicate samples treated with DMSO only.

Table 7.22 Proteins significantly enriched (Log_2(fold change for probe vs DMSO) ≥ 1) for benzophenone probe **4.5** from HL-60 cell lysate.

Gene name	Uniprot ID	qusm	MS1 intensity (log_{10})	TMT 126[a]	TMT 127L[a]	TMT 127H[b]	TMT 128L[b]	TMT 128H[c]	TMT 129L[c]	Log_2(fold change for probe vs competition corrected)	Log_2(fold change for probe vs DMSO corrected)
CD63	P08962	3	7.3344	1.00	1.35	0.65	0.65	0.03	0.02	0.59	4.73
BRD3	Q15059	5	7.3745	1.00	1.36	0.14	0.14	0.05	0.05	2.78	3.55
RAB21	Q9UL25	2	6.7039	1.00	1.37	0.84	0.95	0.08	0.06	0.15	3.12
ARL6IP1	Q15041	3	7.0868	1.00	1.41	0.99	0.79	0.15	0.03	0.18	2.78
RRP12	Q5JTH9	2	6.7951	1.00	1.51	0.89	0.85	0.17	0.03	0.28	2.74
VAMP8	Q9BV40	2	7.3850	1.00	1.33	0.75	0.73	0.09	0.09	0.40	2.71
BRD2	P25440	2	6.7192	1.00	1.23	0.13	0.14	0.11	0.10	2.77	2.49
CPT2	P23786	2	6.6746	1.00	1.38	0.78	0.72	0.14	0.08	0.40	2.47
TM9SF3	Q9HD45	3	6.7096	1.00	1.25	0.97	0.94	0.11	0.11	-0.02	2.41
ARL8B	Q9NVJ2	2	6.2671	1.00	1.41	0.93	0.92	0.15	0.12	0.12	2.22
ACADVL	P49748	5	7.3079	1.00	1.43	0.63	0.61	0.14	0.14	0.71	2.16
TRIP13	Q15645	3	6.3132	1.00	1.32	0.80	0.92	0.17	0.13	0.18	2.03
RAB5B	P61020	2	6.5355	1.00	1.39	1.06	1.02	0.15	0.15	-0.06	2.03
IPO7	O95373	4	6.6759	1.00	1.38	0.82	0.81	0.17	0.14	0.29	1.99
PRTN3	P24158	2	7.5703	1.00	1.39	0.74	0.65	0.18	0.14	0.52	1.93
SLC29A1	Q99808	2	6.3639	1.00	1.45	0.89	0.93	0.22	0.11	0.17	1.91
TROVE2	P10155	2	6.6421	1.00	1.27	1.15	1.15	0.21	0.13	-0.28	1.77
PDS5B	Q9NTI5	2	6.5209	1.00	1.33	0.83	0.87	0.20	0.16	0.19	1.75
WASHC5	Q12768	2	6.1706	1.00	1.40	0.76	0.86	0.20	0.18	0.31	1.72
LAMP1	P11279	2	6.8367	1.00	1.30	0.90	0.84	0.19	0.17	0.14	1.71
NUP155	O75694	2	6.2004	1.00	1.43	0.82	0.88	0.28	0.14	0.26	1.59
RAB5C	P51148	2	6.9260	1.00	1.41	1.14	1.23	0.21	0.21	-0.23	1.55
SLC16A3	O15427	3	6.6234	1.00	1.39	0.83	0.76	0.27	0.15	0.33	1.54
DHRS7	Q9Y394	3	6.6261	1.00	1.35	0.74	0.81	0.31	0.11	0.34	1.53
RAB10	P61026	2	6.4565	1.00	1.56	0.96	0.92	0.28	0.18	0.19	1.52
CAB39	Q9Y376	2	6.1253	1.00	1.22	0.85	0.84	0.25	0.15	0.14	1.50
ACAD9	Q9H845	2	6.3917	1.00	1.36	0.83	0.74	0.24	0.21	0.33	1.46
DDOST	P39656	2	6.0894	1.00	1.18	0.99	0.77	0.26	0.15	0.05	1.45
CANX	P27824	10	7.5715	1.00	1.39	0.78	0.80	0.24	0.21	0.35	1.45

LBR	Q14739	3	6.1070	1.00	1.36	0.80	0.85	0.34	0.11	0.26	1.44
M6PR	P20645	3	6.6702	1.00	1.33	0.77	0.80	0.24	0.21	0.32	1.44
PRKDC	P78527	37	7.3672	1.00	1.30	0.83	0.86	0.26	0.19	0.19	1.43
LMAN2	Q12907	2	6.4355	1.00	1.40	0.84	0.85	0.24	0.24	0.26	1.37
TSPAN14	Q8NG11	2	6.9107	1.00	1.36	0.81	0.80	0.25	0.23	0.30	1.35
RAP1B	P61224	4	7.1817	1.00	1.32	1.02	1.01	0.25	0.22	-0.07	1.34
FDFT1	P37268	6	6.8385	1.00	1.31	0.67	0.73	0.25	0.23	0.47	1.33
CYBB	P04839	2	6.3081	1.00	1.44	0.88	0.87	0.35	0.18	0.22	1.26
RAB3D	O95716	3	6.2671	1.00	1.36	0.96	0.99	0.26	0.25	0.02	1.25
NOC3L	Q8WTT2	3	6.4452	1.00	1.36	1.01	0.99	0.26	0.26	-0.01	1.23
IPO5	O00410	9	7.1227	1.00	1.27	0.80	0.83	0.26	0.26	0.22	1.18
ATP1A1	P05023	2	6.6510	1.00	1.29	0.76	0.86	0.31	0.22	0.24	1.16
EEF1G	P26641	9	7.4500	1.00	1.27	0.83	0.90	0.29	0.24	0.14	1.16
TFRC	P02786	4	6.6087	1.00	1.31	0.71	0.74	0.28	0.25	0.42	1.16
MAT2A	P31153	4	6.9774	1.00	1.23	0.93	1.02	0.28	0.24	-0.07	1.15
TUBA1B	P68363	10	7.8475	1.00	1.26	0.78	0.84	0.27	0.26	0.23	1.14
SLC25A3	Q00325	4	7.2187	1.00	1.34	0.87	0.85	0.38	0.18	0.18	1.11
RHOA	P61586	5	7.0874	1.00	1.20	0.83	0.91	0.30	0.23	0.08	1.10
SURF4	O15260	3	6.6572	1.00	1.45	0.91	0.87	0.37	0.24	0.20	1.06
MYBBP1A	Q9BQG0	2	6.1377	1.00	1.34	0.87	0.84	0.37	0.21	0.20	1.06
PREX1	Q8TCU6	2	6.1856	1.00	1.38	0.81	0.97	0.39	0.20	0.16	1.05
SLC25A5	P05141	5	7.0492	1.00	1.39	0.84	0.91	0.34	0.26	0.19	1.04
UBR4	Q5T4S7	3	5.9487	1.00	1.37	0.75	0.90	0.27	0.34	0.27	1.02

[a] Duplicate samples treated with PAL probe **4.5** + DMSO. [b] Duplicate samples treated with PAL probe **4.5** + (+)-JQ1. [c] Duplicate samples treated with DMSO only.

7.4.2.9 PAL MS-based proteomics experiment in live HL-60 cells with PAL probes 4.3, 4.4 and 4.5

HL-60 cells were cultured by Jürgen Stuhlfauth and Nilma Garcia-Altrieth.

Cell treatment, irradiation and lysis

1 L of HL-60 cells (3.6×10^6 cells/mL, 80% viable) were harvested by centrifugation. The cells were washed with Gibco IMDM medium (without Phenol Red), re-centrifuged and the pellet was diluted with IMDM medium (without Phenol Red) to give a final cell concentration of 5×10^6 cells/mL. For each probe, six 10 cm culture dishes were prepared containing:

1. PAL probe (400 µM, 37.5 µL in DMSO, FAC = 1 µM) + DMSO (37.5 µL)
2. PAL probe (400 µM, 37.5 µL in DMSO, FAC = 1 µM) + DMSO (37.5 µL)
3. PAL probe (400 µM, 37.5 µL in DMSO, FAC = 1 µM) + Competitor **4.2** (20 mM, 37.5 µL, FAC = 50 µM)

4. PAL probe (400 µM, 37.5 µL in DMSO, FAC = 1 µM) + Competitor **4.2** (20 mM, 37.5 µL, FAC = 50 µM)
5. DMSO (75 µL)
6. DMSO (75 µL)

Cells (5 × 10^6 cells/mL, 15 mL) were added to each dish and incubated (37 °C, 1 h, 5% CO_2, 50 rpm rotary shaking). The plates were then irradiated on ice (302 nm, 10 min). The UV lamp was warmed for 2 min prior to sample irradiation. The contents of each irradiated dish were transferred to a 15 mL tube and centrifuged (1500 rpm, 5 min, rt) and the supernatant was aspirated. Lysis buffer containing 50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 1% IGEPAL CA-630, 1x PIM5 (in-house protease inhibitor mix containing 500 nM Aprotinin, 40 µM Bestatin, 100 µM Leupeptin, 1.46 µM Pepstatin, 10 µM Phosphoramidon) and Benzonase nuclease (Sigma, E1014) (1 µL/mL) was prepared. Lysis buffer (5x cell pellet volume) was added and the samples were mixed by rotary shaking (550 rpm, 30 min, rt). The lysate was centrifuged (20000 g, 20 min, 4 °C) and the supernatants were transferred to 1.5 mL tubes and diluted (1 in 2.5) with buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF and 1x PIM5) to reduce the IGEPAL-CA630 concentration from 1% to 0.4%. The protein concentration was determined by BCA assay (Pierce BCA protein assay kit, Thermo Scientific, 23225, using Quick Start BSA Standard Set, Bio-Rad, 500-0207). The protein concentration in each sample was adjusted to 2.4 mg/mL by diluting with buffer (50 mM HEPES, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 25 mM NaF, 0.4% IGEPAL CA-630 and 1x PIM5). 1 mL of each sample was transferred to a 15 mL tube.

Click ligation, pulldown and data treatment

CuAAC-enabled pulldown was performed following General Procedure 5 described in Section 7.4.1.3. Further sample preparation and LC-MSMS analysis (2 injections, data combined) was performed following General Procedure 4 (Section 7.4.1.2). After filtering for ≥2 qusm, the relative TMT values for each replicate were averaged and the fold-changes were calculated for PAL probe *vs* DMSO control and PAL probe *vs* competition. The proteins that were significantly enriched by probes **4.3**, **4.4** and **4.5** (log_2(PAL probe *vs* DMSO) ≥ 1) are listed in Table 7.23, Table 7.24 and Table 7.25,

respectively. XY scatter plots were created in GraphPad Prism (version 5.04 for Windows) for probes **4.3**, **4.4** and **4.5** (Figure 4.18). Gene ontology analysis was performed using DAVID (version 6.8) bioinformatics resources.[252,253] The 66 proteins significantly enriched by both probes **4.4** and **4.5** were used as the gene list and all proteins identified by both probes (≥2 qusm, duplicates removed, 656 proteins) was used as the background list. Biological process (GOTERM_BP_3) and Molecular function (GOTERM_MF_3) terms were obtained with an EASE cut-off of 0.05. These terms and their associated p-values are shown in Figure 4.19. The 18 genes contained within the biological process GO term "Transmembrane transport" and the associated $\log_2$(PAL probe *vs* DMSO) values for probes **4.4** and **4.5** are shown in Table 4.3.

Table 7.23 Proteins significantly enriched (Log_2(fold change for probe vs DMSO) ≥ 1) for alkyl diazirine **4.3** from HL-60 cells.

Gene name	Uniprot ID	qusm	MS1 intensity ($\log_{10}$)	TMT 126[a]	TMT 127L[a]	TMT 127H[b]	TMT 128L[b]	TMT 128H[c]	TMT 129L[c]	Log_2(fold change for probe vs competition corrected)	Log_2(fold change for probe vs DMSO corrected)
SLC16A3	O15427	5	7.0232	1.00	1.18	0.94	0.94	0.12	0.11	0.61	3.77
SLC25A20	O43772	12	7.7766	1.00	1.15	1.34	1.45	0.16	0.14	0.03	3.35
ERMP1	Q7Z2K6	15	7.8098	1.00	1.07	1.11	1.38	0.32	0.36	0.14	2.16

[a] Duplicate samples treated with PAL probe **4.3** + DMSO. [b] Duplicate samples treated with PAL probe **4.3** + parent affinity function **4.2**. [c] Duplicate samples treated with DMSO only.

