Development of macrocyclic peptides as multifunctional probes in chemical biology



Grace Roper

Linacre College
University of Oxford
Michaelmas Term 2022

A thesis submitted to the board of the Faculty of Physical Sciences of the University of Oxford in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

Abstract

Multifunctional probes combine a target engagement motif with at least one additional application, and have been used in a myriad of ways in chemical biology to identify the roles of their target in normal and disease settings. In this thesis, multifunctionality is installed to selective cyclic peptide inhibitors of epigenetic proteins: **CP2** which inhibits histone demethylases 4A-C (KDM4A-C), and **TiP1** which inhibits ten eleven translocation enzymes (TETs).

Multifunctional peptide probes were synthesised for epigenetic targets by addition of unnatural amino acids as click chemistry handles, or through installing photoreactive groups. Cu(I)-catalysed azide-alkyne cycloaddition click chemistry was used to synthesise peptide conjugates. Both **CP2**- and **TiP1**-based affinity peptide probes were produced by conjugating biotin to the peptides, and target engagement of biotinylated peptides was tested in cell lysates. Photoreactive multifunctional peptides were also assessed, for both KDM4A and TET activity, however these did not specifically crosslink their targets under the conditions tested.

E3 ligase binders were also installed to produce cyclic peptide proteolysis recruiting chimeras (PROTACs), to promote target protein degradation through the ubiquitin-proteasome system. A panel of ten **CP2**-based PROTACs was synthesised, using different linkers, recruiting either von Hippel-Lindau (VHL) or cereblon (CRBN) E3 ligases. KDM4A protein degradation was established in cells transiently overexpressing HaloTagged and Flag-tagged KDM4A (HaloFlagKDM4A) using HaloTag targeting PROTACs. Conditions established using HaloPROTACs were used to test the panel of **CP2**-PROTACs.

Acknowledgements

I would like to begin by thanking Prof Akane Kawamura for all of her support, encouragement, and friendship throughout my DPhil. Dr Albert Isidro Llobet has been so valuable as my GSK supervisor, his expertise and ideas in peptide chemistry and beyond have been much appreciated. I would additionally like to thank Prof Christopher Schofield for his ideas, as well as for facilitating my studies at Oxford after I moved to Newcastle. I am grateful to all my supervisory team for helping me to develop as a scientist, despite many obstacles throughout the course of my DPhil.

Having completed lab work at both the University of Oxford and Newcastle University, I have received help and technical support from many people. Thanks to Dr Andrew Porter and the team at NUPPA, as well as the Oxford Chemistry department MS team for all their MS-related help. Several members of the Kawamura lab have helped me with experiments, thanks to Dr Joanna Bonnici for help with MS and assays; Dr Chiara Maniaci, Dr Marie- Hélène Ruchaud, and Dr Hilal Sarac for help with mammalian cell culture and protein degradation experiments; Dr Chiara Maniaci for help with recombinant protein expression and purification; Dr John-Paul Bukowski for help with cloning; and to Siddique Amin for help with analytical HPLC. I would also particularly like to thank my student, Valeria Cobiltean, for her enthusiasm in the TET project. Thanks also to Dr Roman Belle and Dr Klemensas Simelis for their support with the TET project. I am very grateful for the numerous scientific discussions with past and present members of the Kawamura lab, as well as the Flashman and Schofield labs, for inspiring my science. I would additionally like to thank all those that helped with this thesis: Prof Akane Kawamura, Dr Marie- Hélène Ruchaud, Dr Roman Belle, Dr Joanna Bonnici, Dr Chiara Maniaci, Dr Tom McAlister, and Dr Klemensas Simelis.

The support of the Doctoral Training Centre has been extremely helpful in allowing me to finish my studies, I would particularly like to thank Nina Stamps.

I am grateful for the friendship of many people, especially throughout the many times when the end of this DPhil seemed a long way away. Special thanks to Joanna, Leah, Karen, Anka, and Chiara for all the moral support! Thanks also to Ros and Martika out of the lab, as well as Loane, Klem, Raphael, Hilal, Siddique, Emma, all the Toms, Tim, Oli, Filo, Roman, Ollie, and to all members of the Kawamura lab.

My husband, Dr Cameron Edward Roper Taylor, has been a constant support outside the science, and with the science. There is no one else I would have wanted by my side through the particularly tough times which coincided with my DPhil, thank you. A special thanks also to my parents, Thea and Nigel, as well as to my entire family on both the Roper and Taylor sides.

Declaration and contribution list

I declare that the work presented in this thesis is entirely my own, unless a reference to a person or a published source has been made, with the exception of the following work:

Chapter 2:

TiP1-based probes were synthesised by Valeria Cobiltean under my supervision.

Ph-3, **Ph-4**, and **Ph-5** were synthesised by Dr Jacob Bush.

RapidFire mass spectrometry to confirm the mass of His_6-$KDM4A_{1-359}$ was performed by Dr Leah Taylor-Kearney.

Chapter 3:

B-1 mass spectrometry-based proteomics were performed with Dr Andrew Porter, at Newcastle University Protein and Proteome Analysis facility (NUPPA). This includes proteomics on an excised gel band, and whole pulldown sample proteomics experiments.

TiP1-based pulldown and photoreactive probe experiments were performed by Valeria Cobiltean under my supervision.

Doxycycline-induced/not-induced FlagTET1 U2OS cell lysates were produced with Dr Hilal Sarac.

Chapter 6:

Some peptide HRMS spectra were acquired by the University of Oxford, Department of Chemistry mass spectrometry service.

Abbreviations

=	Equal to
≈	Approximately equal to
%	Percent
+	Plus
±	Plus or minus
×	Times
>	Greater than
<	Less than
≥	Greater than or equal to
&	And
°C	Degrees Celsius
× *g*	Times gravity
µ	Micro
2OG	2-Oxoglutarate
5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
Å	Angstrom
ABBP	Activity-based protein profiling
Ac	Acetyl
AdPROM	Affinity-Directed Protein Missile
ALP	Autophagy-lysosome pathway
AMP	Adenosine monophosphate
amu	Atomic mass unit
APS	Ammonium persulfate
Ari	ADP ribosylation
ATP	Adenosine triphosphate
b	Base
BCA	Bicinchoninic acid
BER	Base excision repair
BET	Bromodomain and extraterminal domain

Boc	*Tert*-butyloxycarbonyl
Bpa	Benzophenone
BRD	Bromodomain
BRET	Bioluminescence resonance energy transfer
BSA	Bovine serum albumen
Bu	Butyryl
c	Centi
Cas9	CRISPR associated protein 9
cDNA	Complementary deoxyribonucleic acid
Chem-seq	Chemical affinity capture followed by DNA sequencing
ChIP-seq	Chromatin immunoprecipitation followed by DNA sequencing
Cit	Citrullin
CMV	Cytomegalovirus
CPP	Cell penetrating peptide
Cq	Quantification cycle
Cr	Crotonyl
CRBN	Cereblon
CRISPR	Clustered regularly interspaced short palindromic repeats
CuAAC	Cu(I)-catalysed 1,3-dipolar azide-alkyne cycloaddition
CV	Column volume
D	Dextrorotatory relative to dextrorotatory-glyceraldehyde
Da	Dalton
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DC_{50}	Half maximal degradation constant
DMEM	Dulbecco's modified eagle's medium
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyl transferase
dNTP	Deoxynucleotide triphosphate
DUBTAC	Deubiquitinase targeting chimera
E. Coli	*Escherichia coli*
e.g.	*exempli gratia*
E1	Ubiquitin-activating enzyme

E2	E2 ubiquitin-conjugating enzyme
E3 ligase	E3 ubiquitin-ligase
EC_{50}	Half maximal effective concentration
EDTA	Ethylenediaminetetraacetic acid
eq.	Equivalent
ERK	Ras-dependent extracellular signal-regulated kinase
et al.	*et alibī*
FBS	Foetal bovine serum
FDH	Formaldehyde dehydrogenase
Flag-tag	DYKDDDDK-Tag
Fmoc	Fluorenylmethyloxycarbonyl
Fo	Formyl
FPLC	Fast protein liquid chromatography
g	Gram
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
H1, H1.4, H2A, H2B, H3, H3.1, H4	Histone protein 1, 1.4, 2A, 2B, 3, 3.1, 4
H1.4K26Me_3	Tri-methylated lysine-26 on histone 1.4
H3K4Me_3	Tri-methylated lysine-9 on histone 4
H3K20Me_3	Tri-methylated lysine-20 on histone 3
H3K23Me_3	Tri-methylated lysine-23 on histone 3
H3K27	Lysine-27 on histone 3
H3K36$Me_{2/3}$	Di- and tri-methylated lysine-36 on histone 3
H3K9$Me_{2/3}$	Di- and tri-methylated lysine-9 on histone 3
H3K9Me_3(1–15)	Tri-methylated lysine-9 on histone 3 residues 1–15
H3K9Me_3(28–42)	Tri-methylated lysine on histone-36 residues 28–42
HATU	1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium3-oxid hexafluorophosphate
HDAC	Histone deacetylase
HECT	Homologous to E6AP *C*-terminus
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid)
Hib	2-Hydroxyisobutyryl
HIF	Hypoxia-inducible factor
His_6	Hexahistidine

HPLC	High-performance liquid chromatography
hr	Hour
HRMS	High-resolution mass spectrometry
HRP	Horse radish peroxidase
Hz	Hertz
hv	Add energy *using* photons
IAP	Inhibitors of apoptosis proteins
IC_{50}	Half maximal inhibitory concentration
k	Kilo
K_d	Dissociation constant
KDM	Histone lysine demethylase
$KDM4A_{1-359}$	Histone lysine demethylase 4A residues 1–359
KMe_2	Di-methyllysine
KMe_3	Tri-methylated lysine
L	Laevorotatory relative to dextrorotatory-glyceraldehyde
L	Litre
LB	Luria-Bertani
LC-MS	Liquid chromatography-mass spectrometry
LDS	Lithium dodecyl sulphate
LHS	Left-hand side
LYTAC	Lysosome targeting chimera
m	Meter
m	Milli
M	Molar
m/z	Mass-to-charge ratio
MALDI-MS	Matrix-assisted laser desorption/ionisation time of flight mass spectrometry
MDM	Mouse double minute 2 homolog
Me	Methyl
MeCN	Acetonitrile
min	Minute
mol	Mole
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry

MS/MS	Tandem mass spectrometry
mut	Mutant
MW	Molecular weight
MΩ	Megaohms
n	Nano
NAD^+	Nicotinamide adenine dinucleotide
NMP	1-Methyl-2-pyrrolidinone
NMR	Nuclear magnetic resonance spectroscopy
No.	Number
NOG	N-Oxalylglycine
NUPPA	Newcastle University Protein and Proteome Analysis service
OD_{600}	Optical density at 600 nm
Og	*O*-linked-*N*-acetylglucosaminyl
p	Pico
P	P-value
Pbf	2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PET	Positron emission tomography
pH	Potential of hydrogen
PHD	Prolyl hydroxylase domain-containing protein
PHICS	Phosphorylation-inducing chimeric small molecules
Pho	Phosphoryl
PhoRC	Phosphatase recruiting chimera
pK_d^{app}	Negative log of the apparent dissociation constant
POI	Protein of interest
PP_i	Pyrophosphate
PPI	Protein-protein interaction
ppm	Parts per million
Pr	Propionyl
PRC	Polycomb repressive complex
PRMT	Arginine methyltransferase

PROTAC	Proteolysis targeting chimera
PTM	Post translational modification
R	Unspecified organic group
RaPID	Random non-standard peptide integrated discovery
ref.	Reference
RFU	Relative fluorescence unit
RHS	Right-hand side
RIBOTAC	Ribonuclease targeting chimera
RING	Really interesting new gene
RNA	Ribonucleic acid
RPM	Rotations per minute
RT-qPCR	Real time-quantitative polymerase chain reaction
s	Second
S.O.C.	Super optimal broth with catabolite
SAHA	Suberoylanilide hydroxamic acid
SAM	S-Adenosylmethionine
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SPAAC	Strain-promoted azide-alkyne cycloaddition
SPECT	Single-photon emission computed tomography
SPPS	Solid-phase peptide synthesis
Su	Succinyl
TBS	Tris buffered saline
TBST	Tris buffered saline containing 0.1% Tween20
TBTA	Tris[(-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine
*t*Bu	*Tert*-butyl
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TDG	Thymine DNA glycosylase
TET	Ten-eleven translocation enzyme
TFA	Trifluoroacetic acid
THPTA	Tris(3-hydroxypropyltriazolylmethyl)amine
TNBS	2,4,6-Trinitrobenzenesulfonic acid
TPD	Targeted protein degradation
TRAP	Thrombospondin-related anonymous protein

Trt	Trityl
TSP	3-(Trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt
U	Unit
Ub	Ubiquitin
UPS	Ubiquitin-proteasome system
UV	Ultraviolet
v	Version
V	Volt
v/v	Volume per volume
VHL	von Hippel-Lindau protein
w/v	Weight per volume
wt	Wild type

Table of Contents

Abstract iii

Acknowledgements iv

Declaration and contribution list v

Abbreviations vi

Chapter 1 Introduction 1

1.1 Epigenetics 1

1.1.1 Mechanisms of epigenetics 1

1.1.2 2-oxyglutarate and Fe(II) dependent oxygenases in epigenetics 4

1.1.2.1 KDM4s 4

1.1.2.2 TETs 7

1.2 Multifunctional probes in chemical biology 10

1.2.1 Unmet need for selective epigenetic probes 11

1.2.2 Selected categories of multifunctional probes 11

1.2.2.1 Affinity probes and activity-based protein profiling (ABPP) 11

1.2.2.2 Protein dimerization probes 13

1.2.2.3 Protein degradation and PROTACs 14

1.2.2.3.1 Protein degradation 14

1.2.2.3.2 PROTACs 16

1.2.2.4 Probes as imaging tools 17

1.2.2.5 Antibodies as multifunctional probes 18

1.2.3 Characteristics of multifunctional probes for different applications 20

1.2.3.1 Impact of binding-orientation on multifunctional probes 20

1.2.3.2 Impact of linkers in multifunctional probes 21

1.3 Cyclic peptides as chemical probes 23

1.4 Click chemistry for rapid functionalisation of molecules 26

1.4.1 Selected click chemistry examples 27

1.5 Project Aims 30

Chapter 2 Synthesis and characterisation of multifunctional peptide probes 32

2.1 Introduction 32

2.1.1 Multifunctional probes 32

2.1.2 Synthesis of multifunctional probes 35

2.1.3 Aims 36

2.2 Results 38

2.2.1 General cyclic peptide synthesis 38

2.2.2 Design of clickable macrocyclic peptides 39

2.2.2.1 **CP2**-based clickable macrocyclic peptides 39

2.2.2.1.1 **CP2.3-P13K$_{N3}$** synthesis 41
2.2.2.2 **TiP1**-based clickable macrocyclic peptides 45
2.2.3 Click chemistry as a platform to synthesise multifunctional cyclic peptides 46
2.2.3.1 Click reaction optimisation: coupling **CP2-T13K$_{N3}$** to biotin-(PEG)$_4$-alkyne as a model system 46
2.2.3.2 Synthesis of multifunctional macrocyclic peptide conjugates 48
2.2.4 Design and synthesis of photoreactive macrocyclic peptides 51
2.2.5 KDM4A inhibition by **B-1** compared to **CP2** 53
2.2.5.1 Recombinant expression and purification of *N*-His$_6$ KDM4A$_{1-359}$ 53
2.2.5.2 Formaldehyde dehydrogenase coupled assay to compare inhibitory activity of **B-1** to **CP2** 55
2.3 Discussion 56
Chapter 3 Macrocyclic peptide affinity probes 59
3.1 Introduction 59
3.1.1 Aims 63
3.2. Results 64
3.2.1 Biotinylated macrocyclic peptides as affinity probes 64
3.2.1.1 Optimisation of KDM4A pulldown from cell lysates using **B-1** 65
3.2.1.1.1 Optimisation of FlagKDM4A overexpression and lysis for efficient affinity pulldown 66
3.2.1.1.2 Pulldown of KDM4A from HEK293T cell lysates containing overexpressed FlagKDM4A 69
3.2.1.1.3 Pulldown of KDM4A from endogenous HEK293T cell lysates 71
3.2.1.1.4 Work towards capture of endogenous protein-protein interaction partners of KDM4A 75
3.2.1.2 **B-1** as a MS-based proteomics probe 78
3.2.1.3 TET pulldown from cell lysates overexpressing TET1 83
3.2.2 Photoreactive cyclic peptide probes 88
3.2.2.1 **CP2** photoreactive probes 90
3.3.2.2 **TiP1** photoreactive probes 96
3.3 Discussion 98
Chapter 4 Macrocyclic peptide protein degradation 101
4.1 Introduction 101
4.1.1 Mode of action of PROTACs 101
4.1.2 Peptide-based PROTACs 103
4.1.3 Epigenetic PROTACs 104
4.1.4 The potential of macrocyclic peptides PROTACs 106
4.1.5 Chapter aims 106
4.2 Results 107

4.2.1 Establishing KDM4A protein degradation using HaloTag targeting PROTACs 107
4.2.1.1 Cloning FlagKDM4A into a pFN21A vector 109
4.2.1.2 HaloPROTAC protein degradation 111
4.2.1.2.1 HaloPROTAC treatment 48 hours post-transfection 111
4.2.1.2.2 HaloPROTAC treatment 4 hours post-transfection 113
4.2.1.2.3 Investigation of HaloPROTAC-E mode of action 114
4.2.2 Cyclic peptide protein degradation 118
4.2.2.1 Testing a panel of macrocyclic peptide protein degraders 120
4.2.2.2 Detailed investigation of **E3-10** 123
4.2.2.2.1 **E3-10** dose-response 123
4.2.2.2.2 Quantification of mRNA expression in cells treated with **E3-10** and HaloPROTAC-E 124
4.2.2.2.3 **E3-11** – a **CP2.3**-based PROTAC? 125
4.2.2.2.4 Protein degradation is not observed when **E3-10** or **E3-11** are added to non-transfected HEK293T cells 127
4.2.2.2.5 Work towards validation of the mode of action of **E3-10** 129
4.3 Discussion 131
Chapter 5 Discussion and future work 134
Chapter 6 Materials and methods 140
6.1 General information 140
6.2 Peptide Synthesis 140
6.2.1 9-Fluorenylmethoxycarbonyl (Fmoc) solid-phase peptide synthesis 140
6.2.1.1 Fmoc-protected non-proteinogenic amino acids used 142
6.2.2 **CP2.3-P13K$_{N3}$** peptide synthesis 143
6.2.2.1 Manual amino acid coupling 143
6.2.2.2 TNBS test for primary amines 143
6.2.2.3 *p*-Chloranil test for secondary amines 144
6.2.2.4 Synthesis of CP2.3-P13K$_{N3}$ involving manual *N*-methylation 144
6.2.3 Cyclisation 145
6.2.4 Peptide purification 145
6.2.5 Purified peptide characterisation 146
6.3 MALDI-MS 147
6.4 Click reactions 147
6.4.1 Synthesis of peptide conjugates 147
6.4.2 Conjugation of Cy5 dye to peptides after photocrosslinking 148
6.5 General chemical biology methods 149
6.5.1 Protein SDS-PAGE gels 149
6.5.1.1 Hand-cast SDS-PAGE protein gels 149

6.5.1.2 Running pre-cast SDS-PAGE protein gels 150
6.5.2 Western blotting 150
6.5.2.1 Antibodies used to probe western blots 151
6.5.3 DNA agarose gel electrophoresis 152
6.5.4 Transformation of *E. coli* 153
6.5.5 DNA plasmid preparation 153
6.6 Production of recombinant KDM4A 154
6.6.1 Expression of *N*-His_6-$KDM4A_{1-359}$ in *E. Coli* 154
6.6.2 *N*-His_6 $KDM4A_{1-359}$ purification 155
6.7 Formaldehyde dehydrogenase coupled His_6-$KDM4A_{1-359}$ assays 155
6.8 Mammalian cell culture 156
6.8.1 Cell line maintenance and seeding 157
6.8.2 Cell lysis 158
6.8.2.1 Whole cell lysate production 158
6.8.2.2 Nuclear cell lysate production 159
6.8.2.3 Cell lysate protein quantification 159
6.9 Pulldown assays and MS-based proteomics experiments 160
6.9.1 Affinity pulldown using peptide-biotin conjugates 160
6.9.2 Affinity pulldown for global MS-based proteomic analysis 161
6.9.3 Silver stain SDS-PAGE protein gel band extraction, destaining, and MS-analysis 163
6.10 Photoreactive protein labelling 164
6.10.1 Recombinant KDM4A photocrosslinking 164
6.10.2 TET photocrosslinking 165
6.11 Cloning pFN21A HaloFlagKDM4Awt and HaloFlagKDM4Amut plasmids 166
6.11.1 Appending AsiSI and PmeI cutting sites to FlagKDM4Awt/mut 166
6.11.2 Restriction digest of cloning inserts and acceptor pFN21A vector 167
6.11.3 Ligation of PCR products and Flexi Vector, and transformation 167
6.11.4 Colony screening 168
6.12 HEK293T protein degradation experiments 169
6.12.1 mRNA extraction and reverse transcription RT-qPCR 170
Chapter 7 References 172
Appendix 1 Peptide and peptide conjugate characterisation 192
Appendix 1.1 Peptide and peptide conjugate characterisation data 192
Appendix 1.2 Peptide conjugates analytical HPLC data 194
Appendix 1.3 Example HRMS data 196
Appendix 2 FDH coupled assay linear range 197
Appendix 3 Additional cell lysate blots 198
Appendix 3.1 Nuclear and remaining cell lysate western blot 198

Appendix 3.2 Endogenous and overexpressed KDM4 western blot 198
Appendix 4 KDM4A protein stability in LC-MS injector after irradiation 199
Appendix 5 Uncropped western blots from Chapter 3 200
Appendix 5.1 Figure 3.4 200
Appendix 5.2 Figure 3.5 202
Appendix 5.3 Figure 3.6 202
Appendix 5.4 Figure 3.7 203
Appendix 5.5 Figure 3.9 204
Appendix 5.6 Figure 3.10 205
Appendix 6 Uncropped western blots from Chapter 4 206
Appendix 6.1 Figure 4.6 206
Appendix 6.2 Figure 4.7 207
Appendix 6.3 Figure 4.8 208
Appendix 6.4 Figure 4.10 208
Appendix 6.5 Figure 4.12 210
Appendix 6.6 Figure 4.13 213
Appendix 6.7 Figure 4.15 214
Appendix 6.8 Figure 4.16 215
Appendix 6.9 Figure 4.17 216
Appendix 7 Fmoc-protected L-amino acid details 217

Chapter 1 Introduction

1.1 Epigenetics

1.1.1 Mechanisms of epigenetics

Epigenetics is the study of heritable changes to gene expression without altering the sequence of DNA. Epigenetic regulation occurs through covalent modifications to chromatin, which is a complex of DNA and histone proteins that forms chromosomes in eukaryotic cells (**Figure 1.1**).[1,2] Covalent chromatin modifications can occur to DNA itself, or to histone proteins.[3,4] Epigenetic modifications are responsible for regulating gene expression, which is especially important during development, and cell differentiation.[1] Dysregulation of epigenetic processes is strongly associated with disease, including cancer, inflammation, the immune response, and neurodegenerative dideases.[1,5]

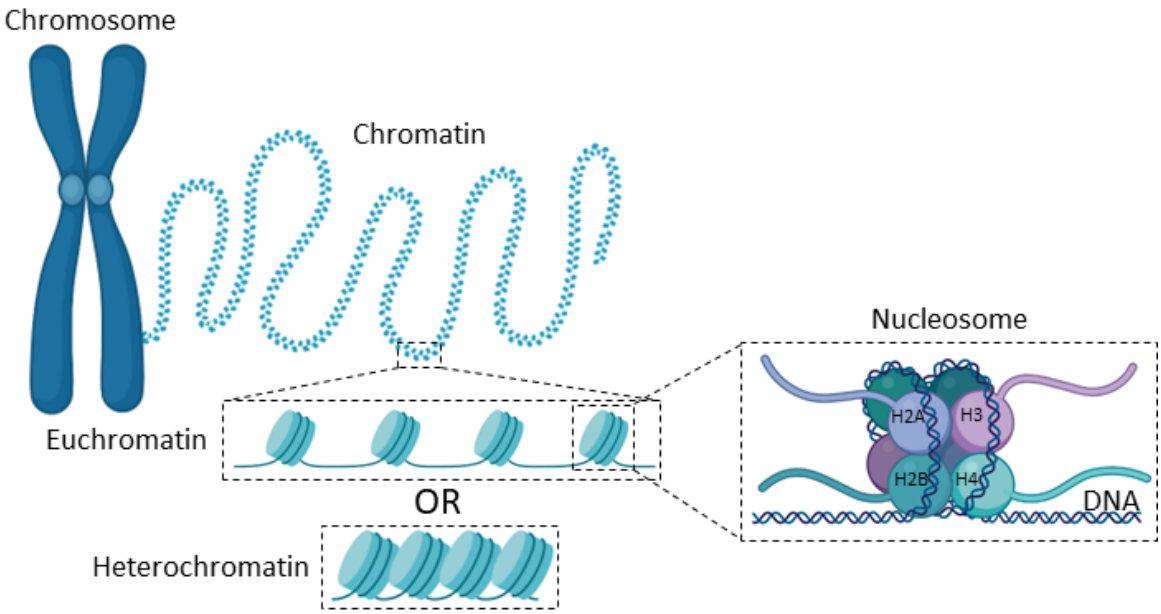


Figure 1.1 In eukaryotes DNA is stored in chromatin, which condenses into chromosomes. DNA is wrapped around nucleosome cores, made up of an octamer of four histone proteins, with their flexible *N*-terminal tails protruding from the surface of the nucleosome. Nucleosomes are linked together by H1, and DNA, to form chromatin. Chromatin can exists in an 'open' form, euchromatin, in which genes are mostly active; or it can exist in a 'closed' form, heterochromatin, in which genes are mostly silent.[6] Figure created with BioRender.com.

