One-Step Synthesis of Photoaffinity Probes for Live-Cell MS-Based Proteomics

David J. Fallon,[a,b] Stephanie Lehmann,[c] Chun-wa Chung,[a] Alex Phillipou,[a] H. Christian Eberl,[c] Ken G. M. Fantom,[a] Francesca Zappacosta,[d] Vipulkumar K. Patel,[a] Marcus Bantscheff,[c] Christopher J. Schofield,[e] Nicholas C. O. Tomkinson,[b] and Jacob T. Bush.[a]*

[a] Dr D. J. Fallon, Dr C.-w. Chung, K. Fantom, A. Phillipou, Dr V. K. Patel, Dr J. T. Bush
GlaxoSmithKline R&D, Gunnels Wood Road
Stevenage, SG1 2NY (UK)
E-mail: Jacob.x.Bush@gmail.com

[b] Dr D. J. Fallon , Prof. N. C. O. Tomkinson
Department of Pure and Applied Chemistry, Thomas Graham Building
University of Strathclyde, Glasgow, G1 1XL (UK)

[c] S. Lehmann, Dr H. C. Eberl, Dr M. Bantscheff
Cellzome GmbH, a GSK company
Meyerhofstraße 1, Heidelberg, 69117 (DE)

[d] Dr F. Zappacosta
GlaxoSmithKline R&D, South Collegeville
PA 19426 (USA)

[e] Prof. C. J. Schofield
Chemistry Research Laboratory, University of Oxford
12 Mansfield Rd, OX1 3TA (UK)

Abstract: We present a one-step Ugi reaction protocol for the expedient synthesis of photoaffinity probes for live-cell MS-based proteomics. The reaction couples an amine affinity function with commonly used photoreactive groups, and a variety of handle functionalities. Using this technology, a series of pan-BET selective bromodomain photoaffinity probes were obtained by parallel synthesis. Studies on the effects of photoreactive group, linker length and irradiation wavelength on photocrosslinking efficiency provide valuable insights into photoaffinity probe design. 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.

Introduction

One of the biggest challenges for the pharmaceutical industry is to lower attrition rates for candidate drug molecules in clinical trials. Safety concerns that arise during late stage clinical investigation are a key cause of attrition, and are often attributed to an insufficient knowledge of off-target activity.[1] Comprehensive characterisation of protein-ligand interactions across the proteome at the pre-clinical stage allows for informed decision making during lead optimisation cycles, and can contribute to reduced attrition rates.[2] Photoaffinity labelling (PAL) coupled with MS-based proteomics has been used to identify the of on- and off-targets of small molecules.[3] PAL probes are typically tri-functional molecules, consisting of an affinity function (small molecule of interest), a photoreactive group and a bio-orthogonal handle or reporter tag. The PAL probe is introduced to a biological system and the samples are irradiated with UV light. This irradiation excites the photoreactive group (most commonly an aryl azide, diazirine or benzophenone) to form a reactive species (a carbene, nitrene or di-radical respectively) which can insert into C–H, N–H or O–H bonds on proximal amino acid residues of the target protein.[4]

When compared to affinity-based or electrophilic enrichment, PAL has specific advantages. Weak/transient interactions can be covalently captured and identified, and the event is relatively unbiased for a reactive partner.[5] Therefore this method has the potential to provide a more accurate representation of non-covalently engaged on- and off-targets for small molecules of interest. To report on these interactions effectively, the PAL probe should; i) retain structural similarity to the parent compound to mimic binding, ii) have appropriate physicochemical properties, and iii) be able to efficiently capture the target protein in yields above the limit of detection.[6] Designing efficient PAL probes that satisfy these three criteria is extremely challenging.[7] 'Minimalist' PAL probes which employ aliphatic diazirines have become popular, as these offer the smallest structural change to the parent compound of interest.[8] However, comparative studies have shown that other photoreactive groups can offer superior photocrosslinking yields.[7a, 9]

Performing PAL studies with a range of probes containing various photoreactive groups should provide a more informed representation of the true target engagement profile of the parent compound. We envisaged that using a multicomponent reaction approach to the synthesis of PAL probes would allow for a range of probes containing a variety of photoreactive groups, handles and affinity functions to be quickly accessed by parallel synthesis. Multi-component reaction approaches for the synthesis of PAL probes have been explored, however a single-step protocol to bring together the three functional components with high interchangeability has not been demonstrated.[9-10] Furthermore, the effect of varying the photoreactive group and the linker length to the affinity function on photocrosslinking efficiency remains poorly understood.

Within this paper we report an Ugi reaction protocol to access a variety of fully functionalised PAL probes in one step (Figure 1).[11]

This is a peer-reviewed, accepted author manuscript of the following article: Fallon, D. J., Lehmann, S., Chung, C-W., Phillipou, A., Eberl, H. C., Fantom, K. G. M., Zappacosta, F., Patel, V. K., Bantscheff, M., Schofield, C. J., Tomkinson, N. C. O., & Bush, J. (2021). One-step synthesis of photoaffinity probes for live-cell MS-based proteomics. Chemistry - A European Journal. https://doi.org/10.1002/chem.202102036

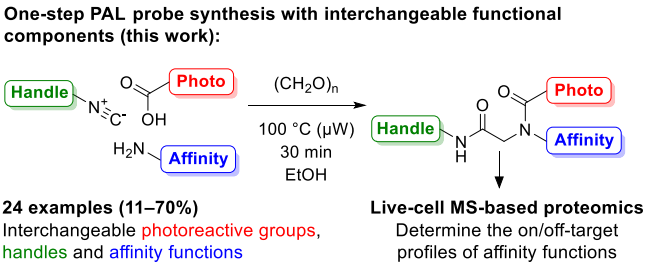


Figure 1. The Ugi reaction can be used to combine the three components of a photoaffinity probe in one green and efficient synthetic step. Each component is interchangeable, allowing access to a diverse library of PAL probes.

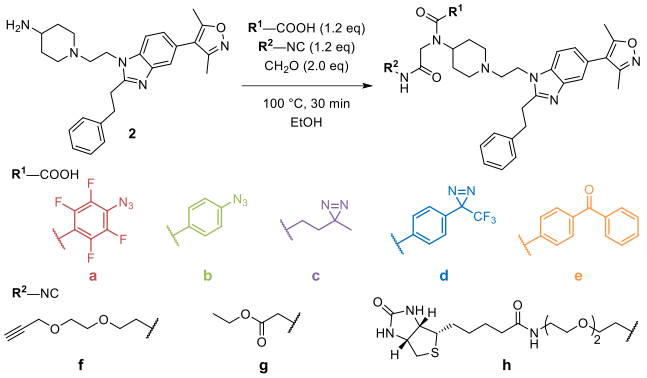
The protocol combines a carboxylic acid photoreactive group, an isonitrile handle, and an amine affinity function with paraformaldehyde. This arrangement allows for each component to be interchanged easily, providing an expedient method for the preparation and optimisation of PAL probes. For proof-of-concept, PAL probes were synthesised from recently reported pan-bromodomain and extra-terminal domain (BET) inhibitors.[12] Subsequent studies on crosslinking yields, physicochemical properties, biochemical and cellular potencies, crystallography and live-cell MS-based proteomics were performed to inform on optimal probe design.