Table 7.24 Proteins significantly enriched (Log_2(fold change for probe vs DMSO) ≥ 1) for tetrafluoromethyl aryl diazirine probe **4.4** from HL-60 cells.

Gene name	Uniprot ID	qusm	MS1 intensity ($\log_{10}$)	TMT 126[a]	TMT 127L[a]	TMT 127H[b]	TMT 128L[b]	TMT 128H[c]	TMT 129L[c]	Log_2(fold change for probe vs competition corrected)	Log_2(fold change for probe vs DMSO corrected)
BRD3	Q15059	6	7.2996	1.00	1.20	0.19	0.19	0.08	0.08	2.51	3.75
BRD4	O60885	5	7.0048	1.00	1.26	0.30	0.31	0.19	0.19	1.90	2.54
BRD2	P25440	4	7.0734	1.00	1.24	0.35	0.34	0.08	0.08	1.71	3.75
ORM1	P02763	2	6.6322	1.00	1.17	1.02	0.88	0.12	0.11	0.20	3.25
ABCB10	Q9NRK6	7	7.2243	1.00	1.93	1.12	1.44	0.09	0.05	0.20	4.41
HM13	Q8TCT9	2	6.7751	1.00	2.19	1.35	1.46	0.40	0.40	0.18	2.00
SLC16A10	Q8TF71	9	7.8068	1.00	1.33	1.07	0.99	0.06	0.05	0.18	4.38
VDAC1	P21796	2	6.6500	1.00	2.23	1.32	1.55	0.66	0.77	0.17	1.18
ADPGK	Q9BRR6	2	6.2225	1.00	1.81	1.24	1.41	0.55	0.50	0.08	1.41
TMEM43	Q9BTV4	3	6.9864	1.00	2.02	1.27	1.59	0.51	0.43	0.08	1.68
DHRS7	Q9Y394	2	6.4826	1.00	1.62	1.12	1.36	0.43	0.46	0.08	1.56
SLC16A3	O15427	19	8.7776	1.00	1.70	1.25	1.30	0.08	0.08	0.08	4.06
MTDH	Q86UE4	2	6.2266	1.00	1.23	1.11	1.01	0.34	0.56	0.08	1.31
NSDHL	Q15738	3	6.4178	1.00	1.39	1.24	1.03	0.44	0.41	0.07	1.50
PRDX4	Q13162	3	6.3933	1.00	1.77	1.30	1.37	0.57	0.68	0.06	1.15
DCTPP1	Q9H773	7	7.4575	1.00	1.40	1.23	1.08	0.28	0.29	0.06	2.07
BSG	P35613	2	7.1640	1.00	1.56	1.21	1.28	0.28	0.29	0.04	2.15
DEGS1	O15121	3	6.3864	1.00	2.00	1.35	1.58	0.51	0.45	0.03	1.64
ALDH18A1	P54886	4	6.8217	1.00	1.18	1.04	1.10	0.15	0.13	0.03	2.93
HACD3	Q9P035	2	6.5015	1.00	1.66	1.21	1.40	0.53	0.42	0.02	1.49
MIF	P14174	4	7.2665	1.00	1.39	1.16	1.20	0.66	0.50	0.02	1.04

CPOX	P36551	4	6.7004	1.00	1.55	1.25	1.28	0.22	0.25	0.01	2.44
IPO4	Q8TEX9	3	6.2829	1.00	1.23	1.05	1.16	0.25	0.17	0.01	2.41
SCARB1	Q8WTV0	4	6.9096	1.00	1.45	1.20	1.24	0.23	0.21	0.01	2.49
HSD17B4	P51659	11	7.1373	1.00	1.36	1.21	1.13	0.38	0.40	0.01	1.61
ESYT1	Q9BSJ8	5	6.6972	1.00	1.52	1.23	1.30	0.57	0.55	-0.01	1.18
SCCPDH	Q8NBX0	8	7.7325	1.00	1.39	1.18	1.23	0.16	0.17	-0.01	2.90
TMPPE	Q6ZT21	5	7.0409	1.00	1.59	1.23	1.38	0.46	0.48	-0.01	1.47
TBXAS1	P24557	3	6.6635	1.00	1.54	1.30	1.29	0.12	0.13	-0.03	3.36
UGGT1	Q9NYU2	3	6.4936	1.00	1.51	1.26	1.31	0.47	0.46	-0.04	1.44
MYDGF	Q969H8	4	6.9171	1.00	1.66	1.31	1.43	0.56	0.63	-0.04	1.16
CALM1	P0DP23	6	7.3064	1.00	1.06	1.12	1.01	0.12	0.14	-0.04	2.99
ERMP1	Q7Z2K6	38	8.3849	1.00	1.62	1.31	1.39	0.15	0.16	-0.04	3.09
ABCC4	O15439	2	6.6195	1.00	1.32	1.17	1.23	0.25	0.25	-0.04	2.21
SCP2	P22307	6	6.9280	1.00	1.27	1.24	1.10	0.49	0.56	-0.05	1.10
P4HB	P07237	87	9.4295	1.00	1.28	1.18	1.18	0.12	0.12	-0.05	3.27
SLC16A7	O60669	2	7.3397	1.00	1.37	1.23	1.22	0.10	0.10	-0.05	3.53
XPO1	O14980	9	7.2433	1.00	1.19	1.19	1.09	0.29	0.29	-0.05	1.92
PSMB2	P49721	14	7.4985	1.00	1.44	1.14	1.40	0.28	0.27	-0.06	2.15
PRCP	P42785	5	7.1294	1.00	1.27	1.22	1.15	0.27	0.25	-0.06	2.15
ATP5A1	P25705	2	6.5462	1.00	1.27	1.16	1.21	0.56	0.53	-0.06	1.06
PDIA6	Q15084	6	7.1621	1.00	1.51	1.28	1.34	0.58	0.54	-0.06	1.17
STT3A	P46977	5	6.7618	1.00	1.54	1.25	1.40	0.38	0.30	-0.06	1.90
ATP5B	P06576	6	7.1445	1.00	1.32	1.22	1.21	0.48	0.46	-0.06	1.31
ACSL3	O95573	7	7.6388	1.00	0.96	1.04	1.00	0.07	0.08	-0.07	3.75
RPN1	P04843	12	7.3431	1.00	1.66	1.29	1.49	0.43	0.47	-0.07	1.56
SLC25A20	O43772	45	9.2684	1.00	1.42	1.25	1.28	0.10	0.10	-0.07	3.59
UNC119B	A6NIH7	7	7.5680	1.00	1.52	1.22	1.43	0.18	0.17	-0.07	2.83
ACOX3	O15254	17	7.5030	1.00	1.31	1.21	1.22	0.10	0.10	-0.08	3.52
SLC25A3	Q00325	9	7.4225	1.00	1.46	1.27	1.33	0.47	0.41	-0.08	1.49
NDUFA4	O00483	3	7.0901	1.00	1.18	1.12	1.19	0.43	0.38	-0.08	1.42
SOAT1	P35610	2	6.5627	1.00	1.75	1.49	1.43	0.30	0.14	-0.08	2.65
ERP44	Q9BS26	9	7.5596	1.00	1.28	1.24	1.17	0.29	0.26	-0.08	2.04
MFSD10	Q14728	7	7.0412	1.00	1.57	1.38	1.34	0.22	0.19	-0.08	2.64
SLC29A1	Q99808	4	6.9250	1.00	1.47	1.36	1.25	0.21	0.19	-0.08	2.65
PSMB6	P28072	3	7.0618	1.00	1.20	1.14	1.20	0.40	0.42	-0.08	1.42
ACADVL	P49748	10	7.7028	1.00	1.30	1.25	1.20	0.13	0.14	-0.09	3.11
CAB39	Q9Y376	3	6.5132	1.00	1.28	1.20	1.23	0.56	0.53	-0.09	1.06
IARS	P41252	4	6.6715	1.00	1.33	1.33	1.15	0.47	0.43	-0.09	1.37
HADHA	P40939	19	7.3194	1.00	1.42	1.27	1.32	0.27	0.27	-0.10	2.18
ZMPSTE24	O75844	26	8.2821	1.00	1.34	1.25	1.26	0.07	0.06	-0.10	4.19
RHOT2	Q8IXI1	4	6.7473	1.00	1.33	1.22	1.28	0.16	0.17	-0.10	2.81
TMEM167A	Q8TBQ9	3	6.3517	1.00	1.44	1.33	1.28	0.45	0.38	-0.10	1.57
PREPL	Q4J6C6	4	6.5176	1.00	1.09	1.13	1.11	0.08	0.08	-0.10	3.73
ATP5C1	P36542	3	7.3474	1.00	1.16	1.17	1.16	0.30	0.34	-0.11	1.77
MPV17	P39210	4	7.1181	1.00	1.44	1.26	1.37	0.04	0.03	-0.11	5.18
CPT1A	P50416	2	6.6857	1.00	1.24	1.20	1.23	0.46	0.43	-0.11	1.34
UBA52	P62987	9	7.7071	1.00	1.35	1.24	1.31	0.57	0.50	-0.11	1.14
BRIX1	Q8TDN6	2	6.3758	1.00	1.41	1.28	1.33	0.57	0.51	-0.12	1.17
DHCR24	Q15392	5	6.7700	1.00	1.36	1.26	1.30	0.46	0.46	-0.12	1.35
TSPO	P30536	4	7.4901	1.00	1.21	1.17	1.23	0.38	0.40	-0.12	1.51
IPO7	O95373	3	6.4200	1.00	1.19	1.31	1.07	0.48	0.58	-0.12	1.04
TM9SF3	Q9HD45	3	6.8291	1.00	1.49	1.30	1.41	0.25	0.26	-0.12	2.28
TMED10	P49755	2	7.2448	1.00	1.78	1.47	1.57	0.41	0.24	-0.13	2.10
COX7A2	P14406	2	6.4626	1.00	1.46	1.27	1.42	0.50	0.34	-0.13	1.55
CES2	O00748	2	7.0367	1.00	1.47	1.38	1.32	0.20	0.18	-0.13	2.69
TYMS	P04818	3	7.0075	1.00	1.41	1.38	1.26	0.16	0.16	-0.13	2.91
GLUD1	P00367	7	7.4283	1.00	1.39	1.27	1.34	0.28	0.31	-0.13	2.00
HPCAL1	P37235	4	7.0306	1.00	1.05	1.17	1.07	0.32	0.44	-0.13	1.43
OCIAD1	Q9NX40	3	6.8216	1.00	1.25	1.27	1.19	0.26	0.26	-0.13	2.10