In eukaryotes, DNA is stored in chromosomes; made up of compacted chromatin, formed of DNA coiled around nucleosome subunits (**Figure 1.1**).[6] Nucleosomes contain an octamer of histone proteins, containing two each of histone H2A, H2B, H3, and H4.[7] Each nucleosome has approximately 146 DNA base pairs wound around it, and nucleosomes are linked together by H1 proteins.[6] The core nucleosome histone proteins are globular, with flexible tails protruding from the surface.[8] Generally, chromatin exists in two forms: (i) euchromatin, which is relatively 'open', and here genes are mostly actively transcribed; or (ii) heterochromatin, which is tightly condensed or 'closed', and genes here are primarily not transcribed.[9]

Histone proteins can have a number of different post translational modifications (PTMs), which are primarily located within their *N*-terminal tails.[10] There are a vast number of potential combinations of histone PTMs, forming a 'histone code' for epigenetic regulation.[8,10] H3.1 modifications are detailed in **Figure 1.2**, they demonstrate the diversity of chemistry involved, and the complexity of the epigenetic landscape.[11,12] Alongside the many epigenetic modifications, there are many proteins involved in regulating histone PTMs, commonly described as readers, writers, and erasers. Epigenetic proteins can be split into several families, made up of multiple related proteins, which often have overlapping (or redundant) functions.[13] Reader proteins bind specific modifications without further modifying them, writer enzymes install histone PTMs, and eraser enzymes remove them.[14,15] Furthermore, epigenetic proteins often form complex multiprotein units, which can impact gene expression through the recruitment of transcription factors, as well as influencing chromatin structure.[16,17]

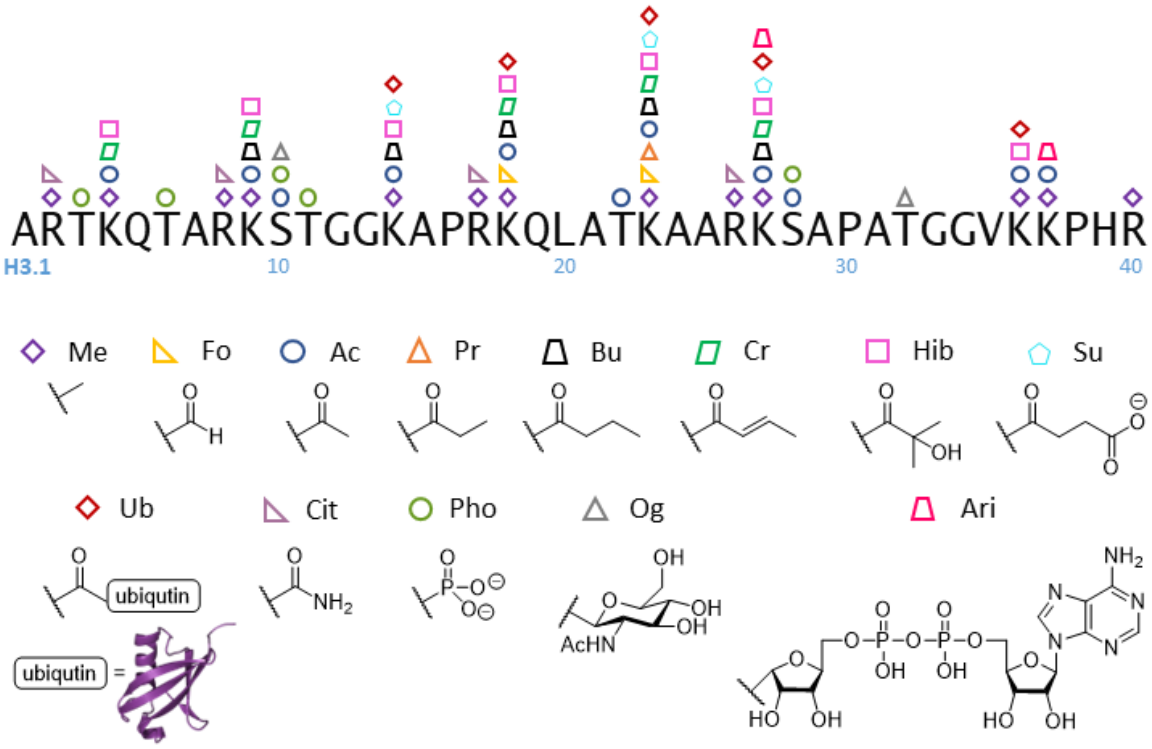


Figure 1.2 PTMs recorded in the *N*-terminal region of Histone H3.1 (amino acids 1–40).[11] Me = methylation, Fo = formylation, Ac = acetylation, Pr = propionylation, Bu = butyrylation, Cr = crotonylation, Hib = 2-hydroxyisobutyrylation, Su = succinylation, Ub = ubiquitination, Cit = citrullination, Pho = phosphorylation, Og = *O*-linked-*N*-acetylglucosaminylation, Ari = ADP ribosylation.

Further to histone modifications, epigenetic regulation can occur through DNA methylation, and subsequent oxidation of methylated DNA.[4] The 5-position of cytosine is most commonly modified.[18] DNA methylation is also regulated by reader, writer, and eraser proteins. DNA methyl transferases (DNMTs) are responsible for methylation of DNA, which can be inherited, or passively diluted during cell division. Methylcytosine oxidations are performed by ten eleven translocation enzymes (TETs), after which DNA methylation can be actively reversed through thymine DNA glycosylase (TDG)-dependent base excision repair (BER), or passively diluted.[19]

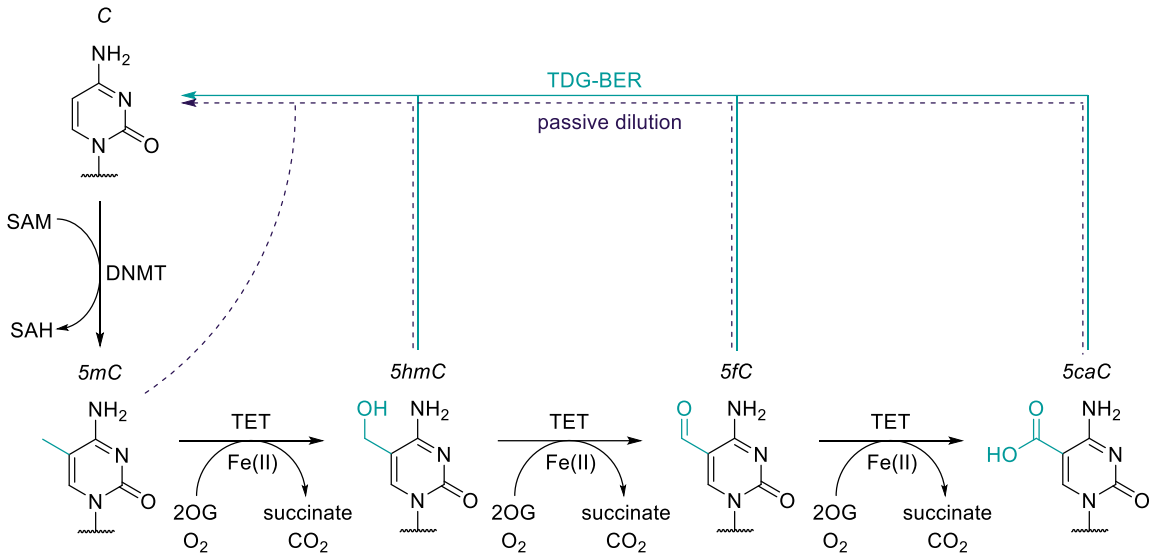


Figure 1.3 The 5-position of cytosine in DNA can be methylated and subsequently oxidised. Methylation is performed by S-adenosylmethionine (SAM)-dependent DNMTs. TETs catalyse oxidation of 5-methylcytosine (5mC), to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (cytosine modifications shown in teal). Removal of (oxidised) methyl groups in cytosine can be performed by passive dilution or actively through TDG-BER.[19]

1.1.2 2-oxyglutarate and Fe(II) dependent oxygenases in epigenetics

The 2-oxyglutarate (2OG) and Fe(II) dependent oxygenases (2OG-oxygenases) are a superfamily of proteins, which catalyse a variety of reactions.[20,21] Different 2OG-oxygenases have a wide substrate scope, oxidising DNA, RNA, proteins, and fatty acids.[21] In the context of epigenetics, 2OG-oxygenases can oxidise both methylated histones with histone demethylases (KDMs), and methylated DNA with TETs.[22] KDMs oxidise the N^{ε}-methyl group in methyllysine residues to form a hemiaminal intermediate, which fragments to form the demethylated lysine product and formaldehyde (**Figure 1.4**).[23] TETs oxidise 5-methylcytosine progressively, to give 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine (**Figure 1.3** & **Figure 1.6**).[19]

1.1.2.1 KDM4s

PTMs to lysine residues include mono-, di-, or tri-methylation.[24] Within histones, the context of lysine methylation impacts whether the PTM is more often associated with gene activation or silencing, and this can include the lysine residue methylated, and the level of methylation.[25] The reversal of lysine methylation in histones is performed by KDMs, and over 20 KDMs have been discovered to date. Furthermore, KDM family members have different (and sometimes

overlapping) substrate profiles with respect to specific lysine residues and methylation states. The KDMs can be split into nine subfamilies, KDM1–9, based upon their substrates, and sequence similarities.[26,27] KDMs use two mechanisms for demethylation, the KDM1 subfamily are flavin-dependent monoamine oxidases, while the KDM2–9 subfamilies are 2OG-dependent oxygenases (**Figure 1.4**).[26] 2OG-dependent KDMs all have Jumonji-C catalytic domains, and they are able to demethylate mono- di-, or tri-methylated lysine.[23] Furthermore, whilst the Jumonji-C KDMs are best characterised for demethylation of methyllysine in histones,[25] methylarginine demethylation activity has also been reported on peptides.[28]

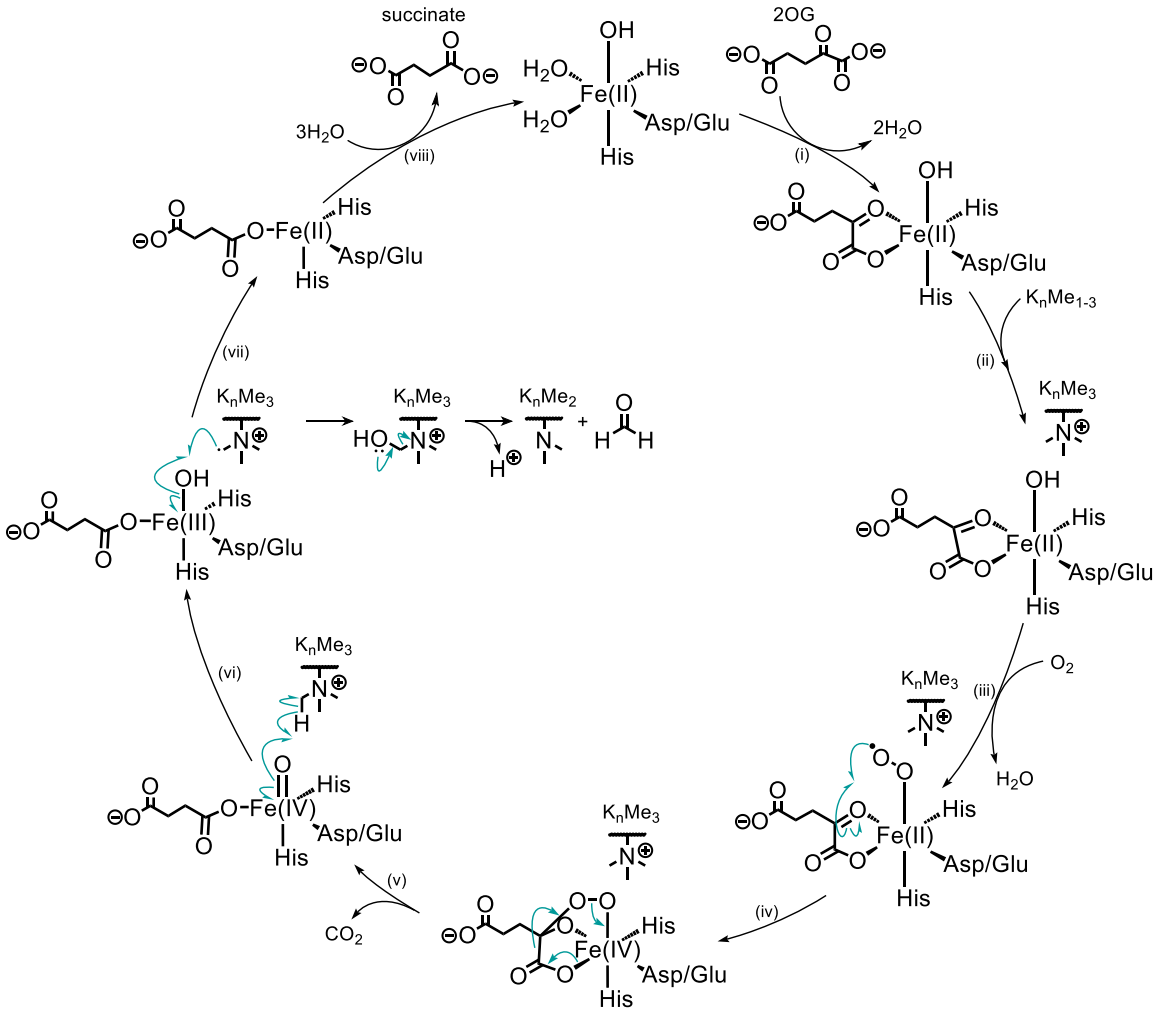


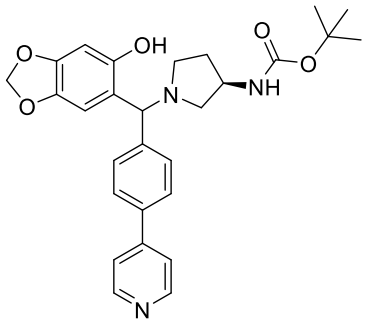
Figure 1.4 Proposed mechanism of methyllysine demethylation by Jumonji-C KDMs. (i) 2OG is bound by Fe(II) within the enzyme active site. (ii) The substrate enters the active site, (iii) followed by O_2. (iv) In complex with Fe(II), radical addition of O_2 to 2OG forms an unstable intermediate. (v) Decarboxylation of the unstable intermediate gives an Fe(IV)-oxo species, succinate, and CO_2. (vi) A radical rebound mechanism for oxidation of methyllysine likely occurs, in which the Fe(IV)-oxo species abstracts hydrogen from a methyl group, (vii) prior to addition of OH to the methyl radical. The hydroxymethyl group leaves (methyl)lysine as formaldehyde, resulting in demethylated lysine. (viii) In the final step, succinate is displaced with water in the enzyme active site.[23,29]

The KDM4 subfamily is split into six members, KDM4A–F.[30] Within the KDM4 family, KDM4A–D have been widely shown to be actively transcribed, KDM4E has recently been updated as an actively transcribed gene (UniProt accession number B2RXH2),[31] and KDM4F is currently thought to be a pseudogene.[30,32] KDM4A–E all demethylate di- and tri-methylated lysine 9 in histone H3 (H3K9$Me_{2/3}$), KDM4A–C also demethylate H3K36$Me_{2/3}$, and KDM4A–D demethylate H1.4K26Me_3.[33–35] All KDM4s share Jumonji-C and Jumonji-N domains, but KDM4A–C have additional reader domains: a double Tudor domain, and two plant homeodomains.[30]

Extensive research has been performed which links KDM4 overexpression and dysregulation with several cancers, including breast, lung, prostate, and colorectal cancers, and leukaemia.[33] Several mouse models have been used to investigate the role of KDM4s in development and cancer.[36] For example, the progression of acute myeloid leukaemia is perturbed in *Kdm4A–C* triple-knockout mice, but not in *Kdm4C* single-knockout mice. This shows that KDM4 activity is required for the survival of acute myeloid leukaemia cells, and demonstrates redundancy within the KDM4 family.[37] Additionally, the KDM4s have been implicated in cardiovascular disease and mental retardation, making them attractive therapeutic targets.[36]

Only a handful of selective inhibitors for the KDM4 subfamily, or members of the subfamily, are available (**Figure 1.5**).[36,38] Recently, a potent and selective KDM4D inhibitor was published by Fang *et al.* (2021),[39] compound **24s** (**Figure 1.5**), which was reported to have greater than 1000-fold selectivity for KDM4D over KDM4A. When colorectal cancer cells were treated with compound **24s**, reduced proliferation and migration was observed. A KDM4A–C selective macrocyclic peptide inhibitor was discovered using genetically encoded cyclic peptide selection technology, **CP2** (**Figure 1.5**). Structural studies demonstrated that arginine-6 in **CP2** is key for potent inhibition, and showed that it binds to the methyllysine binding site of KDM4A. Extensive **CP2** structure activity relationship studies yielded a cell active analogue, **CP2.3**, capable of increasing global H3K9/36Me_3 methylation in HeLa cells.[40] Celgene developed a KDM4 selective, cell active

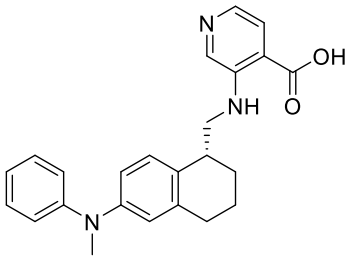
inhibitor, **QC6352** (**Figure 1.5**). **QC6352** has antiproliferative activity in cancer cell lines, and it achieves a minimum of 7-fold greater potency for KDM4 enzymes over KDM5s, however KDM5 activity in cells may still be contributing to the biological effect.[41] Cell active KDM4/5 selective inhibitors have also been developed, based upon two separate scaffolds: 3-amino-4-pyridine carboxylate derivatives (Compound **34**, **Figure 1.5**),[42] and 8-substituted pyrido[3,4-d]pyrimidin-4(3H)-one derivatives (compound **54j**, **Figure 1.5**).[43,44] Other KDM4 inhibitors have been reported, but many of them also inhibit several Jumonji-C KDMs, or even several 2OG-oxygenases.[23]



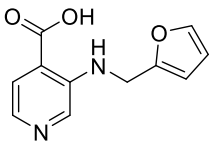
Compound **24s**
KDM4D selective
IC_{50}(KDM4D) = 23 nM
IC_{50}(KDM4A) = >100 μM



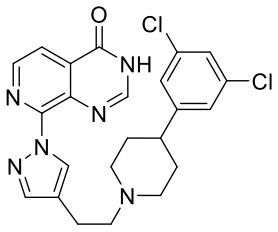
CP2
KDM4A-C selective
IC_{50}(KDM4A) = 42 nM
IC_{50}(KDM4B) = 33 nM
IC_{50}(KDM4C) = 39 nM
IC_{50}(KDM4D) = 6.3 μM
IC_{50}(KDM4E) = 9.2 μM



QC6352
KDM4 selective
IC_{50}(KDM4A) = 104 nM
IC_{50}(KDM4B) = 56 nM
IC_{50}(KDM4C) = 35 nM
IC_{50}(KDM4D) = 104 nM
IC_{50}(KDM5B) = 750 nM



Compound **34**
3-Amino-4-pyridine carboxylate derivative
KDM4/5 selective
IC_{50}(KDM4A) = 100 nM
IC_{50}(KDM4C) = 80 nM
IC_{50}(KDM4D) = 100 nM
IC_{50}(KDM4E) = 40 nM
IC_{50}(KDM5B) = 100 nM



Compound **54j**
8-substituted pyrido[3,4-d]pyrimidin-4(3H-one derivative
KDM4/5 selective
IC_{50}(KDM4A) = 80 nM
IC_{50}(KDM4B) = 17 nM
IC_{50}(KDM5B) = 14 nM
IC_{50}(KDM5C) = 51 nM

Figure 1.5 Structure, selectivity information, and IC_{50} data of some of the most selective KDM4 inhibitors.[39–44]

1.1.2.2 TETs

In humans there are three TET proteins, TET1–3, which oxidise methylcytosine and promote demethylation (**Figure 1.6**).[45–47] All TET proteins contain a catalytic double-stranded β-helix, and

cysteine rich domain at their *C*-terminus, and TET1/3 contain an additional CXXC domain at their *N*-terminus.[19] Truncated isoforms, without the *N*-terminal CXXC domain, of both TET1, and TET3, are also expressed.[48–50] TETs contribute to the control of gene expression through both their enzymatic role, and through interaction with transcriptional regulators.[51,52]

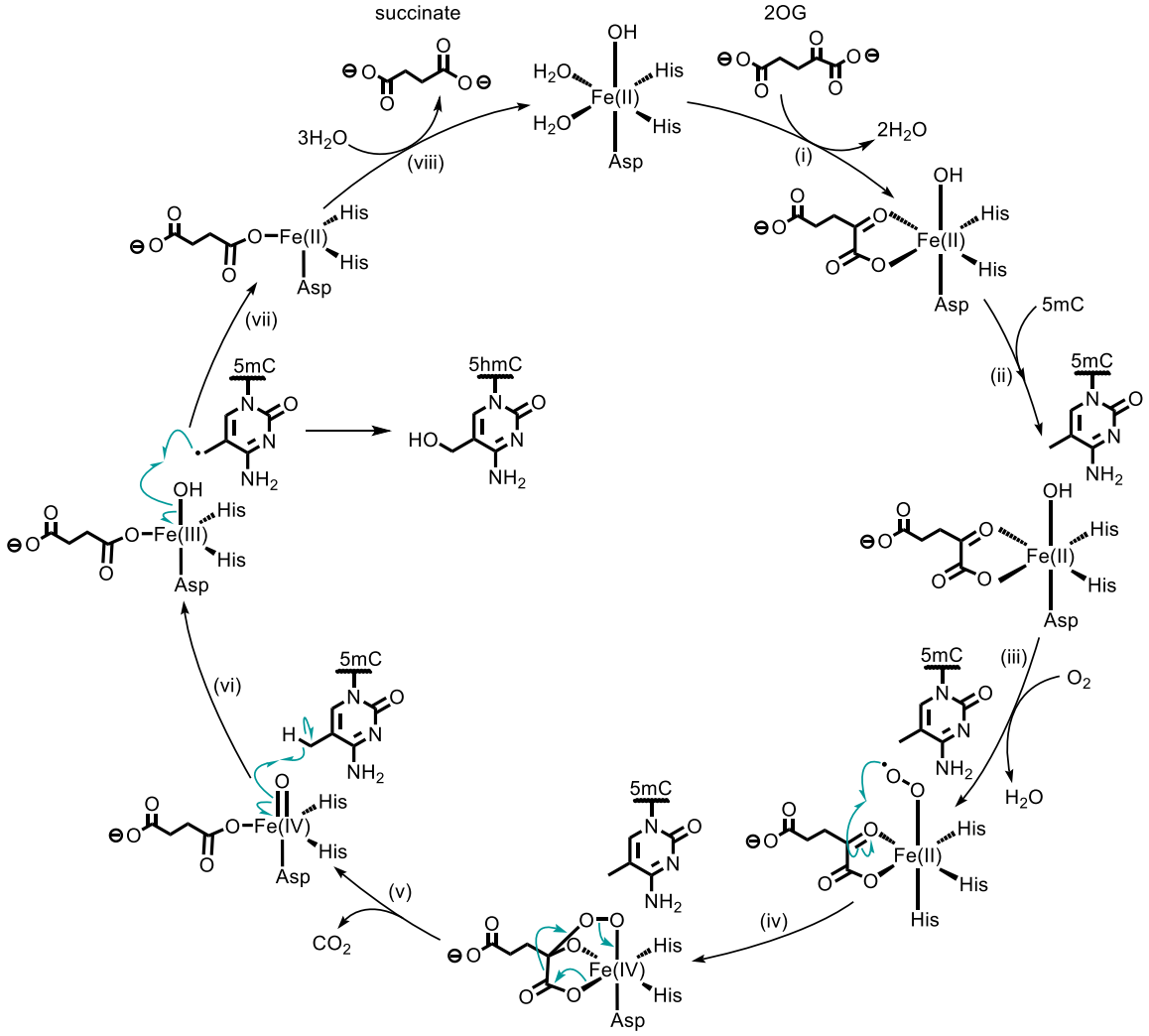


Figure 1.6 Proposed mechanism for the oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) by TETs. (i-viii) Oxidation of 5mC occurs *via* an analogous mechanism to Jumonji-C KDMs (**Figure 1.4**).[29,47]

Many cancers are associated with TETs, through both TET mutations (inactivating and not), and decreased expression.[46,53] Decreased 5-hydroxymethylcytosine levels are also associated with many cancers, but little is known about the contribution of TET enzymes, *versus* passive dilution through replication, in these changes.[46] Conversely, TET1 is overexpressed in MLL-rearranged leukaemia, and TET knockdown inhibits proliferation of MLL-rearranged leukaemia cells.[54] The

opposing role TETs play in different cancers makes them interesting probe targets, however, the prevalence of TET mutations in cancers may make them difficult to target using traditional inhibition-based strategies.[46]

The development of selective TET inhibitors has been challenging, and many inhibitors used are 2OG-mimics/Fe(II) chelators, which tend to have low affinity for TET enzymes.[55,56] However, a handful of TET-selective inhibitors have been reported (**Figure 1.7**). Possibly the most selective TET inhibitor found to date has been a macrocyclic peptide selective for TET1, discovered using genetically encoded cyclic peptide selection technology, **TiP1** (**Figure 1.7**).[55] **TiP1** is 7-fold selective for TET1 over TET2, however selectivity compared to other 2OG oxygenases was not determined. **TETi76** (**Figure 1.7**) is a TET specific analogue of 2OG, which was designed using a TET2-N-oxalylglycine (NOG, 2OG-mimic) co-crystal structure. **TETi76** can induce cell death in TET-dependent leukaemia cells, and inhibits TET1–3 in enzymatic assays. Furthermore, **TETi76** was not active when cross screened against 16 other 2OG-oxygenases using a single concentration.[57] Another TET selective inhibitor, compound **C35** (**Figure 1.7**), was identified using virtual screening, and was found to inhibit TET1–3, although cross screening against other 2OG-oxygenases was not performed. **C35** was able to suppress 5-hydroxymethylcytosine formation in human and mouse cells, and could promote somatic cell reprogramming (in which differentiated cells are reverted to pluripotent stem cells).[58] A cytosine-based TET inhibitor was also developed, **Bobcat339** (**Figure 1.7**), which was reported as TET1/2 selective inhibitor.[59] However a later study attributed the inhibitory activity to the presence of copper(II) contaminants in the compound.[60]

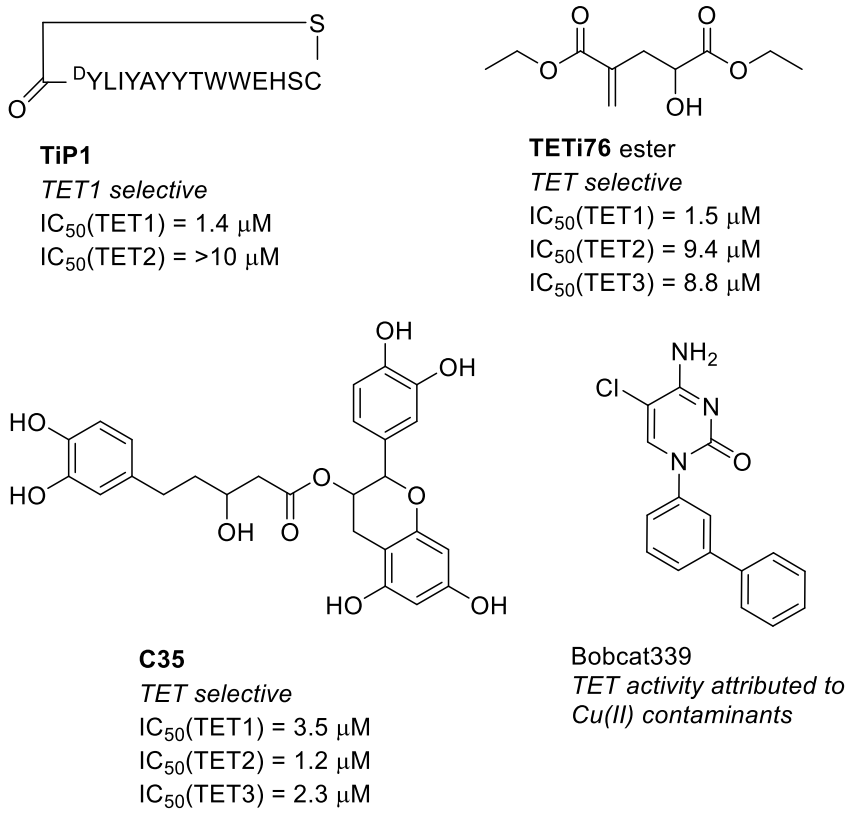


Figure 1.7 Structure, selectivity information, and IC_{50} data for some of the most selective TET inhibitors reported to date.[55,57–59]

1.2 Multifunctional probes in chemical biology

Probes are reagents which, upon binding to their target, can be used to learn about the function of the target, or its role in a disease setting.[61–64] The importance of probes has been acknowledged by many initiatives; such as Target 2035, which aims to develop a chemical and/or biological probe for every protein in the human proteome by 2035.[65] Many different types of molecules can be used as probes in chemical biology, for a variety of applications, including; small molecules,[64] peptides,[66] DNA/RNA,[67] and antibodies.[68] Multifunctional probes combine target engagement with at least one additional application/function, expanding the capabilities of the probe.[69] Furthermore, multifunctional probes offer alternative modes of action for the modulation of proteins compared to traditional small molecule inhibitors; this allows multifunctional probes to regulate an expanded proportion of the proteome, of which only ≈20% contains suitable small molecule binding sites for occupancy-based inhibition.[70]

1.2.1 Unmet need for selective epigenetic probes

During target validation, selective tool molecules are complementary to genetic knockdown/knockout techniques to learn about the role of a target in disease settings.[63,71,72] Indeed, for some epigenetic targets with established tool compounds, chemical inhibitors have entered the clinic, including histone deacetylases (HDACs), and DNMTs.[73] However, many epigenetic enzymes would benefit from more selective inhibitors or probes, to facilitate further characterisation, including for the KDM4 and TET families discussed in Section 1.1.2.[13,74]

The use of multifunctional tool compounds could widen the scope of target validation studies and expand the number of ligandable proteins (Section 1.2.2). Multifunctional probes may be particularly useful for epigenetic proteins, to study multiprotein complexes at chromatin modifications.[13] These studies will help us to learn more about the role of these enzymes,[75] and proteomics has been particularly important for identification of protein-protein interaction (PPI) partners.[76]

1.2.2 Selected categories of multifunctional probes

1.2.2.1 Affinity probes and activity-based protein profiling (ABPP)

Affinity probes are used for functional analysis of biological targets of interest, they contain a ligand which binds the target (*e.g.* chemical probe), connected to a reporter group (**Figure 1.8**).[77] Small molecules are often employed as affinity probes, and alkyne or azide handles may be used to allow the reporter group to be conjugated after probe-protein binding using click chemistry (Section 1.4).[78,79] Bio-compatible reactions, such as click chemistry, provide the opportunity of connecting the reporter group after binding to the target, keeping the binding kinetics of the affinity probe as similar to the parent molecule as possible.[80] Highly selective antibodies can also be used as affinity probes; they can either be conjugated to reporter groups directly, or secondary antibodies can be used to conjugate reporter groups after antibody binding.[81] Antibodies will be

expanded upon in Section 1.2.2.5, and small molecule based affinity probes will be the focus of this section.