Results and Discussion

Initially, a recently described BET inhibitor **1** with a known binding mode was selected as the affinity function (Figure 2A).[13] Derivative **2** containing a primary amine in a solvent exposed vector (determined by X-ray crystallography, see Figure 3A) was synthesised (see Supplementary Information Scheme S1 for synthesis).[13-14] Amine **2**, isonitrile-alkyne handle **f** and paraformaldehyde were combined with a series of commercially available carboxylic acids containing photoreactive groups (**a–e**), then reacted under microwave irradiation (100 °C, 30 min, Table 1). Diazirines (**c** and **d**) and benzophenone (**e**) afforded the products (**2cf**, **2df** and **2ef**) in moderate to good yields. Good conversions were achieved with aryl azides (**a** and **b**), however, the products were found to be unstable upon purification, potentially due to Huisgen cycloaddition of the azide and alkyne groups upon concentration (Supplementary Information Figure S1–S4). Introduction of the ethyl ester isonitrile **g**, as an alternative handle, afforded the PAL probes (**2ag** and **2bg**) in good yields. These Ugi conditions were also compatible with more complex handles. For example, a reaction incorporating the isonitrile-biotin handle **h** (Supplementary Information Scheme S3 for synthetic route) gave the biotinylated PAL probe **2ch**. Satisfied with the utility of the Ugi protocol for the synthesis of PAL probes, we examined the effect of the Ugi scaffold on target affinity.

The affinities of the probes for the N-terminal and C-terminal recombinant domains of BRD4 (BD1 and BD2 respectively) were measured in a TR-FRET assay (Table 1). For BD1, all probes were found to have similar or improved potency over the parent amine **2** (pIC_{50} +0.1 to +0.6). The effect on BD2 was more diverse, with probes having lower as well as higher affinity over **2** (pIC_{50} -0.8 to +0.8), resulting in some BD1 biased probes (e.g. **2ef**, >63-fold selective).

Table 1 Synthetic yields and biochemical potencies (TR-FRET) for BRD4 BD1 and BD2 of PAL probes synthesised using the Ugi multicomponent reaction.



R^1	R^2	Compound	(%)[a]	pIC_{50} BD1[b]	pIC_{50} BD2[c]
–	–	**2**	–	5.7	5.0
a	**g**	**2ag**	(67)	5.9	<4.3
b	**g**	**2bg**	(66)	6.3	5.7
c	**f**	**2cf**	(54)	5.9	5.4
d	**f**	**2df**	(49)	6.0	5.9
e	**f**	**2ef**	(64)	6.1	<4.3
c	**h**	**2ch**	(28)	5.8	5.3

[a] Isolated synthetic yield. [b,c] TR-FRET assay with recombinant BRD4 BD1/BD2.

To compare the photocrosslinking efficiencies of the five different photoreactive groups, probes **2ag**, **2bg**, **2cf**, **2df** and **1ef** (20 μM) were irradiated (302 nm, 3.8 mW cm^{-2}) in the presence of recombinant BRD4-BD1 (3 μM) over a range of timepoints and analysed by intact protein LCMS (Figure 2C). Figure 2B shows the intact protein LCMS spectra (e.g. **2bg**) used to construct these photolabelling timecourses. All five probes afforded appreciable photocrosslinking yields (>15%), with aryl azide (**2bg**), alkyl diazirine (**2cf**) and benzophenone (**2ef**) reaching *ca.* 40% labelling. The rate of activation was fastest for the aryl azide probes (*ca.* 2 min to complete activation), while the diazirine and benzophenone probes required longer irradiation times (5–10 min). The timecourse was repeated with 365 nm irradiation (7.6 mW cm^{-2}), another commonly used wavelength in PAL studies.[6, 8a, 15] The rate of photoactivation at this wavelength was considerably slower, with <10% photolabelling after 20 min for all five probes, despite the 365 nm lamp delivering twice the irradiance of the 302 nm lamp (see Supplementary Information Figure S5 for further discussion).

It was postulated that by variation of the vector and the linker length from the selectivity function to the photoreactive group, an optimal positioning could be identified to achieve higher levels of labelling

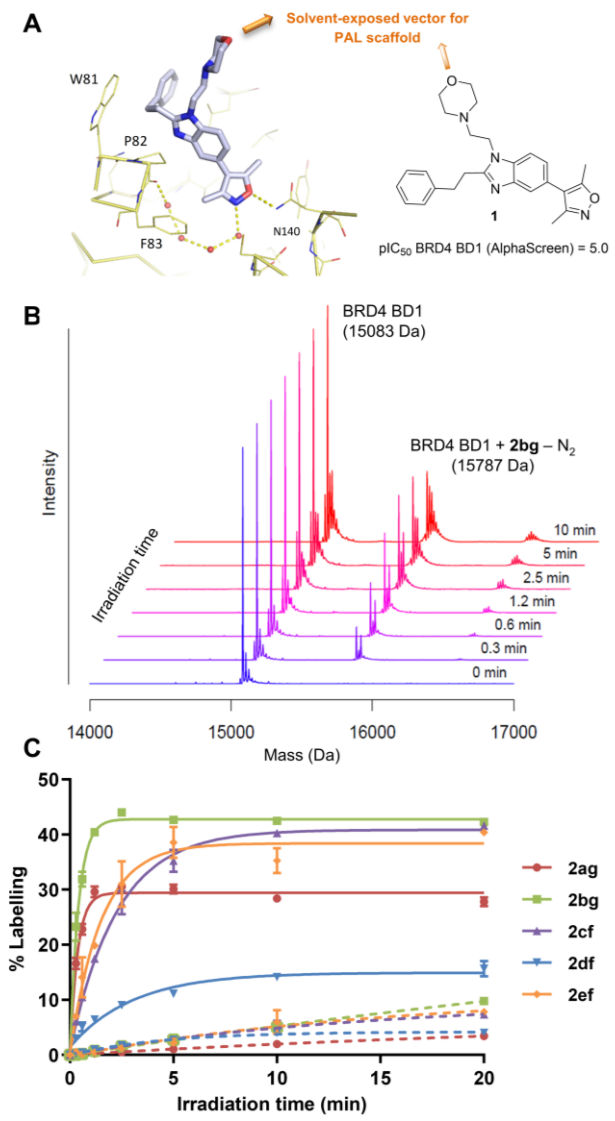


Figure 2 A) BRD4 inhibitor **1** reported by Hay *et al.* co-crystallised with BRD4 BD1 (PDB4NR8).[13] B) Synthetic yields and biochemical potencies (TR-FRET) for BRD4 BD1 and BD2 of PAL probes synthesised using the Ugi multicomponent reaction. C) Example of intact protein LCMS analysis of BRD4-BD1 upon irradiation (302 nm) with probe **2bg**. D) Photocrosslinking timecourses for probes at 302 nm, 3.8 mW cm^{-2} (solid line) and 365 nm, 7.6 mW cm^{-2} (dashed line).