MTAP	Q13126	3	6.2704	1.00	1.44	1.31	1.37	0.58	0.59	-0.13	1.06
PMPCA	Q10713	3	6.6662	1.00	1.63	1.35	1.54	0.45	0.42	-0.13	1.59
PSMB9	P28065	2	6.9173	1.00	1.03	1.11	1.11	0.34	0.31	-0.13	1.64
HMOX2	P30519	5	6.6917	1.00	1.27	1.27	1.22	0.56	0.54	-0.13	1.03
CYP51A1	Q16850	6	7.2952	1.00	1.31	1.27	1.27	0.09	0.08	-0.13	3.70
CPT2	P23786	3	6.5636	1.00	1.44	1.35	1.33	0.35	0.36	-0.14	1.79
DPP8	Q6V1X1	7	6.9635	1.00	1.37	1.26	1.34	0.36	0.35	-0.14	1.74
VAMP8	Q9BV40	2	6.8595	1.00	1.43	1.31	1.37	0.57	0.55	-0.14	1.12
NUCB2	P80303	19	7.4859	1.00	1.20	1.26	1.16	0.16	0.17	-0.14	2.74
TXNDC5	Q8NBS9	8	7.1121	1.00	1.47	1.32	1.40	0.49	0.50	-0.14	1.32
GANAB	Q14697	4	6.7092	1.00	1.43	1.34	1.34	0.51	0.59	-0.14	1.15
MT-ND4	P03905	3	6.5771	1.00	1.49	1.46	1.29	0.36	0.24	-0.15	2.06
PSMA6	P60900	11	7.7052	1.00	1.08	1.12	1.18	0.35	0.33	-0.15	1.63
ABHD10	Q9NUJ1	3	6.8624	1.00	1.21	1.24	1.21	0.47	0.44	-0.15	1.28
MT-ATP6	P00846	4	6.4984	1.00	1.40	1.30	1.36	0.25	0.18	-0.15	2.46
CD81	P60033	2	6.2833	1.00	1.49	1.36	1.41	0.42	0.41	-0.15	1.58
LONP1	P36776	4	6.5154	1.00	1.33	1.32	1.28	0.61	0.52	-0.15	1.05
CANX	P27824	17	7.6647	1.00	1.32	1.27	1.33	0.51	0.55	-0.16	1.13
VDAC2	P45880	4	6.5083	1.00	1.44	1.38	1.35	0.39	0.45	-0.16	1.54
PRTN3	P24158	10	9.0477	1.00	1.19	1.31	1.14	0.12	0.13	-0.16	3.11
ACAD9	Q9H845	5	6.7195	1.00	1.39	1.37	1.31	0.35	0.43	-0.16	1.62
TMEM33	P57088	4	6.9386	1.00	1.50	1.35	1.45	0.43	0.41	-0.17	1.58
FDFT1	P37268	64	8.7229	1.00	1.33	1.32	1.29	0.12	0.13	-0.17	3.23
PSMB1	P20618	11	7.4492	1.00	1.31	1.29	1.32	0.36	0.31	-0.17	1.78
PPOX	P50336	6	6.8895	1.00	1.38	1.33	1.35	0.61	0.57	-0.17	1.01
LPCAT3	Q6P1A2	3	6.4430	1.00	1.47	1.37	1.41	0.49	0.41	-0.17	1.45
ACSL4	O60488	8	7.1390	1.00	1.11	1.25	1.13	0.23	0.21	-0.17	2.26
ADK	P55263	8	7.1603	1.00	1.30	1.33	1.26	0.33	0.34	-0.18	1.78
PSMB8	P28062	16	7.9686	1.00	1.34	1.28	1.37	0.19	0.20	-0.18	2.60
PPIB	P23284	3	6.5441	1.00	1.32	1.31	1.32	0.62	0.53	-0.18	1.01
TXNDC12	O95881	3	6.8711	1.00	1.46	1.41	1.38	0.50	0.41	-0.18	1.43
EPHX1	P07099	2	6.5205	1.00	1.65	1.41	1.60	0.61	0.59	-0.18	1.15
SLC22A18	Q96BI1	3	6.8928	1.00	1.34	1.35	1.31	0.19	0.11	-0.18	3.00
HEBP1	Q9NRV9	2	7.0002	1.00	1.31	1.29	1.34	0.35	0.32	-0.19	1.77
HSP90B1	P14625	16	7.6806	1.00	1.42	1.33	1.44	0.55	0.55	-0.20	1.14
ERO1A	Q96HE7	4	6.6210	1.00	1.35	1.35	1.35	0.52	0.61	-0.20	1.05
BRI3BP	Q8WY22	2	6.7707	1.00	1.40	1.40	1.35	0.17	0.13	-0.20	2.99
PREP	P48147	3	6.5721	1.00	1.56	1.45	1.50	0.58	0.67	-0.21	1.02
BCCIP	Q9P287	2	6.3135	1.00	1.02	1.20	1.14	0.31	0.40	-0.22	1.49
SLC16A1	P53985	7	7.1811	1.00	1.24	1.38	1.22	0.25	0.25	-0.22	2.15
VAT1	Q99536	4	7.4158	1.00	1.26	1.38	1.25	0.47	0.58	-0.22	1.10
SURF4	O15260	4	6.6072	1.00	1.39	1.43	1.37	0.42	0.40	-0.23	1.54
G6PD	P11413	2	7.1085	1.00	1.30	1.37	1.35	0.44	0.47	-0.24	1.34
PITRM1	Q5JRX3	13	7.2730	1.00	1.29	1.32	1.38	0.12	0.12	-0.24	3.25
ASNA1	O43681	2	6.9117	1.00	1.14	1.26	1.28	0.44	0.42	-0.25	1.33
PLAUR	Q03405	3	6.5025	1.00	1.14	1.25	1.30	0.09	0.09	-0.25	3.57
SLC33A1	O00400	6	7.2394	1.00	1.34	1.43	1.36	0.12	0.09	-0.25	3.49
PSAP	P07602	4	6.4790	1.00	1.18	1.39	1.26	0.47	0.58	-0.28	1.05
MLEC	Q14165	4	6.8328	1.00	1.27	1.41	1.36	0.53	0.53	-0.28	1.09
GPR183	P32249	2	6.2799	1.00	1.31	1.56	1.29	0.25	0.10	-0.30	2.75
MFSD14B	Q5SR56	4	7.4524	1.00	1.26	1.44	1.35	0.13	0.14	-0.31	3.08
RECK	O95980	7	6.8304	1.00	1.20	1.43	1.30	0.25	0.23	-0.31	2.19
RTN4	Q9NQC3	2	6.3912	1.00	1.33	1.51	1.41	0.45	0.45	-0.33	1.37
COPG2	Q9UBF2	2	5.7602	1.00	1.72	1.66	1.89	0.54	0.76	-0.38	1.06
SMARCA1	P28370	2	6.4374	1.00	1.24	2.02	1.78	0.43	0.51	-0.76	1.26

[a] Duplicate samples treated with PAL probe **4.4** + DMSO. [b] Duplicate samples treated with PAL probe **4.4** + parent affinity function **4.2**. [c] Duplicate samples treated with DMSO only.

Table 7.25 Proteins significantly enriched (Log_2(fold change for probe vs DMSO) ≥ 1) for benzophenone probe **4.5** from HL-60 cells.

Gene name	Uniprot ID	qusm	MS1 intensity (log10)	TMT 126[a]	TMT 127L[a]	TMT 127H[b]	TMT 128L[b]	TMT 128H[c]	TMT 129L[c]	Log2(fold change for probe vs competition corrected)	Log2(fold change for probe vs DMSO corrected)
BRD4	O60885	4	6.7365	1.00	1.00	0.12	0.13	0.11	0.12	2.99	3.10
BRD3	Q15059	4	6.7762	1.00	1.06	0.14	0.16	0.11	0.12	2.75	3.15
BRD2	P25440	2	6.9058	1.00	1.02	0.16	0.17	0.14	0.25	2.62	2.39
SUPT16H	Q9Y5B9	3	6.7037	1.00	1.08	0.27	0.32	0.27	0.29	1.83	1.90
ZMYM1	Q5SVZ6	5	6.8349	1.00	1.10	0.39	0.39	0.04	0.04	1.42	4.69
ANXA11	P50995	3	6.1429	1.00	0.96	0.60	0.78	0.39	0.40	0.51	1.31
ALDH3A2	P51648	3	6.4276	1.00	0.84	0.52	0.81	0.17	0.20	0.46	2.33
RAB11B	Q15907	2	6.8296	1.00	1.02	0.71	0.78	0.42	0.57	0.45	1.04
UBA52	P62987	7	7.5450	1.00	1.26	0.74	0.93	0.44	0.47	0.43	1.32
SLC16A10	Q8TF71	3	6.5186	1.00	1.08	0.80	0.78	0.21	0.18	0.39	2.44
NDUFA5	Q16718	2	6.8435	1.00	1.14	0.76	0.90	0.45	0.56	0.36	1.08
DCTPP1	Q9H773	3	7.1425	1.00	0.91	0.73	0.77	0.09	0.12	0.35	3.18
SCARB1	Q8WTV0	5	6.9433	1.00	1.04	0.69	0.90	0.14	0.18	0.35	2.64
CALR	P27797	13	7.9226	1.00	1.26	0.80	1.01	0.49	0.54	0.33	1.14
ATP5B	P06576	5	6.7907	1.00	1.22	0.82	0.97	0.41	0.46	0.31	1.35
LMAN1	P49257	4	6.5138	1.00	1.25	0.82	1.00	0.50	0.62	0.31	1.00
ALDH18A1	P54886	3	6.7243	1.00	1.02	0.74	0.91	0.15	0.11	0.29	2.93
MIF	P14174	2	7.5193	1.00	1.13	0.78	0.96	0.37	0.33	0.29	1.62
MPO	P05164	49	8.2462	1.00	1.17	0.81	0.97	0.38	0.41	0.28	1.45
TMEM43	Q9BTV4	2	6.9433	1.00	1.00	0.72	0.93	0.20	0.27	0.28	2.10
TMEM167A	Q8TBQ9	2	6.5078	1.00	1.44	1.00	1.02	0.50	0.50	0.28	1.28
SCP2	P22307	6	6.8966	1.00	1.03	0.83	0.86	0.36	0.42	0.27	1.39
ATP5C1	P36542	13	8.3442	1.00	0.85	0.63	0.90	0.08	0.08	0.27	3.51
EPDR1	Q9UM22	3	6.3493	1.00	1.04	0.81	0.89	0.40	0.43	0.26	1.30
LONP1	P36776	3	6.5263	1.00	1.27	0.89	1.01	0.34	0.37	0.25	1.68
RAB21	Q9UL25	2	6.8752	1.00	1.19	0.87	1.01	0.30	0.28	0.22	1.90
PDIA6	Q15084	6	7.1149	1.00	1.17	0.81	1.06	0.41	0.45	0.22	1.35
SOAT1	P35610	2	6.2991	1.00	1.02	0.78	0.96	0.30	0.18	0.21	2.09
RPN1	P04843	8	7.1249	1.00	1.16	0.85	1.02	0.35	0.40	0.21	1.53
CANX	P27824	14	7.6267	1.00	1.13	0.84	1.00	0.43	0.55	0.21	1.13
MT-ND4	P03905	3	6.3930	1.00	1.12	0.86	0.98	0.33	0.27	0.20	1.80
BSG	P35613	8	7.0821	1.00	0.97	0.81	0.92	0.23	0.23	0.19	2.10
ABCB10	Q9NRK6	4	6.9539	1.00	0.84	0.66	0.96	0.10	0.08	0.19	3.38
CES2	O00748	2	6.7455	1.00	0.98	0.82	0.93	0.16	0.16	0.18	2.61
SLC29A2	Q14542	2	6.7401	1.00	1.18	0.98	0.95	0.53	0.21	0.18	1.56
VDAC2	P45880	2	7.1903	1.00	1.13	0.87	1.02	0.40	0.47	0.18	1.29
CTSD	P07339	3	6.5461	1.00	1.11	0.91	0.96	0.32	0.44	0.17	1.46
SLC25A5	P05141	9	7.2451	1.00	1.16	0.89	1.03	0.50	0.54	0.17	1.05
ATP1A1	P05023	3	6.6861	1.00	1.10	0.87	1.00	0.42	0.42	0.16	1.31
CPT2	P23786	9	7.3628	1.00	0.80	0.71	0.90	0.09	0.10	0.16	3.25
ERO1A	Q96HE7	3	6.7922	1.00	1.23	0.87	1.12	0.50	0.49	0.16	1.17
SCCPDH	Q8NBX0	11	7.7686	1.00	0.69	0.65	0.86	0.08	0.08	0.16	3.46
MYDGF	Q969H8	3	7.1808	1.00	0.99	0.80	0.99	0.39	0.50	0.16	1.16
MPV17	P39210	4	7.2048	1.00	1.00	0.87	0.93	0.05	0.05	0.15	4.30
B2M	P61769	2	6.0573	1.00	1.42	0.98	1.20	0.67	0.50	0.15	1.05
ERP29	P30040	10	7.3776	1.00	1.08	0.83	1.05	0.39	0.50	0.15	1.24
ACTR3	P61158	3	6.9604	1.00	1.07	0.91	0.96	0.44	0.57	0.15	1.04
P4HB	P07237	56	8.6033	1.00	1.03	0.86	0.98	0.25	0.31	0.15	1.84
NOC3L	Q8WTT2	2	6.2186	1.00	1.38	1.08	1.08	0.53	0.53	0.14	1.17
SLC22A18	Q96BI1	3	7.0969	1.00	0.99	0.88	0.93	0.03	0.02	0.13	5.25
HSP90B1	P14625	17	7.6490	1.00	1.11	0.88	1.05	0.47	0.56	0.13	1.03
SLC25A3	Q00325	6	7.1142	1.00	1.07	0.86	1.03	0.41	0.46	0.13	1.25
LBR	Q14739	3	6.3191	1.00	1.19	0.93	1.07	0.42	0.49	0.13	1.27