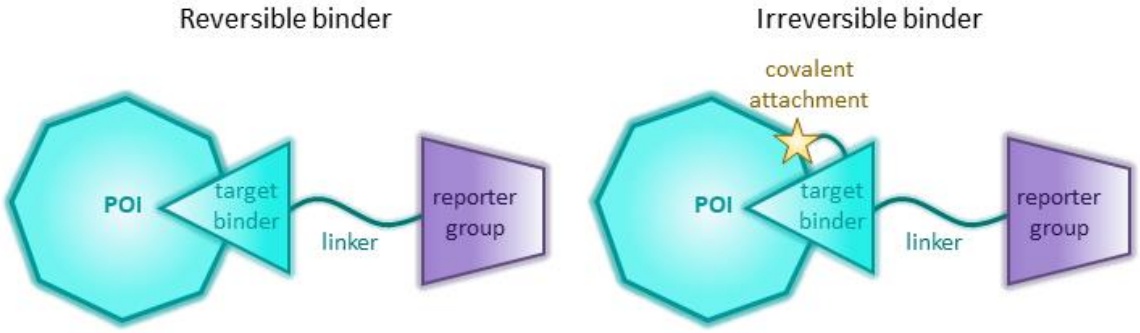


Figure 1.8 Affinity probes can bind reversibly or irreversibly. They contain both a target binder and reporter group. Bioorthogonal handles may be installed instead of a reporter group, allowing the reporter to be conjugated after probe-protein binding.[82]

Small molecule affinity probes can bind reversibly, but covalent attachment to the target is often preferred.[82] Covalent linkage ensures the probe and target remain engaged during sample analysis, which is especially helpful with small molecule affinity probes as they typically bind their targets less potently than other probe classes (Section 1.3).[83] Active site targeting affinity probes may be designed to covalently react with suitable amino acids in the active site, and these probes are often used for activity-based protein profiling (ABPP). Photoreactive groups can also be installed, which can react with any proximal group within a protein upon activation; when combined with an affinity group/handle photoreactive probes form photoaffinity probes. Photoaffinity probes are useful where covalent linkage is desired but suitable reactive amino acids are absent from the target binding site, or when structural information to elucidate probe-target interactions is unavailable.[84]

Affinity probes often use biotin as a reporter group, allowing streptavidin to be used to 'pull' the affinity probe, and proteins bound to the probe, out of a complex mixture.[76] Mass spectrometry (MS)-based proteomics can be used to profile unknown binding partners,[83] and western blotting can be used to confirm the presence of known/predicted binding partners.[85] Small molecule biotinylated-probes can also be used to map the probe's chromatin interactions, using a technique

called Chem-seq,[86] which works similarly to ChIP-seq.[87] Chromatin complexes bound to the biotinylated-probe are isolated, and next generation sequencing is used to profile the DNA associated.[88] Other reporter groups may also be used, such as fluorescent groups for imaging applications (Section 1.2.2.4). Affinity probes are expanded upon in Section 3.1, including examples of their utility to study epigenetic targets.

1.2.2.2 Protein dimerization probes

Bifunctional probes can induce PPIs (**Figure 1.9**A), either by stabilising known PPIs, or through inducing *de novo* interactions.[89] Protein dimerization probes may be formed of two chemical entities joined with a linker, or one chemical entity which can simultaneously bind two proteins, known as a molecular glue.

Many protein dimerization probes are derived from natural products; including cyclosporin A (**Figure 1.9**B), which binds calcineurin, and inhibits its activity through recruitment of another protein called cyclophilin.[90,91] However, synthetic dimerization probes have been developed both by design and serendipity.[92] Computer aided drug design identified compounds that could stabilise a PPI between thrombospondin-related anonymous protein (TRAP), and an aldolase from a malaria-causing parasite (*Plasmodium palcifarum*). The TRAP/aldolase stabilisers identified all contained an *N*-(benzylideneamino) benzamide ring system based scaffold (**Figure 1.9**B).[93] Another protein dimerization probe was identified through the serendipitous discovery that novel androgen receptor inhibitors were more potent inhibitors of bromodomain 4 (BRD4). Structural studies demonstrated that a BRD4 protein dimerizer had been discovered. Further structure-activity relationship studies led to the discovery of **biBET** (**Figure 1.9**B), which was reported as the most potent cellular probe for bromodomain and extraterminal domain (BET) proteins to date.[94] Proteolysis targeting chimeras (PROTACs) are perhaps the most common dimerization probes developed currently, which recruit E3 ligases to promote protein degradation.[89] Protein degraders

and PROTACs have evolved into a diverse field of research over the past 20 years,[95] and they are discussed in more detail below (Section 1.2.2.3.2).

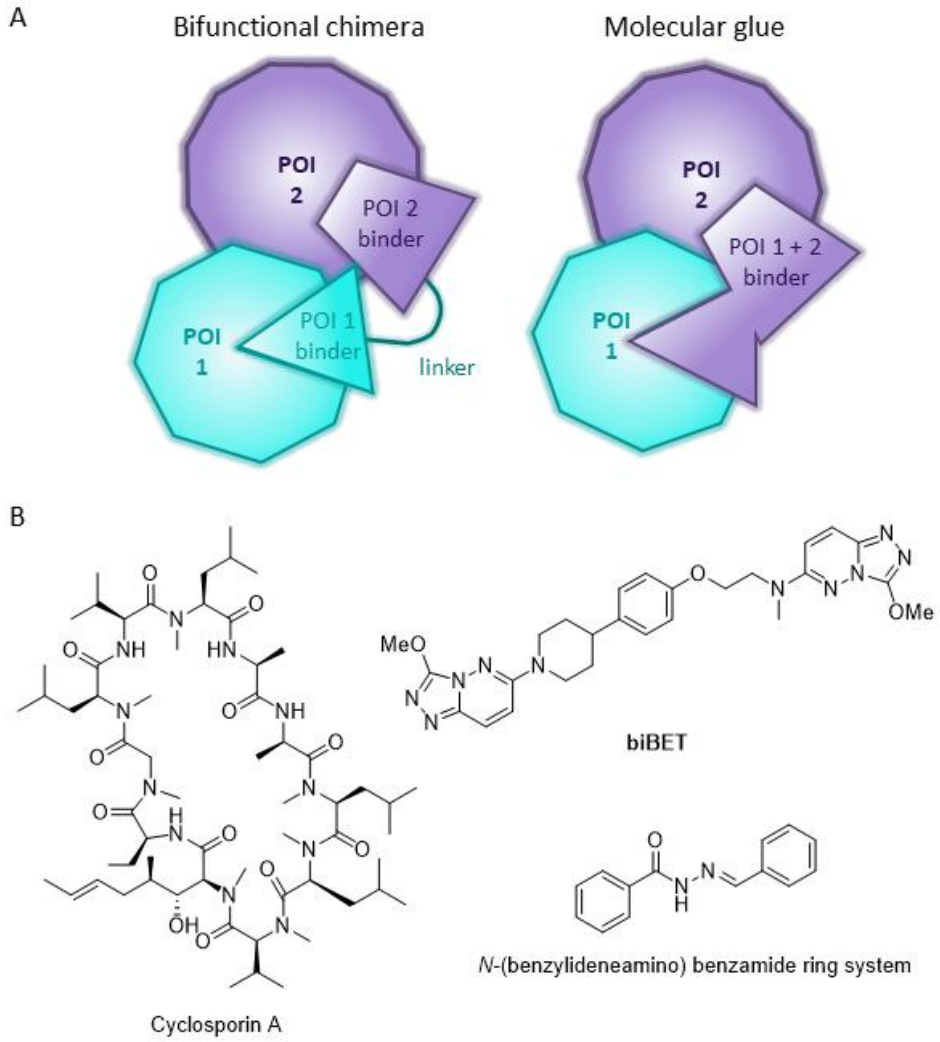


Figure 1.9 Protein dimerization probes can be formed from two molecules joined with a linker, or one chemical entity which induces the PPI.[89] (A) Schematic representation of protein dimerization probes. (B) Structures of the dimerization probes discussed.[90,93,94]

1.2.2.3 Protein degradation and PROTACs

1.2.2.3.1 Protein degradation

Protein homeostasis involves regulation of proteins, including: protein synthesis, protein folding, PTMs, protein trafficking, and eventually protein degradation. Many diseases are associated with improper protein homeostasis, including cancer and neurodegenerative diseases.[96,97] Protein degradation is modulated by two systems, the ubiquitin-proteasome system (UPS), and the

autophagy-lysosome pathway (ALP). Small molecules which can 'hijack' either pathway have been identified, including molecules targeting different degradation mechanisms within the ALP, although UPS modulators are the best studied.[98]

Within the UPS, proteins are marked for degradation through post translational modification with ubiquitin. A cascade is used to transfer ubiquitin to the target protein, initiated through ubiquitin-activating enzymes (E1s), which use ATP to form a thioester bond between a cysteine residue in the E1 and the *C*-terminus of ubiquitin. Next, the ubiquitin molecule is transferred from the E1 to an E2 ubiquitin-conjugating enzyme, using trans-thioesterification. In the final step, an E3 ubiquitin-ligase (E3 ligase) facilitates transfer of the ubiquitin molecule to a lysine side chain within the protein target (**Figure 1.10**A). Different E3 ligases use different mechanisms for ubiquitination of proteins; really interesting new gene (RING) E3 ligases facilitate E2 enzymes to directly transfer ubiquitin to the target, whereas homologous to E6AP *C*-terminus (HECT) E3 ligases covalently bind ubiquitin prior to ubiquitination of the target (**Figure 1.10**A).[99] The ubiquitin chain may later be extended, targeting lysine residues within the protein-conjugated ubiquitin for further ubiquitination. A 'ubiquitin code' is formed, which can signal different outcomes, and a polyubiquitin chain linked through ubiquitin lysine-48 residues often signals for protein degradation *via* the UPS.[100] Protein degradation most commonly occurs by the 26S proteasome, which is a complex multiprotein structure. The 26S proteasome is formed of two sub-complexes, the 20S core particle, and the 19S regulatory particle (**Figure 1.10**B).[101] The 20S core particle has a barrel shaped structure, inside which protein degradation takes place by multiple proteases.[102] Docking of either the 19S regulatory particle, or alternative regulators, helps coordinate the opening and closing of the 'gate' of the proteasome core particle. Additionally, the 19S regulatory particle is responsible for protein identification, binding, deubiquitination, protein unfolding, and translocation of the substrates into the core particle.[101]

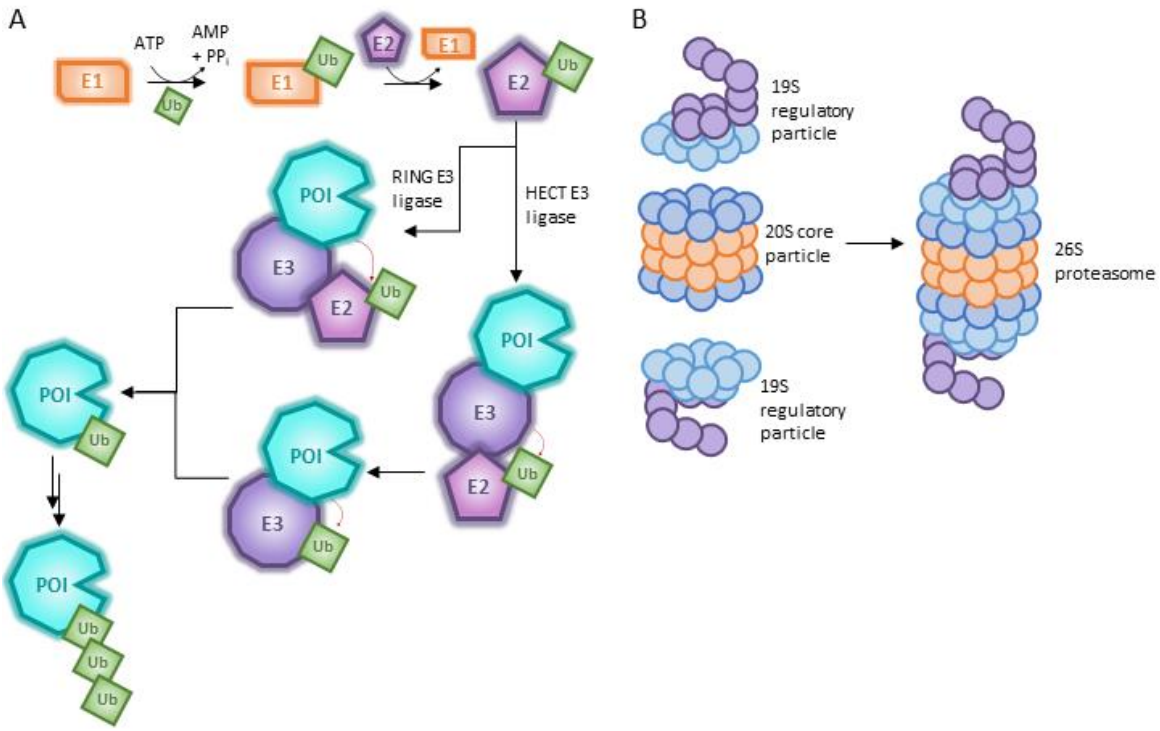


Figure 1.10 Components of the UPS degradation machinery. (A) Ubiquitin is transferred to a protein through an ATP-dependent enzyme cascade. Initially ubiquitin-activating E1 enzymes bind ubiquitin, which is subsequently transferred to an E2 ubiquitin-conjugating enzyme. E3 ligases regulate ubiquitination of the protein of interest (POI), either by bringing the E2 into proximity of the POI to facilitate the transfer (RING E3 ligases), or by binding the ubiquitin itself prior to transfer (HECT E3 ligases).[99] (B) The 26S proteasome degrades polyubiquitinylated proteins, it is a multiprotein structure formed of a 20S core particle capped by two 19S regulatory particles. Protein degradation occurs in the pore in the 20S particle, and proteins entering the proteasome are regulated by the 19S regulatory particle.[102]

1.2.2.3.2 PROTACs

PROTACs contain a protein of interest (POI) binder, and an E3 ligase binder, connected by a linker. A ternary complex is formed between the POI, PROTAC, and E3 ligase; this brings the ubiquitination machinery into proximity of the POI, and the POI is ubiquitinated. The ubiquitin chain is extended, marking the POI of interest for degradation.[103] Provided the PROTAC binds the POI reversibly, the process is catalytic, and the PROTAC and E3 ligase can be released upon binding the proteasome complex, allowing them to degrade further POI molecules.[104]

The first PROTAC was reported in 2001,[105] and a small number PROTAC of publications were released per year for 15 years, prior to exponential growth in the number of PROTAC publications recently (2016–2021).[106] The rapid growth in PROTAC research output in recent years has been facilitated by many discoveries, including the advent of small molecule E3 ligase binders

(Section 4.1.1),[107] and a report of *in vivo* protein knockdown using PROTACs.[108] More recently protein degradation was expanded beyond the UPS, with the discovery of ALP protein degraders using two separate mechanisms: lysosome targeting chimeras (LYTACs),[109] and macroautophagy degradation targeting chimeras.[110–112]

1.2.2.4 Probes as imaging tools

Imaging probes have many applications, including for cellular imaging, diagnostic medicine, and identifying the distribution of drug compounds in cells or *in vivo*.[113,114] Multifunctional probes are often used as imaging agents, generally by linking a target selective motif and an imaging readout (**Figure 1.11**).[115]



Figure 1.11 Imaging probes comprise of a target binding motif and an imaging readout. Reporter groups may include a fluorescent group (left) or radioactive isotope (right).[115,116]

To image a specific cellular target, a selective chemical probe against the target can be conjugated to a fluorescent group. This approach has been particularly successful for live-cell imaging, for example, fluorescent tubulin and actin probes were used for live cell imaging of cytoskeletal dynamics in neurons.[117] However, extensive optimisation is necessary for each imaging agent, and many parameters come into play, including: cell permeability, solubility, stability, brightness, and emission wavelength.[118] An alternative approach is to express the target either directly with a fluorescent protein, or with an epitope tag. Many epitope-directed imaging probes are available, *e.g.* fluorescent HaloTag ligands,[119] and these fusion proteins can be endogenously expressed using relatively recently developed genome editing technologies, such as CRISPR/Cas9.[120] Furthermore, fusion protein expression has enabled bioluminescence resonance energy transfer (BRET) assays to be developed, which can interrogate PPIs, giving a more detailed understanding

of the role of the proteins than imaging them alone.[121] Conversely, fluorescent proteins and epitope tags are often large (>20 kDa), and may alter the behaviour of the target protein.[118] Antibodies are also extensively employed as imaging agents, usually on fixed cells, and these will be discussed in Section 1.2.2.5.

Imaging may also be used to assess cell permeability of a parent compound, for which a fluorescent reporter group is often conjugated to the parent probe. However addition of the fluorescent group can significantly alter probe properties and lead to misleading results.[122] Alternatively, bioorthogonal chemistry can be used to tether the fluorescent group after probe-target binding, which should have lower impact on the probe properties (Section 1.4).[123]

Radioactive isotopes can be installed in probes for quantitative positron emission tomography (PET) on live organisms. PET can be used to investigate biodistribution of drug molecules, or for diagnostic imaging. Installing PET active radioisotopes (such as $^{11}C/^{18}F$) causes little change to the parent probe, and so minimally impacts the pharmacokinetics.[124]

1.2.2.5 Antibodies as multifunctional probes

Antibodies are large proteins, made up of multiple subunits, usually forming a 'Y' shaped structure (**Figure 1.12**). B cells generate antibodies against disease-generated antigens as part of the immune response to fight infection.[125] For research purposes, antibodies are often generated against a POI by immunising animals, and both polyclonal mixtures of antibodies, and monoclonal, single construct antibodies can be produced using this technique.[126] However, further structural engineering may be necessary to produce highly specific antibodies, especially for antibody therapeutics.[127]

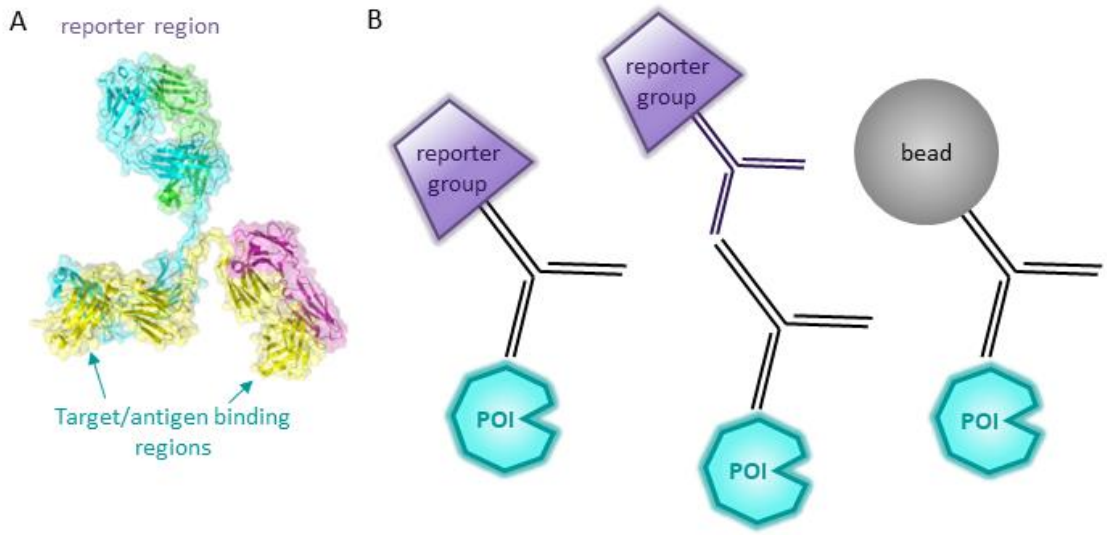


Figure 1.12 Antibodies are useful as multifunctional probes. (A) Crystal structure of an antibody (immunoglobulin), demonstrating the typical 'Y' structure. The antigen/target binding regions, and the reporter region are labelled. (PDB 1IGT).[128] (B) Antibodies are often used to analyse a protein of interest (POI). Both reporter groups (directly / *via* a secondary antibody), or solid supports are frequently used.[129]

Antibodies can be used as multifunctional probes in a variety of settings, such as immunoassays. Conjugating reporter groups to antibodies allows quantification of the amount of antibody bound to the target, and reporters may include: fluorescence, chemiluminescence, or occasionally radioactive isotopes. Reporter groups can be conjugated to the antibody directly, but greater flexibility is achieved when reporters are introduced using secondary antibodies.[130] Antibody-based imaging is used in western blotting to identify the presence of a target protein in complex mixtures.[131] Analogously, antibodies are used in immunofluorescence experiments to quantify the amount of their targets in fixed cells or tissues.[129] Immunoprecipitation experiments use antibodies to affinity purify their targets, and binding partners, out of a complex mixture; target-specific antibodies are immobilised on solid support to allow the target to be 'fished' out of the mixture. Coupled with MS-based proteomics, immunoprecipitation can be used to identify interacting proteins; alternatively immunoprecipitation can be coupled with next generation sequencing to identify genomic DNA interacting with the target (ChIP-seq).[87,132] Immunoprecipitation is often used in target validation within drug discovery, and here the specificity of the antibody is extremely important to ensure accuracy of the results.[130]

1.2.3 Characteristics of multifunctional probes for different applications

Each multifunctional probe has features which adapt it for its' purpose, which can include the protein region targeted and the linkage used for the introduction of additional functionality.

1.2.3.1 Impact of binding-orientation on multifunctional probes

The binding site, and binding kinetics, of multifunctional probes is important in determining the stability of the probe-protein complex formed,[133] and the accessibility of reporter groups.[134] Certain applications require probes which target a specific binding site, and structural information may be required to achieve this, such as from co-crystal structures, biophysical studies, or from known protein interactors.[135] Antibodies can offer an alternative solution for selective binding to a desired protein region, or even to a PTM. To generate antibodies which bind to a specific region of a protein, a truncated peptide antigen spanning only the desired section of the target is used for immunisation. Even so, achieving selectivity for the modified target over the unmodified target is challenging, and further antibody engineering is ideal.[136]

Non-inhibitory, allosteric site binding probes can be used for target-substrate capture studies, to elucidate the role of a target whilst in complex with its substrate. An exemplar study developed a non-competitive, and non-inhibitory biotinylated macrocyclic peptide binder of prolyl hydroxylase domain-containing protein 2 (PHD2) (**3CBt**, **Figure 1.13**), which was used to capture PHD2 bound to hypoxia-inducible factor 1-α (HIF1-α) in cell lysates. **3CBt** is much smaller than alternative antibody reagents (**3CBt** ≈ 3 kDa, antibody ≈ 150 kDa), and thus should have less impact upon the PPIs captured.[85] Other multifunctional probes require active site binding, such as the photoreactive histone H3K4Me$_3$ substrates used to identify novel histone readers and erasers (**Figure 1.13**).[137] Conversely, for protein degradation, the specific protein site targeted is not important for activity, which allows any ligandable site to be used for PROTAC development.[104] However, multiple linker attachment points, and linker lengths are often tested in PROTACs to find degraders which form a cooperative ternary complex.[133]

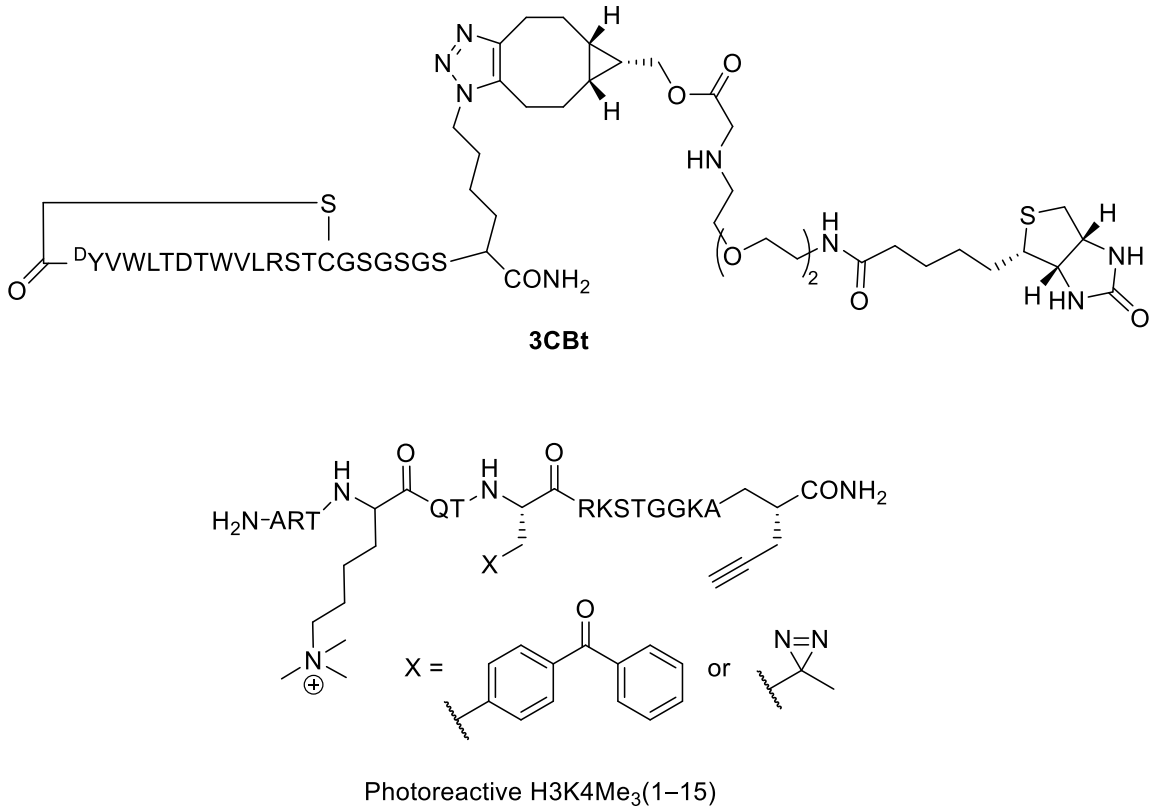


Figure 1.13 Structure of a non-competitive biotinylated peptide against PHD2 (3CBt),[85] and a photoreactive histone peptide probe (H3K4Me$_3$).[137]

1.2.3.2 Impact of linkers in multifunctional probes

A linker is often used to conjugate different molecules together to form a multifunctional probe. The length and composition of the linker can have profound impacts on the activity of the resulting probe, and this area of research is receiving increasing interest.[138] There are many examples across the multifunctional probe spectrum which demonstrate the importance of linkers, a selection of which is highlighted here.