To investigate the effect on photocrosslinking efficiency when the linker-length between the affinity function and photoreactive group was varied, three amine derivatives (**3**–**5**) of the more potent pan-BET inhibitor **6** (TR-FRET pIC_{50} BD1/BD2 = 7.1/6.3) were synthesised (Figure 3A–B).[12] These contained a one (Short), four (Medium) and nine (Long) atom linker respectively (Supplementary Information Schemes S4–S6 for synthetic routes). These three affinity functions were combined with the five photoreactive groups (**a**–**e**) *via* parallel synthesis using the Ugi protocol to access 15 novel photoaffinity probes in a single step. Reasonable synthetic yields were obtained (11–67%, Table 2), demonstrating synthetic tolerance for alternative affinity functions. The alkyne handle (**f**) was incorporated into all probes except for the aryl azide (**a** and **b**) probes, where ester-isonitrile (**g**) was employed to avoid unwanted cycloaddition reactions.

To establish how the various photoreactive groups impact on the cellular behaviour of the probes, the physicochemical (physchem) properties of each probe were measured (Table 2). Overall, good aqueous solubility was maintained upon addition of the PAL scaffold to the parent compound **6**, while permeability was reduced. All probes showed an increase in chromLogD relative to **6**. This increase in lipophilic character was reflected in an increase in percentage-binding to human serum albumin (HSA), which may correlate with non-specific binding should the probes be used in whole cells or lysates. Overall, probes containing the alkyl diazirine (**3cf**–**5cf**) showed the lowest deviation in properties from parent **6**.

In general, 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 each domain (TR-FRET, pIC_{50} ±0.5) with a 10-fold increase over the parent compound **6**. 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 LPS, which causes the downstream production of the pro-inflammatory cytokine MCP-1. Well-characterised BET targeting compounds (such as I-BET762) have been shown to inhibit the release of this cytokine, an important component of many inflammatory diseases.[16] All 15 probes were found to have activity in the assay, with slightly reduced potency (pIC_{50} = 6.1–6.6) compared to **6** (pIC_{50} = 7.1). This was likely due to a balance between increased target affinity (*cf.* TR-FRET biochemical assays) and reduced permeability, as suggested from the artificial membrane permeability (AMP).

Timecourse experiments were conducted with BRD4 BD1 to assess the effect of the three linker-lengths on photocrosslinking yield (Figure 3C–E). Comparing the photoreactive groups, the aryl azide probes gave the highest levels of labelling and rate of photoactivation for all three linker lengths. Probe **3ag** gave the highest levels of labelling (*ca.* 80%) and showed an extremely fast rate of photoactivation, reaching full activation within 15 seconds (Supplementary Information Figure S6). The exceptional level and rate of BRD4 BD1 labelling by **3ag** makes it a useful tool for studies with recombinant protein. However, the applications of aryl azides in cell-based PAL studies are limited due to their incompatibility with alkyne handles. Of the remaining three photoreactive groups, benzophenone gave the highest crosslinking yields. Interestingly, these probes showed high crosslinking yields for the short (**3ef**) and long linker (**5ef**) (*ca.* 40% and 60% respectively), while the medium linker (**4ef**) gave poor crosslinking (<5%). The aryl and alkyl diazirines (**3df–5df**, **3cf–5cf**) showed moderate levels of photolabelling, best exemplified by the short linker probe **3df** (15%). The alkyl diazirine probes were slower to photoactivate, requiring *ca.* 20 min irradiation versus *ca.* 2 min for the aryl-CF_3 diazirine probes. 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.

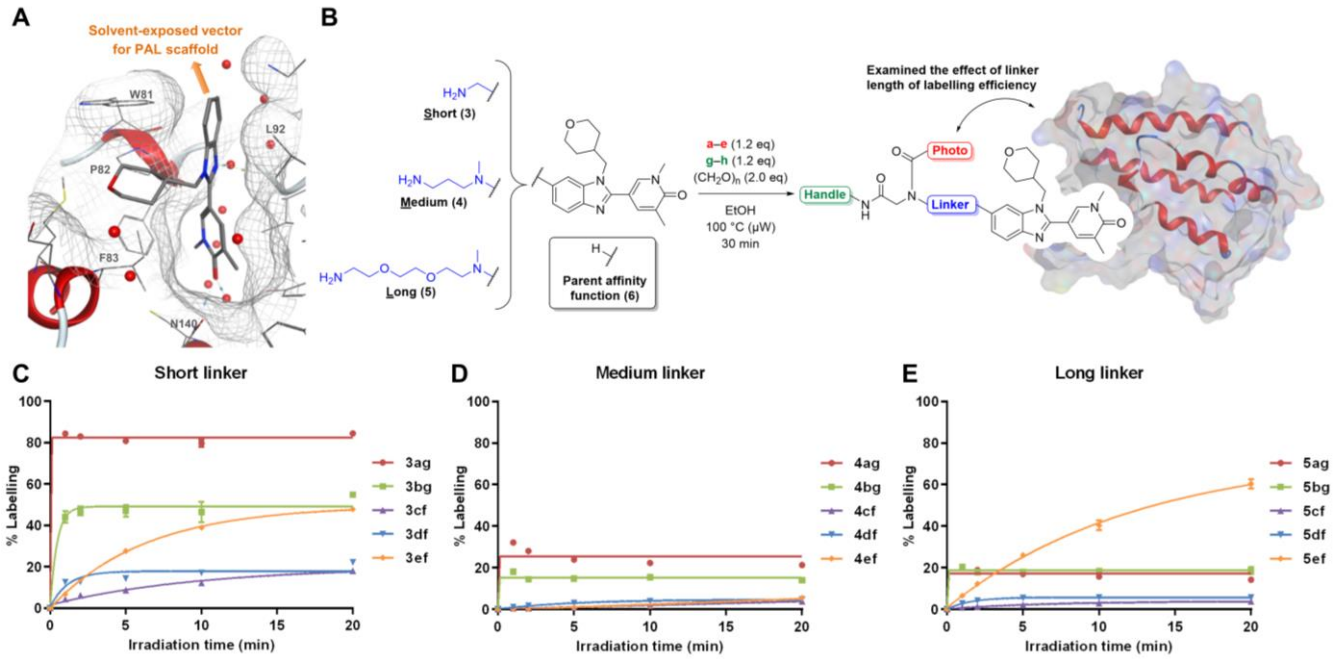
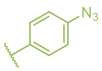
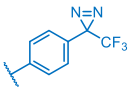
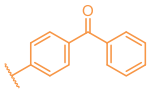