PRTN3	P24158	7	8.1461	1.00	1.09	0.91	1.00	0.21	0.22	0.13	2.27
HEBP1	Q9NRV9	3	6.6601	1.00	0.92	0.84	0.92	0.10	0.13	0.13	3.01
EEF1G	P26641	21	7.7351	1.00	1.19	1.00	1.01	0.43	0.56	0.13	1.15
IPO7	O95373	6	6.4075	1.00	1.07	0.96	0.94	0.35	0.36	0.12	1.54
NUCB2	P80303	2	6.0361	1.00	1.02	0.84	1.01	0.33	0.39	0.12	1.49
CPOX	P36551	8	6.9107	1.00	0.95	0.87	0.93	0.18	0.20	0.11	2.34
LPCAT3	Q6P1A2	2	6.1522	1.00	1.24	0.97	1.10	0.38	0.34	0.11	1.63
OCIAD1	Q9NX40	3	6.6704	1.00	1.08	0.95	0.98	0.35	0.40	0.11	1.47
VDAC3	Q9Y277	4	6.6605	1.00	1.15	0.84	1.15	0.29	0.34	0.11	1.76
UGGT1	Q9NYU2	2	6.4499	1.00	0.93	0.79	1.02	0.28	0.30	0.10	1.75
DHCR24	Q15392	4	6.7329	1.00	1.04	0.88	1.04	0.37	0.42	0.09	1.37
GANAB	Q14697	3	6.7608	1.00	1.14	0.87	1.13	0.47	0.52	0.09	1.11
ERMP1	Q7Z2K6	18	7.5550	1.00	0.87	0.82	0.94	0.16	0.18	0.09	2.49
CPT1A	P50416	4	6.3116	1.00	1.00	0.87	1.01	0.31	0.32	0.09	1.67
SLC25A1	P53007	2	6.3977	1.00	1.25	1.00	1.11	0.46	0.41	0.09	1.37
ADK	P55263	2	7.0129	1.00	1.10	0.97	1.01	0.39	0.47	0.08	1.29
SURF4	O15260	4	6.5761	1.00	1.10	0.91	1.07	0.46	0.45	0.08	1.21
PSMB1	P20618	4	7.1214	1.00	1.15	0.94	1.10	0.28	0.33	0.07	1.81
ACAD9	Q9H845	15	7.8090	1.00	1.01	0.93	0.98	0.10	0.11	0.07	3.23
MFSD10	Q14728	8	7.6146	1.00	1.04	0.94	1.00	0.21	0.22	0.07	2.24
MTAP	Q13126	10	7.4417	1.00	0.83	0.82	0.93	0.21	0.23	0.06	2.07
ACOX3	O15254	9	7.1815	1.00	0.98	0.88	1.02	0.11	0.13	0.06	2.99
ZMPSTE24	O75844	14	7.6163	1.00	0.95	0.90	0.98	0.15	0.16	0.05	2.68
SLC3A2	P08195	4	6.8357	1.00	1.11	0.95	1.09	0.27	0.29	0.05	1.91
FDFT1	P37268	14	7.4249	1.00	1.01	0.94	1.01	0.25	0.29	0.05	1.92
SLC16A7	O60669	2	6.9127	1.00	0.98	0.93	0.99	0.06	0.06	0.04	3.99
HADHA	P40939	16	7.4252	1.00	0.98	0.91	1.03	0.14	0.14	0.03	2.80
PSMB2	P49721	9	7.0909	1.00	0.94	0.87	1.04	0.30	0.35	0.03	1.58
SLC29A1	Q99808	9	7.9698	1.00	1.01	0.99	0.98	0.14	0.14	0.03	2.86
TBXAS1	P24557	2	6.4491	1.00	0.92	0.94	0.95	0.20	0.35	0.03	1.80
SLC25A20	O43772	42	8.9802	1.00	0.93	0.90	1.00	0.06	0.06	0.02	3.93
PPIB	P23284	3	6.7450	1.00	1.13	0.95	1.15	0.42	0.46	0.02	1.27
PSMB4	P28070	3	7.0284	1.00	1.06	1.02	1.02	0.49	0.42	0.01	1.18
SLC16A3	O15427	28	8.9232	1.00	0.92	0.99	0.95	0.03	0.03	-0.01	5.13
HADHB	P55084	8	7.1970	1.00	1.05	1.00	1.07	0.16	0.16	-0.02	2.68
ACADVL	P49748	38	8.7034	1.00	0.91	0.98	0.98	0.10	0.11	-0.03	3.17
MBOAT7	Q96N66	3	6.7770	1.00	1.03	1.08	1.03	0.35	0.36	-0.06	1.50
PSMB8	P28062	8	7.4795	1.00	1.05	1.03	1.13	0.27	0.30	-0.07	1.85
SLC16A1	P53985	7	7.3853	1.00	0.98	1.10	0.99	0.22	0.22	-0.08	2.19
FNTB	P49356	2	7.0334	1.00	0.94	1.03	1.02	0.07	0.05	-0.08	3.95
PITRM1	Q5JRX3	20	7.8793	1.00	1.02	1.11	1.08	0.08	0.08	-0.12	3.72
PCTP	Q9UKL6	2	6.7099	1.00	1.14	2.10	1.42	0.16	0.25	-0.72	2.38

[a] Duplicate samples treated with PAL probe **4.5** + DMSO. [b] Duplicate samples treated with PAL probe **4.5** + parent affinity function **4.2**. [c] Duplicate samples treated with DMSO only.

Chapter 8 References

(1) Bunnage, M. E. *Nat. Chem. Biol.* **2011**, *7*, 335.
(2) Samsdodd, F. *Drug Discov. Today* **2005**, *10*, 139.
(3) Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A. *Nat. Rev. Drug Discov.* **2015**, *14*, 475.
(4) Hay, M.; Thomas, D. W.; Craighead, J. L.; Economides, C.; Rosenthal, J. *Nat. Biotechnol.* **2014**, *32*, 40.
(5) Cook, D.; Brown, D.; Alexander, R.; March, R.; Morgan, P.; Satterthwaite, G.; Pangalos, M. N. *Nat. Rev. Drug Discov.* **2014**, *13*, 419.
(6) Mohs, R. C.; Greig, N. H. *Alzheimers Dement (N Y)* **2017**, *3*, 651.
(7) Smalley, E. *Nat. Biotechnol.* **2017**, *35*, 604.
(8) Comess, K. M.; McLoughlin, S. M.; Oyer, J. A.; Richardson, P. L.; Stockmann, H.; Vasudevan, A.; Warder, S. E. *J. Med. Chem.* **2018**, *61*, 8504.
(9) Janzen, W. P. *Chem. Biol.* **2014**, *21*, 1162.
(10) Castaldi, M. P.; Zuhl, A.; Ricchiuto, P.; Hendricks, J. A. In *Annu. Rep. Med. Chem.*; Goodnow, R. A., Ed.; Academic Press: 2017; Vol. 50, p 335.
(11) Schenone, M.; Dancik, V.; Wagner, B. K.; Clemons, P. A. *Nat. Chem. Biol.* **2013**, *9*, 232.
(12) Holenz, J.; Stoy, P. *Bioorg. Med. Chem. Lett.* **2019**, *29*, 517.
(13) Muranaka, H.; Momose, T.; Handa, C.; Ozawa, T. In *Photoaffinity Labeling for Structural Probing Within Protein*; Hatanaka, Y., Hashimoto, M., Eds.; Springer Japan: Tokyo, 2017, p 241.
(14) Smith, E.; Collins, I. *FutureMedChem* **2015**, *7*, 159.
(15) Lapinsky, D. J.; Johnson, D. S. *FutureMedChem* **2015**, *7*, 2143.
(16) Sumranjit, J.; Chung, S. J. *Molecules* **2013**, *18*, 10425.
(17) Su, Y.; Ge, J.; Zhu, B.; Zheng, Y. G.; Zhu, Q.; Yao, S. Q. *Curr. Opin. Chem. Biol.* **2013**, *17*, 768.
(18) Parker, C. G.; Galmozzi, A.; Wang, Y.; Correia, B. E.; Sasaki, K.; Joslyn, C. M.; Kim, A. S.; Cavallaro, C. L.; Lawrence, R. M.; Johnson, S. R.; Narvaiza, I.; Saez, E.; Cravatt, B. F. *Cell* **2017**, *168*, 527.
(19) Denton, K. E.; Krusemark, C. J. *Med. Chem. Commun.* **2016**, *7*, 2020.
(20) Benjamin, C.; Yujia, W.; Melissa, D.; Jarrett, R.; Hsin-yu, L.; Marian, K.; Steven, G.; Gregory, V.; Michael, L.; Christopher, P. *ChemRxiv* **2019**, preprint.
(21) Pan, S.; Jang, S. Y.; Wang, D.; Liew, S. S.; Li, Z.; Lee, J. S.; Yao, S. Q. *Angew. Chem. Int. Ed.* **2017**, *56*, 11816.
(22) Li, Z.; Hao, P.; Li, L.; Tan, C. Y.; Cheng, X.; Chen, G. Y.; Sze, S. K.; Shen, H. M.; Yao, S. Q. *Angew. Chem. Int. Ed.* **2013**, *52*, 8551.
(23) Shi, H.; Zhang, C.-J.; Chen, G. Y. J.; Yao, S. Q. *J. Am. Chem. Soc.* **2012**, *134*, 3001.
(24) Nguyen, C.; West, G. M.; Geoghegan, K. F. In *Cancer Gene Networks*; Kasid, U., Clarke, R., Eds.; Springer New York: New York, NY, 2017, p 11.
(25) Jones, L. H.; Neubert, H. *Sci. Transl. Med.* **2017**, *9*, eaaf7951.
(26) Geurink, P. P.; Prely, L. M.; van der Marel, G. A.; Bischoff, R.; Overkleeft, H. S. In *Activity-Based Protein Profiling*; Sieber, S. A., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012, p 85.

(27) Sadaghiani, A. M.; Verhelst, S. H.; Bogyo, M. *Curr. Opin. Chem. Biol.* **2007**, *11*, 20.
(28) McClure, R. A.; Williams, J. D. *ACS Med. Chem. Lett.* **2018**, *9*, 785.
(29) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Anal. Chem.* **2006**, *78*, 2113.
(30) Simon, G. M.; Niphakis, M. J.; Cravatt, B. F. *Nat. Chem. Biol.* **2013**, *9*, 200.
(31) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd *Chem. Rev.* **2013**, *113*, 2343.
(32) Bantscheff, M.; Eberhard, D.; Abraham, Y.; Bastuck, S.; Boesche, M.; Hobson, S.; Mathieson, T.; Perrin, J.; Raida, M.; Rau, C.; Reader, V.; Sweetman, G.; Bauer, A.; Bouwmeester, T.; Hopf, C.; Kruse, U.; Neubauer, G.; Ramsden, N.; Rick, J.; Kuster, B.; Drewes, G. *Nat. Biotechnol.* **2007**, *25*, 1035.
(33) Kawatani, M.; Osada, H. *Med. Chem. Commun.* **2014**, *5*, 277.
(34) Rutkowska, A.; Thomson, D. W.; Vappiani, J.; Werner, T.; Mueller, K. M.; Dittus, L.; Krause, J.; Muelbaier, M.; Bergamini, G.; Bantscheff, M. *ACS Chem. Biol.* **2016**, *11*, 2541.
(35) Yang, Y.; Yang, X.; Verhelst, S. H. *Molecules* **2013**, *18*, 12599.
(36) Guo, H.; Li, Z. *Med. Chem. Commun.* **2017**, *8*, 1585.
(37) Kang, K.; Park, J.; Kim, E. *Proteome Sci.* **2016**, *15*, 15.
(38) Martell, J.; Weerapana, E. *Molecules* **2014**, *19*, 1378.
(39) Mackinnon, A. L.; Taunton, J. *Curr. Protoc. Chem. Biol.* **2009**, *1*, 55.
(40) Ravasco, J. M. J. M.; Monteiro, C. M.; Trindade, A. F. *Org. Chem. Front.* **2017**, *4*, 1167.
(41) Kleiner, P.; Heydenreuter, W.; Stahl, M.; Korotkov, V. S.; Sieber, S. A. *Angew. Chem. Int. Ed.* **2017**, *56*, 1396.
(42) Bantscheff, M.; Hopf, C.; Savitski, M. M.; Dittmann, A.; Grandi, P.; Michon, A. M.; Schlegl, J.; Abraham, Y.; Becher, I.; Bergamini, G.; Boesche, M.; Delling, M.; Dumpelfeld, B.; Eberhard, D.; Huthmacher, C.; Mathieson, T.; Poeckel, D.; Reader, V.; Strunk, K.; Sweetman, G.; Kruse, U.; Neubauer, G.; Ramsden, N. G.; Drewes, G. *Nat. Biotechnol.* **2011**, *29*, 255.
(43) Zheng, W.; Li, G.; Li, X. *Arch. Pharm. Res.* **2015**, *38*, 1661.
(44) Wang, S.; Tian, Y.; Wang, M.; Wang, M.; Sun, G. B.; Sun, X. B. *Front. Pharmacol.* **2018**, *9*, 353.
(45) Tuley, A.; Fast, W. *Biochemistry* **2018**, *57*, 3326.
(46) Gehringer, M.; Laufer, S. A. *J. Med. Chem.* **2019**, *62*, 5673.
(47) Yu, J.; Yang, X.; Sun, Y.; Yin, Z. *Angew. Chem. Int. Ed.* **2018**, *57*, 11598.
(48) Martin, J. S.; MacKenzie, C. J.; Fletcher, D.; Gilbert, I. H. *Bioorg. Med. Chem.* **2019**, *27*, 2066.
(49) Lin, S.; Yang, X.; Jia, S.; Weeks, A. M.; Hornsby, M.; Lee, P. S.; Nichiporuk, R. V.; Iavarone, A. T.; Wells, J. A.; Toste, F. D.; Chang, C. J. *Science* **2017**, *355*, 597.
(50) Taylor, M. T.; Nelson, J. E.; Suero, M. G.; Gaunt, M. J. *Nature* **2018**, *562*, 563.
(51) D'Ascenzio, M.; Pugh, K. M.; Konietzny, R.; Berridge, G.; Tallant, C.; Hashem, S.; Monteiro, O.; Thomas, J. R.; Schirle, M.; Knapp, S.; Marsden, B.; Fedorov, O.; Bountra, C.; Kessler, B. M.; Brennan, P. E. *Angew. Chem.* **2018**, *131*, 1019.
(52) Simon, G. M.; Cravatt, B. F. *J. Biol. Chem.* **2010**, *285*, 11051.