Within protein degradation, optimisation of linker length is often crucial.[139] Recently, a systematic analysis of the impact of linkers on both degradation and permeability of PROTACs was performed by Hendrick *et al.* (2022),[140] in a 'direct-to-biology' approach; 91 linkers were selected, and two PROTAC series utilising different CRBN ligands were synthesised without chromatographic purification. Interestingly, linker trends differed between the two series despite targeting the same E3 ligase, suggesting that independent linker optimisation may be needed for every PROTAC

series. PROTAC linkers have been used to improve target selectivity of the parent small molecules used in PROTAC development. This has been widely reported for **MZ1** (**Figure 1.14**), which selectively degrades BRD4 but was developed from a pan-BRD inhibitor (JQ1).[141] Linker effects can be extremely subtle, as is the case for mitogen-activated protein kinase isoform selective PROTACs developed by Smith *et al.* (2019),[133] which were isoform selective for p38α/p38δ when using 13/10 atom linkers respectively, but did not maintain isoform selectivity when linkers one atom shorter/longer were tested (**SJFα**, selective for p38α in **Figure 1.14**).

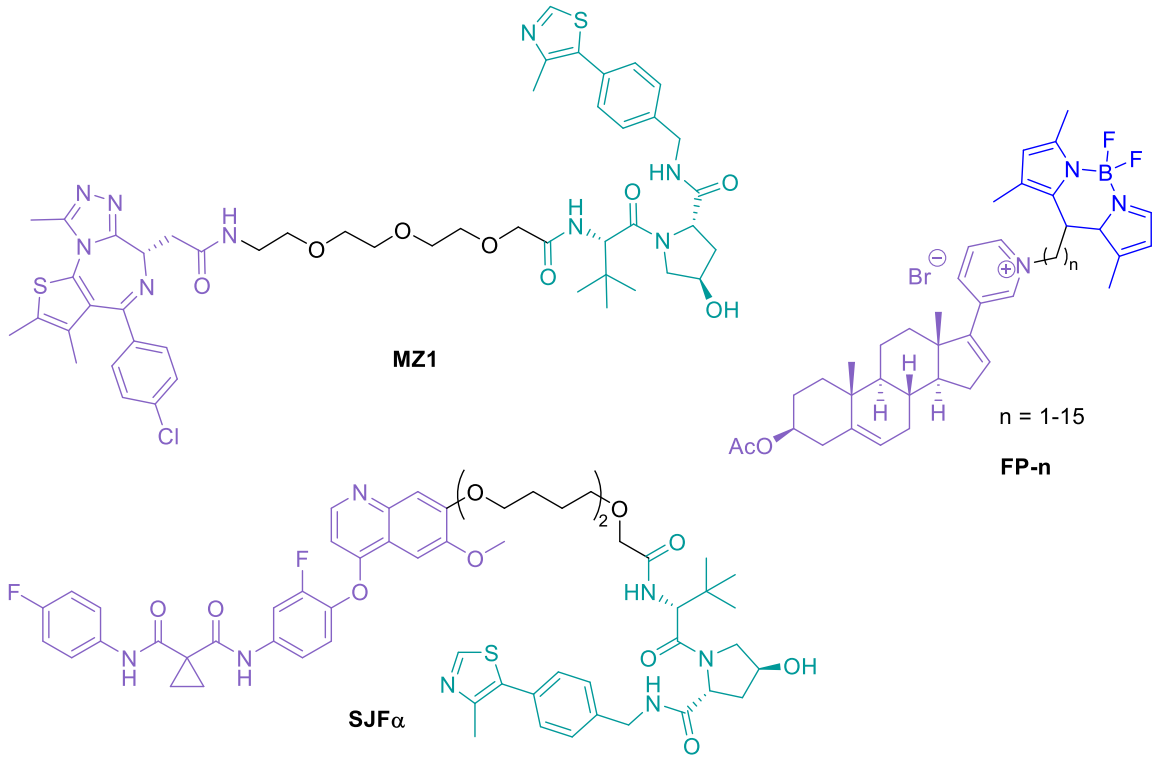


Figure 1.14 Structures of PROTACs MZ1[141] & SFJα,[133] and sterol imaging probes, FP-n.[142] Target binding region of molecules are coloured purple, VHL E3 ligase binders in teal, and fluorescent moiety in blue.

For fluorescent imaging probes linker length is also important, as the linker must be sufficiently long such that the fluorescent group does not sterically hinder the warhead.[138] Linker length was found to impact cell permeability, cellular localisation, and the kinetics of cholesterol labelling for a fluorescent sterol probe (**FP-n**, **Figure 1.14**).[142]

Multifunctional probes can be used for drug delivery to a cell/tissue, often by conjugating the cargo molecule to cell penetrating peptides (CPPs), tumour homing peptides, or antibodies. In

these scenarios, cleavable linkers, such as disulfides, may be exploited to release the cargo upon delivery. In particular, antibody-drug conjugate linkers require fine tuning for timely release of the drug molecule at the antibody binding site, while maintaining a stable linkage until antibody binding occurs. The linker composition, and position can have an impact on whether the antibody-drug conjugate is successful.[143]

However, not all multifunctional probes include a linker. Photoreactive probes are often developed without a linker to the photoreactive group, instead modifying the parent molecule at a position which does not significantly impact target binding.[144] Furthermore, molecular glues are a single small molecule entity which join two proteins together (Section 1.2.2.2). Molecular glues are attractive therapeutic agents as they tend to fit within Lipinski's rule-of-five for cell permeability.[145] However, they are challenging to design and the majority of molecular glues known to date were discovered serendipitously.[146]

1.3 Cyclic peptides as chemical probes

Probes in chemical biology come in a wide range of sizes, from small molecule fragments (<300 Da), to multi-subunit antibodies (≈150 kDa), and each has advantages and disadvantages. Peptides sit between small molecules and biologicals (such as antibodies) in terms of their molecular weight, selectivity, potency, and other properties (**Figure 1.15**).[147,148] Peptide probes can be found in nature, such as the hormone oxytocin which can be used to induce labour.[149] Peptide probes may also be inspired by nature, often by truncating the PPI portion of a protein, such as stapled peptides which are commonly used to mimic α-helix containing PPIs.[150] Furthermore, *de novo* peptide probes may be discovered using genetically encoded screening technologies, or through synthetic library screening.[149] Cyclic peptides can offer advantages over linear peptides, with increased stability, and potency, through a lower entropic binding penalty resulting from lower conformational freedom.[147]

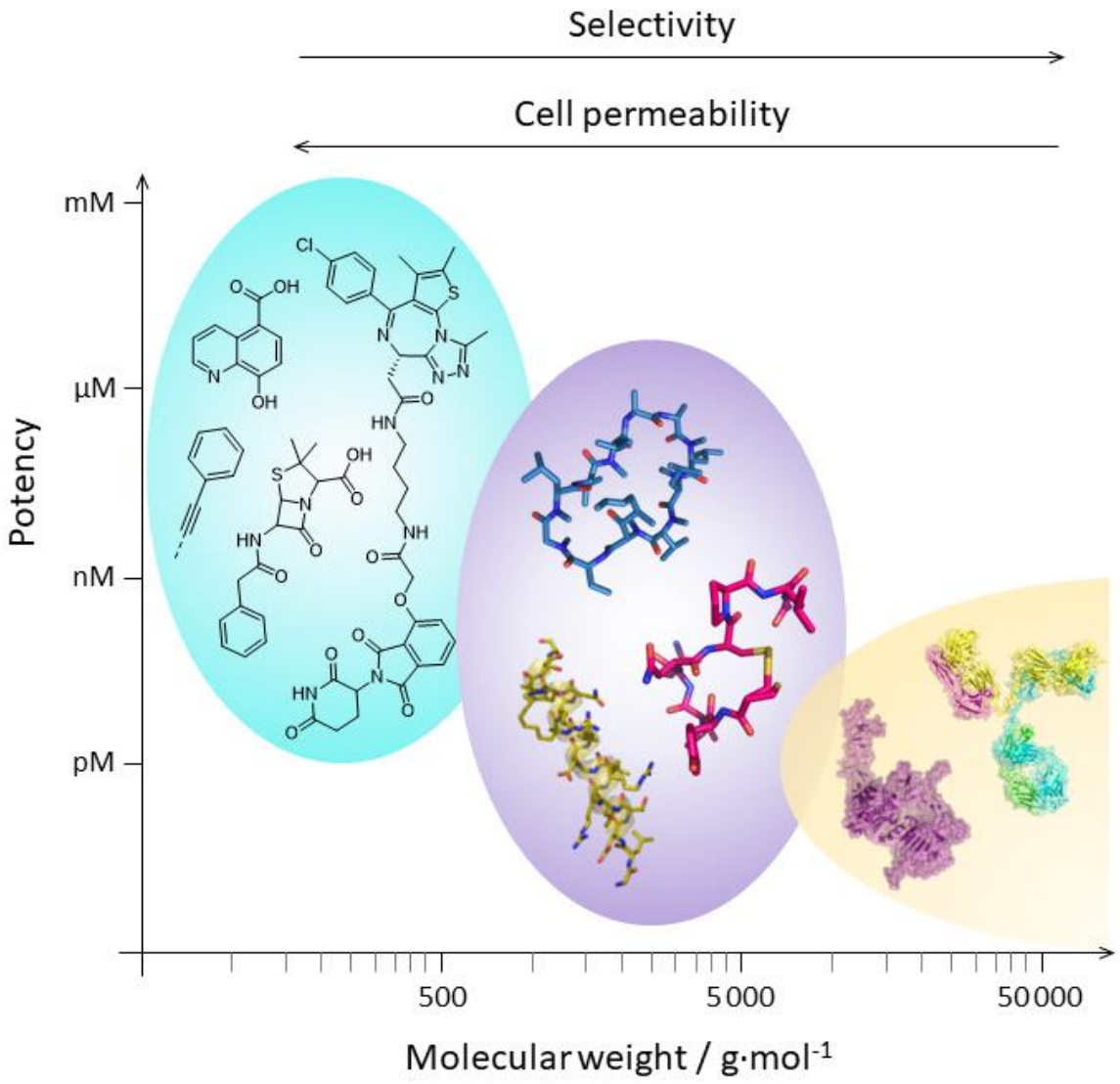


Figure 1.15 Peptides fall between small molecules and biologics (including antibodies) in terms of many properties. Peptides are larger than small molecules, giving them a greater interaction surface area, which often results in improved selectivity. Peptides are also smaller than antibodies, making them more likely to be cell permeabile.[147] Small molecules: a poised fragment,[151] IOX1,[152] penicillin G,[153] dBET1.[154] Peptides displayed: oxytocin (PDB 1XY2),[155] cyclosporin A (PDB 1CWA),[156] MCL (PDB 3MK8).[157] Biologics: HER2 antibody (PDB 6J71),[158] immunoglobulin (PDB 1IGT).[128]

Small molecules are attractive therapeutics, partly because they are often orally available, and they are also extensively used in target validation.[64] Small molecules have proven to be extremely useful tools for targeting proteins with smaller pockets of roughly 300–1000 Å^2, often termed druggable sites. However, they have suffered a lack of selectivity when targeting larger sites, often termed undruggable sites (including PPIs), generally covering surface areas of 1500–3000 Å^2.[159] Furthermore, off-target activity is a common issue with small molecule chemical probes, partly because cross screening is frequently limited to known similar proteins.[64] To combat these

problems, larger reagents (such as peptides) may be preferrable for targeting PPIs.[66] Antibodies are even larger than peptides, and can offer superior selectivity for their targets. However, highly selective antibodies often require engineering after antibody discovery, which requires expertise, and adds to the expense of antibody discovery.[136] As a result, many antibodies on the market have specificity and reproducibility issues.[160] One study tested over 200 commercially available histone modification antibodies to find that more than 25% were not specific.[161] Conversely, peptide screening, synthesis, and development is relatively quick compared to antibody development.[147,162]

Peptide therapeutics offer high potency and selectivity, as discussed above; this reduces off-target effects, and thus minimises side effects.[163] Another advantage of peptide drugs is that the degradation products tend to have low toxicity, which is a major concern with small molecule drugs.[149] However, peptides are susceptible to degradation by proteases, and can suffer from high hepatic/renal clearance.[164] Intracellular targets can be challenging for peptides, as the majority of peptides have low cell permeability. But there are bioactive peptides, such as cyclosporin A, which have inspired strategies to improve membrane permeability, as well as improve proteolytic stability.[149] These strategies often use non-canonical amino acids (such as D- and backbone *N*-methyl amino acids), and peptide cyclisation.[149,163] Additionally, cyclisation of peptides can shield hydrogen bond donors and acceptors, by encouraging intramolecular interactions, which can improve cell permeability.[165] Another strategy commonly employed involves conjugating CPPs, however this strategy has not seen much success in the clinic so far.[166] Conjugation of polyethylene glycol (PEG) units to peptides has seen success at increasing clearance times in approved peptide drugs, through increasing the molecular weight of the molecule; although this strategy is best for cell surface targets where cell penetration is not required.[163] Furthermore, the low proteolytic stability of peptides means they frequently need to injected, as with antibodies.

The ease, and repeatability of peptide synthesis is a key advantage of peptide probes.[147] The development of solid-phase peptide synthesis (SPPS) using 9-fluorenylmethoxycarbonyl (Fmoc) and *tert*-butyl (*t*Bu) chemistry facilitated the development of automated synthesisers, which reduced the time required for peptide synthesis from weeks to hours.[147,167] Chemical synthesis minimises batch-to-batch variability, one of the major problems with antibody probes,[160] and the speed of automated SPPS facilitates rapid structure activity relationship optimisation.[168] The potential structural diversity of natural peptides is huge, with 20 canonical amino acids, and variable peptide length. Furthermore, including non-canonical amino acids in peptides expands the structural diversity even further.[168] Some genetically encoded peptide discovery technologies allow the inclusion of non-coding amino acids at the peptide discovery stage, facilitating rapid discovery of diverse, potent, and selective peptides.[162,169]

1.4 Click chemistry for rapid functionalisation of molecules

Click chemistry enables stereospecific and regiospecific conjugation of modular blocks in high yielding biocompatible reactions that can be performed on both small, and large scales. Click reactions must have: a wide substrate scope, reagents that are stable in water, reagents that are stable in the presence of oxygen, and byproducts should be non-toxic.[170] Many functional groups can be used in click reactions, including: azide, alkyne, alkene, and thiol (**Figure 1.16**).[171] Click chemistry was awarded the 2022 Nobel Prize in Chemistry for bringing about "the era of functionalism",[172] highlighting how important multifunctional probes are for scientific research (Section 1.2).

1,3-Dipolar cycloadditions have been the most widely used click reactions, but a range of chemistries have been developed within the click reaction criteria that can be used to ligate two or more functional groups together (**Figure 1.16**).[171,173] Click chemistry can be particularly useful in the synthesis of multifunctional probes, as it is a reliable method to connect separate chemical entities, and click reactions have been extensively studied since they were first proposed by Sharpless in 2001.[170,174] Examples of the utility of click reactions can be found in the synthesis of

peptide conjugates,[171] PROTAC libraries,[175] imaging,[116] protein labelling,[123] for drug discovery,[176,177] as well as in materials science.[178]

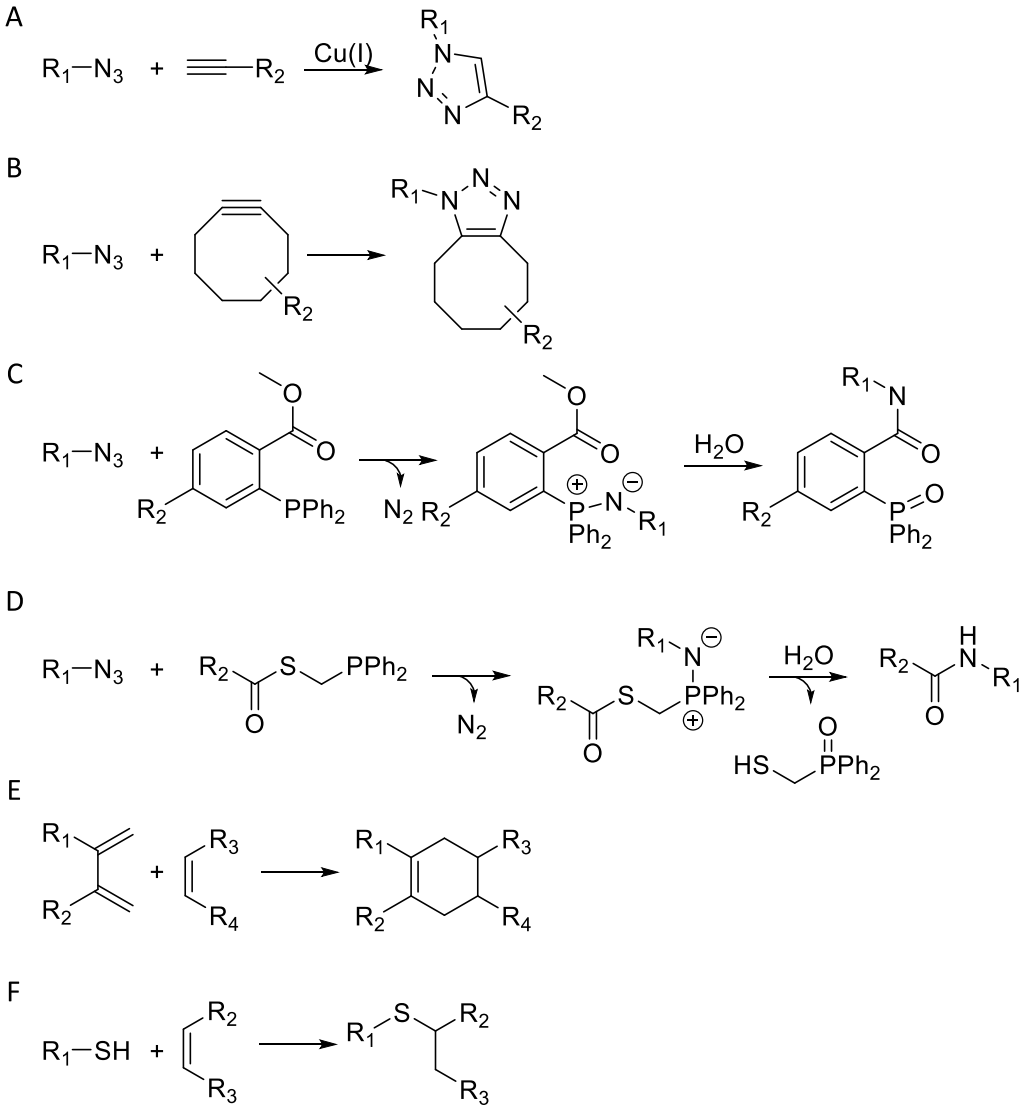


Figure 1.16 A number of click reactions are displayed which functionalise molecules using a range of linkages. (A) Cu(I)-catalysed 1,3-dipolar azide-alkyne cycloaddition (CuAAC) reactions form a 1,4-disubstituted 1,2,3-triazole.[179] (B) Strain-promoted azide-alkyne cycloaddition (SPAAC) reactions use strained cyclooctynes, allowing the cycloaddition to occur without a catalyst, and a 1,2,3-triazole is also formed.[180] (C) The Staudinger ligation connects chemical groups across a phenyl ring, and loss of N_2 is an entropic driving force for this reaction.[181] (D) The traceless Staudinger ligation connects two functional groups across an amide bond in a bioorthogonal reaction. Loss of phosphine oxide allows a two atom linker to hold the molecules together, and N_2 is also an entropic driving force for this reaction.[182] (E) Diels Alder [4 + 2] cycloadditions can be used to connect up to four separate functional groups, from two molecules, across a cyclohexene.[183,184] (F) Thiol-ene chemistry can be used to functionalise thiol groups, and thus is particularly useful for functionalisation of cysteine.[185]

1.4.1 Selected click chemistry examples

Click chemistry has been widely applied for the synthesis of peptide conjugates.[171] 'Clickable' groups are often introduced to peptides during SPPS, and azides are commonly used as they are

reagents in multiple click reactions (**Figure 1.16**), including: Cu(I)-catalysed azide-alkyne cycloaddition (CuAAC), strain-promoted azide-alkyne cycloaddition (SPAAC), and the (traceless) Staudinger ligation.[171] Peptide conjugates formed with click chemistry have linked functionalised peptides with wide ranging groups, including: biological molecules, small molecules, polymers, and surfaces.[171] A double CuAAC click chemistry approach has been developed to synthesise stapled peptides from two components, in which cyclisation is used to stabilise α-helical peptide conformations.[186] The double-click approach allowed a library of mouse double minute 2 homolog (MDM2)-binding peptides with varying linkers to be synthesised, cellular uptake, and activity of the resulting peptides was assessed (**Figure 1.17**A). A total of 19 stapled peptides were synthesised, and increased activity was found when small molecule cell delivery agents, or peptide nuclear localisation signals were included in the peptide staple.[187] For the synthesis of peptide conjugates containing ^{18}F for PET imaging, SPAAC was preferred to CuAAC. The resultant probes could be used for imaging directly (**Figure 1.17**B), without purification, whereas Cu from CuAAC reactions would need to be removed prior to *in vivo* imaging.[188] The thiol-ene reaction is also commonly used in the synthesis of peptide conjugates, it is initiated using UV light, and can occur on cysteine, a naturally occurring amino acid. However, greater fine-tuning may be needed for thiol-ene chemistry than other click reactions to avoid issues with alkene polymerisation, or disulfide bond formation, which would prevent formation of the desired peptide conjugate.[171] Thiol-ene chemistry was used to produce highly affordable peptide vaccine candidates, by making peptide-lipid conjugates (**Figure 1.17**C), in which the lipid enhances the immune response. Bioactivity studies demonstrated upregulation of cluster of differentiation 80 protein within human monocytes in fresh blood samples for two of three peptide-lipid conjugates tested.[189]

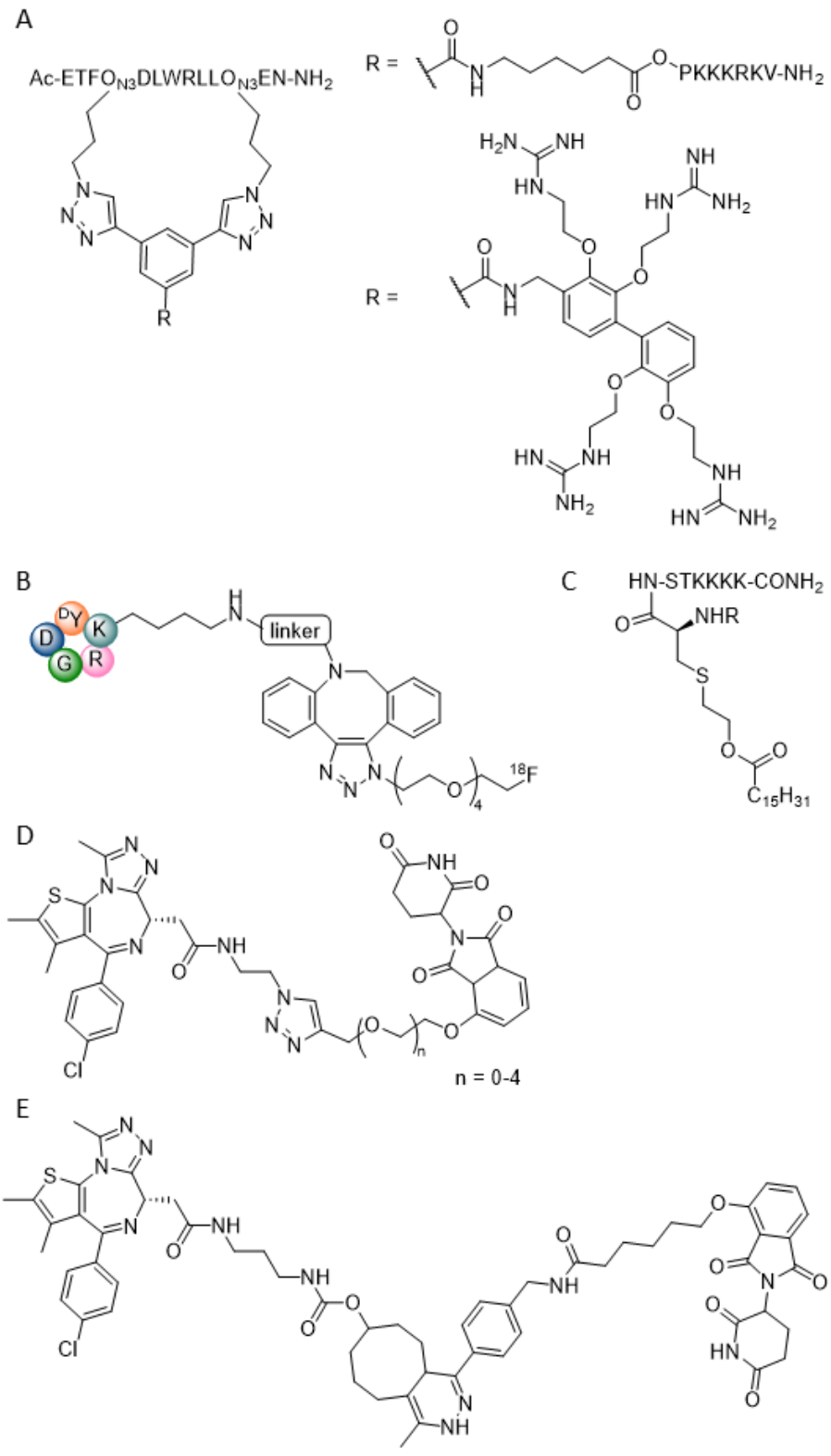


Figure 1.17 Structures of compounds synthesised using click chemistry described. (A) MDM2-binding peptides were stapled using a double CuAAC click chemistry approach. Cell activity of the peptide was improved by including cell delivery agents in the staple.[187] (B) SPAAC was used to form a ^{18}F labelled peptide for PET imaging, which could be safely injected without requirement for purification.[188] (C) Thiol-ene click chemistry was used to form peptide-lipid conjugate vaccine candidates.[189] (D) CuAAC click chemistry was used to synthesise a small library of VHL and CRBN linked BRD4 degraders, the CRBN degrader series is shown here.[175] (E) In-cell assembly of a BRD4 targeting PROTAC was made possible using inverse electron demand Diels Alder chemistry.[190]

Click chemistry has also been used in the synthesis of series of PROTACs, as a convenient way to link a variety of linkers / E3 ligase binders to a POI ligand. CuAAC click chemistry was used to synthesise ten new von Hippel-Lindau protein (VHL) and cereblon (CRBN) recruiting BRD4 PROTACs, and the effect of linker length on activity was assessed using an enzymatic ternary complex formation assay, as well as using cellular assays.[175] In-cell assembly of PROTACs has also been achieved using inverse electron demand Diels Alder chemistry, inducing degradation of either BRD4, or Ras-dependent extracellular signal-regulated kinase 1 (ERK1) / ERK2.[190] Formation of PROTACs inside a cell may circumvent any cell permeability issues associated with the larger size of small molecule-based PROTACs (700–1500 Da) compared to traditional small molecule inhibitors (<500 Da).[145,191]

1.5 Project Aims

Cyclic peptide inhibitors of epigenetic enzymes KDM4A–C, and TET1, have excellent selectivity, especially when compared to many small molecule inhibitors of these targets (Section 1.1.2).[40,55] This thesis aims to bring these peptide agents into "the era of functionalism",[172] inspired work such as McAllister *et al.* (2018),[85] in which a peptide-biotin conjugate was used to engage PHD2 in complex with HIF1-α substrate (Section 1.2.3.1). Multifunctional cyclic peptide probes may be particularly useful for these epigenetic targets, as selective probes to assess and/or modulate the proteins while they are part of multiprotein units.[76,192]

The overall aim of this thesis is to carry out proof-of-concept studies to demonstrate the scope and utility of multifunctional cyclic peptides. Starting from highly selective and potent cyclic peptides as the target engagement motif for epigenetic targets, various functionality will be installed using click chemistry and SPPS. Multifunctional cyclic peptide probes will be explored as affinity probes, and as modulators of target protein levels inside cells. The methodology developed can be easily applied to other peptides.