Figure 3 A) A more potent pan-BET inhibitor **6** with a known solvent-exposed vector was chosen as the affinity function (PDB 6TPY). B) A range of PAL probes with various linker lengths and photoreactive groups were synthesised in parallel using the Ugi protocol, enabling comparative assessment of crosslinking efficiency. C–E) Photocrosslinking timecourses (302 nm) of 15 PAL probes (6 μM) with recombinant BRD4 BD1 (3 μM), grouped by linker length.

Table 2 Measured affinity and physicochemical properties for the 15 PAL probes synthesised in parallel

Photoreactive group																
Linker length		S	M	L	S	M	L	S	M	L	S	M	L	S	M	L
Compound	**6**	**3ag**	**4ag**	**5ag**	**3bg**	**4bg**	**5bg**	**3cf**	**4cf**	**5cf**	**3df**	**4df**	**5df**	**3ef**	**4ef**	**5ef**
Yield (%)[a]	–	61	60	53	31	44	41	15	33	39	19	11	19	37	47	52
pIC_{50} BD1[b]	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[c]	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[d]	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 (%)[e]	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
Sol. (μM)[f]	452	293	115	244	254	232	231	519	398	444	379	422	491	303	453	857
AMP (nm/sec)[g]	223	98	27	76	23	19	21	<3	<3	<3	52	37	35	10	<3	7.9

[a] Isolated yield. [b,c] TR-FRET assay with recombinant BRD4 BD1/BD2. [d] Inhibition of MCP-1 cytokine production from monocytes in human whole blood. [e] Human Serum Albumin binding. [f] Solubility measured by Charged Aerosol Detection (CAD). [g] Artificial Membrane Permeability (AMP).

Overall, the short linker probes gave the highest levels of photolabelling (mean of three replicates). The large reduction in photocrosslinking yield between the short and medium linkers was surprising, since the linker length had increased by only three atoms. This trend of reduced crosslinking with increasing linker length continued for the long linker probes, except for **5ef**, demonstrating a more complex relationship between linker length, photoreactive group and labelling efficiency. The Ugi protocol described here can be used to quickly optimise the positioning of the photoreactive group to achieve high levels of photolabelling.

To further investigate the relationship between photoaffinity and linker length, selected probes that gave the highest levels of labelling were co-crystallised with BRD4 BD1. For the aryl azides **3ag** (PDB 6TPY) and **3bg** (PDB 7P6W), the interactions of the affinity function were conserved with those of parent **6** (Figure 4A, B). Both photoreactive groups show an edge to face interaction with W81 of the WPF shelf. Despite these similar binding modes, **3ag** resulted in higher levels of labelling *vs* **3bg** (80% *vs* 50% respectively), which may be explained through the lifetimes of their reactive intermediates. Extensive studies have shown that upon irradiation, the tetrafluoroaryl azide (**a**) not only produces a longer-lived singlet nitrene, but also that the re-arranged ketenimine intermediate is more electrophilic than the non-fluorinated analogue produced from **b**.[17]

For the long linker that incorporates a benzophenone moiety **5ef**, which was the only probe with a long linker to give high levels of labelling (*ca*. 60%), it was proposed that the long and flexible PEG chain may allow for the hydrophobic benzophenone (and the diradical produced upon irradiation) to sample transient interactions and crosslink to residues distal from the acetyl-lysine pocket. A co-crystal structure of **5ef** in BRD4 BD1 was obtained (PDB 7P6Y). The benzophenone photoreactive group was modelled in different conformations within the two chains in the asymmetric unit of this crystal structure, indicating its high flexibility (Supplementary Information Figure S7). One modelled conformation places the benzophenone in an energetically

unfavourable out-of-plane conformation, close to the protein surface as it forms a hydrogen bond with W81 (Figure 4C). In the second conformation, the benzophenone is less tightly associated with the protein and adopts more relaxed dihedral angles either side of the carbonyl group. The first may represent a conformation that leads to protein labelling, and the second may be quenched by buffer components. Both are likely to be two of many conformations dynamically sampled in solution. Benzophenone has been shown to photolabel methionine residues regioselectively.[18] 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 identify the residue that was covalently modified by **5ef** in solution, tryptic digestion and LC-MSMS experiments were performed, and **5ef** was found to label M149 exclusively (Figure 4D).

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 the benzophenone group 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, achieving this moderate level of methionine-selective labelling would be extremely challenging to predict, and may hinder accurate determination of the true binding site of the affinity function.[4d]