(53) Cravatt, B. F.; Wright, A. T.; Kozarich, J. W. *Annu. Rev. Biochem.* **2008**, *77*, 383.
(54) Kathman, S. G.; Xu, Z.; Statsyuk, A. V. *J. Med. Chem.* **2014**, *57*, 4969.
(55) Kharenko, O. A.; Patel, R. G.; Brown, S. D.; Calosing, C.; White, A.; Lakshminarasimhan, D.; Suto, R. K.; Duffy, B. C.; Kitchen, D. B.; McLure, K. G.; Hansen, H. C.; van der Horst, E. H.; Young, P. R. *J. Med. Chem.* **2018**, *61*, 8202.
(56) Tolmachova, K. A.; Moroz, Y. S.; Konovets, A.; Platonov, M. O.; Vasylchenko, O. V.; Borysko, P.; Zozulya, S.; Gryniukova, A.; Bogolubsky, A. V.; Pipko, S.; Mykhailiuk, P. K.; Brovarets, V. S.; Grygorenko, O. O. *ACS Comb. Sci.* **2018**, *20*, 672.
(57) Mortenson, D. E.; Brighty, G. J.; Plate, L.; Bare, G.; Chen, W.; Li, S.; Wang, H.; Cravatt, B. F.; Forli, S.; Powers, E. T.; Sharpless, K. B.; Wilson, I. A.; Kelly, J. W. *J. Am. Chem. Soc.* **2018**, *140*, 200.
(58) Shannon, D. A.; Weerapana, E. *Curr. Opin. Chem. Biol.* **2015**, *24*, 18.
(59) Eddleston, M.; Cohen, A. F.; Webb, D. J. *Br. J. Clin. Pharmacol* **2016**, *81*, 582.
(60) van Esbroeck, A. C. M.; Janssen, A. P. A.; Cognetta, A. B., 3rd; Ogasawara, D.; Shpak, G.; van der Kroeg, M.; Kantae, V.; Baggelaar, M. P.; de Vrij, F. M. S.; Deng, H.; Allara, M.; Fezza, F.; Lin, Z.; van der Wel, T.; Soethoudt, M.; Mock, E. D.; den Dulk, H.; Baak, I. L.; Florea, B. I.; Hendriks, G.; De Petrocellis, L.; Overkleeft, H. S.; Hankemeier, T.; De Zeeuw, C. I.; Di Marzo, V.; Maccarrone, M.; Cravatt, B. F.; Kushner, S. A.; van der Stelt, M. *Science* **2017**, *356*, 1084.
(61) Zhao, Q.; Ouyang, X.; Wan, X.; Gajiwala, K. S.; Kath, J. C.; Jones, L. H.; Burlingame, A. L.; Taunton, J. *J. Am. Chem. Soc.* **2017**, *139*, 680.
(62) Das, J. *Chem. Rev.* **2011**, *111*, 4405.
(63) Preston, G. W.; Wilson, A. J. *Chem. Soc. Rev.* **2013**, *42*, 3289.
(64) Shah, N. P. *Science* **2004**, *305*, 399.
(65) Schirle, M.; Jenkins, J. L. *Drug Discov. Today* **2016**, *21*, 82.
(66) Wright, M. H.; Sieber, S. A. *Nat. Prod. Rep.* **2016**, *33*, 681.
(67) Parker, C. G.; Kuttruff, C. A.; Galmozzi, A.; Jorgensen, L.; Yeh, C. H.; Hermanson, D. J.; Wang, Y.; Artola, M.; McKerrall, S. J.; Josyln, C. M.; Norremark, B.; Dunstl, G.; Felding, J.; Saez, E.; Baran, P. S.; Cravatt, B. F. *ACS Cent. Sci.* **2017**, *3*, 1276.
(68) Wang, J.; Gao, L.; Lee, Y. M.; Kalesh, K. A.; Ong, Y. S.; Lim, J.; Jee, J. E.; Sun, H.; Lee, S. S.; Hua, Z. C.; Lin, Q. *Pharmacol. Ther.* **2016**, *162*, 10.
(69) Bush, J. T.; Walport, L. J.; McGouran, J. F.; Leung, I. K. H.; Berridge, G.; van Berkel, S. S.; Basak, A.; Kessler, B. M.; Schofield, C. J. *Chem. Sci.* **2013**, *4*, 4115.
(70) Li, G.; Liu, Y.; Yu, X.; Li, X. *Bioconjugate Chem.* **2014**, *25*, 1172.
(71) Preston, G. W.; Radford, S. E.; Ashcroft, A. E.; Wilson, A. J. *ACS Chem. Biol.* **2014**, *9*, 761.
(72) Sakurai, K.; Yasui, T.; Mizuno, S. *Asian J. Org. Chem.* **2015**, *4*, 724.
(73) Sakurai, K.; Ozawa, S.; Yamada, R.; Yasui, T.; Mizuno, S. *Chembiochem* **2014**, *15*, 1399.
(74) Kotzyba-Hibert, F.; Kapfer, I.; Goeldner, M. *Angew. Chem. Int. Ed.* **1995**, *34*, 1296.
(75) A. Fleming, S. *Tetrahedron* **1995**, *51*, 12479.

(76) Fillion, D.; Deraet, M.; Holleran, B. J.; Escher, E. *J. Med. Chem.* **2006**, *49*, 2200.
(77) Backus, K. M.; Correia, B. E.; Lum, K. M.; Forli, S.; Horning, B. D.; Gonzalez-Paez, G. E.; Chatterjee, S.; Lanning, B. R.; Teijaro, J. R.; Olson, A. J.; Wolan, D. W.; Cravatt, B. F. *Nature* **2016**, *534*, 570.
(78) Craven, G. B.; Affron, D. P.; Allen, C. E.; Matthies, S.; Greener, J. G.; Morgan, R. M. L.; Tate, E. W.; Armstrong, A.; Mann, D. J. *Angew. Chem. Int. Ed.* **2018**, *57*, 5257.
(79) Johansson, H.; Isabella Tsai, Y. C.; Fantom, K.; Chung, C. W.; Kumper, S.; Martino, L.; Thomas, D. A.; Eberl, H. C.; Muelbaier, M.; House, D.; Rittinger, K. *J. Am. Chem. Soc.* **2019**, *141*, 2703.
(80) Ziegler, S.; Pries, V.; Hedberg, C.; Waldmann, H. *Angew. Chem. Int. Ed.* **2013**, *52*, 2744.
(81) Savitski, M. M.; Reinhard, F. B.; Franken, H.; Werner, T.; Savitski, M. F.; Eberhard, D.; Martinez Molina, D.; Jafari, R.; Dovega, R. B.; Klaeger, S.; Kuster, B.; Nordlund, P.; Bantscheff, M.; Drewes, G. *Science* **2014**, *346*, 55.
(82) West, G. M.; Tang, L.; Fitzgerald, M. C. *Anal. Chem.* **2008**, *80*, 4175.
(83) Lomenick, B.; Hao, R.; Jonai, N.; Chin, R. M.; Aghajan, M.; Warburton, S.; Wang, J.; Wu, R. P.; Gomez, F.; Loo, J. A.; Wohlschlegel, J. A.; Vondriska, T. M.; Pelletier, J.; Herschman, H. R.; Clardy, J.; Clarke, C. F.; Huang, J. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 21984.
(84) Becher, I.; Werner, T.; Doce, C.; Zaal, E. A.; Tögel, I.; Khan, C. A.; Rueger, A.; Muelbaier, M.; Salzer, E.; Berkers, C. R.; Fitzpatrick, P. F.; Bantscheff, M.; Savitski, M. M. *Nat. Chem. Biol.* **2016**, *12*, 908.
(85) West, G. M.; Tucker, C. L.; Xu, T.; Park, S. K.; Han, X.; Yates, J. R., 3rd; Fitzgerald, M. C. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 9078.
(86) Lomenick, B.; Olsen, R. W.; Huang, J. *ACS Chem. Biol.* **2011**, *6*, 34.
(87) Feng, Y.; De Franceschi, G.; Kahraman, A.; Soste, M.; Melnik, A.; Boersema, P. J.; de Laureto, P. P.; Nikolaev, Y.; Oliveira, A. P.; Picotti, P. *Nat. Biotechnol.* **2014**, *32*, 1036.
(88) Singh, A.; Thornton, E. R.; Westheimer, F. H. *J. Biol. Chem.* **1962**, *237*, 3006.
(89) Fleet, G. W. J.; Porter, R. R.; Knowles, J. R. *Nature* **1969**, *224*, 511.
(90) Platz, M. S.; Schuster, G. B. In *Advances in Photochemistry*; Volman, D., Hammond, G., Neckers, D., Eds.; John Wiley and Sons: 1992; Vol. 17, p 69.
(91) Buchmueller, K. L.; Hill, B. T.; Platz, M. S.; Weeks, K. M. *J. Am. Chem. Soc.* **2003**, *125*, 10850.
(92) Li, Y. Z.; Kirby, J. P.; George, M. W.; Poliakoff, M.; Schuster, G. B. *J. Am. Chem. Soc.* **1988**, *110*, 8092.
(93) Leyva, E.; Platz, M. S.; Persy, G.; Wirz, J. *J. Am. Chem. Soc.* **1986**, *108*, 3783.
(94) Chen, T.; Michalak, J.; Platz, M. S. *Photochem. Photobiol.* **1995**, *61*, 600.
(95) Belanger, J. M.; Raviv, Y.; Viard, M.; de la Cruz, M. J.; Nagashima, K.; Blumenthal, R. *Virology* **2011**, *417*, 221.
(96) Keana, J. F. W.; Cai, S. X. *J. Org. Chem.* **1990**, *55*, 3640.
(97) Rizk, M. S.; Shi, X.; Platz, M. S. *Biochemistry* **2006**, *45*, 543.
(98) Schnapp, K. A.; Poe, R.; Leyva, E.; Soundararajan, N.; Platz, M. S. *Bioconjugate Chem.* **1993**, *4*, 172.
(99) Smith, R. A.; Knowles, J. R. *J. Am. Chem. Soc.* **1973**, *95*, 5072.