Chapter 2 details the synthesis of multifunctional peptide probes targeting KDM4A, and TET1. Azide-variants of **CP2** (ref.[40]), and **TiP1** (ref.[55]) are synthesised to allow peptide conjugates to be accessed using CuAAC click chemistry. Biotinylated **CP2**- and **TiP1**-based peptides were synthesised, to be tested as affinity probes in Chapter 3. The peptide-biotin conjugates synthesised are likely to bind reversibly, but for some affinity probes irreversible binding may be desirable (Section 1.2.2.1). Therefore, photoreactive probes were also synthesised based on both **CP2** and **TiP1**. **CP2**-based conjugates for targeted protein degradation (TPD) were also synthesised. Both VHL and CRBN recruiting ligands were connected to **CP2**, using a range of linker chemistry, and these probes were tested in Chapter 4.

Affinity cyclic peptide probe target engagement is assessed in Chapter 3. Reversibly binding **CP2**- and **TiP1**-biotin conjugates are tested as affinity enrichment probes in cell lysates. Preliminary MS-based proteomics are performed using **CP2**-biotin conjugate, **B-1**, after it was demonstrated to enrich KDM4A from cell lysates expressing endogenous levels of proteins using western blots. Covalent target engagement of photoreactive **CP2**- and **TiP1**-based probes is also assessed, although follow up studies are required.

In Chapter 4, KDM4A protein degradation using cyclic peptide-based PROTACs is explored. Prior to testing **CP2**-PROTACs, the degradability of HaloTagged and Flag-tagged KDM4A (HaloFlagKDM4A) protein is assessed using HaloTag targeting PROTACs (HaloPROTACs).[193,194] Once degradation conditions are established in HEK293T cells transiently overexpressing HaloFlagKDM4A, the **CP2**-PROTAC panel is tested. The best VHL-recruiting **CP2**-PROTAC is followed up with more detailed protein degradation profiling (**E3-10**).

Chapter 2 Synthesis and characterisation of multifunctional peptide probes

2.1 Introduction

2.1.1 Multifunctional probes

Multifunctional probes may be made of multiple components with different functionality covalently linked together, or the probe may be made up of one chemical entity which has more than one biological effect (Section 1.2, **Figure 2.1**).[69] Multifunctional probes are especially useful in chemical biology, and the most common utilities include: isolation of a POI from a complex mixture,[195] imaging applications,[115] and bringing two or more proteins into proximity.[196]

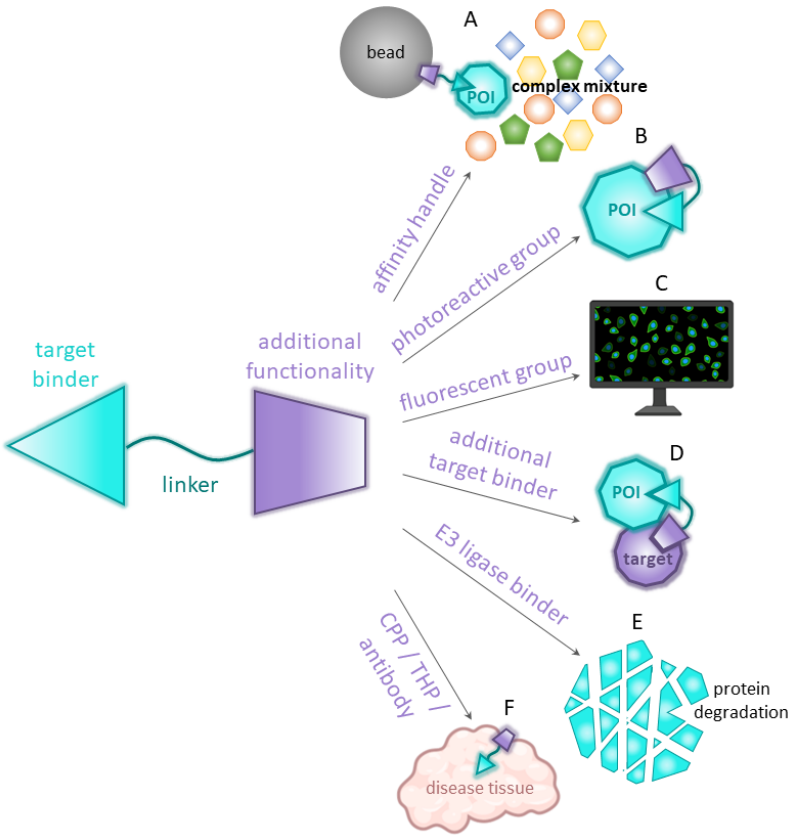
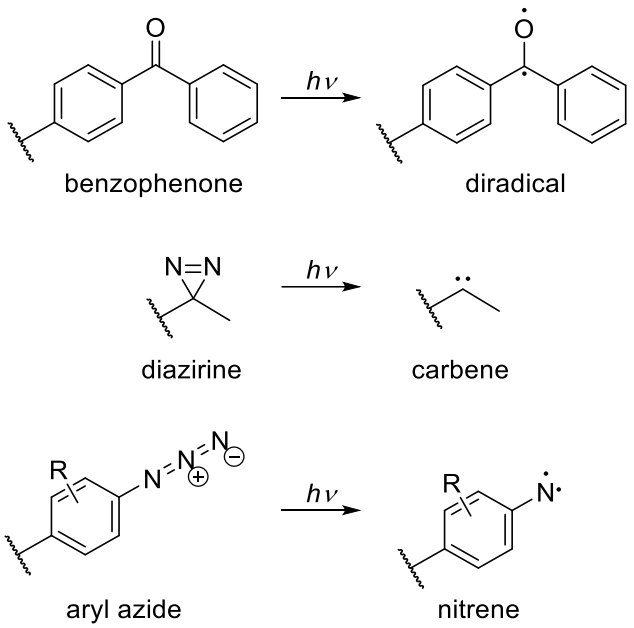


Figure 2.1 Some u**ses of multifunctional (heterobifunctional) probes.** (A) Affinity handles allow the target of probes to be pulled out of complex mixtures; these probes are often used in proteomics to study protein interaction partners.[82] (B) Photoreactive probes covalently label a protein of interest (POI) upon activation with light.[78] (C) Fluorescent groups are often installed for imaging applications.[115] (D) Two proteins can be recruited using one multifunctional probe, to dimerise proteins that would otherwise not interact, and potentially induce protein modifications (*e.g.* PROTACs).[89] (E) PROTACs are made up of a probe for the POI linked to an E3 ligase binder. PROTACs degrade the POI *via* the UPS (Section 1.2.2.3).[197] (F) Groups may be appended to a molecule to improve cell permeability (*e.g.* a cell penetrating peptide (CPP)),[166] or to selectively deliver the molecule to diseased tissues (*e.g.* a tumour-homing peptide (THP)).[198] Fluorescent imaging and diseased tissue diagrams were created with BioRender.com.

Biotin-conjugated small molecules and antibodies are commonly used as multifunctional affinity probes, and are particularly useful in proteomics studies, especially for pre-clinical target validation and drug selectivity studies (Section 1.2.2.1 & 1.2.2.5).[83] Biotin offers an exceptionally high affinity for streptavidin and avidin (K_d = 10^{14}– 10^{15} M), which is one of the strongest non-covalent interactions found in nature.[199] The biotin-(strept)avidin complex forms rapidly, is stable under a wide pH/temperature range, and is resistant to denaturing agents or organic solvents.[200,201] Biotin is thus commonly added as a tag to (bio)molecules for purification, or detection of binding partners from complex mixtures (*e.g.* cells), using (strept)avidin for capture. Such applications include co-precipitation experiments, where biotinylated molecules and their binding partners can be affinity purified, after stringent washes, from complex mixtures using (strept)avidin beads (**Figure 2.1**A).[201]

Activity-based probes and photoaffinity probes covalently bind their targets, either directly (*e.g.* 2-chloroacetamides react with cysteine side chains) for activity-based probes, or by activation of photoreactive groups using light for photoaffinity probes.[82] Suitable reactive groups must be available in the target active site to design activity-based probes, with good proximity to the probe binding interface. Whereas, upon activation, photoreactive groups react with any molecule in close proximity, such as the target or bulk solvent. The main advantage of photoaffinity probes is that they react with a broader range of amino acids, and can potentially be applied to a wider range of targets.[82,138] Furthermore, a number of photoreactive groups are available for crosslinking to proteins which form highly reactive species when activated with light (**Scheme 2.1**), including: diazirines, benzophenones, and aryl azides.[78] Photoreactive probes without reporter groups are also used for alternative applications, such as mapping the binding region of a probe to the POI.[202]



Scheme 2.1 Common photoreactive groups used in chemical probes. Benzophenones are activated to a diradical species, diazirines are activated to carbenes, and aryl azides are activated to nitrenes.[203]

Functionalisation of probes for imaging applications is a major field within life sciences (Section 1.2.2.4). In diagnostic medicine, multifunctional probes may be used in PET and single-photon emission computed tomography (SPECT) to deliver unstable radionuclides to specific parts of the body for imaging.[204] Similarly, fluorescent groups may be incorporated in multifunctional probes for cellular imaging applications.

Another major class of multifunctional probes is TPDs, where the POI ligand is conjugated to a ligand of an enzyme that can degrade the POI. The most studied TPDs are PROTACs, which induce proximity of a POI and E3 ligase, causing degradation of the POI through the UPS (Section 1.2.2.3). TPD by multifunctional PROTACs has been used to target 'undruggable' proteins, and often benefits from increased selectivity compared to the parent target binder.[205,206] Furthermore, successful PROTACs have often been derivatised from molecules with lower affinity for their targets than the best-in-class inhibitor, demonstrating that multifunctional probes can be useful for targets for which high potency is a challenge.[207] PROTACs have inspired a number of other strategies to modulate the stability of target proteins in cells, which include: i) deubiquitinase-targeting chimeras (DUBTACs) which stabilise the POI,[208] ii) phosphorylation-inducing chimeric

small molecules (PHICS) / phosphatase recruiting chimeras (PhoRCs) which phosphorylate/dephosphorylate a POI,[209,210] iii) LYTACs which degrade a POI through the lysosomal degradation system,[109] and iv) ribonuclease targeting chimeras (RIBOTACs) which induce nuclease-mediated degradation of RNA.[211]

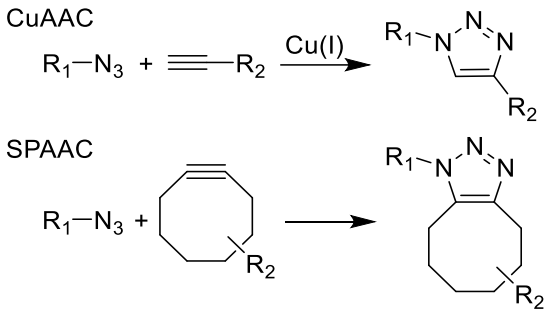
Multifunctional probes may also be used to deliver a molecule to a desired location (Section 1.2.3.2). Tumour-homing peptides, and antibody-drug conjugates deliver pharmaceuticals directly to diseased tissues, reducing off-target effects (**Figure 2.1**F).[212] While CPPs are often conjugated to impermeable peptide cargos to deliver them into cells. CPPs are generally positively charged and interact with negative surface charges within cell membranes to promote cell permeability.[213]

In preclinical drug discovery, selective multifunctional probes are highly desirable during target validation (Section 1.2).[61,214] Selective probes are used to demonstrate the effect of modulating a potential drug target in disease models, and to assess the safety of modulating the proposed target.[71,215] High probe selectivity is also an advantage for imaging probes, such that signal can be attributed to the target (Section 1.2.2.4).[117] However, in some instances a multifunctional probe targeting multiple proteins is preferred. Many ABPP probes have been developed to profile whole families of enzymes. Activity-based probes only bind to active enzymes, and MS-based proteomics has been used to identify PPIs of these enzymes, and to profile the selectivity of inhibitors targeting members of the family of enzymes.[216] ABPP of serine hydrolases has been used to investigate the serine hydrolase proteome of different human cancers,[217] and to identify selective inhibitors of uncharacterised serine hydrolases through library screening.[218]

2.1.2 Synthesis of multifunctional probes

One of the most commonly used biorthogonal methods of rapidly and efficiently linking two entities together is the click reaction.[219] The range of click chemistry available, and examples of the utility of this reaction for synthesising multifunctional probes is discussed in Section 1.4. A

particularly useful click reaction is the Huisgen 1,3-dipolar cycloaddition, in which an alkyne and azide form a 1,2,3-triazole.[220] Often, terminal alkynes and azides are coupled using Cu(I) as a catalyst in CuAAC reactions.[176] CuAAC reactions are quick, reliable, and tolerate water; they were chosen for this project due to the wide availability of groups conjugated to terminal alkynes for coupling to azide-containing peptides. Conversely, the main disadvantage of using CuAAC is the necessity to remove Cu after the reaction, as it highly cytotoxic, and this can be achieved through chromatography, dialysis, or addition of Cu chelators.[79] To circumvent the requirement for Cu(I), SPAAC reactions can be used to couple azides to strained cyclooctynes (**Scheme 2.2**).[180] SPAAC offers a copper-free approach, removing the need for any work-up after the reaction, but extensive optimisation may be required for each reaction, in part due to solubility issues associated with cyclooctynes.[174]



Scheme 2.2 Huisgen 1,3-dipolar cycloaddition joining an azide and alkyne is one of the most commonly used bioorthogonal click reactions. Cu(I)-catalysed azide-alkyne cycloadditions (CuAAC) use Cu(I) as a catalyst, with a ligand for Cu(I) often included in the reaction, along with a reducing agent to stabilise Cu(I) ions. Cu(I) is highly toxic and must be removed before downstream biological applications. Conversely, strain-promoted azide-alkyne cycloadditions (SPAAC) use strained cyclooctynes, and do not require a catalyst. SPAAC reactions can be performed inside cells, and even living organisms.[79]

2.1.3 Aims

The aim of this chapter is to synthesise macrocyclic peptide multifunctional probes against epigenetic targets. KDM4 and TET enzymes engage histone proteins and genomic DNA respectively. These proteins have large interaction sites, with high similarity to related enzymes, which small molecule probes often cannot selectively target (Section 1.1.2).[38,60] However, high selectivity can be extremely helpful in multifunctional probes when elucidating the role of a specific target in a biological setting. Peptides are larger than small molecules, offering a greater

probe-target interaction surface area, and consequently greater selectively compared to small molecules for these targets.[147] **CP2** (ref. [40]) and **TiP1** (ref. [55]) are macrocyclic peptides targeting KDM4A–C and TETs respectively. They were both discovered using genetically encoded macrocyclic peptide selection technology, through the random nonstandard peptides integrated discovery platform (RaPID).[162,169] Both **CP2** and **TiP1** offer improved selectivity for their respective targets over currently known small molecule inhibitors (Section 1.1.2), and so were chosen as promising probes for multifunctionalisation.

Multifunctional macrocyclic peptide probes with a range of applications were synthesised using SPPS. Additional functionality was installed to cyclic peptides through an amino acid containing a chemical 'handle', which was later used to link a range of functional groups to the peptide *via* click chemistry, creating a peptide conjugate (Section 1.4). This route would also facilitate the addition of groups that are not stable under peptide synthesis conditions.[171]

A schematic detailing considerations in the synthesis of multifunctional peptide probes for this chapter is detailed in **Figure 2.2**, key aims for this chapter are as follows:

- Synthesis of clickable macrocyclic peptides against KDM4A–C and TETs.
- Optimisation of CuAAC click reaction conditions for coupling of azide containing macrocyclic peptides to alkyne containing functional groups.
- Synthesis of a range of macrocyclic peptide multifunctional probes. CuAAC click was used to create **CP2**-based peptide conjugates containing biotin and E3 ligase binders, and photoreactive groups were installed in macrocyclic peptides during SPPS. Methodology developed using **CP2** was later applied to **TiP1**.
- Validation of target engagement of multifunctional probes against the POI.

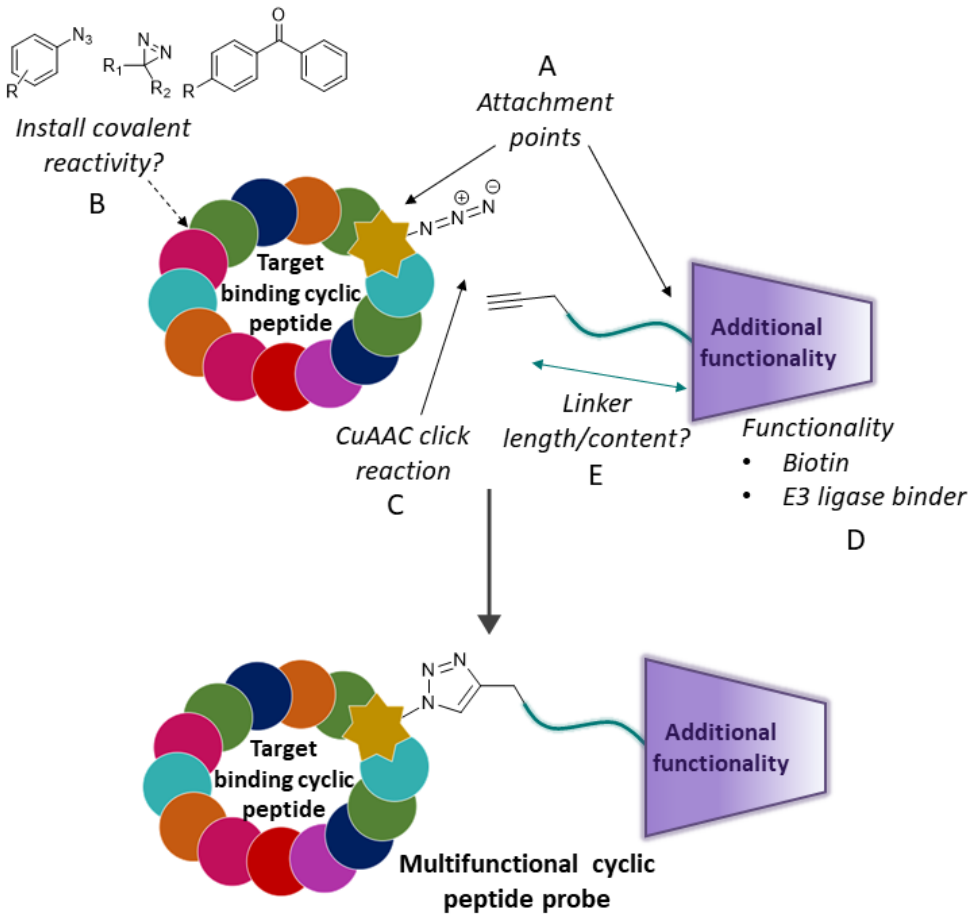


Figure 2.2 Considerations for multifunctional macrocyclic peptide probe development. (A) Attachment points for linkers between cyclic peptide and additional functional group. (B) Whether, and where, to install reactive groups for covalent linkage to the target. Photoreactive amino acids were tested in this project, which react upon activation with light. (C) Chemistry for making cyclic peptide conjugates, here CuAAC click reactions were used. (D) Many groups can be linked to the peptide to develop multifunctional probes. In this project biotin was used for afffinity (tested in Chapter 3), and E3 ligase binders were used to develop macrocyclic peptide protein degraders (tested in Chapter 4). (E) Linker length and type.

2.2 Results

2.2.1 General cyclic peptide synthesis

Cyclic peptides were synthesised using Fmoc/*t*Bu SPPS and purified using reverse phase high-performance liquid chromatography (HPLC), using well-established methodology similar to Kawamura *et al.* (2017)[40] (Section 6.2). Macrocyclisation was induced using a thioether linkage between the *C*-terminal cysteine and the chloroacetylated *N*-terminus.

Both microwave assisted (CEM Corporation, Liberty Blue) and heat induction (Gyros Protein Technologies, PurePep Chorus) peptide synthesisers were used in this project, producing similar results. Linear precursors to cyclic peptides were *N*-terminally chloroacetylated before cleavage

from the resin and deprotection. Peptides were subsequently cyclised in a basic water/acetonitrile solution (pH ≈ 9) at 37 °C for 1 hour, prior to purification using HPLC across a gradient of 0.1% trifluoroacetic acid (TFA) in water/acetonitrile. Pure fractions were identified using matrix-assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-MS), and the purest fractions were pooled (Section 6.2).

2.2.2 Design of clickable macrocyclic peptides

*2.2.2.1 **CP2**-based clickable macrocyclic peptides*

CP2 was selected for initial proof-of-concept studies due to the availability of both a KDM4A-**CP2** co-crystal structure, and extensive structure activity relationship data. Furthermore, structural information was used to develop a more cell active variant in the original study, **CP2.3**.[40] Threonine-13 was chosen as the position to install a clickable azide handle, as it is remote from the active site, and other peptide-KDM4A interactions in the co-crystal structure (**Figure 2.3**); also, T13 is not involved in the cyclisation of **CP2**. A variant of **CP2.3** was synthesised using proline-13 as the site for click handle incorporation, with the assumption that it will be positioned similarly to T13 in **CP2**. ε-Azidolysine was installed as the click chemistry handle, resulting in a flexible system for further functionalisation using either CuAAC or SPAAC. ε-Azidolysine contains a four carbon linker between the peptide backbone and azide group, it was hypothesised inclusion of this spacer would position reporter groups further from the peptide-target binding interface, and therefore impact peptide-target binding less than shorter linkers.

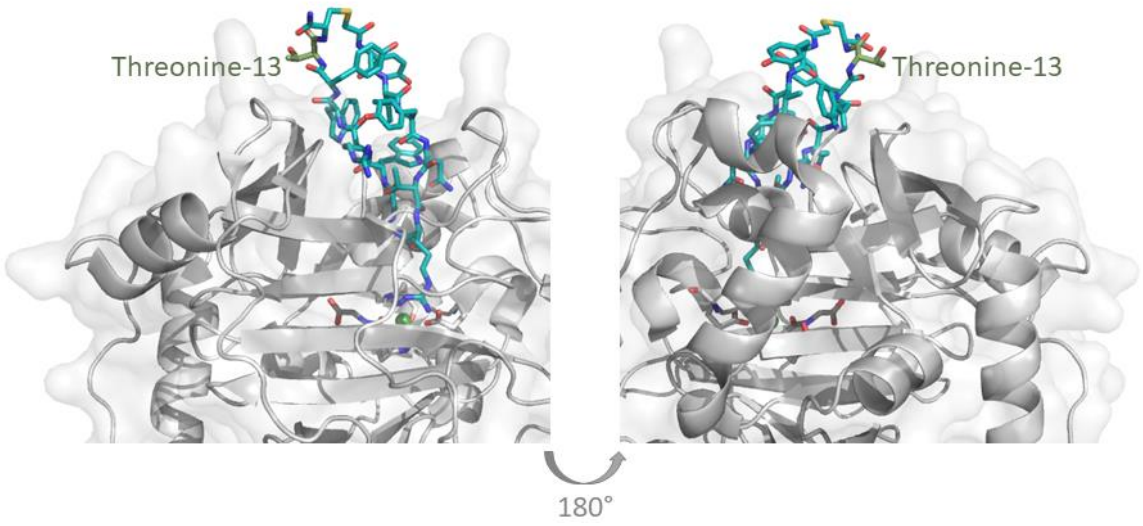


Figure 2.3 Threonine-13 (green) was chosen as the site to incorporate an azide in CP2 (teal). Threonine-13 is remote from the active site, and interacts minimally with KDM4A (grey) in the **CP2**-KDM4A co-crystal structure (PDB 5LY1).[40] Heteroatoms: red = O, blue = N, yellow = S.

The clickable **CP2** variants synthesised are detailed in **Table 2.1**, and most were obtained in good purity after standard SPPS, as detailed in Section 6.2.1. However, extensive optimisation was needed for **CP2.3-P13K$_{N3}$** (Section 2.2.2.1.1).

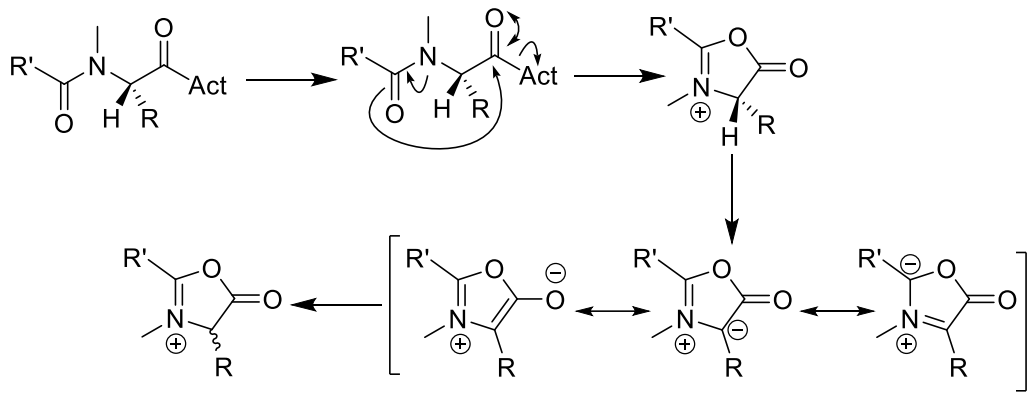
Peptide	Parent peptide	Variations to parent	Amino acid sequence	Molecular weight / g·mol^{-1}
CP2	**CP2**	-	^{D}YVYNTRSGWRWYTC	1894.10
CP2-T13K$_{N3}$	**CP2**	T13K$_{N3}$	^{D}YVYNTRSGWRWYK$_{N3}$C	1947.17
CP2-R6A-T13K$_{N3}$	**CP2**	R6A T13K$_{N3}$	^{D}YVYNTASGWRWYK$_{N3}$C	1862.06
CP2.3-P13K$_{N3}$	**CP2.3**	P13K$_{N3}$	^{D}Y^{Me}VYNTRS^{D}AWRWF$_{4F}$K$_{N3}$^{Me}C	1992.23

Table 2.1 Clickable CP2-based cyclic peptides synthesised. Azides were incorporated in the 13th position, which is unlikely to reduce the binding affinity of the resulting peptide. Peptides are cyclised by a thioether bond between the *C*-terminal cysteine, and an *N*-terminal acetyl group.

Identifying the purest fractions for azide containing peptides was more challenging than other peptides, as azides are susceptible to degradation during MALDI-MS, which uses a high power laser to take molecules into the gas phase.[221] Peaks at –26 Da were consistently observed (**Figure 2.4**E), which is attributed to loss of N_2 with gain of 2H. Azide peptides were later coupled to reporter groups using CuAAC, and purified using HPLC, which gave final products with excellent purity (Section 2.2.3).

2.2.2.1.1 CP2.3-P13K$_{N3}$ synthesis

The presence of two backbone *N*-methyl amino acids in **CP2.3-P13K$_{N3}$** (**Table 2.1**) complicated the synthesis, and optimisation was required to obtain the purified product. Coupling after *N*-Me amino acids is challenging, due to steric hinderance of the methylated amino group.[222] Additionally, coupling *N*-Me amino acids themselves can also be challenging, and racemisation can take place during activation of the amino acid by tautomerisation of the oxazolonium ion, leading to pseudo aromatic structures (**Scheme 2.3**).[223]



Scheme 2.3 Racemisation during *N*-me peptide synthesis mainly occurs via tautomerisation of the oxazolonium ion, analogously to the oxazolone formed in standard peptide synthesis.[223,224] Act = activator

Initially, **CP2.3-P13K$_{N3}$** synthesis was attempted using microwave assisted SPPS with commercially sourced *N*-Me cysteine/valine; the desired product was observed as a minor peak by MALDI-MS, but the results were not repeatable (**Figure 2.4**A). Extended couplings and double couplings were tried for *N*-Me amino acids without noticeable improvement in the yield. In order to identify the problematic couplings, a truncated version of the peptide (7-mer ^{D}Ala[8]–^{Me}Cys[14]) was synthesised and a test cleavage was performed to analyse the resulting peptide (**Figure 2.4**B). A minor peak in the MALDI-MS spectrum contained the expected product, and another minor peak contained the K$_{N3}$ deletion product. However, a number of major unidentified side product peaks were also present, demonstrating that optimisation was required within the first seven amino acid couplings.