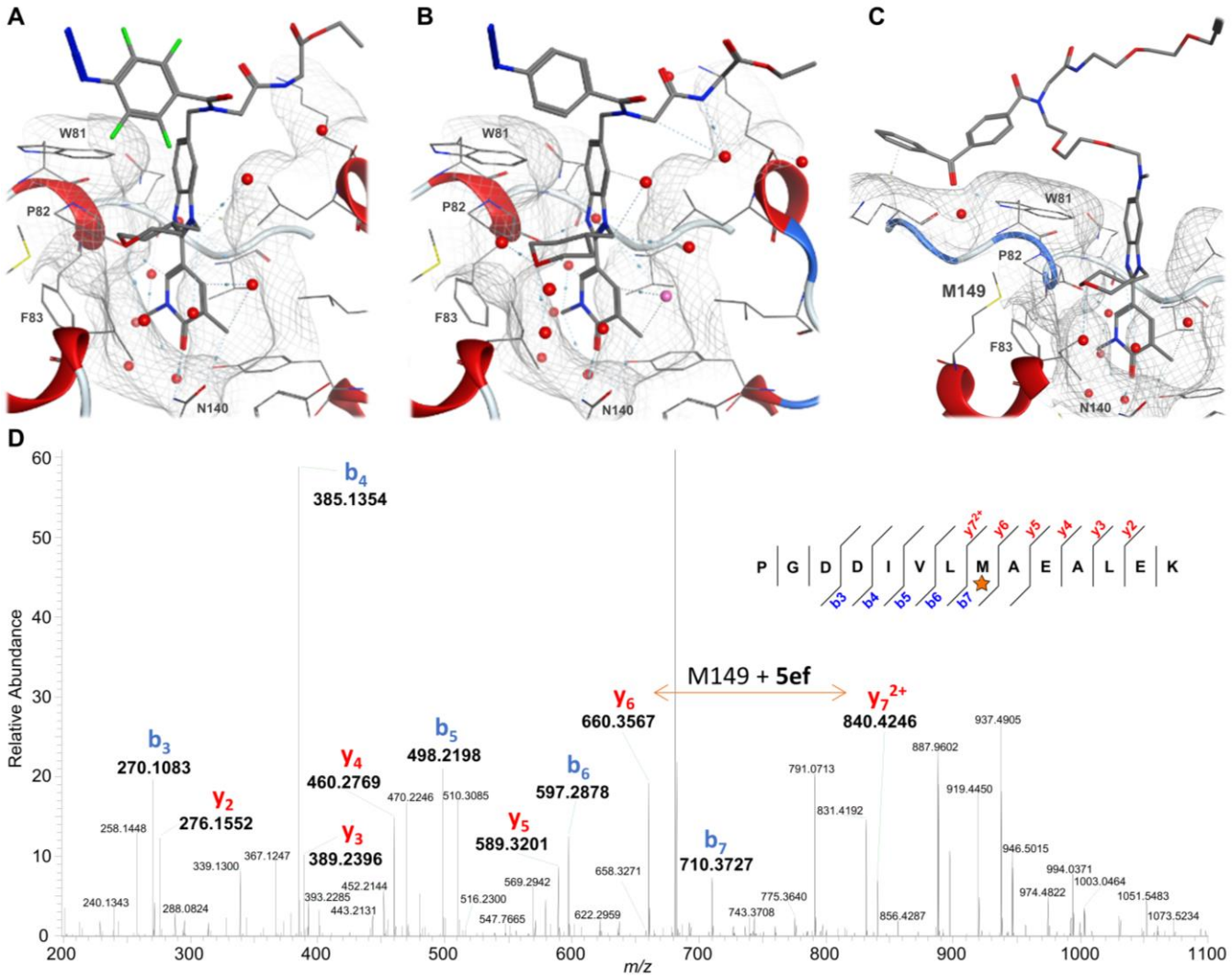


Figure 4 A–C) Views from co-crystal structures of BRD4 BD1 with **3ag** (PDB 7P6V), **3bg** (PDB 7P6W) and **5ef** (PDB 7P6Y) respectively. D) Tryptic digestion and LC-MSMS analysis determined that **5ef** covalently modified M149 by identification of the y_6 and y_7^{2+} fragment ions.

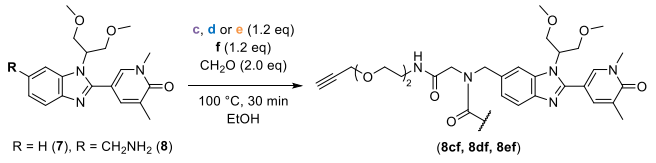
For all probes (6 μM), no labelling was observed in the presence of a competitive ligand (100 μM), indicating that the photolabelling was due to specific binding to the active site of BRD4 BD1 (Supplementary Information Figure S8), consistent with the crystallographic observations.

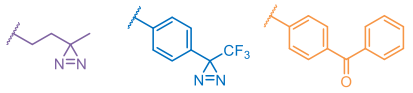
Timecourse studies were also carried out with 365 nm irradiation (Supplementary Information Figure S9). At this wavelength, aryl azide probes (**4bg** and **5bg**) afforded high levels of labelling, however, their rate of photoactivation was slower (*ca*. 10 min for complete activation). All diazirine and benzophenone probes gave low levels of photolabelling (<20% after 80 min). These studies indicate that efficient labelling is highly dependent on the precise nature of the photoreactive group, linker length, irradiation wavelength, and on topology of the protein's active site. Thus, the

design of PAL probes with high levels of labelling is non-trivial. Furthermore, if the probe is to be used in lysates or whole cells, appropriate physchem properties must also be considered. Therefore, the Ugi protocol enables the highly efficient preparation of multiple chemically diverse PAL probes in parallel, to enable a rapid assessment of labelling efficiency and physchem profiles.

We next assessed the use of the Ugi-derived PAL probes to capture endogenous bromodomains in live-cell MS-based proteomics. Probes **8cf**, **8df** and **8ef** containing an alkyl diazirine (**c**), aromatic diazirine (**d**) and benzophenone (**e**) respectively were synthesised from parent compound **8**, a short linker amine derivative of **7**, which had been found to have improved cellular potency over **6** (Figure 2B).[12] Upon Ugi reaction, the three probes (**8cf**, **8df** and **8ef**) showed similar trends in physchem properties, with cellular potency being maintained when compared to the parent **7**. Photocrosslinking yields with recombinant BRD4 BD1 and BD2 after 10 min irradiation (302 nm) were obtained and showed that the aromatic diazirine probe **8df** gave the highest average photolabelling across both domains (50%), followed by the benzophenone probe **8ef** (22%). The alkyl diazirine probe **8cf** gave poor levels of labelling at both domains (4%). Despite the high sequence homology between BD1 and BD2, **8df** and **8ef** showed high variations in crosslinking efficiencies, further demonstrating that subtle changes in protein topology surrounding the active site can have a significant impact on crosslinking yield.

Table 3 Ugi-derived PAL probes used in live-cell MS based proteomics.





Compound	7	8cf	8df	8ef
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
Sol. (μM)[e]	529	493	519	489
AMP (nm/sec)[e]	308	8.5	100	19
HSA (%)[g]	36	39	89	88
BD1 labelling (%)[h]	–	6	23	40
BD2 labelling (%)[i]	–	3	78	4

[a] Isolated yield from Ugi reaction. [b, c] TR-FRET assay with recombinant BRD4 BD1/BD2. [d] Inhibition of MCP-1 cytokine production from monocytes in human whole blood. [e] Solubility data measured by CAD (μM). [f] Artificial membrane permeability. [g] Binding to human serum albumin. [h, i] Photocrosslinking yield to recombinant BRD4 BD1 and BD2 respectively (302 nm, 10 min).