(100) Horne, J. E.; Walko, M.; Calabrese, A. N.; Levenstein, M. A.; Brockwell, D. J.; Kapur, N.; Wilson, A. J.; Radford, S. E. *Angew. Chem. Int. Ed.* **2018**, *57*, 16688.
(101) Brunner, J.; Senn, H.; Richards, F. M. *J. Biol. Chem.* **1980**, *255*, 3313.
(102) Halloran, M. W.; Lumb, J. P. *Chem. Eur. J.* **2019**, *25*, 4885.
(103) Platz, M.; Admasu, A. S.; Kwiatkowski, S.; Crocker, P. J.; Imai, N.; Watt, D. S. *Bioconjugate Chem.* **1991**, *2*, 337.
(104) Noller, B.; Poisson, L.; Maksimenka, R.; Gobert, O.; Fischer, I.; Mestdagh, J. M. *J. Phys. Chem. A* **2009**, *113*, 3041.
(105) Dubinsky, L.; Krom, B. P.; Meijler, M. M. *Bioorg. Med. Chem.* **2012**, *20*, 554.
(106) Ford, F.; Yuzawa, T.; Platz, M. S.; Matzinger, S.; Fülscher, M. *J. Am. Chem. Soc.* **1998**, *120*, 4430.
(107) Hill, J. R.; Robertson, A. A. B. *J. Med. Chem.* **2018**, *61*, 6945.
(108) Ge, S.-S.; Chen, B.; Wu, Y.-Y.; Long, Q.-S.; Zhao, Y.-L.; Wang, P.-Y.; Yang, S. *RSC Advances* **2018**, *8*, 29428.
(109) Lang, W.; Yuan, C.; Zhu, B.; Pan, S.; Liu, J.; Luo, J.; Nie, S.; Zhu, Q.; Lee, J. S.; Ge, J. *Org. Biomol. Chem.* **2019**, *17*, 3010.
(110) Li, Z.; Wang, D.; Li, L.; Pan, S.; Na, Z.; Tan, C. Y.; Yao, S. Q. *J. Am. Chem. Soc.* **2014**, *136*, 9990.
(111) Jackson, P.; Lapinsky, D. J. *J. Org. Chem.* **2018**, *83*, 11245.
(112) Ziemianowicz, D. S.; Bomgarden, R.; Etienne, C.; Schriemer, D. C. *J. Am. Soc. Mass. Spectrom.* **2017**, *28*, 2011.
(113) Manzi, L.; Barrow, A. S.; Scott, D.; Layfield, R.; Wright, T. G.; Moses, J. E.; Oldham, N. J. *Nat. Commun.* **2016**, *7*, 13288.
(114) Jumper, C. C.; Bomgarden, R.; Rogers, J.; Etienne, C.; Schriemer, D. C. *Anal. Chem.* **2012**, *84*, 4411.
(115) Dorman, G.; Nakamura, H.; Pulsipher, A.; Prestwich, G. D. *Chem. Rev.* **2016**, *116*, 15284.
(116) Marazzi, M.; Mai, S.; Roca-Sanjuan, D.; Delcey, M. G.; Lindh, R.; Gonzalez, L.; Monari, A. *J. Phys. Chem. Lett.* **2016**, *7*, 622.
(117) Demeter, A.; Horvath, K.; Boor, K.; Molnar, L.; Soos, T.; Lendvay, G. *J. Phys. Chem. A* **2013**, *117*, 10196.
(118) Wittelsberger, A.; Thomas, B. E.; Mierke, D. F.; Rosenblatt, M. *FEBS Lett.* **2006**, *580*, 1872.
(119) Deseke, E.; Nakatani, Y.; Ourisson, G. *Eur. J. Org. Chem.* **1998**, *1998*, 243.
(120) Lewandowska-Andralojc, A.; Kazmierczak, F.; Hug, G. L.; Horner, G.; Marciniak, B. *Photochem. Photobiol.* **2013**, *89*, 14.
(121) Sigrist, H.; Mühlemann, M.; Dolder, M. *J. Photoch. Photobio. B* **1990**, *7*, 277.
(122) Battenberg, O. A.; Nodwell, M. B.; Sieber, S. A. *J. Org. Chem.* **2011**, *76*, 6075.
(123) Li, Z.; Qian, L.; Li, L.; Bernhammer, J. C.; Huynh, H. V.; Lee, J. S.; Yao, S. Q. *Angew. Chem. Int. Ed.* **2016**, *55*, 2002.
(124) Zhao, S.; Dai, J.; Hu, M.; Liu, C.; Meng, R.; Liu, X.; Wang, C.; Luo, T. *Chem. Commun.* **2016**, *52*, 4702.
(125) Herner, A.; Marjanovic, J.; Lewandowski, T. M.; Marin, V.; Patterson, M.; Miesbauer, L.; Ready, D.; Williams, J.; Vasudevan, A.; Lin, Q. *J. Am. Chem. Soc.* **2016**, *138*, 14609.
(126) Ota, E.; Usui, K.; Oonuma, K.; Koshino, H.; Nishiyama, S.; Hirai, G.; Sodeoka, M. *ACS Chem. Biol.* **2018**, *13*, 876.

(127) Chowdhry, V.; Westheimer, F. H. *Annu. Rev. Biochem.* **1979**, *48*, 293.
(128) Presolski, S. I.; Hong, V. P.; Finn, M. G. *Curr. Protoc. Chem. Biol.* **2011**, *3*, 153.
(129) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. *Angew. Chem.* **2009**, *121*, 10063.
(130) Chivers, C. E.; Koner, A. L.; Lowe, E. D.; Howarth, M. *Biochem. J.* **2011**, *435*, 55.
(131) Srebalus Barnes, C. A.; Lim, A. *Mass Spectrom. Rev.* **2007**, *26*, 370.
(132) Zhang, G.; Annan, R. S.; Carr, S. A.; Neubert, T. A. *Curr. Protoc. Mol. Biol.* **2014**, *108*, 10.21.1.
(133) Gaskell, S. J. *J. Mass Spectrom.* **1997**, *32*, 677.
(134) Carr, S. A.; Annan, R. S. *Curr. Protoc. Protein. Sci.* **2001**, *16*, 16.1.1.
(135) Ferrige, A. G.; Seddon, M. J.; Green, B. N.; Jarvis, S. A.; Skilling, J.; Staunton, J. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 707.
(136) Ferrige, A. G.; Seddon, M. J.; Jarvis, S.; Skilling, J.; Aplin, R. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 374.
(137) Xu, G.; Stupak, J.; Yang, L.; Hu, L.; Guo, B.; Li, J. *Rapid Commun. Mass Spectrom.* **2018**, *32*, 763.
(138) Yates, J. R. *J. Mass Spectrom.* **1998**, *33*, 1.
(139) Steen, H.; Mann, M. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 699.
(140) Makarov, A. *Anal. Chem.* **2000**, *72*, 1156.
(141) Hardman, M.; Makarov, A. A. *Anal. Chem.* **2003**, *75*, 1699.
(142) Scigelova, M.; Makarov, A. *Proteomics* **2006**, *6*, 16.
(143) Makarov, A.; Denisov, E.; Lange, O.; Horning, S. *J. Am. Soc. Mass. Spectrom.* **2006**, *17*, 977.
(144) Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; Wouters, E. R.; Senko, M.; Makarov, A.; Mann, M.; Horning, S. *Mol. Cell Proteomics* **2009**, *8*, 2759.
(145) Michalski, A.; Damoc, E.; Hauschild, J. P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. *Mol. Cell Proteomics* **2011**, *10*, M111.011015.
(146) Nolting, D.; Malek, R.; Makarov, A. *Mass Spectrom. Rev.* **2019**, *38*, 150.
(147) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass. Spectrom.* **1994**, *5*, 976.
(148) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. *J. Proteome Res.* **2011**, *10*, 1794.
(149) Cox, J.; Matic, I.; Hilger, M.; Nagaraj, N.; Selbach, M.; Olsen, J. V.; Mann, M. *Nat. Protoc.* **2009**, *4*, 698.
(150) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551.
(151) Elias, J. E.; Gygi, S. P. *Nat. Methods* **2007**, *4*, 207.
(152) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994.
(153) Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. *Anal. Chem.* **2003**, *75*, 6843.
(154) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 2836.
(155) McClatchy, D. B.; Liao, L.; Park, S. K.; Xu, T.; Lu, B.; Yates Iii, J. R. *PloS One* **2011**, *6*, e16039.

(156) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. *Mol. Cell Proteomics* **2004**, *3*, 1154.
(157) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol. Cell Proteomics* **2002**, *1*, 376.
(158) Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C. *Anal. Chem.* **2003**, *75*, 1895.
(159) Rauniyar, N.; Yates, J. R., 3rd *J. Proteome Res.* **2014**, *13*, 5293.
(160) Werner, T.; Sweetman, G.; Savitski, M. F.; Mathieson, T.; Bantscheff, M.; Savitski, M. M. *Anal. Chem.* **2014**, *86*, 3594.
(161) Arul, A. B.; Robinson, R. A. S. *Anal. Chem.* **2019**, *91*, 178.
(162) Muranaka, H.; Momose, T.; Handa, C.; Ozawa, T. *ACS Med. Chem. Lett.* **2017**, *8*, 660.
(163) *Angew. Chem.* **1959**, *71*, 373.
(164) Reguera, L.; Mendez, Y.; Humpierre, A. R.; Valdes, O.; Rivera, D. G. *Acc. Chem. Res.* **2018**, *51*, 1475.
(165) Bienaymé, H.; Hulme, C.; Oddon, G.; Schmitt, P. *Chem. Eur. J.* **2000**, *6*, 3321.
(166) Ugi, I.; Werner, B.; Domling, A. *Molecules* **2003**, *8*, 53.
(167) Dömling, A.; Ugi, I. *Angew. Chem.* **2000**, *39*, 3168.
(168) Medeiros, G. A.; da Silva, W. A.; Bataglion, G. A.; Ferreira, D. A.; de Oliveira, H. C.; Eberlin, M. N.; Neto, B. A. *Chem. Commun.* **2014**, *50*, 338.
(169) Cheron, N.; Ramozzi, R.; El Kaim, L.; Grimaud, L.; Fleurat-Lessard, P. *J. Org. Chem.* **2012**, *77*, 1361.
(170) Ge, J.; Cheng, X.; Tan, L. P.; Yao, S. Q. *Chem. Commun.* **2012**, *48*, 4453.
(171) Kambe, T.; Correia, B. E.; Niphakis, M. J.; Cravatt, B. F. *J. Am. Chem. Soc.* **2014**, *136*, 10777.
(172) Rader, A. F. B.; Reichart, F.; Weinmuller, M.; Kessler, H. *Bioorg. Med. Chem.* **2018**, *26*, 2766.
(173) Berger, S. L.; Kouzarides, T.; Shiekhattar, R.; Shilatifard, A. *Genes Dev.* **2009**, *23*, 781.
(174) Schonheit, J.; Leutz, A.; Rosenbauer, F. *J. Mol. Biol.* **2015**, *427*, 670.
(175) Skinner, M. K. *Birth Defects Res C Embryo Today* **2011**, *93*, 51.
(176) Bonasio, R.; Tu, S.; Reinberg, D. *Science* **2010**, *330*, 612.
(177) Bannister, A. J.; Kouzarides, T. *Cell Res.* **2011**, *21*, 381.
(178) Jenuwein, T.; Allis, C. D. *Science* **2001**, *293*, 1074.
(179) Prinjha, R. K.; Witherington, J.; Lee, K. *Trends Pharmacol. Sci.* **2012**, *33*, 146.
(180) Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J. P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Muller, S.; Pawson, T.; Gingras, A. C.; Arrowsmith, C. H.; Knapp, S. *Cell* **2012**, *149*, 214.
(181) Filippakopoulos, P.; Knapp, S. *Nat. Rev. Drug Discov.* **2014**, *13*, 337.
(182) LeRoy, G.; Rickards, B.; Flint, S. J. *Mol. Cell* **2008**, *30*, 51.
(183) Liu, Z.; Wang, P.; Chen, H.; Wold, E. A.; Tian, B.; Brasier, A. R.; Zhou, J. *J. Med. Chem.* **2017**, *60*, 4533.
(184) Matzuk, M. M.; McKeown, M. R.; Filippakopoulos, P.; Li, Q.; Ma, L.; Agno, J. E.; Lemieux, M. E.; Picaud, S.; Yu, R. N.; Qi, J.; Knapp, S.; Bradner, J. E. *Cell* **2012**, *150*, 673.