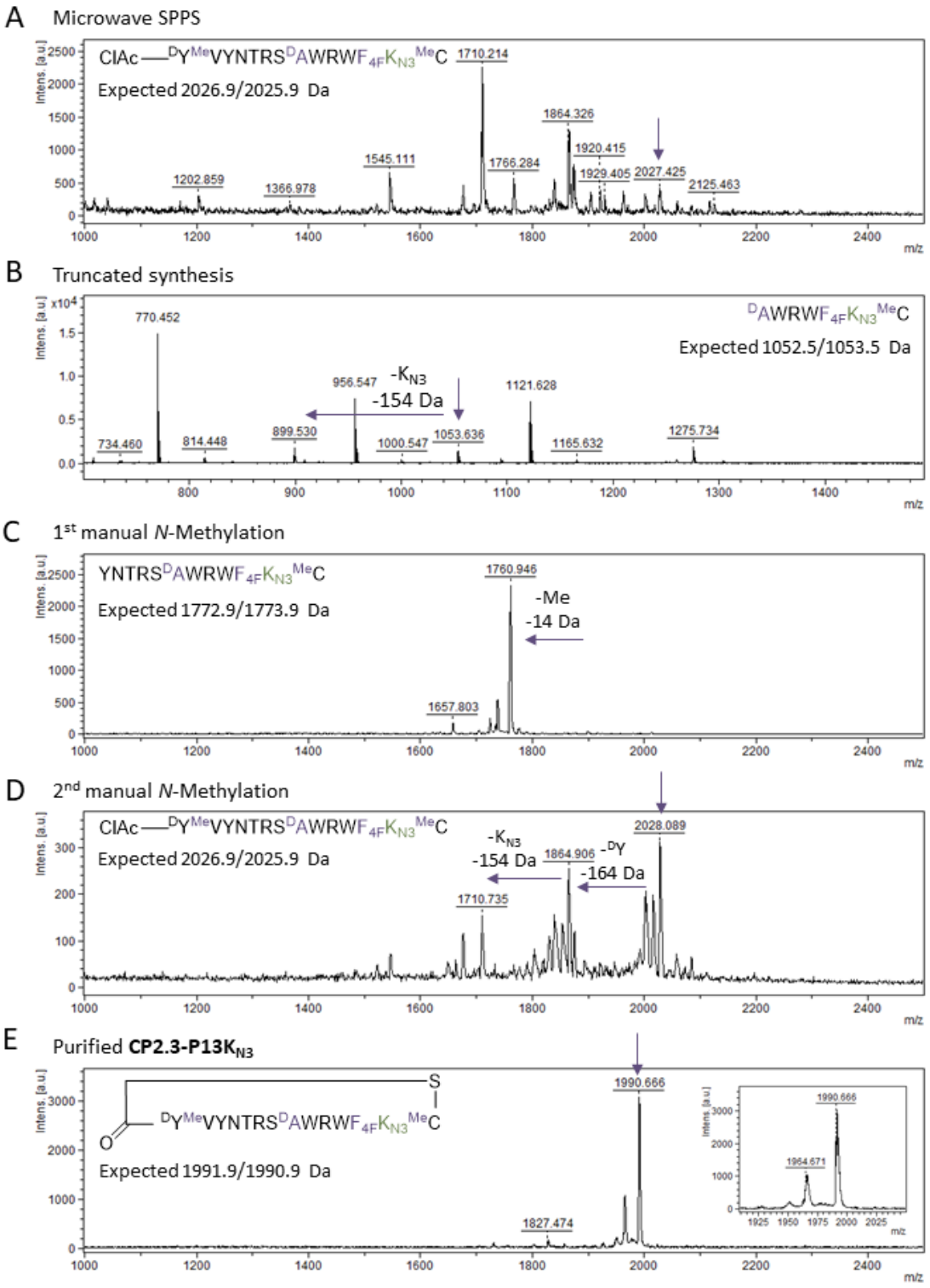
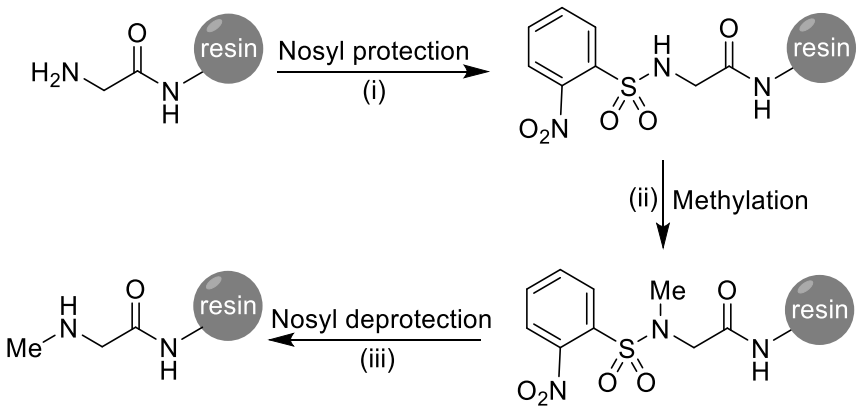


Figure 2.4 MALDI-MS spectra showing the optimisation of CP2.3-P13K$_{N3}$ synthesis. (A) Microwave-assisted SPPS produced little desired product. (B) Synthesis of a truncated *C*-terminal 7-mer peptide demonstrated that optimisation was needed within the first seven couplings; desired product is a minor peak in the spectrum and most peaks were difficult to assign. (C) Initial *N*-methylation reaction was unsuccessful, but high purity of the main product suggested this route was promising. (D) Second manual *N*-methylation conditions tested successfully produced the desired product as the major peak. The final coupling was particularly low yielding, so in the fully optimised method a triple coupling was used. (E) Optimised synthesis successfully produced **CP2.3-P13K$_{N3}$** in sufficient purity by MALDI-MS; –26 Da peak is attributed to loss of N_2 during laser desorption (Section 2.2.2.1).

In order to reduce the extent of side product formation, capping was introduced, which is the process of adding an acetyl group to the *N*-terminus of peptides which have not successfully coupled to the next amino acid in the sequence. After a coupling reaction, the resin is incubated with acetic anhydride, which reacts with the free *N*-termini of uncoupled peptides. Capping prevents unsuccessful peptides from extending throughout the synthesis, avoiding variants with deleted amino acids. Single amino acid deletion variants can be difficult to separate from the parent peptide, as they are closer in size to the product than most truncated peptides. The peptide synthesiser set up at the time did not support automated capping, so manual SPPS was carried out for coupling of *N*-Me amino acids, capping, and the subsequent coupling. 1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium3-oxide hexafluorophosphate (HATU) alongside Oxyma Pure were used as the coupling reagents, as HATU is an established coupling reagent for *N*-Me amino acids.[225] Couplings were monitored using 2,4,6-trinitrobenzenesulfonic acid (TNBS) tests for primary amines, and *p*-chloranil tests for secondary amines. However, following acidic cleavage of the peptide from the solid support, no peptide precipitate was observed upon addition of ice cold diethyl ether, and no product was found by MALDI-MS.

The next strategy tested involved manual backbone *N*-methylation of unmodified cysteine/valine. It was envisaged that this might increase peptide yield, as coupling *N*-methylated amino acids to resin can be challenging, while peptides with *C*-terminal unmodified cysteine are routinely synthesised. *N*-Methylation was performed on Fmoc-deprotected amino acids on resin; the *N*-terminus was nosyl protected, prior to being methylated with dimethyl sulfate, and subsequently nosyl deprotected (**Scheme 2.4**). Nosyl protection ensures only a single methyl group is added to the *N*-terminus. All couplings were performed on a peptide synthesiser, removing the resin for manual *N*-methylation reactions, and double couplings were carried out after *N*-methyl amino acids. Initial methylation reaction conditions involved incubation of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) and dimethyl sulfate simultaneously,[226] but very little

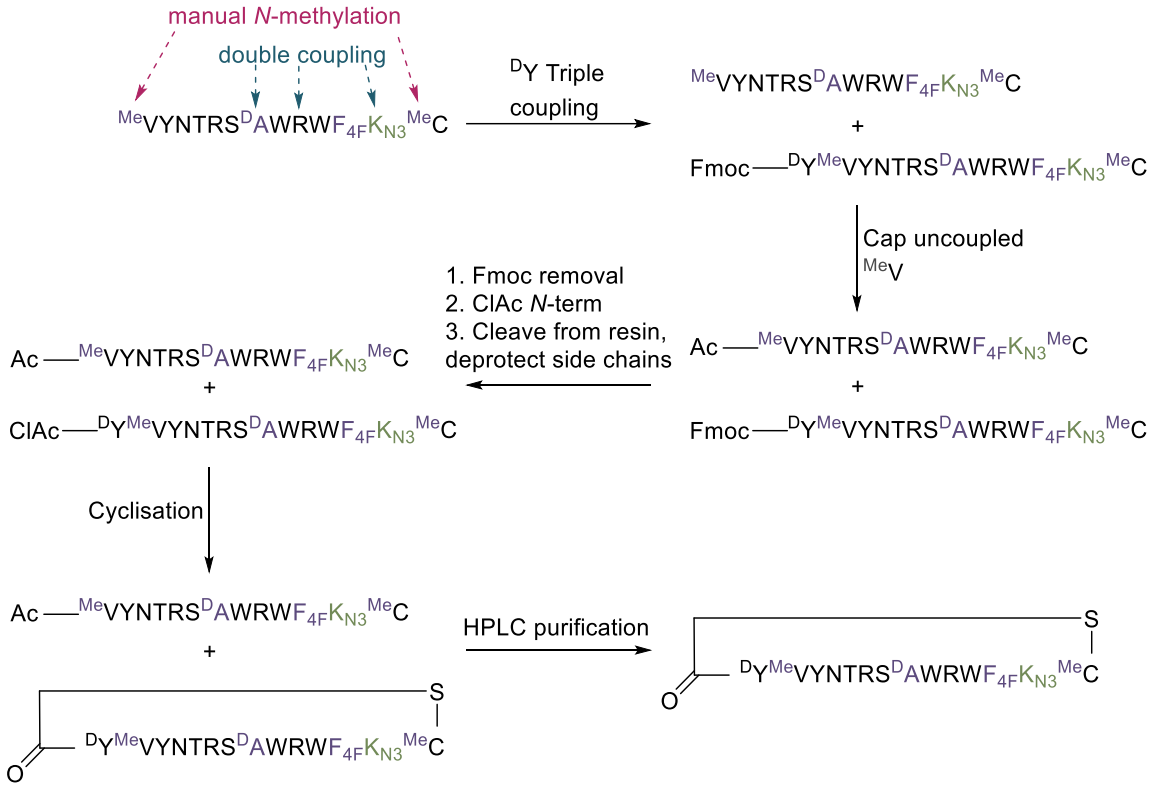
methylated product was produced (**Figure 2.4**C), ascertained using a test cleavage prior to the second *N*-methylation reaction on valine (13th coupling). However, the fact that the major peak in the MALDI-MS spectrum was the unmethylated 13-mer did suggest that, upon optimisation of the methylation reaction, a successful synthesis may be possible. Subsequently, DBU was incubated with the resin for 3 minutes prior to the addition of dimethyl sulfate for 4 minutes (**Scheme 2.4**), and the doubly *N*-methylated 13-mer peptide was produced. Conditions were based upon Biron *et al.* (2006),[227] although dimethyl sulfate incubation time was doubled to encourage methylation, since very little methylated product had been synthesised using the first reaction conditions tested. While the desired product was formed as the major species using this approach, it was noted that the 14th coupling produced only 60% conversion (**Figure 2.4**D).



Scheme 2.4 Backbone *N*-methylation strategy. All steps were performed at room temperature, with end-over-end rotation; five 1-methyl-2-pyrrolidinone (NMP) washes (3–5 mL, 5 minutes) were performed after each step. Reagents and conditions for 0.05 mmol scale: (i) 2-nitrobenzenesulfonyl chloride (44 mg, 0.2 mmol) and 2,4,6-collidine (66 µL, 0.5 mmol) in 1 mL NMP for 15 minutes. (ii) 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU) (45 µL, 0.3 mmol) in 1 mL NMP for 3 minutes, followed by addition of dimethyl sulfate (95 µL, 1 mmol) for 4 minutes. Mixture drained and procedure repeated. The initial reaction conditions added DBU and dimethyl sulfate simultaneously for 5 minutes, but little methylated product was produced. (iii) DBU (37.5 µL, 0.25 mmol) and 2-mercaptoethanol (35 uL, 0.5 mmol) in 1 mL NMP for 5 minutes. Mixture drained and procedure repeated.

Purification using HPLC could not separate incomplete 13-mer and 14-mer both pre- and post-cyclisation. To improve separation, undesired 13-mer was capped after the 14th coupling, but before *N*-terminal Fmoc-removal and chloroacetylation. This produced a mixture of peptides within which only the 14-mer could be cyclised, and the cyclic 14-mer could be easily separated from the capped linear 13-mer by HPLC (**Figure 2.4**E). The final synthesis is detailed in **Scheme 2.5**. In the final synthesis a triple coupling was used for ^{D}Y, and double couplings were used where

amino acid coupling might be less efficient, including: after *N*-Me amino acids, for ^{D}A as it is positioned after multiple bulky residues, and for Arginine(2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl) due to the bulkiness of the side chain and protecting group (**Figure 2.4**E).[228,229]



Scheme 2.5 Optimised synthesis strategy for CP2.3-P13K$_{N3}$. Manual *N*-methylation was performed as described in **Scheme 2.4**, and double/triple couplings were used where coupling efficiency may be suboptimal. Capping was used to prevent 13-mer side products from cyclisation, which allowed cyclic 14-mer product to be separated from linear 13-mer using HPLC. **CP2.3** modifications compared to **CP2** are shown in purple, proline-13 has been replaced with ε-azidolysine (green) to form **CP2.3-P13K$_{N3}$**.

2.2.2.2 TiP1-based clickable macrocyclic peptides

Having obtained clickable peptides from **CP2** and its analogues, clickable TET targeting peptides were developed using **TiP1** (ref. [55]) as the parent peptide.[a] Limited structural information is available on **TiP1**-TET binding, and currently there is no co-crystal structure, therefore results from an alanine scan were used to identify residues that were unlikely to reduce **TiP1** binding affinity.[230] Three separate positions in **TiP1** were selected for incorporation of an azide handle, to explore

[a] **TiP1**-based multifunctional probes were synthesised by Valeria Cobiltean, a Newcastle University final year MChem student, under my supervision.

the accessibility of reporter groups in different positions (**Table 2.2**). **TiP1**, **TiP1-Y6K$_{N3}$'** and **TiP1-S13K$_{N3}$** were all successfully purified. However, obtaining pure **TiP1-H12K$_{N3}$** was not possible due to solubility issues.

Peptide	Parent peptide	Variations to parent	Amino acid sequence	Molecular weight / g·mol^{-1}
TiP1	**TiP1**	-	^{D}YLIYAYYTWWEHSC	1937.16
TiP1-Y6K$_{N3}$	**TiP1**	Y6K$_{N3}$	^{D}YLIYAK$_{N3}$YTWWEHSC	1928.16
TiP1-H12K$_{N3}$	**TiP1**	H12K$_{N3}$	^{D}YLIYAYYTWWEK$_{N3}$SC	1954.19
TiP1-S13K$_{N3}$	**TiP1**	S13K$_{N3}$	^{D}YLIYAYYTWWEHK$_{N3}$C	2004.26

Table 2.2 Clickable TiP1-based cyclic peptides synthesised. Azides were incorporated in positions which had minimal impact on binding affinity of **TiP1** in an alanine scan. **TiP1-H12K$_{N3}$** could not be purified by HPLC due to solubility issues. Peptides were synthesised by Valeria Cobiltean, under my supervision.

2.2.3 Click chemistry as a platform to synthesise multifunctional cyclic peptides

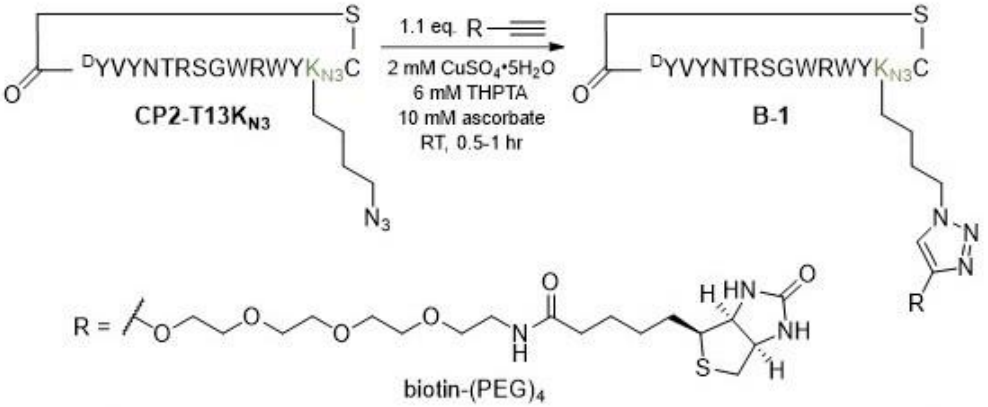
Following successful synthesis of azide, clickable cyclic peptides (Section 2.2.2), CuAAC was used to couple cyclic peptides to biotin, and E3 ligase binders. CuAAC is more robust than SPAAC,[171] likely facilitating conjugation of various functional groups with minimal optimisation. HPLC purification was performed following click chemistry to remove copper ions, and other click reagents; it also ensured high purity of the final product.

*2.2.3.1 Click reaction optimisation: coupling **CP2-T13K$_{N3}$** to biotin-(PEG)$_4$-alkyne as a model system*

CuAAC click reaction conditions can be optimised by altering many factors, including: the ratio of alkyne/azide, the copper source, the reducing agent to convert Cu(II) to Cu(I), use of a Cu(I) chelating ligand to stabilise Cu(I) ions, and the solvent system.[219,231]

A solvent screen was performed to identify optimal CuAAC click conditions (**Table 2.3**), while the ratio of azide/alkyne was kept close to 1:1, so that key reagents were used efficiently. Cu(I) was generated by reducing Cu(II) with excess sodium ascorbate. Cu(I) species were stabilised using tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) (commercially available), which had greater water solubility compared to alternative Cu(I) chelators.[219] Additionally, THPTA has been shown

to react with reactive oxygen species produced by oxygen/copper/ascorbate, protecting peptides from oxidation.[232]



Entry	Purified peptide?	Solvent	% DMSO	Solubility	Conversion after 1 hour / %	Yield after HPLC purification / %
1	No	1:1 MeCN:H_2O	2	High	0	-
2	Yes	1:2 MeCN:H_2O	2.5	High	5	-
3	Yes	1:3 MeCN:H_2O	2.5	High	5	-
4	Yes	H_2O	2	Low	100	0
5	Yes	PBS buffer	9.6	Medium	100	8
6	Yes	PBS buffer	20	Medium-High	100	40
7	Yes	PBS buffer	50	High	100 (0.5 hr)	48

Table 2.3 Optimisation of a CuAAC click reaction to couple azide cyclic peptides to alkyne-linked groups. CP2-T13K$_{N3}$ was coupled to biotin-(PEG)$_4$-alkyne to give **B-1**. Ratio of azide/alkyne was close to 1:1 to use reagents efficiently. Conditions tested are displayed in the table along with yields after HPLC purification. Conversion of click reactions was monitored by MALDI-MS. THPTA = Tris(3-hydroxypropyltriazolylmethyl)amine.

Initially, conjugation was carried out on a crude peptide (**Table 2.3**, Entry 1), to minimise chromatography steps, and maximise product yield. However, this reaction did not yield any product. Conversely, reactions using purified peptide (**Table 2.3**, Entry 2/3) produced a small amount of product, so further optimisation was performed using purified peptide. Acetonitrile has been reported to slow the rate of some CuAAC click reactions, by coordinating Cu(I) *via* π-acceptor interactions.[232] When acetonitrile was removed from the CuAAC reaction solvent mixture, quantitative conversion occurred within 1 hour (**Table 2.3**, Entries 4-7). As dimethyl sulfoxide (DMSO) content was increased, the yield of product increased, likely due to increased solubility of the peptide. Phosphate buffered saline (PBS, pH 7.4) was also employed, as changes in pH alter

the solubility of peptides. Increasing the percentage of DMSO to 20–50% (**Table 2.3**, Entry 6-7) resulted in an increase in yield after HPLC purification to a satisfactory 40–48%.

Once the click reaction was complete, acetonitrile was added to the reaction mixture (25–100% of initial reaction volume), which increased solubility of reagents, and reduced risk of samples precipitating in the water/acetonitrile HPLC system prior to reaching the column. Good solubility of HPLC samples was important as samples were filtered prior to injection, so insoluble product could not be purified.

2.2.3.2 Synthesis of multifunctional macrocyclic peptide conjugates

The previously described, optimised, CuAAC click reaction conditions (Section 2.2.3.1) were used to couple macrocyclic peptides to either biotin, or E3 ligase binders. Biotin was linked to both **CP2**-based and **TiP1**-based peptides (**B-n, Figure 2.5** & **Figure 2.6**), to test the resultant peptide conjugates as affinity probes (Chapter 3). **CP2** was also linked to E3 ligase binders (**E3-n**) as TPD probes (**Figure 2.5**). **E3-n** probes used different E3 ligase binders, different linkers, and different parent peptides to improve the likelihood of identifying a KDM4A protein degrader (Chapter 4).[197]

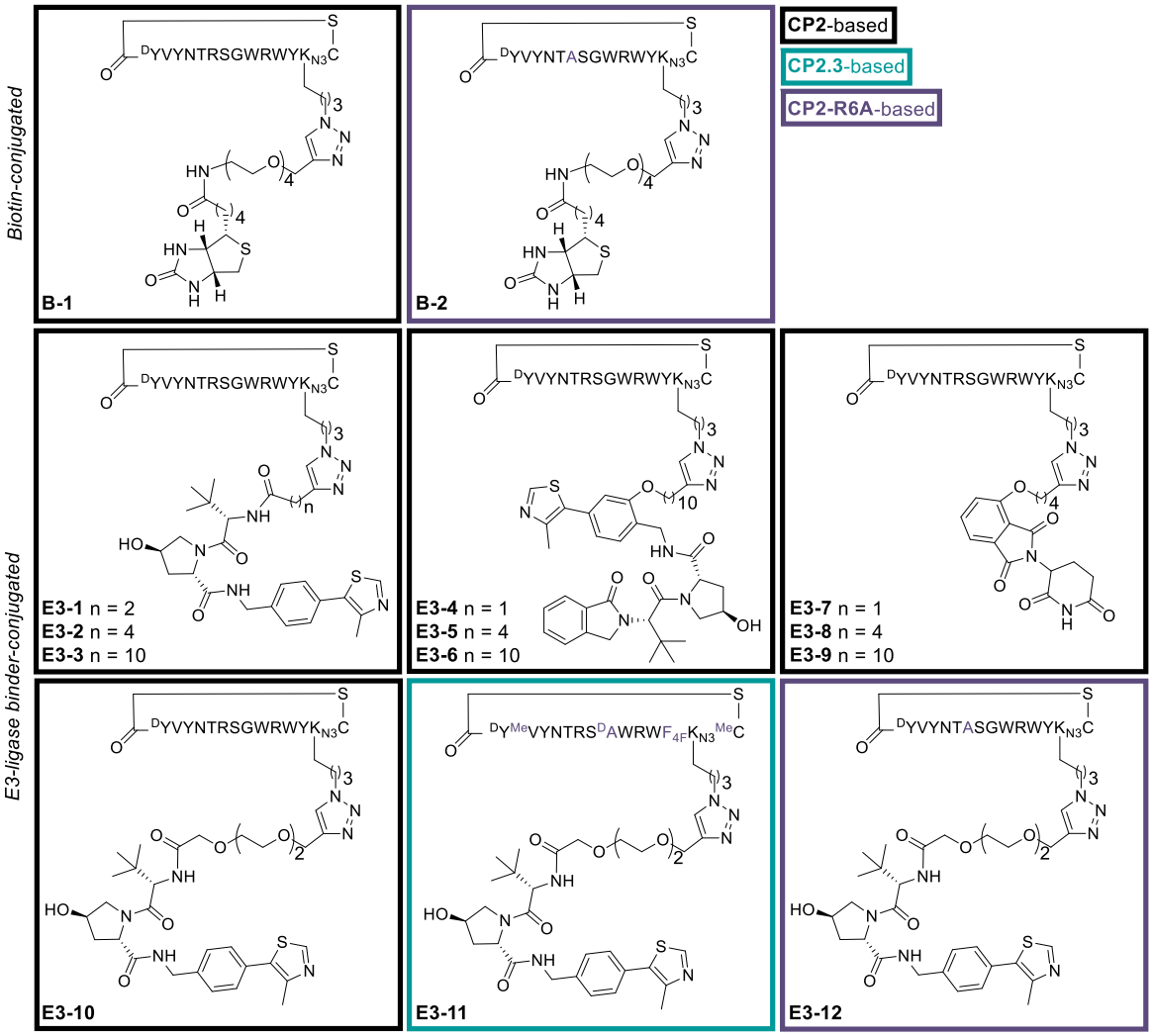


Figure 2.5 Structures of CP2-based multifunctional cyclic peptides synthesised. CP2 is a KDM4A–C targeting cyclic peptide, **CP2.3** is a more cell active variant of **CP2**, and **CP2-R6A** is a **CP2**-variant with >50-fold lower inhibitory activity.[40] **B-1** and **B-2** are peptide-biotin conjugates, and. **E3-1** – **E3-12** are peptide-E3 ligase binder conjugates.

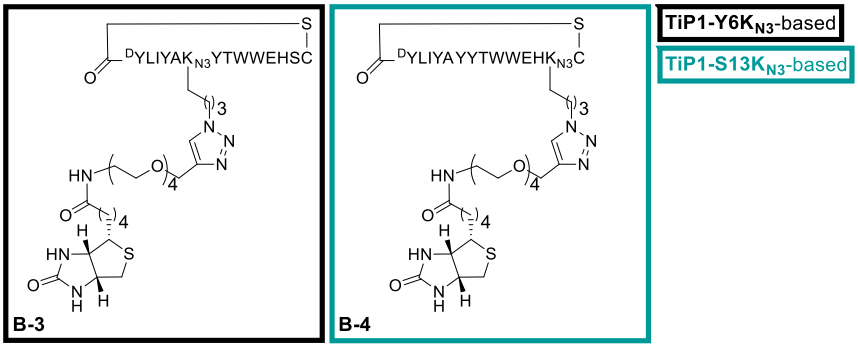
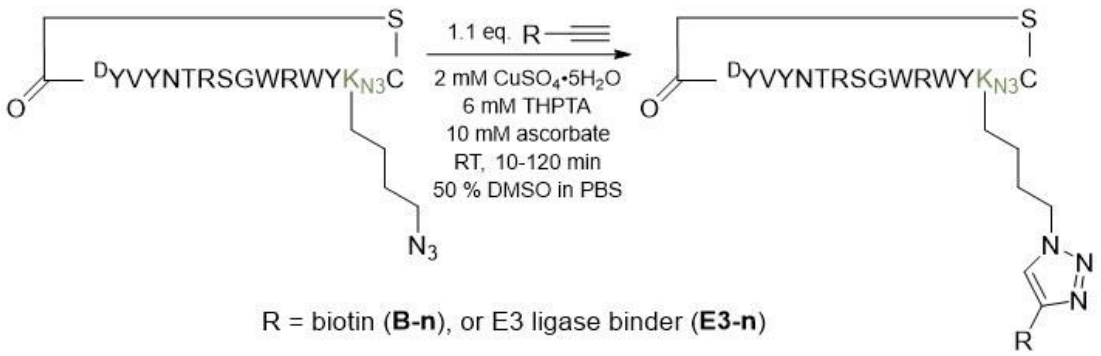


Figure 2.6 Structures of TiP1-based biotinylated cyclic peptides. Biotin functionality was installed in multiple positions as it was unknown if biotin would be accessible at the positions tested.