To compare the performance of the three photoreactive groups in MS-based proteomic experiments, HL-60 cells were treated with three conditions in duplicate; (i) PAL probe (1 μM); (ii) PAL probe + competition with parent amine **8** (50 μM); (iii) DMSO control. The samples we incubated for 1 h before irradiation (302 nm, 10 min). The cells were lysed, and the probe linked to biotin by CuAAC chemistry.[5b] Acetone protein precipitation was performed to remove the excess CuAAC reagents, and the proteins were resolubilised before enrichment of biotinylated proteins by neutravidin bead pulldown (Figure 5A). Captured proteins were subjected to on-bead digestion and the peptides were labelled with TMT reagents before analysis by tandem mass spectrometry.[19]

In the experiments containing CF_3-diazirine **8df** and benzophenone **8ef**, 135 and 95 proteins were ≥2-fold enriched by the presence of PAL probe *vs* DMSO control respectively, with 66 proteins commonly identified (Figure 5B–D). For the proteins ≥2-fold enriched by **8df**, only the BET family (BRD2, 3 and 4) were ≥2-fold competed by the presence of **8** (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 **8ef**, the BET family was enriched and competed greater than 4-fold. **8ef** 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.[20] Capture of these off-target proteins could arise from them 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, thus the longer linker probes synthesised here might prove particularly effective in sampling the local protein environment.[3c] For the 66 proteins that were enriched (≥2-fold) by both probes, but not competed by parent **8**, gene ontology (GO) term enrichment analysis was performed using DAVID bioinformatics resources using the total number of proteins identified (656) as background. The biological process and molecular function GO terms were grouped with a statistical cut-off of p-value ≤ 0.05 (Supplementary Information Figure S10). This analysis revealed a high representation of mitochondrial transport proteins, specifically the solute carrier (SLC) family of transporter proteins (Supplementary Information Table S1).[21] Non-competitive enrichment by both probes suggested that these proteins may have been captured by binding events through the groups other than the affinity function present in both probes, such as the alkyne handle or core Ugi scaffold. Many 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 (Supplementary Information Figure S11). The alkyl diazirine probe **8cf** did not afford robust identification of any BET family members from live cells, consistent with the low photocrosslinking yields observed with recombinant protein (Supplementary Information Figure S12, S13). This suggests that while the alkyl diazirine is often favoured in photolabelling studies due to its small size, it is not always optimal for protein capture, and these probes reported may have the potential to be optimised further.[3a]

These target engagement studies showed that for a PAL probe to successfully inform on non-covalent interactions formed by the parent compound, several criteria must be satisfied. The PAL probe should have good cell permeability and maintain potency at the same targets engaged by the parent compound, while being able to sufficiently label those proteins for successful identification. *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 comprehensive picture of the true ligand-protein interactions formed. Two of the three probes studied here (**8df** and **8ef**) showed successful competitive enrichment of the BET family, however, if the minimalist (and frequently preferred) alkyl diazirine was the only probe (**8cf**) to be

synthesised, the BET family members would have not been identified in the live cell experiments performed.

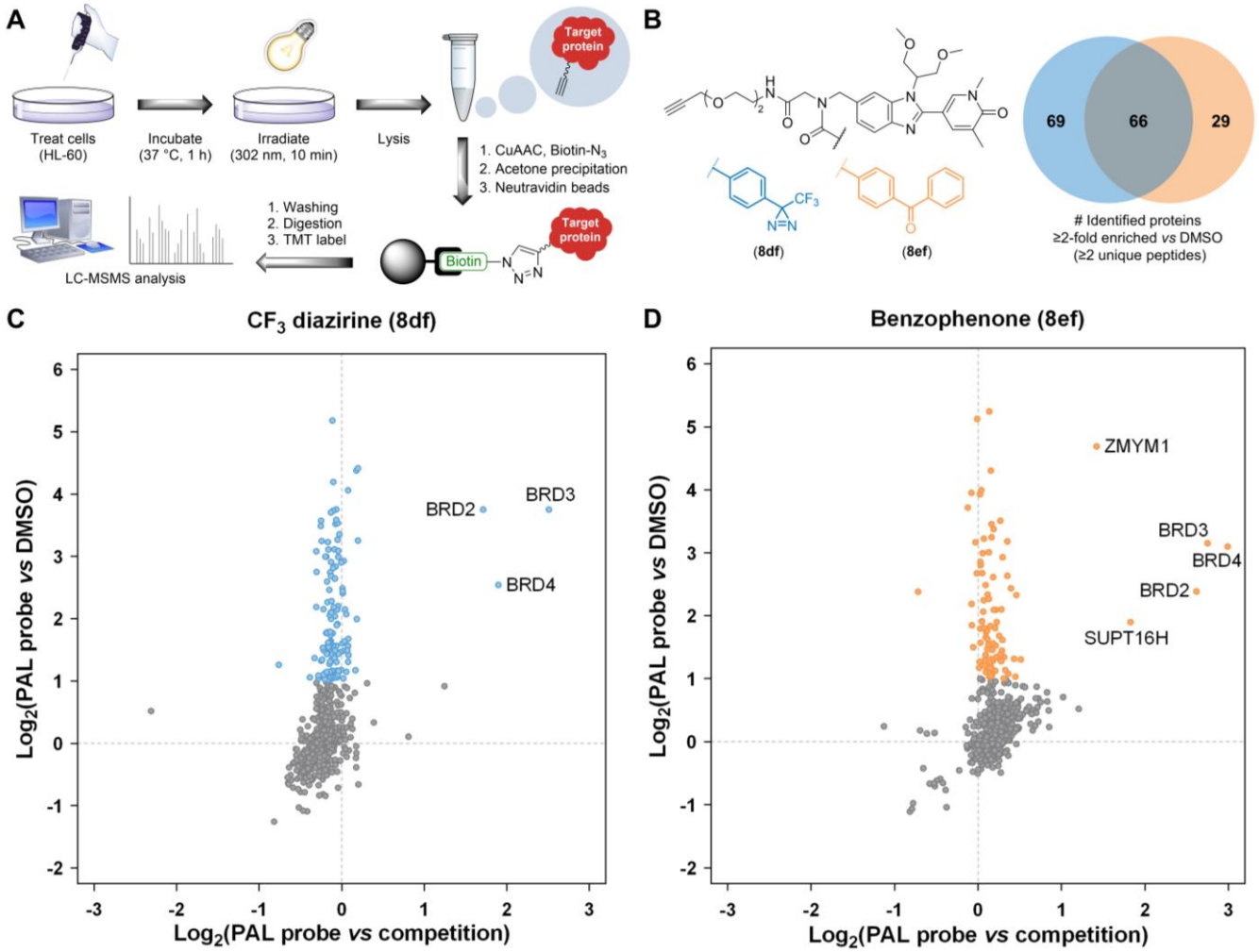


Figure 5 A) PAL MS-based proteomic workflow to identify proteins enriched by photolabelling. B) Number of proteins enriched ≥2-fold by probes **8df** (blue) and **8ef** (orange). C) Log2 representation of fold-difference between probe **8df** (1 μM) *vs* DMSO control (proteins ≥2-fold enriched in blue) and *vs* competitor **8** (50 μM). The BET family were enriched and specifically competed by the parent compound. D) Log2 representation of fold-difference between probe **8ef** (1 μM) *vs* DMSO control (proteins ≥2-fold enriched in orange) and *vs* competitor **8** (50 μM). The BET family were enriched and specifically competed by the parent compound, along with a subunit of the FACT complex (SUPT16H) a nuclear zinc finger protein (ZMYM1). Values are the mean of two replicates (see Supplementary Information Figure S13).