(185) Belkina, A. C.; Nikolajczyk, B. S.; Denis, G. V. *J. Immunol.* **2013**, *190*, 3670.
(186) Klein, K.; Kabala, P. A.; Grabiec, A. M.; Gay, R. E.; Kolling, C.; Lin, L. L.; Gay, S.; Tak, P. P.; Prinjha, R. K.; Ospelt, C.; Reedquist, K. A. *Ann. Rheum. Dis.* **2016**, *75*, 422.
(187) Zhang, Q. G.; Qian, J.; Zhu, Y. C. *Immunol. Lett.* **2015**, *166*, 103.
(188) Park-Min, K. H.; Lim, E.; Lee, M. J.; Park, S. H.; Giannopoulou, E.; Yarilina, A.; van der Meulen, M.; Zhao, B.; Smithers, N.; Witherington, J.; Lee, K.; Tak, P. P.; Prinjha, R. K.; Ivashkiv, L. B. *Nat. Commun.* **2014**, *5*, 5418.
(189) Chaidos, A.; Caputo, V.; Gouvedenou, K.; Liu, B.; Marigo, I.; Chaudhry, M. S.; Rotolo, A.; Tough, D. F.; Smithers, N. N.; Bassil, A. K.; Chapman, T. D.; Harker, N. R.; Barbash, O.; Tummino, P.; Al-Mahdi, N.; Haynes, A. C.; Cutler, L.; Le, B.; Rahemtulla, A.; Roberts, I.; Kleijnen, M.; Witherington, J. J.; Parr, N. J.; Prinjha, R. K.; Karadimitris, A. *Blood* **2014**, *123*, 697.
(190) Cheng, Z.; Gong, Y.; Ma, Y.; Lu, K.; Lu, X.; Pierce, L. A.; Thompson, R. C.; Muller, S.; Knapp, S.; Wang, J. *Clin. Cancer Res.* **2013**, *19*, 1748.
(191) Zuber, J.; Shi, J.; Wang, E.; Rappaport, A. R.; Herrmann, H.; Sison, E. A.; Magoon, D.; Qi, J.; Blatt, K.; Wunderlich, M.; Taylor, M. J.; Johns, C.; Chicas, A.; Mulloy, J. C.; Kogan, S. C.; Brown, P.; Valent, P.; Bradner, J. E.; Lowe, S. W.; Vakoc, C. R. *Nature* **2011**, *478*, 524.
(192) da Motta, L. L.; Ledaki, I.; Purshouse, K.; Haider, S.; De Bastiani, M. A.; Baban, D.; Morotti, M.; Steers, G.; Wigfield, S.; Bridges, E.; Li, J. L.; Knapp, S.; Ebner, D.; Klamt, F.; Harris, A. L.; McIntyre, A. *Oncogene* **2017**, *36*, 122.
(193) Hay, D. A.; Fedorov, O.; Martin, S.; Singleton, D. C.; Tallant, C.; Wells, C.; Picaud, S.; Philpott, M.; Monteiro, O. P.; Rogers, C. M.; Conway, S. J.; Rooney, T. P.; Tumber, A.; Yapp, C.; Filippakopoulos, P.; Bunnage, M. E.; Muller, S.; Knapp, S.; Schofield, C. J.; Brennan, P. E. *J. Am. Chem. Soc.* **2014**, *136*, 9308.
(194) Aldeghi, M.; Ross, G. A.; Bodkin, M. J.; Essex, J. W.; Knapp, S.; Biggin, P. C. *Commun. Chem.* **2018**, *1*, 19.
(195) Park, J.; Koh, M.; Koo, J. Y.; Lee, S.; Park, S. B. *ACS Chem. Biol.* **2016**, *11*, 44.
(196) Engelhardt, H.; Martin, L.; Smethurst, C.; GmbH, B. I. I., Ed.; Boehringer Ingelheim International GmbH: United States, 2016; Vol. US 9296748 B2.
(197) Bit, R. A.; (GSK), G., Ed. United States, 2018; Vol. US 20180044317A1.
(198) Deshmane, S. L.; Kremlev, S.; Amini, S.; Sawaya, B. E. *J. Interferon Cytokine Res.* **2009**, *29*, 313.
(199) Nguyen, T. H.; Maltby, S.; Eyers, F.; Foster, P. S.; Yang, M. *PloS One* **2016**, *11*, e0163392.
(200) Lamos, S. M.; Krusemark, C. J.; McGee, C. J.; Scalf, M.; Smith, L. M.; Belshaw, P. J. *Angew. Chem. Int. Ed.* **2006**, *45*, 4329.
(201) Wolfram, S.; Wurfel, H.; Habenicht, S. H.; Lembke, C.; Richter, P.; Birckner, E.; Beckert, R.; Pohnert, G. *Beilstein J. Org. Chem.* **2014**, *10*, 2470.
(202) Yang, L.; Chumsae, C.; Kaplan, J. B.; Moulton, K. R.; Wang, D.; Lee, D. H.; Zhou, Z. S. *Bioconjugate Chem.* **2017**, *28*, 2302.
(203) Song, Z.; Huang, W.; Zhang, Q. *Chem. Commun.* **2012**, *48*, 3339.
(204) Wacker, S. A.; Kashyap, S.; Li, X.; Kapoor, T. M. *J. Am. Chem. Soc.* **2011**, *133*, 12386.
(205) Croston, G. E. *Expert Opin. Drug Discov.* **2017**, *12*, 427.

(206) Eder, J.; Sedrani, R.; Wiesmann, C. *Nat. Rev. Drug. Discov.* **2014**, *13*, 577.
(207) Renaud, J. P.; Chung, C. W.; Danielson, U. H.; Egner, U.; Hennig, M.; Hubbard, R. E.; Nar, H. *Nat. Rev. Drug Discov.* **2016**, *15*, 679.
(208) Rosano, G. L.; Ceccarelli, E. A. *Front. Microbiol.* **2014**, *5*, 172.
(209) Genick, C. C.; Wright, S. K. *Expert Opin. Drug Discov.* **2017**, *12*, 897.
(210) Mayr, L. M.; Bojanic, D. *Curr. Opin. Pharmacol.* **2009**, *9*, 580.
(211) Wildey, M. J.; Haunso, A.; Tudor, M.; Webb, M.; Connick, J. H. In *Annu. Rep. Med. Chem.*; Goodnow, R. A., Ed.; Academic Press: 2017; Vol. 50, p 149.
(212) Inglese, J.; Johnson, R. L.; Simeonov, A.; Xia, M.; Zheng, W.; Austin, C. P.; Auld, D. S. *Nat. Chem. Biol.* **2007**, *3*, 466.
(213) Fang, X.; Zheng, Y.; Duan, Y.; Liu, Y.; Zhong, W. *Anal. Chem.* **2019**, *91*, 482.
(214) Eglen, R. M.; Reisine, T.; Roby, P.; Rouleau, N.; Illy, C.; Bosse, R.; Bielefeld, M. *Curr. Chem. Genomics.* **2008**, *1*, 2.
(215) Yasgar, A.; Jadhav, A.; Simeonov, A.; Coussens, N. P. In *High Throughput Screening: Methods and Protocols*; Janzen, W. P., Ed.; Springer New York: New York, NY, 2016, p 77.
(216) Degorce, F.; Card, A.; Soh, S.; Trinquet, E.; Knapik, G. P.; Xie, B. *Curr. Chem. Genomics.* **2009**, *3*, 22.
(217) Ergin, E.; Dogan, A.; Parmaksiz, M.; Elcin, A. E.; Elcin, Y. M. *Curr. Pharm. Biotechnol.* **2016**, *17*, 1222.
(218) Lea, W. A.; Simeonov, A. *Expert Opin. Drug Discov.* **2011**, *6*, 17.
(219) Hall, M. D.; Yasgar, A.; Peryea, T.; Braisted, J. C.; Jadhav, A.; Simeonov, A.; Coussens, N. P. *Methods Appl. Fluoresc.* **2016**, *4*, 022001.
(220) Förster, T. *Annalen der Physik* **1948**, *437*, 55.
(221) Zhang, H.; Wu, Q.; Berezin, M. Y. *Expert Opin. Drug Discov.* **2015**, *10*, 1145.
(222) Brogan, K. L.; Shin, J. H.; Schoenfisch, M. H. *Langmuir* **2004**, *20*, 9729.
(223) Loo, R. R.; Dales, N.; Andrews, P. C. *Protein Sci.* **1994**, *3*, 1975.
(224) Annesley, T. M. *Clin. Chem.* **2003**, *49*, 1041.
(225) Mo, J.; Yan, Q.; So, C. K.; Soden, T.; Lewis, M. J.; Hu, P. *Anal. Chem.* **2016**, *88*, 9495.
(226) Zeder-Lutz, G.; Benito, A.; Van Regenmortel, M. H. V. *J. Mol. Recognit.* **1999**, *12*, 300.
(227) Gosmini, R.; Nguyen, V. L.; Toum, J.; Simon, C.; Brusq, J. M.; Krysa, G.; Mirguet, O.; Riou-Eymard, A. M.; Boursier, E. V.; Trottet, L.; Bamborough, P.; Clark, H.; Chung, C. W.; Cutler, L.; Demont, E. H.; Kaur, R.; Lewis, A. J.; Schilling, M. B.; Soden, P. E.; Taylor, S.; Walker, A. L.; Walker, M. D.; Prinjha, R. K.; Nicodeme, E. *J. Med. Chem.* **2014**, *57*, 8111.
(228) Mirguet, O.; Lamotte, Y.; Donche, F.; Toum, J.; Gellibert, F.; Bouillot, A.; Gosmini, R.; Nguyen, V. L.; Delannee, D.; Seal, J.; Blandel, F.; Boullay, A. B.; Boursier, E.; Martin, S.; Brusq, J. M.; Krysa, G.; Riou, A.; Tellier, R.; Costaz, A.; Huet, P.; Dudit, Y.; Trottet, L.; Kirilovsky, J.; Nicodeme, E. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 2963.
(229) McDaniel, K. F.; Wang, L.; Soltwedel, T.; Fidanze, S. D.; Hasvold, L. A.; Liu, D.; Mantei, R. A.; Pratt, J. K.; Sheppard, G. S.; Bui, M. H.; Faivre, E. J.; Huang, X.; Li, L.; Lin, X.; Wang, R.; Warder, S. E.; Wilcox, D.; Albert, D. H.; Magoc, T. J.; Rajaraman, G.; Park, C. H.; Hutchins, C. W.; Shen, J. J.; Edalji, R. P.; Sun, C. C.; Martin, R.; Gao, W.; Wong, S.; Fang, G.; Elmore, S. W.; Shen, Y.; Kati, W. M. *J. Med. Chem.* **2017**, *60*, 8369.