In total, 16 peptide conjugates were synthesised (**Table 2.4**). The length of time taken to reach full conversion was recorded following reaction monitoring by MALDI-MS. Most reactions reached full

conversion within 1 hour; however, in cases where full conversion was not observed, additional alkyne was added, and the reaction left longer. However, **E3-6** could not be pushed to full conversion, despite addition of more alkyne (1.0 equivalent), and extended reaction time. Purification of each reaction was performed *via* HPLC. Purity of the products was analysed and found to be >92% for all but one peptide conjugate, **E3-11**. The reaction yields varied considerably across the peptide conjugates compared to the 48% yield obtained for **B-1** in reaction optimisation. Notably, the lowest yields were for **B-2** and **E3-12**, the two multifunctional probes employing **CP2-R6A** as the parent peptide; possibly due to lower solubility of **CP2-R6A** in the reaction solvent used. A similarly low yield was also obtained for **B-3**, using **TiP1-Y6K$_{N3}$**. A lower yield was obtained for **E3-11**, the only multifunctional probe employing **CP2.3** as the parent peptide, due to lower purity of the product. Overall, despite variable reaction yields the reaction developed was successful at coupling a range of parent peptides (**CP2**- and **TiP1**-varitants), to a range of functionalised alkynes.



R = biotin (**B-n**), or E3 ligase binder (**E3-n**)

Compound	Parent Peptide	Molecular weight / g·mol^{-1}	Time/conditions to full conversion	Purity (Analytical HPLC) / %	Yield / %
B-1	**CP2**	2404.75	30 min	93.2	48.0
B-2	**CP2-R6A**	2319.64	10 min	96.8	8.9
E3-1	**CP2**	2457.82	120 min + 0.5 eq.	99.9	28.8
E3-2	**CP2**	2485.88	50 min	99.0	38.4
E3-3	**CP2**	2570.04	10 min	98.7	49.4
E3-4	**CP2**	2547.90	90 min + 0.5 eq.	98.7	28.9
E3-5	**CP2**	2589.99	90 min + 0.5 eq.	98.3	53.5
E3-6	**CP2**	2674.15	120 min + 1 eq.*	97.8	49.6
E3-7	**CP2**	2259.45	30 min	96.8	60.5
E3-8	**CP2**	2301.53	30 min	96.8	55.1
E3-9	**CP2**	2385.69	120 min + 0.5 eq.	99.4	73.0
E3-10	**CP2**	2561.93	10 min	97.8	61.8
E3-11	**CP2.3**	2605.20	10 min	76.8	23.4
E3-12	**CP2-R6A**	2476.82	20 min	98.6	9.5
B-3	**TiP1 (Y6)**	2385.75	60 min	92.6	9.9
B-4	**TiP1 (S13)**	2461.84	60 min	98.9	32.4

Table 2.4 Multifunctional cyclic peptides synthesised using optimised CuAAC click reaction conditions (Table 2.3). CuAAC conditions used are shown, with **CP2-T13K$_{N3}$** as an exemplar peptide. Reactions were monitored using MALDI-MS, and most reactions reached full conversion within 1 hour. In cases of incomplete conversion, an additional 0.5–1.0 equivalents (eq.) of E3 ligase binder were added to the reaction mixture, and the reaction was incubated for a further 30–60 minutes. ***E3-6** did not reach full conversion, but upon HPLC purification product was obtained in good purity. **B-3** and **B-4** were synthesised by Valeria Cobiltean, under my supervision. **B-n** products contain biotin, and **E3-n** products contain von-Hippel Lindau/CRBN E3 ligase ligands; structures of all compounds are shown in **Figure 2.5** & **Figure 2.6**.

2.2.4 Design and synthesis of photoreactive macrocyclic peptides

Covalent linkage between a target and probe can be desirable in the development of affinity probes (Section 2.1.1), to ensure the probe-target complex remains engaged during sample processing, and washing procedures. To facilitate irreversible peptide-protein linkage, photoreactive **CP2**- and **TiP1**-variants were synthesised. Different photoreactive groups can have different crosslinking activity, and/or influence on binding.[78] Hence, both benzophenone and diazirine photoactivatable groups were tested in photoreactive cyclic peptides. Photoreactive

amino acids were incorporated into multiple positions in an effort to locate sites where covalent crosslinking to the protein occurred while maintaining binding affinity (**Table 2.5**). Some of the peptides had azide or alkyne click handles, which could be used for further functionalisation after crosslinking. Other peptides did not contain click handles, but clickable variants could be made in the future if they are able to specifically label their target.

Peptide	Parent peptide	Variations to parent	Amino acid sequence	Molecular weight / g·mol^{-1}
Ph-1	**CP2**	R6(Bpa)	^{D}YVYNT(Bpa)SGWRWYTC	1987.84
Ph-2	**CP2**	Y12(Bpa)	^{D}YVYNTRSGWRW(Bpa)TC	1980.87
Ph-3	**CP2**	R6(Bpa) T13(Pra)	^{D}YVYNT(Bpa)SGWRWY(Pra)C	2023.88
Ph-4	**CP2**	W9(Bpa) T13(Pra)	^{D}YVYNTRSG(Bpa)RWY(Pra)C	1981.82
Ph-5	**CP2**	W11(Bpa) T13(Pra)	^{D}YVYNTRSGWR(Bpa)Y(Pra)C	1951.85
Ph-6	**TiP1**	Y6(Bpa) S13K_{N3}	^{D}YLIYA(Bpa)YTWWEHK$_{N3}$C	2092.37
Ph-7	**TiP1**	Y6L_p S13K_{N3}	^{D}YLIYAL$_p$YTWWEHK$_{N3}$C	1966.21
Ph-8	**TiP1**	S13K_{N3} *C*-term-L_p	^{D}YLIYAYYTWWEHK$_{N3}$C-L$_p$	2129.39
Ph-9	**TiP1**	S13K_{N3} *C*-term-K_p	^{D}YLIYAYYTWWEHK$_{N3}$C-K$_p$	2158.43

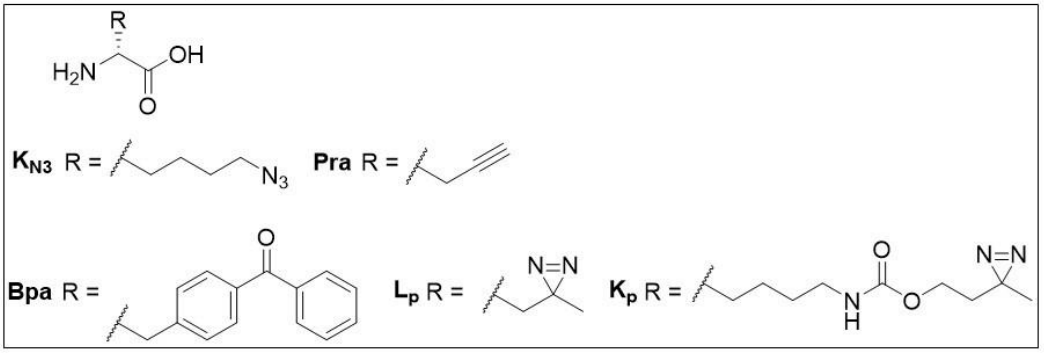


Table 2.5 Photoreactive cyclic peptides synthesised. A range of **CP2**-variants were made containing benzophenone (Bpa) photoreactive amino acids (**Ph-1** – **Ph-5**), as well as a propargyl glycine (Pra) as an alkyne click handle in **Ph-3** – **Ph-5**. **Ph-6** – **Ph-9**, **TiP1**-based photoreactive probes, contained benzophenone or diazirine groups, an azide handle was installed in each peptide. Structures of non-proteinogenic amino acids are presented below the Table. **Ph-6** – **Ph-9** were synthesised by Valeria Cobiltean, under my supervision.

Photoreactive peptides were synthesised as in Section 2.2.1, without requirement for optimisation. When using commercially sourced photo-leucine and photo-lysine derivatives the number equivalents of Fmoc amino acid used was reduced compared to natural amino acids (1.5 eq. *versus* 5 eq.), due to the high costs associated with the materials. Following photo-leucine/lysine coupling, peptides were capped with acetic anhydride before Fmoc removal (Section 6.2.1).

2.2.5 KDM4A inhibition by B-1 compared to CP2

Alteration of **CP2** to produce multifunctional variants is not expected to significantly alter their activity (Section 2.2.2.1), however it was important to establish that **CP2** modification at threonine-13 does not affect KDM4A binding. Investigation of the inhibitory activity of **B-1** (a biotinylated **CP2** peptide conjugate) compared to parent **CP2** was carried out, using recombinantly expressed KDM4A.

2.2.5.1 Recombinant expression and purification of N-His_6 $KDM4A_{1-359}$

The catalytic domain of KDM4A with an *N*-terminal His_6-tag was expressed in *Escherichia coli* (*E. coli*), and purified as previously described by Ng *et al.* (2007)[233] (Section 6.6). A plasmid for bacterial expression of KDM4A (pNIC28A-Bsa4-*N*-His_6KDM4A_{1-359}) was transformed into BL21 DE3 *E. coli* cells, and a single colony was used for a starter culture, from which a large scale culture was produced. Overexpression of KDM4A was induced using isopropyl β-D-1-thiogalactopyranoside and cells were incubated at 18 °C overnight before harvesting the cell pellet. Cell pellet was stored at -80 °C until purification, when the pellet was suspended in binding buffer (Section 6.6.2), and lysed by sonication. Purification was performed using fast protein liquid chromatography (FPLC), *via* a Ni(II) HisTrap column followed by size exclusion chromatography. The first stage of purification used a Ni(II) column to capture His_6-tagged protein. The column was subsequently washed prior to protein elution using a high imidazole buffer. Fractions containing KDM4A were identified using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and were pooled and concentrated before size exclusion chromatography. A Superdex S200 column was used for size exclusion chromatography in a buffer containing 10 mM HEPES (pH 7.5), 300 mM NaCl, 5% glycerol. Fractions containing KDM4A with the highest purities (>90% by SDS-PAGE, **Figure 2.7**) were pooled and concentrated to approximately 200 µM. Samples were aliquoted and flash frozen in liquid nitrogen before storage at -80 °C. The identity of the isolated protein was

confirmed using RapidFire MS; the observed mass, 44.265 kDa, is the same as the theoretical mass.[b]

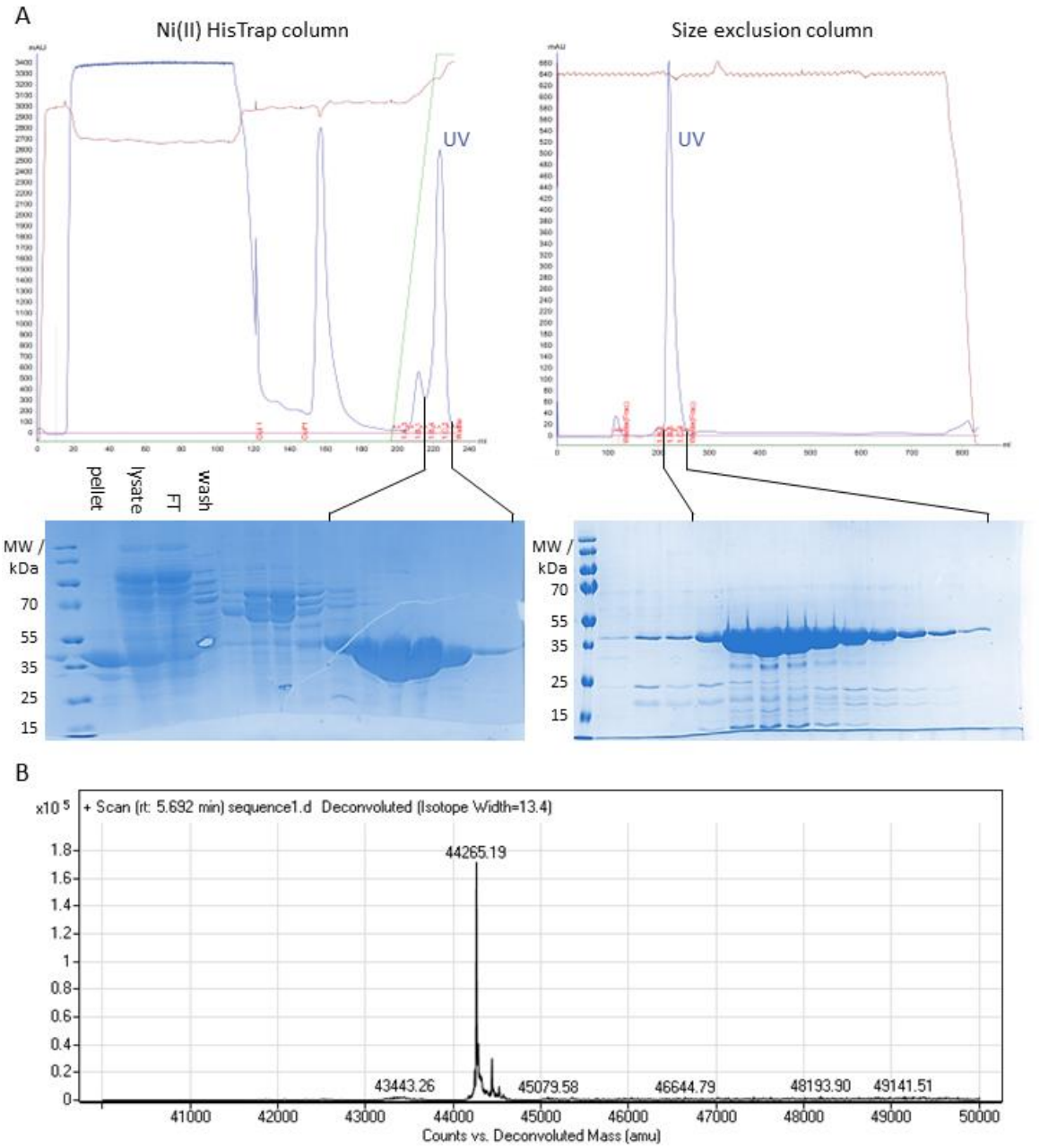


Figure 2.7 His_6-$KDM4A_{1-359}$ protein purification and MS analysis. (A) KDM4A was purified using a Ni(II) HisTrap column (left), followed by size exclusion chromatography (right); purity of fractions after each column were analysed using SDS-PAGE, and the purest fractions pooled. The HisTrap column gel shows cell pellet, suspended lysate, column flow through (FT), and wash buffer from the HisTrap column, as well as elution fractions. (B) Deconvoluted RapidFire MS shows protein obtained was in accordance with the theoretical mass of the KDM4A construct (44.265 kDa). RapidFire MS performed by Dr Leah Taylor-Kearney.

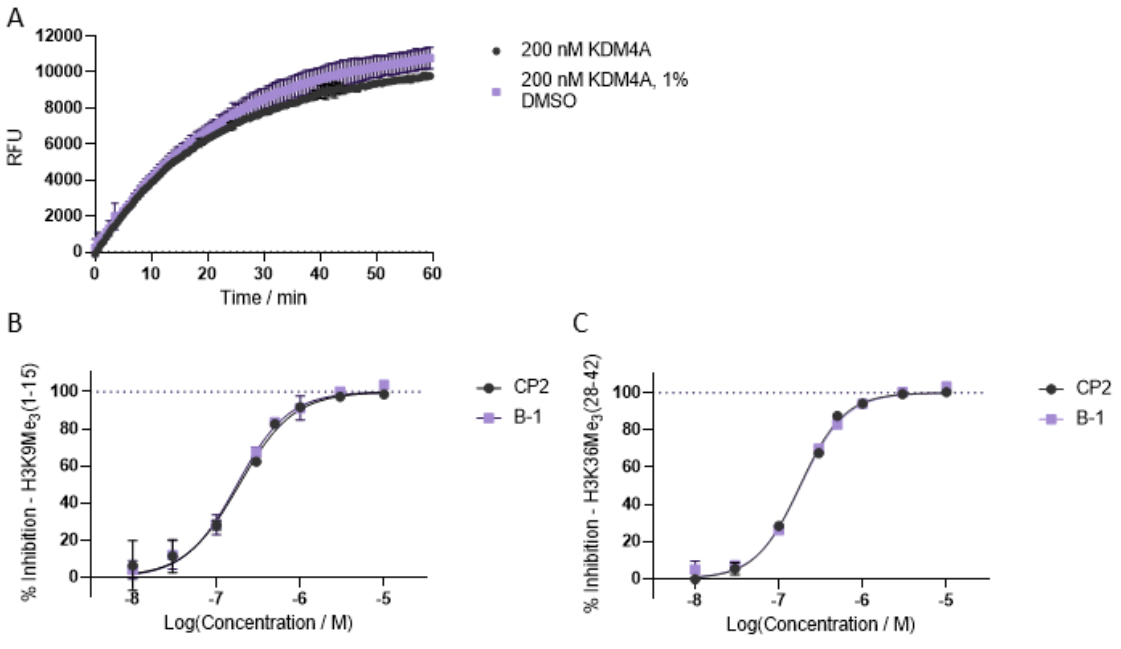
[b] RapidFire MS performed by Dr Leah Taylor-Kearney.

*2.2.5.2 Formaldehyde dehydrogenase coupled assay to compare inhibitory activity of **B-1** to **CP2***

An enzyme assay was used to confirm the inhibitory activities of **CP2** and **B-1**. There are several enzyme assays available for measuring the activities of KDMs, and for this application a real time, formaldehyde dehydrogenase (FDH) coupled, fluorescence-based assay was used. Couture *et al.* (2007)[234] coupled KDM demethylation activity with oxidation of formaldehyde to formate by FDH, and the protocol was later adapted for high-throughput screening of KDM4 inhibitors.[235] FDH reduces nicotinamide adenine dinucleotide (NAD^+) to NADH during oxidation of formaldehyde, NADH is fluorescent and can be monitored using a plate reader. The rate of formaldehyde dehydrogenase turnover is faster than KDM turnover, and thus the rate limiting variable in the assay is KDM turnover. Furthermore, inhibitor kinetics identified using this assay correlate to those identified using alternative assays.[236,237]

KDM4A demethylates H3K9$Me_{3/2}$ as well as H3K36$Me_{3/2}$ (Section 1.1.2.1), and inhibition by **B-1** compared to **CP2** was tested for both H3K9Me_3 and H3K36Me_3 substrates. Activity of KDM4A (200 nM) was tested using H3K9Me_3(1–15) substrate (**Figure 2.8**A), prior to measuring IC_{50} values for **B-1** and **CP2**. Enzyme activity was analysed in the presence and absence of 1% DMSO, and very little difference in the activity was observed. Inhibition data was collected within the linear range of enzyme activity for each substrate; for H3K9Me_3(1–15) 15 minutes was used (**Figure 2.8**A, Appendix 2), and for H3K36Me_3(28–42) 30 minutes was used (Appendix 2).

Inhibition data shows **CP2** and **B-1** have very similar potencies using both substrates (**Figure 2.8**B-C). This indicates that functionalisation of **CP2** at the threonine-13 position does not impact the activity of the resulting **CP2**-variants, and suggests that **CP2-T13K$_{N3}$**-based peptide conjugates are appropriate for study of KDM4A–C in biological systems.



Substrate	CP2 IC_{50} / nM	B-1 IC_{50} / nM
$H3K9Me_3$	187	176
$H3K36Me_3$	178	182

Figure 2.8 KDM4A inhibition activity of CP2 and B-1 was analysed using an FDH coupled assay. (A) Activity of 200 nM KDM4A using $H3K9Me_3$(1–15) substrate is similar with and without 1% DMSO. (B/C) IC_{50} data using $H3K9Me_3$(1–15)/$H3K36Me_3$(28–42) substrates, IC_{50} values are displayed in the table and show that there is little difference in inhibitory activity between **CP2** and **B-1**. (n = 3, mean ± standard deviation).

2.3 Discussion

This chapter details the synthesis and characterisation of multifunctional peptides, which will be used as affinity probes in Chapter 3, and for KDM4A protein degradation studies in Chapter 4.

Click handles and/or photoreactive amino acids were installed into the peptides, and most peptides were isolated using standard SPPS protocols. However, extensive optimisation was required to isolate and purify **CP2.3-P13K$_{N3}$**, employing manual backbone *N*-methylation in the final synthesis. CuAAC click chemistry was used to produce peptide conjugates, installing biotin or E3 ligase binders. Finally, the activity of **B-1** was compared to the parent cyclic peptide, **CP2**, and inhibitory activity was maintained upon conjugation of $(PEG)_4$-biotin. This validates threonine-13 as an appropriate attachment point for the functionalisation of **CP2**.

CuAAC click reaction conditions optimised for **B-1** were robustly applied to other multifunctional probes synthesised, and all click reactions tried using these conditions produced the desired

products. Obtaining sufficient cell permeability may be challenging for the peptide conjugates, which are heavier than the parent peptides and contain more hydrogen bonding groups. To circumvent this issue, SPAAC could be used to couple promising multifunctional peptides inside biological systems in the future. **CP2.3-P13K$_{N3}$** is likely to be best candidate for the development of *in cellulo* SPAAC click chemistry, as the parent peptide is cell active.[40] Furthermore, optimised synthesis conditions for **CP2.3-P13K$_{N3}$** (Section 2.2.2.1.1) allowed isolation of pure peptide on a double-digit milligram scale.

The structure of peptides synthesised was inferred using mass spectrometry, based upon the theoretical mass for the amino acid sequence. Other techniques could be employed to confirm the amino acids are in the intended order, such as tandem mass spectrometry.[238] Further information on the structure of the peptides could be gained using both NMR and X-ray crystallography.[239,240]

Techniques used to synthesise multifunctional variants of **CP2** are broadly applicable to other peptides, which was demonstrated through successful production of **TiP1**-based multifunctional probes. Structure activity relationship data is extremely helpful in identifying suitable sites for functionalisation; while a co-crystal structure can give a very detailed map of probe-target binding, it was not necessary for the methods of multifunctionalisation used, and structure activity relationship data gained from an alanine scan facilitated synthesis of **TiP1**-based multifunctional probes. This project could be diversified by installing other types of functionality to the parent probes for a myriad of applications. Furthermore, interest in bifunctional probes is rapidly increasing, along with an expanding range of functionalities which can be installed.[89,241,242] An interesting expansion to this project could explore opposing functionalities installed in a single parent probe, for example, investigating protein degradation using PROTACs, compared to protein stabilisation using DUBTACs. This would be especially interesting using TET-targeting probes, as

conflicting data on the roles of TETs in cancers has been found, with evidence for both the upregulation and downregulation of TETs in cancer (Section 1.1.2.2).[53,54]

Chapter 3 Macrocyclic peptide affinity probes

3.1 Introduction

One of the main reasons drugs fail in clinical trials is insufficient understanding of the role of the pharmaceutical target in disease pathology.[71] Tools such as affinity probes are essential for rigorous target validation in medicinal chemistry. Affinity probes can be used to learn about the biological role of a target, and for drug selectivity studies.[83]

In the context of this project, affinity probes are referred to as molecules containing a chemical or biological entity that allows the probe to be captured while engaged with the target. Affinity modality can take different forms, for example, biotin, IgG, or a chemical handle; whereby chemical handles are used to conjugate a reporter group after binding.[243] By this definition, covalent activity-based probes, reversible binding biotin-conjugated probes, and antibodies are all affinity probes.[126,243] Affinity probes can be used in a wide variety of contexts to enrich their targets, often with other protein binding partners,[243,244] or interacting genomic DNA.[87,88] A similar workflow is used for target enrichment by many affinity probes, sometimes referred to as a pulldown assay. Briefly, affinity probes are incubated with a complex mixture containing the target; a solid support is used to extract the probe-target complex (*e.g.* magnetic beads); and a read out is used to identify presence of the target and/or target interactions (**Figure 3.1**). There are a number of more detailed considerations for each experiment, such as: whether the probe can be dosed on whole cells prior to lysis and affinity purification;[214] or whether to use formaldehyde crosslinking to aid preservation of protein-DNA and protein-protein interactions.[88,244]

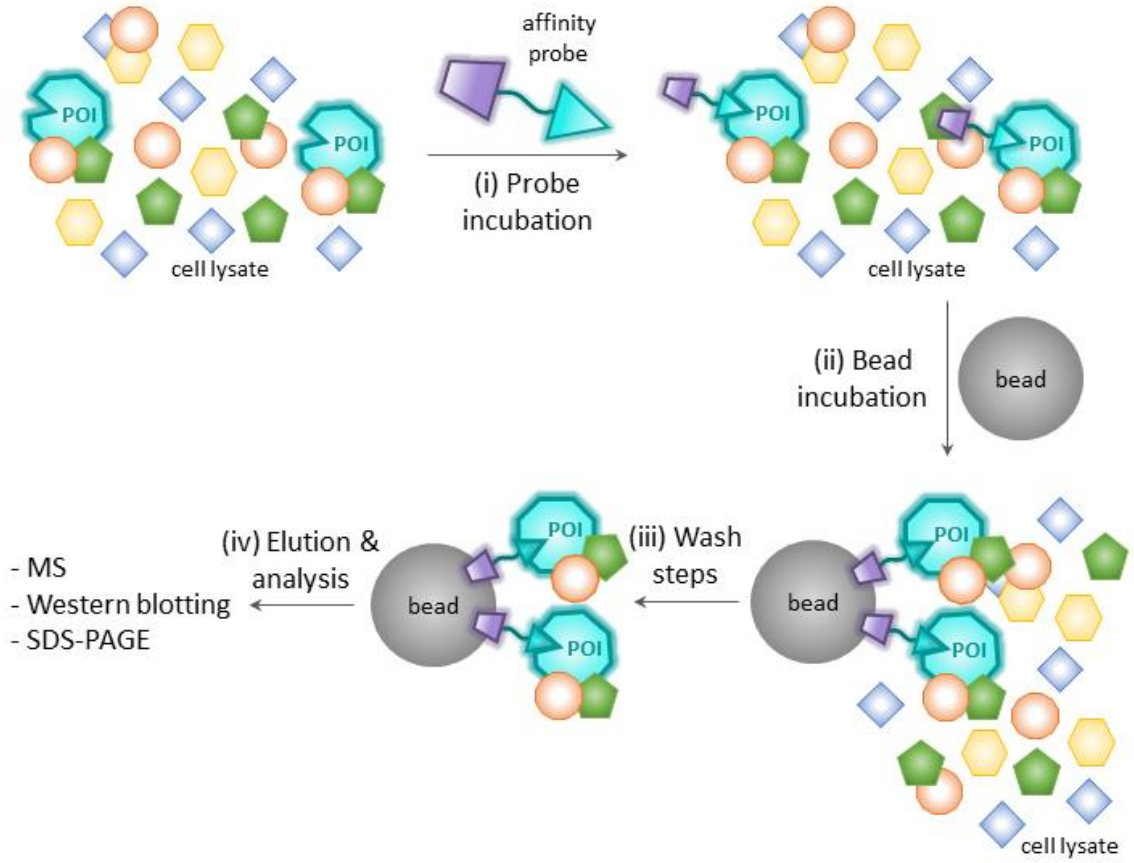


Figure 3.1 Schematic representation of a protein of interest (POI) pulldown from cell lysate using a bifunctional probe. (i) The probe is incubated with the cell lysate, allowing it to bind the POI. (ii) Affinity capture beads are incubated with the probe-lysate mixture, and the beads bind the bifunctional molecule. (iii) Beads are washed, removing unbound proteins. (iv) Finally, the proteins bound to the bead are eluted and analysed. Some common experimental readouts are listed: MS may be used to identify proteins eluted from the beads,[214] western blotting can be used to confirm if a specific protein was eluted,[76] and SDS-PAGE protein gels can be used to visualise proteins eluted based on molecular weight (however, ≥0.5 ng protein needs to be present to be visible using most protein stains).[245]