Conclusion

In summary, we describe a highly versatile protocol for the synthesis of PAL probes, which is compatible with commonly used photoreactive groups, handles and amine containing affinity functions. A series of pan-BET PAL probes were synthesised and the effects of the photoreactive group, linker length and irradiation wavelength on photocrosslinking yield were investigated, with the aid of crystallography and LC-MSMS studies. It was found that subtle variation in probe structure and protein topology around the active site can have a significant impact on photocrosslinking yield. An important finding is the observed instability of probes containing both alkyne and azide functional groups, which limits their application in live cells. We applied a selection of Ugi-derived probes in live-cell MS-based proteomic experiments, two of which (**8df** and **8ef**) afforded selective and robust capture of the BET family (BRD2, 3 and 4). Furthermore, these two probes potently inhibited MCP-1 cytokine release in human whole blood, thus are suitable for use as BET-targeting PAL tools *in vitro*. Many factors need to be balanced to achieve a PAL probe suitable for live-cell MS-based proteomics, including cell permeability, appropriate physicochemical properties and high photocrosslinking yields, and the trends observed here should inform future probe design. Given the observed variability in photocrosslinking yields, we propose it is advisable to synthesise a variety of probes to maximise target capture. The Ugi protocol provides a unique and convenient method to synthesise an array of PAL probes from an amine-functionalised compound of interest in a single step, thus facilitating the application of live-cell MS-based proteomics to provide on/off-target information, crucial in drug development.

Acknowledgements

We thank the GlaxoSmithKline/University of Strathclyde Collaborative PhD programme for funding this work. We thank the EPSRC for funding via Prosperity Partnership EP/S035990/1, along with a Postgraduate and Early Career Researcher Exchanges (PECRE) grant from the Scottish Funding Council

(H17014). We thank Samuel Dalton for his insight into CuAAC chemistry, Jürgen Stuhlfauth and Nilma Garcia-Altrieth for cell preparation, Emily Lowndes and Jennifer Love for crystallisation support, Kerstin Kammerer for LC-MSMS instrument support, and the GSK core profiling team for physicochemical measurements. CJS thanks the Wellcome Trust, the EPSRC and Cancer Research UK for funding.

Keywords: Photoaffinity labelling • Multi-component reactions • MS-based proteomics • Bromodomains • Epigenetics

[1] a) H.-P. Shih, X. Zhang, A. M. Aronov, *Nat. Rev. Drug. Discov.* **2017**, *17*, 19-33; b) M. J. Waring, J. Arrowsmith, A. R. Leach, P. D. Leeson, S. Mandrell, R. M. Owen, G. Pairaudeau, W. D. Pennie, S. D. Pickett, J. Wang, O. Wallace, A. Weir, *Nat. Rev. Drug Discov.* **2015**, *14*, 475-486.

[2] K. M. Comess, S. M. McLoughlin, J. A. Oyer, P. L. Richardson, H. Stockmann, A. Vasudevan, S. E. Warder, *J. Med. Chem.* **2018**, *61*, 8504-8535.

[3] a) H. Guo, Z. Li, *Med. Chem. Commun.* **2017**, *8*, 1585-1591; b) D. J. Lapinsky, D. S. Johnson, *FutureMedChem* **2015**, *7*, 2143-2171; c) D. P. Murale, S. C. Hong, M. M. Haque, J. S. Lee, *Proteome Sci.* **2016**, *15*, 14; d) G. W. Preston, A. J. Wilson, *Chem. Soc. Rev.* **2013**, *42*, 3289-3301; e) E. K. Grant, D. J. Fallon, H. C. Eberl, K. G. M. Fantom, F. Zappacosta, C. Messenger, N. C. O. Tomkinson, J. T. Bush, *Angew. Chem. Int. Ed. Engl.* **2019**, *58*, 17322-17327.

[4] a) J. Das, *Chem. Rev.* **2011**, *111*, 4405-4417; b) L. Dubinsky, B. P. Krom, M. M. Meijler, *Bioorg. Med. Chem.* **2012**, *20*, 554-570; c) J. R. Hill, A. A. B. Robertson, *J. Med. Chem.* **2018**, *61*, 6945-6963; d) G. Dorman, H. Nakamura, A. Pulsipher, G. D. Prestwich, *Chem. Rev.* **2016**, *116*, 15284-15398; e) E. Smith, I. Collins, *FutureMedChem* **2015**, *7*, 159-183.

[5] a) D. A. Shannon, E. Weerapana, *Curr. Opin. Chem. Biol.* **2015**, *24*, 18-26; b) C. G. Parker, A. Galmozzi, Y. Wang, B. E. Correia, K. Sasaki, C. M. Joslyn, A. S. Kim, C. L. Cavallaro, R. M. Lawrence, S. R. Johnson, I. Narvaiza, E. Saez, B. F. Cravatt, *Cell* **2017**, *168*, 527-541 e529; c) E. K. Grant, D. J. Fallon, M. M. Hann, K. G. M. Fantom, C. Quinn, F. Zappacosta, R. S. Annan, C. W. Chung, P. Bamborough, D. P. Dixon, P. Stacey, D. House, V. K. Patel, N. C. O. Tomkinson, J. T. Bush, *Angew. Chem. Int. Ed. Engl.* **2020**, *59*, 21096-21105.

[6] P. Kleiner, W. Heydenreuter, M. Stahl, V. S. Korotkov, S. A. Sieber, *Angew. Chem. Int. Ed.* **2017**, *56*, 1396-1401.

[7] a) K. Sakurai, S. Ozawa, R. Yamada, T. Yasui, S. Mizuno, *Chembiochem* **2014**, *15*, 1399-1403; b) G. W. Preston, S. E. Radford, A. E. Ashcroft, A. J. Wilson, *ACS Chem. Biol.* **2014**, *9*, 761-768.