(230) Atkinson, S. J.; Soden, P. E.; Angell, D. C.; Bantscheff, M.; Chung, C.-w.; Giblin, K. A.; Smithers, N.; Furze, R. C.; Gordon, L.; Drewes, G.; Rioja, I.; Witherington, J.; Parr, N. J.; Prinjha, R. K. *Med. Chem. Commun.* **2014**, *5*, 342.
(231) Ember, S. W.; Zhu, J. Y.; Olesen, S. H.; Martin, M. P.; Becker, A.; Berndt, N.; Georg, G. I.; Schonbrunn, E. *ACS Chem. Biol.* **2014**, *9*, 1160.
(232) Bamborough, P.; Diallo, H.; Goodacre, J. D.; Gordon, L.; Lewis, A.; Seal, J. T.; Wilson, D. M.; Woodrow, M. D.; Chung, C. W. *J. Med. Chem.* **2012**, *55*, 587.
(233) Bamborough, P.; Barnett, H. A.; Becher, I.; Bird, M. J.; Chung, C. W.; Craggs, P. D.; Demont, E. H.; Diallo, H.; Fallon, D. J.; Gordon, L. J.; Grandi, P.; Hobbs, C. I.; Hooper-Greenhill, E.; Jones, E. J.; Law, R. P.; Le Gall, A.; Lugo, D.; Michon, A. M.; Mitchell, D. J.; Prinjha, R. K.; Sheppard, R. J.; Watson, A. J.; Watson, R. J. *ACS Med. Chem. Lett.* **2016**, *7*, 552.
(234) Demont, E. H.; Bamborough, P.; Chung, C. W.; Craggs, P. D.; Fallon, D.; Gordon, L. J.; Grandi, P.; Hobbs, C. I.; Hussain, J.; Jones, E. J.; Le Gall, A.; Michon, A. M.; Mitchell, D. J.; Prinjha, R. K.; Roberts, A. D.; Sheppard, R. J.; Watson, R. J. *ACS Med. Chem. Lett.* **2014**, *5*, 1190.
(235) Munson, P. J.; Rodbard, D. *J. Recept. Res.* **2008**, *8*, 533.
(236) Mirguet, O.; Gosmini, R.; Toum, J.; Clement, C. A.; Barnathan, M.; Brusq, J. M.; Mordaunt, J. E.; Grimes, R. M.; Crowe, M.; Pineau, O.; Ajakane, M.; Daugan, A.; Jeffrey, P.; Cutler, L.; Haynes, A. C.; Smithers, N. N.; Chung, C. W.; Bamborough, P.; Uings, I. J.; Lewis, A.; Witherington, J.; Parr, N.; Prinjha, R. K.; Nicodeme, E. *J. Med. Chem.* **2013**, *56*, 7501.
(237) Picaud, S.; Leonards, K.; Lambert, J. P.; Dovey, O.; Wells, C.; Fedorov, O.; Monteiro, O.; Fujisawa, T.; Wang, C. Y.; Lingard, H.; Tallant, C.; Nikbin, N.; Guetzoyan, L.; Ingham, R.; Ley, S. V.; Brennan, P.; Muller, S.; Samsonova, A.; Gingras, A. C.; Schwaller, J.; Vassiliou, G.; Knapp, S.; Filippakopoulos, P. *Sci. Adv.* **2016**, *2*, e1600760.
(238) Munoz, L. *Nat. Rev. Drug Discov.* **2017**, *16*, 424.
(239) Xue, X.; Zhang, Y.; Liu, Z.; Song, M.; Xing, Y.; Xiang, Q.; Wang, Z.; Tu, Z.; Zhou, Y.; Ding, K.; Xu, Y. *J. Med. Chem.* **2016**, *59*, 1565.
(240) Leveridge, M.; Buxton, R.; Argyrou, A.; Francis, P.; Leavens, B.; West, A.; Rees, M.; Hardwicke, P.; Bridges, A.; Ratcliffe, S.; Chung, C. W. *J. Biomol. Screen.* **2014**, *19*, 278.
(241) De Cesare, V.; Johnson, C.; Barlow, V.; Hastie, J.; Knebel, A.; Trost, M. *Cell Chem. Biol.* **2018**, *25*, 1117.
(242) Weigt, D.; Hopf, C.; Medard, G. *Clin. Epigenetics* **2016**, *8*, 76.
(243) Dalton, S. E.; Dittus, L.; Thomas, D. A.; Convery, M. A.; Nunes, J.; Bush, J. T.; Evans, J. P.; Werner, T.; Bantscheff, M.; Murphy, J. A.; Campos, S. *J. Am. Chem. Soc.* **2018**, *140*, 932.
(244) Colombo, M. F.; Austrilino, L.; Nascimento, O. R.; Castellano, E. E.; Tabak, M. *Can. J. Chem.* **1987**, *65*, 821.
(245) Huyer, G.; Liu, S.; Kelly, J.; Moffat, J.; Payette, P.; Kennedy, B.; Tsaprailis, G.; Gresser, M. J.; Ramachandran, C. *J. Biol. Chem.* **1997**, *272*, 843.
(246) Filippakopoulos, P.; Qi, J.; Picaud, S.; Shen, Y.; Smith, W. B.; Fedorov, O.; Morse, E. M.; Keates, T.; Hickman, T. T.; Felletar, I.; Philpott, M.; Munro, S.; McKeown, M. R.; Wang, Y.; Christie, A. L.; West, N.; Cameron, M. J.;

Schwartz, B.; Heightman, T. D.; La Thangue, N.; French, C. A.; Wiest, O.; Kung, A. L.; Knapp, S.; Bradner, J. E. *Nature* **2010**, *468*, 1067.
(247) Burgermeister, W.; Hekman, M.; Helmreich, E. J. *J. Biol. Chem.* **1982**, *257*, 5306.
(248) Li, G.; Liu, Y.; Liu, Y.; Chen, L.; Wu, S.; Liu, Y.; Li, X. *Angew. Chem. Int. Ed.* **2013**, *52*, 9544.
(249) Worrell, B. T.; Malik, J. A.; Fokin, V. V. *Science* **2013**, *340*, 457.
(250) Murale, D. P.; Hong, S. C.; Haque, M. M.; Lee, J. S. *Proteome Sci.* **2016**, *15*, 14.
(251) Belotserkovskaya, R.; Oh, S.; Bondarenko, V. A.; Orphanides, G.; Studitsky, V. M.; Reinberg, D. *Science* **2003**, *301*, 1090.
(252) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. *Nat. Protoc.* **2008**, *4*, 44.
(253) Huang da, W.; Sherman, B. T.; Lempicki, R. A. *Nucleic Acids Res.* **2009**, *37*, 1.
(254) Lin, L.; Yee, S. W.; Kim, R. B.; Giacomini, K. M. *Nat. Rev. Drug. Discov.* **2015**, *14*, 543.
(255) Tonazzi, A.; Giangregorio, N.; Indiveri, C.; Palmieri, F. *J. Biol. Chem.* **2005**, *280*, 19607.
(256) Iacobazzi, V.; Convertini, P.; Infantino, V.; Scarcia, P.; Todisco, S.; Palmieri, F. *Biochem. Biophys. Res. Commun.* **2009**, *388*, 643.
(257) Palmieri, F. *Biochim. Biophys. Acta* **2008**, *1777*, 564.
(258) Lebwohl, M.; Swanson, N.; Anderson, L. L.; Melgaard, A.; Xu, Z.; Berman, B. *N. Engl. J. Med.* **2012**, *366*, 1010.
(259) Patricelli, M. P.; Nomanbhoy, T. K.; Wu, J.; Brown, H.; Zhou, D.; Zhang, J.; Jagannathan, S.; Aban, A.; Okerberg, E.; Herring, C.; Nordin, B.; Weissig, H.; Yang, Q.; Lee, J. D.; Gray, N. S.; Kozarich, J. W. *Chem. Biol.* **2011**, *18*, 699.
(260) Besanceney-Webler, C.; Jiang, H.; Zheng, T.; Feng, L.; Soriano del Amo, D.; Wang, W.; Klivansky, L. M.; Marlow, F. L.; Liu, Y.; Wu, P. *Angew. Chem. Int. Ed.* **2011**, *50*, 8051.
(261) Mortenson, P. N.; Erlanson, D. A.; de Esch, I. J. P.; Jahnke, W.; Johnson, C. N. *J. Med. Chem.* **2019**, *62*, 3857.
(262) Johnson, C. N.; Erlanson, D. A.; Jahnke, W.; Mortenson, P. N.; Rees, D. C. *J. Med. Chem.* **2018**, *61*, 1774.
(263) Johnson, C. N.; Erlanson, D. A.; Murray, C. W.; Rees, D. C. *J. Med. Chem.* **2017**, *60*, 89.
(264) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Advanced Drug Delivery Reviews* **2001**, *46*, 3.
(265) Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *J. Med. Chem.* **2002**, *45*, 2615.
(266) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. *Drug Discov. Today* **2003**, *8*, 876.
(267) Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. *Nat. Rev. Drug Discov.* **2004**, *3*, 660.
(268) Murray, C. W.; Rees, D. C. *Nat. Chem.* **2009**, *1*, 187.
(269) Keseru, G. M.; Erlanson, D. A.; Ferenczy, G. G.; Hann, M. M.; Murray, C. W.; Pickett, S. D. *J. Med. Chem.* **2016**, *59*, 8189.
(270) Hall, R. J.; Mortenson, P. N.; Murray, C. W. *Prog. Biophys. Mol. Biol.* **2014**, *116*, 82.

(271) Boyd, S. M.; Turnbull, A. P.; Walse, B. *WIREs Comput. Mol. Sci.* **2012**, *2*, 868.
(272) Schuffenhauer, A.; Ruedisser, S.; Marzinzik, A. L.; Jahnke, W.; Blommers, M.; Selzer, P.; Jacoby, E. *Curr. Top. Med. Chem.* **2005**, *5*, 751.
(273) Lamoree, B.; Hubbard, R. E. *Essays Biochem.* **2017**, *61*, 453.
(274) Jacquemard, C.; Kellenberger, E. *Expert Opin. Drug Discov.* **2019**, *14*, 413.
(275) Gossert, A. D.; Jahnke, W. *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *97*, 82.
(276) Giannetti, A. M. In *Methods Enzymol.*; Kuo, L. C., Ed.; Academic Press: 2011; Vol. 493, p 169.
(277) Davies, D. R.; Begley, D. W.; Hartley, R. C.; Staker, B. L.; Stewart, L. J. In *Methods Enzymol.*; Kuo, L. C., Ed.; Academic Press: 2011; Vol. 493, p 91.
(278) Hubbard, R. E.; Murray, J. B. In *Methods Enzymol.*; Kuo, L. C., Ed.; Academic Press: 2011; Vol. 493, p 509.
(279) Miller, R. M.; Paavilainen, V. O.; Krishnan, S.; Serafimova, I. M.; Taunton, J. *J. Am. Chem. Soc.* **2013**, *135*, 5298.
(280) Bamborough, P.; Chung, C.-w. *Med. Chem. Commun.* **2015**, *6*, 1587.
(281) Chung, C. W.; Dean, A. W.; Woolven, J. M.; Bamborough, P. *J. Med. Chem.* **2012**, *55*, 576.
(282) Ahad, A. M.; Jensen, S. M.; Jewett, J. C. *Org. Lett.* **2013**, *15*, 5060.
(283) Chung, C. W.; Coste, H.; White, J. H.; Mirguet, O.; Wilde, J.; Gosmini, R. L.; Delves, C.; Magny, S. M.; Woodward, R.; Hughes, S. A.; Boursier, E. V.; Flynn, H.; Bouillot, A. M.; Bamborough, P.; Brusq, J. M.; Gellibert, F. J.; Jones, E. J.; Riou, A. M.; Homes, P.; Martin, S. L.; Uings, I. J.; Toum, J.; Clement, C. A.; Boullay, A. B.; Grimley, R. L.; Blandel, F. M.; Prinjha, R. K.; Lee, K.; Kirilovsky, J.; Nicodeme, E. *J. Med. Chem.* **2011**, *54*, 3827.
(284) Law, R. P.; Atkinson, S. J.; Bamborough, P.; Chung, C. W.; Demont, E. H.; Gordon, L. J.; Lindon, M.; Prinjha, R. K.; Watson, A. J. B.; Hirst, D. J. *J. Med. Chem.* **2018**, *61*, 4317.
(285) Camurri, G.; Zaramella, A. *Anal. Chem.* **2001**, *73*, 3716.
(286) Valko, K.; Nunhuck, S.; Bevan, C.; Abraham, M. H.; Reynolds, D. P. *J. Pharm. Sci.* **2003**, *92*, 2236.
(287) Young, R. J.; Green, D. V. S.; Luscombe, C. N.; Hill, A. P. *Drug Discov. Today* **2011**, *16*, 822.
(288) Robinson, M. W.; Hill, A. P.; Readshaw, S. A.; Hollerton, J. C.; Upton, R. J.; Lynn, S. M.; Besley, S. C.; Boughtflower, B. J. *Anal. Chem.* **2017**, *89*, 1772.
(289) Kansy, M.; Senner, F.; Gubernator, K. *J. Med. Chem.* **1998**, *41*, 1007.
(290) Otwinowski, Z.; Minor, W. In *Methods Enzymol.*; Academic Press: 1997; Vol. 276, p 307.
(291) Evans, P. R.; Murshudov, G. N. *Acta Crystallogr D Biol Crystallogr* **2013**, *69*, 1204.
(292) Emsley, P.; Cowtan, K. *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 2126.
(293) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr D Biol Crystallogr* **1997**, *53*, 240.
(294) Collaborative Computational Project, N. *Acta Crystallogr D Biol Crystallogr* **1994**, *50*, 760.
(295) Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.; Bricogne, G. *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 293.
(296) Kabsch, W. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125.

(297) Franken, H.; Mathieson, T.; Childs, D.; Sweetman, G. M. A.; Werner, T.; Tögel, I.; Doce, C.; Gade, S.; Bantscheff, M.; Drewes, G.; Reinhard, F. B. M.; Huber, W.; Savitski, M. M. *Nat. Protoc.* **2015**, *10*, 1567.
(298) Savitski, M. M.; Sweetman, G.; Askenazi, M.; Marto, J. A.; Lang, M.; Zinn, N.; Bantscheff, M. *Anal. Chem.* **2011**, *83*, 8959.
(299) Savitski, M. M.; Mathieson, T.; Zinn, N.; Sweetman, G.; Doce, C.; Becher, I.; Pachl, F.; Kuster, B.; Bantscheff, M. *J. Proteome Res.* **2013**, *12*, 3586.
(300) Silva, J. C.; Gorenstein, M. V.; Li, G. Z.; Vissers, J. P.; Geromanos, S. J. *Mol. Cell Proteomics* **2006**, *5*, 144.