[8] a) S. Pan, S. Y. Jang, D. Wang, S. S. Liew, Z. Li, J. S. Lee, S. Q. Yao, *Angew. Chem. Int. Ed.* **2017**, *56*, 11816-11821; b) Z. Li, D. Wang, L. Li, S. Pan, Z. Na, C. Y. Tan, S. Q. Yao, *J. Am. Chem. Soc.* **2014**, *136*, 9990-9998; c) Z. Li, P. Hao, L. Li, C. Y. Tan, X. Cheng, G. Y. Chen, S. K. Sze, H. M. Shen, S. Q. Yao, *Angew. Chem. Int. Ed.* **2013**, *52*, 8551-8556.

[9] J. T. Bush, L. J. Walport, J. F. McGouran, I. K. H. Leung, G. Berridge, S. S. van Berkel, A. Basak, B. M. Kessler, C. J. Schofield, *Chem. Sci.* **2013**, *4*, 4115.

[10] a) T. Kambe, B. E. Correia, M. J. Niphakis, B. F. Cravatt, *J. Am. Chem. Soc.* **2014**, *136*, 10777-10782; b) J. Ge, X. Cheng, L. P. Tan, S. Q. Yao, *Chem. Commun.* **2012**, *48*, 4453-4455; c) P. Jackson, D. J. Lapinsky, *J. Org. Chem.* **2018**, *83*, 11245-11253.

[11] I. Ugi, R. Meyr, U. Fetzer, C. Steinbruekner, *Angew. Chem.* **1959**, *71*, 386.

[12] C. R. Wellaway, D. Amans, P. Bamborough, H. Barnett, R. A. Bit, J. A. Brown, N. R. Carlson, C.-w. Chung, A. W. J. Cooper, P. D. Craggs, R. P. Davis, T. W. Dean, J. P. Evans, L. Gordon, I. L. Harada, D. J. Hirst, P. G. Humphreys, K. L. Jones, A. J. Lewis, M. J. Lindon, D. Lugo, M. Mahmood, S. McCleary, P. Medeiros, D. J. Mitchell, M. O'Sullivan, A. Le Gall, V. K. Patel, C. Patten, D. L. Poole, R. R. Shah, J. E. Smith, K. A. J. Stafford, P. J. Thomas, M. Vimal, I. D. Wall, R. J. Watson, N. Wellaway, G. Yao, R. K. Prinjha, *J. Med. Chem.* **2020**, *63*, 714-746.

[13] a) D. A. Hay, O. Fedorov, S. Martin, D. C. Singleton, C. Tallant, C. Wells, S. Picaud, M. Philpott, O. P. Monteiro, C. M. Rogers, S. J. Conway, T. P. Rooney, A. Tumber, C. Yapp, P. Filippakopoulos, M. E. Bunnage, S. Muller, S. Knapp, C. J. Schofield, P. E. Brennan, *J. Am. Chem. Soc.* **2014**, *136*, 9308-9319.

[14] R. A. Bit, *US 20180044317A1* (GSK), United States, **2018**.

[15] J. E. Horne, M. Walko, A. N. Calabrese, M. A. Levenstein, D. J. Brockwell, N. Kapur, A. J. Wilson, S. E. Radford, *Angew. Chem. Int. Ed.* **2018**, *57*, 16688-16692.

[16] a) T. H. Nguyen, S. Maltby, F. Eyers, P. S. Foster, M. Yang, *PloS One* **2016**, *11*, e0163392; b) S. L. Deshmane, S. Kremlev, S. Amini, B. E. Sawaya, *J. Interferon Cytokine Res.* **2009**, *29*, 313-326.

[17] K. A. Schnapp, R. Poe, E. Leyva, N. Soundararajan, M. S. Platz, *Bioconjugate Chem.* **1993**, *4*, 172-177.

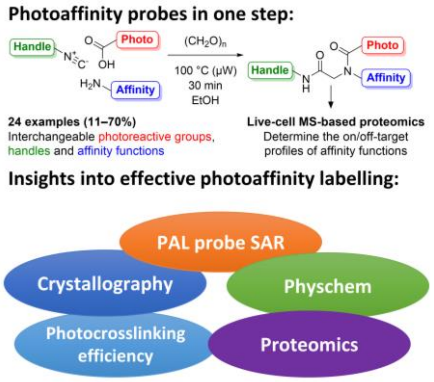
[18] a) A. Wittelsberger, B. E. Thomas, D. F. Mierke, M. Rosenblatt, *FEBS Lett.* **2006**, *580*, 1872-1876; b) E. Deseke, Y. Nakatani, G. Ourisson, *Eur. J. Org. Chem.* **1998**, *1998*, 243-251.

[19] L. Dayon, A. Hainard, V. Licker, N. Turck, K. Kuhn, D. F. Hochstrasser, P. R. Burkhard, J. C. Sanchez, *Anal. Chem.* **2008**, *80*, 2921-2931.

[20] a) R. Belotserkovskaya, S. Oh, V. A. Bondarenko, G. Orphanides, V. M. Studitsky, D. Reinberg, *Science* **2003**, *301*, 1090-1093; b) M. D'Ascenzio, K. M. Pugh, R. Konietzny, G. Berridge, C. Tallant, S. Hashem, O. Monteiro, J. R. Thomas, M. Schirle, S. Knapp, B. Marsden, O. Fedorov, C. Bountra, B. M. Kessler, P. E. Brennan, *Angew. Chem.* **2018**, *131*, 1019-1024.

[21] a) D. W. Huang, B. T. Sherman, R. A. Lempicki, *Nat. Protoc.* **2008**, *4*, 44; b) W. Huang da, B. T. Sherman, R. A. Lempicki, *Nucleic Acids Res.* **2009**, *37*, 1-13; c) L. Lin, S. W. Yee, R. B. Kim, K. M. Giacomini, *Nat. Rev. Drug. Discov.* **2015**, *14*, 543.

Entry for the Table of Contents



Get PALs quick: A multicomponent reaction for the one-step synthesis of PAL probes. A library of diverse bromodomain (BET)-targeting PAL probes was synthesised and profiled extensively (physicochemical properties, biochemical and cellular activities, photolabelling efficiencies and crystallographic studies). Optimal probes were profiled by MS-based proteomics to capture BET proteins from live cells and reveal their potential on/off-target profiles. We envisage that this methodology can be used to quickly access PAL probes for other compounds-of-interest and obtain their target profiles.

Institute and/or researcher Twitter usernames: @Jake_T_Bush