Proteomics is an important component of target validation in drug discovery.[214] MS-based proteomics can be used to compare the proteome of diseased tissues to healthy tissues, or probes may be used to extract the POI and binding partners from biological material. Furthermore, drug selectivity studies can be performed to establish off-target effects of a potential drug molecule, or even for repurposing known drugs.[83,246] Both antibody (Section 1.2.2.5) and small molecule affinity probes (Section 1.2.2.1) have been employed in proteomics to identify the protein interactions of drug targets.[83,132] When using small molecules, linking the affinity moiety after probe binding reduces the impact of the affinity group on the binding kinetics of the probe. Many small molecule chemoproteomic probes bind irreversibly to their target, by linking to reactive

groups within proteins (such as cysteine side chains) or by activation of photoreactive groups using light.[82]

Affinity probe-based proteomics has been widely applied to the study of epigenetic complexes, epigenetic inhibitor selectivity studies, and for identification of novel epigenetic targets.[76] The HDAC inhibitor vorinostat (also known as suberoylanilide hydroxamic acid (SAHA)) has been the subject of many chemoproteomics experiments. Addition of benzophenone and a terminal alkyne facilitated development of a photoreactive affinity probe based upon vorinostat, SAHA-BPyne (**Figure 3.2**). SAHA-BPyne was used to identify different HDACs present in disease models, and their protein-interactions.[247,248] Another study immobilised vorinostat and givinostat, both pan selective HDAC inhibitors, on Sepharose beads allowing MS-based proteomics to profile pK_d^{app} for 16 HDAC inhibitors in K562 myelogenous leukaemia cell extracts. The technique was further developed, replacing proteomics as the readout with a dot blot array visualised using multiplexed fluorescent antibodies, allowing a small library of HDAC inhibitors to be tested in a high throughput work flow.[249] MS-based proteomic selectivity studies demonstrated their value in identifying drug interactions when BET family protein activity was identified for LY294002 (**Figure 3.2**), a commonly used phosphoinositide 3-kinases probe.[250] Standard selectivity studies tend not to extend to protein classes beyond the target, but MS-based proteomics allow the drug interactions across the proteome to be mapped in one experiment.[214] Proteomics was also key in identifying novel and known histone modification interactors, using photoreactive histone peptide baits in multiple studies (example probe in **Figure 3.2**). Photoreactive modified histone tail variants were incubated with cell lysates and irradiated to initiate crosslinking, both benzophenone and diazirine photoreactive groups have been used (Section 1.2.3.1).[137,251,252] Antibodies have also been employed to study epigenetic complexes, but there is less literature describing their use in MS-based proteomics, possibly due to: high costs associated with raising antibodies against specific epitopes, lack of specificity and/or batch-to-batch variability of

antibodies, and the potential for antibodies to interfere with the protein interactions they are being used to eluciadate.[76,160]

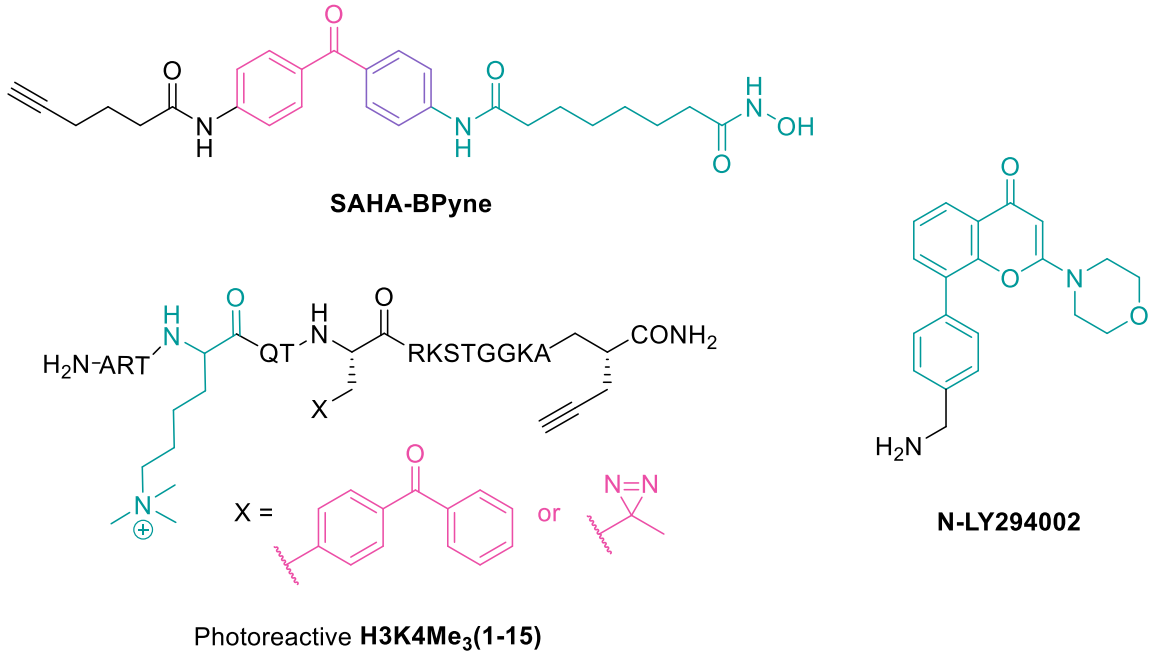


Figure 3.2 Structures of selected epigenetic affinity probes discussed. SAHA-BPyne is a photoaffinity probe which has been used to identify HDACs in different disease models. The target binding moiety (teal) in SAHA-BPyne overlaps with the photoreactive benzophenone group (pink), and the overlap is indicated (purple); an alkyne affinity handle is included (black).[247] N-LY294002 is an amine variant of LY294002 (teal), the amine (black) was added to allow the compound to be immobilised on beads.[250] Photoreactive H3K4Me$_3$(1–15) variants were used to identify novel epigenetic reader enzymes, KMe$_3$ (teal), benzophenone/aziridine photoreactive groups (pink), and an alkyne handle are included.[137]

The high selectivity and affinity of antibodies for their targets means irreversible binding is not required when used in immunoprecipitation experiments.[126] However, due to their large size (≈150 kDa) they may interrupt PPIs,[81] and they are rarely cell permeable.[253] Conversely, small molecule affinity probes (<500 Da) are more likely to benefit from good cell permeability, allowing them to be used directly in cellular disease models,[243] avoiding issues caused by interruption of PPIs during cell lysis.[254] However, small molecules tend to have lower affinity for their targets than peptides or antibodies.[147] Therefore, covalent linkage to the target protein may be required for successful small molecule affinity purification, to ensure the probe and POI remain engaged during sample processing.[214] Macrocyclic peptides occupy a goldilocks space between small molecules and antibodies (Section 1.3).[147] They benefit from high selectivity for their targets, improved binding affinity over small molecules, and the reduced size of peptides (<5 kDa) compared to

antibodies (≈150 kDa) means peptides should have less impact on PPIs.[85] There are examples of peptides making excellent affinity reagents; including McAllister *et al.* (2018),[85] who discovered a non-competitive PHD2 binding cyclic peptide (≈3 kDa) that captures PHD2 (46 kDa) bound to HIF1-α (93 kDa) in cell lysates (Section 1.2.3.1). The enrichment of PHD2 in complex with substrate, suggests other PPIs will be preserved, presenting an opportunity for detailed investigation of the protein interaction partners of PHD2.[85] Furthermore, while the development of cell-permeable peptides frequently requires optimisation,[147] it is generally less challenging than preparing cell permeable antibodies.[253] This suggests that peptides could make promising affinity agents, with high potency and selectivity, and without interrupting the protein interactions investigated.

3.1.1 Aims

In this chapter cyclic peptides are tested as affinity probes for different epigenetic targets. Biotin-conjugated affinity cyclic peptide probes synthesised in Chapter 2, targeting KDM4A-C and TET enzymes (based upon **CP2** (ref. [40]) and **TiP1** (ref. [55])), are profiled in cell lysates, exploring the efficiency and selectivity of reversibly binding peptide affinity probes. Furthermore, covalent attachment of peptides is explored using photoreactive **CP2**- and **TiP1**-derivatives. The aims of the work can be summarised as follows:

- Optimisation of the production of HEK293T cell lysates overexpressing FlagKDM4A, to identify an appropriate system to assess target engagement by KDM4A-targeting peptide-biotin conjugate, **B-1**, in cell lysates.
- KDM4A-targeting and TET1-targeting peptide-biotin conjugates are tested as pulldown probes in cell lysates. Target engagement is assessed using western blots, and SDS-PAGE protein gels.
- Preliminary MS-based proteomics experiments will be used to assess the suitability of the biotinylated cyclic peptides as chemoproteomic probes.

- Photocrosslinking of photoreactive cyclic peptide variants will be explored, in work towards comparison of reversible and irreversible target engagement.

3.2. Results

3.2.1 Biotinylated macrocyclic peptides as affinity probes

Biotin has an exceptionally high affinity for (strep)avidin. This interaction has been exploited for affinity purification, and detection of (bio)molecules (Section 2.1.1).[200] In this Chapter, biotin was exploited as an affinity motif to develop reversibly binding affinity cyclic peptides.

Biotinylated macrocyclic peptides targeting KDM4A-C and TET enzymes were synthesised, **B-1** – **B-4** (**Figure 3.3**), by conjugating azide containing macrocyclic peptides to biotin-$(PEG)_4$-alkyne *via* CuAAC click chemistry (synthesis described in Section 2.2.3). These probes were used as reversibly binding affinity agents, whereby target enrichment relies on sufficiently high affinity of the probe-target complex to remain engaged throughout wash steps. **B-1** – **B-4** were tested in pulldown studies against their targets (Section 3.2.1), and **B-1** was further used in chemoproteomics experiments (Section 3.2.1.2).

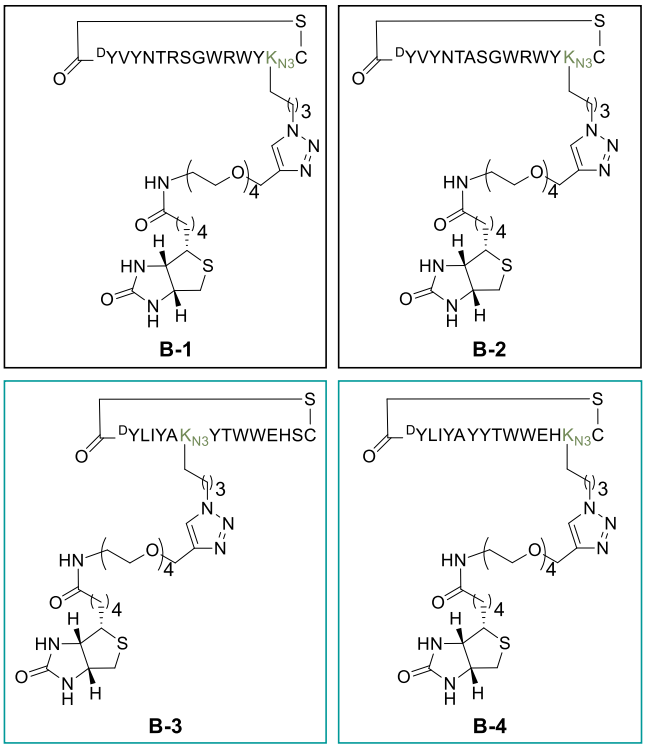


Figure 3.3 Structures of biotinylated macrocyclic peptides. B-1 is derived from **CP2** which binds KDM4A-C, and **B-2** from **CP2-R6A** which has >50-fold lower KDM4A inhibitory activity compared to **CP**2.[40] **B-3**/**B-4** (teal boxes) are derived from **TiP1** which inhibits TET1/2, with greatest affinity for TET1.[55]

*3.2.1.1 Optimisation of KDM4A pulldown from cell lysates using **B-1***

To confirm that biotinylated macrocyclic peptides can engage with their targets in complex mixtures, an efficient and robust method to detect and quantify the target protein is needed. Western blots are the method of choice for confirming the presence of a POI in a sample (Section 6.5.2), for which high affinity, selective antibodies are required. High quality and well validated antibodies against epitope tags can be extremely useful for detecting low levels of protein, or for cases where antibodies against the POI are less well validated/selective.[255] Therefore, in order to establish target-pulldown by **B-1**, affinity pulldown was initially assessed in lysates overexpressing FlagKDM4A. The monoclonal Flag antibody used to detect FlagKDM4A (Merck, F1804) has been tested in many publications,[255–257] and can detect very small levels of protein (<10 ng).[258]

CP2 cyclic peptide selection was performed on the recombinantly expressed Jumonji-C catalytic domain of KDM4A. However, for a biologically relevant affinity probe, engagement with full length

proteins from mammalian cells was desired. The parent peptide of **B-1** (**CP2**) has been shown to inhibit purified full length KDM4A from HEK293T cells,[40] which made **B-1** an excellent choice for preliminary peptide-biotin conjugate target engagement studies. Cell lysates were used for these experiments to ensure cell permeability issues did not impact target engagement of the peptides.

3.2.1.1.1 Optimisation of FlagKDM4A overexpression and lysis for efficient affinity pulldown

Initially, production of cell lysates overexpressing FlagKDM4A was optimised to ensure sufficient protein was expressed in cells to observe target engagement by **B-1** (Section 3.2.1.1.2). Both protein expression and lysis methods were optimised for robust detection of FlagKDM4A using western blots. Experiments were performed in 6-well plates, seeded with HEK293T cells (0.2×10^6 cells per well), 24 hours prior to transfection. A 3:1 ratio of FuGENE HD transfection reagent to DNA was used, which had previously been identified as a suitable reagent, and an optimal ratio of transfection reagent to DNA.[259] Cells were then washed, lysed, and the lysate collected for western blot analysis. Several factors were investigated for optimisation, including the lysis buffer, the amount of DNA transfected, and the length of transfection. Both FlagKDM4A wild type (FlagKDM4Awt), and FlagKDM4A mutant (FlagKDM4Amut) constructs were tested in optimisation experiments, although only FlagKDM4Awt was used in subsequent experiments. The FlagKDM4Amut plasmid contains an H188A inactivating mutation, and including this plasmid in optimisation experiments will allow for direct comparison of active/inactive target engagement by **B-1** in future work.

(i) Lysis buffer optimisation: Efficient cell lysis and protein extraction is a prerequisite for affinity pull down assays. Various lysis buffers are available for lysis of cultured mammalian cells, but they can differ in efficiency based on cell type and protein target (as well as downstream applications).[260] Initially, nuclear lysate extraction (Section 6.8.2.2) was tested, to see whether yield of the nuclear protein KDM4A would be enhanced via nuclear fractionation. FlagKDM4A was present in both the nuclear fraction and the remaining lysate (Appendix 3.1), indicating that whole

cell lysates were needed for accurate quantitation of KDM4A in cells. Whole cell lysate was therefore used in subsequent experiments. In the initial whole cell lysate optimisation experiment (**Figure 3.4**A), RIPA buffer (Merck, **Table 3.2**) was used, which contains a mixture of denaturing and non-denaturing detergents: IGEPAL, sodium dodecyl sulfate (SDS), and sodium deoxycholate. In subsequent optimisation experiments an alternative lysis buffer was tested using only IGEPAL detergent (**Figure 3.4**B/C), 'IGEPAL buffer'. IGEPAL buffer (**Table 3.2**) uses only one detergent (IGEPAL), which is non-denaturing, and this may preserve more PPIs during cell lysis as fewer proteins should misfold. Both RIPA and IGEPAL buffers were successful at solubilising KDM4A, and IGEPAL buffer was chosen for larger scale cell lysate production as it is a mild and non-denaturing lysis buffer (**Figure 3.4**).

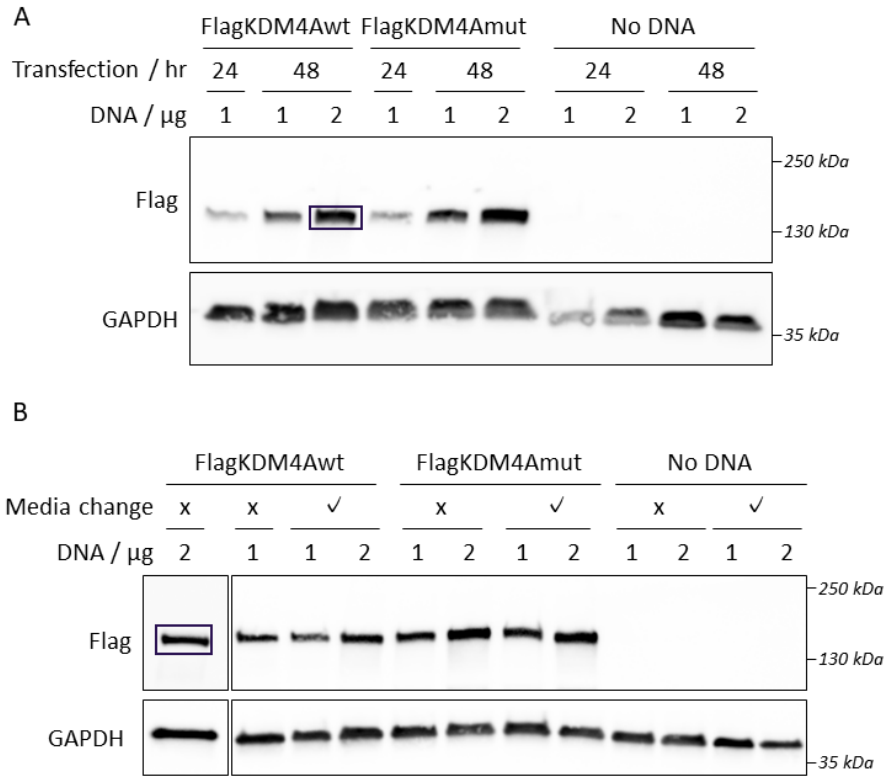


Figure 3.4 Optimisation of conditions for production of HEK293T cell lysate overexpressing FlagKDM4A. (A) Comparison of 1 µg *versus* 2 µg DNA, and cell lysis 24 hours *versus* 48 hours post-transfection using RIPA buffer. For both wildtype (wt) and mutant (mut) plasmids, greater expression was obtained using 2 µg DNA with cell lysis 48 hours post-transfection. (B) Effect of media change 24 hours after transfection, prior to cell lysis 48 hours post-transfection using IGEPAL buffer (✓ = media change, x = no media change). Media change was found to slightly reduce the amount of KDM4A protein. The left hand sample was assessed in a separate experiment, as sufficient protein content for this sample was not obtained in the main experiment. Purple boxes indicate the conditions used in optimised lysis: 2 µg DNA, cell lysis 48 hours post-transfection, and no media change. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is used as a loading control. FlagKDM4A = 122 kDa, GAPDH = 36 kDa.

(ii) DNA transfection optimisation: Two FlagKDM4Awt and FlagKDM4Amut DNA concentrations were tested, 1 μg and 2 μg per well (0.2×10^6 cells per well, 6-well plate). FlagKDM4Awt and FlagKDM4Amut protein expression was observed at each concentration tested, but more protein expression was found when 2 μg DNA compared to 1 μg DNA was added per well (**Figure 3.4**). Greater protein expression was also observed when cells were lysed 48 hours post-transfection compared to 24 hours (**Figure 3.4**). Experimenting with changing the media, to remove transfection reagent 24 hours post-transfection and 24 hours prior to lysis, did not increase FlagKDM4A protein expression (**Figure 3.4**B). Strong KDM4A protein expression was observed transfecting 2 μg DNA, 48 hours prior to lysis, without changing media. Thus, these were used as optimal conditions for subsequent pulldown assays.

The protein content of cell lysate samples obtained in these experiments was variable, as technical challenges were experienced during wash steps, due to the weak attachment of HEK293T cells compared to other adherent cell types.[261–263] Samples from some of the transfection conditions tested did not contain sufficient protein content to be analysed by western blot (**Figure 3.4**), and were omitted from the analysis. An independent experiment was required to assess protein expression using the optimal cell lysate production conditions identified (**Figure 3.4**B, LHS), and results are shown alongside the original experiment (**Figure 3.4**B, RHS). Note that despite the different experiments, LHS and RHS gels contain the same protein loading, as indicated by comparative GAPDH band intensities. Final optimal conditions were: transfection with 2 μg DNA, lysis 48 hours post-transfection with IGEPAL buffer, and no cell culture media change. Subsequently, the wash procedure of cells prior to lysis was optimised, such that samples were rarely obtained with insufficient protein content in future experiments.

Slightly higher FlagKDM4Amut protein expression levels were observed when compared to FlagKDM4Awt protein expression (**Figure 3.4**). This could be because overexpression of active FlagKDM4Awt altered global histone methylation status such that cells were less viable. Precedent

for this is established in findings that the overexpression of KDM4s is linked to several diseases – including multiple cancers.[30] However, further investigations would be required to validate whether overexpression of FlagKDM4Awt reduced cell viability and/or altered histone methylation status.

3.2.1.1.2 Pulldown of KDM4A from HEK293T cell lysates containing overexpressed FlagKDM4A

Having established an optimal transfection and lysis method, **B-1** was next used as an affinity capture reagent for KDM4A in cell lysates containing overexpressed FlagKDM4A. The affinity purification set up was inspired by immunoprecipitation methods using antibodies.[81] Briefly, **B-1**, **B-1** + **CP2**, or biotin (various concentrations) were incubated with cell lysate for 30 minutes at 4 °C. Lysates were subsequently incubated with streptavidin magnetic beads with end-over-end rotation for 60 minutes at 4 °C. Beads were washed three times using tris buffered saline containing 0.1% Tween20 (TBST), followed by the elution of proteins bound to the beads using SDS-PAGE sample loading buffer. The immunoprecipitated samples were then separated from the beads. Samples from the original lysate ('input'), pulldown purification ('pulldown'), and the remaining lysate after pulldown ('unbound fraction') were analysed by western blot probed with Flag antibody and/or silver-stained SDS-PAGE protein gels.

Capture of KDM4A by **B-1** (1 µM), but not biotin (1 µM), could be seen by both a western blot probed with Flag antibody, and by a silver-stained SDS-PAGE protein gel (**Figure 3.5**, FlagKDM4A = 122 kDa). A band at approximately 140 kDa from the **B-1** pulldown was cut out from the protein gel (purple box, **Figure 3.5**A), and proteomics MS was used to confirm the presence of KDM4A in the band (**Table 3.1**).[c] The presence of non-specific pulldown bands in both biotin and **B-1** pulldowns demonstrate the importance of negative controls in these affinity purification experiments. Gel samples from the three wash steps were included in the protein gel, and by the

[c]Dr Andrew Porter within Newcastle University Protein and Proteome Analysis service (NUPPA) performed MS and assigned the proteins in the destained gel band.

third wash no visible protein bands were detected, suggesting that three washes are appropriate for these experiments.

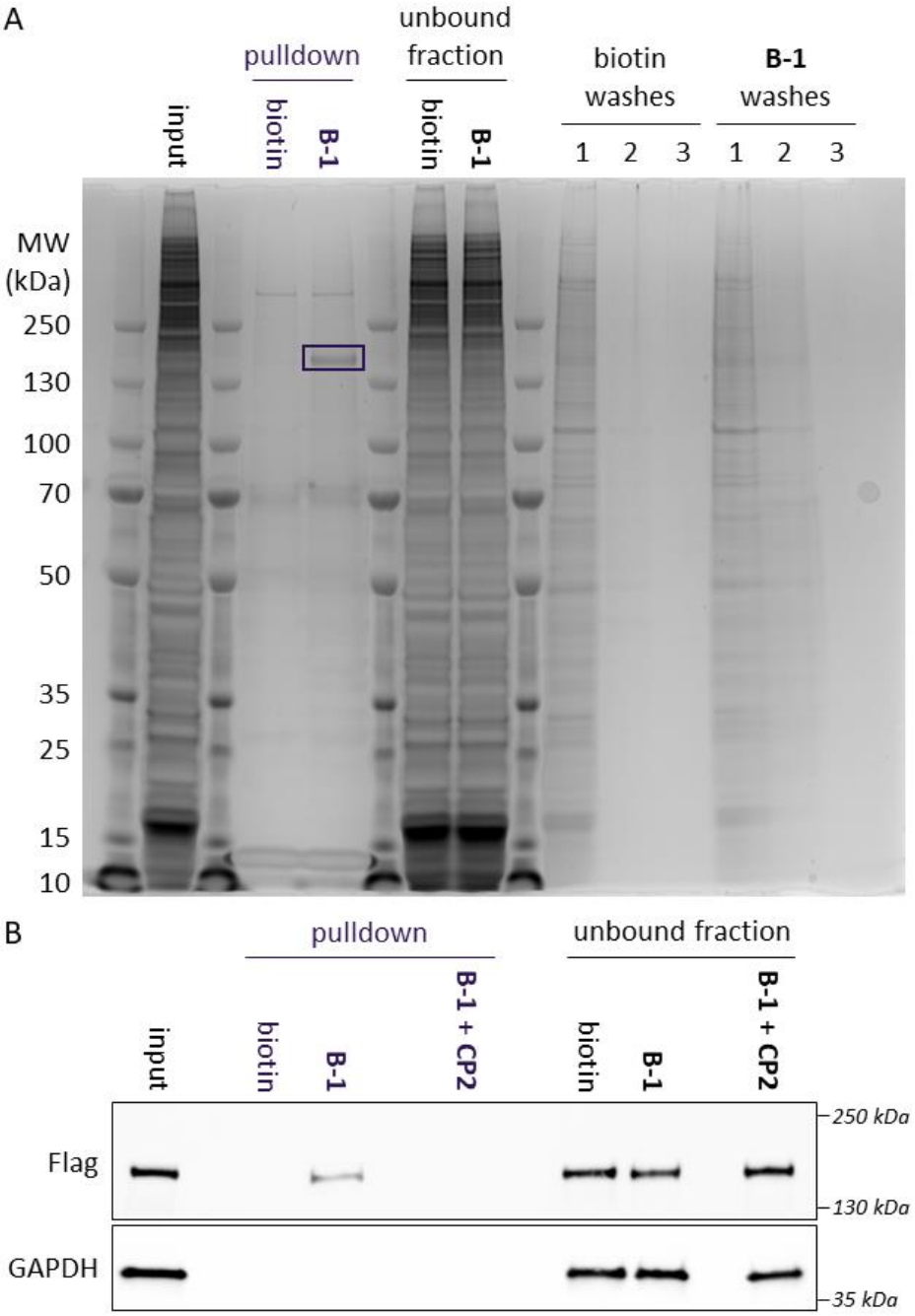


Figure 3.5 B-1 captures KDM4A from HEK293T cell lysates overexpressing FlagKDM4A. (A) Silver stained SDS-PAGE protein gel showing pulldown using **B-1** (1 μM) *versus* biotin (1 μM). Purple box shows band containing KDM4A, as confirmed by MS-based proteomics (**Table 3.1**). Representative silver stain from three biological replicates. (B) Analysis using western blot probed with Flag antibody from pulldown of **B-1** (1 μM), biotin (1 μM), or **B-1** + 100-fold **CP2** competition (1 μM **B-1** + 100 μM **CP2**). KDM4A was not detected when competition was included, indicating that **B-1** pulldown is specific for the same KDM4A binding site as **CP2**. One biological replicate was performed. 'Input' contains unmodified cell lysate used in the experiment, 'pulldown' (purple text) comprises samples eluted from affinity capture beads, 'unbound fraction' contains lysate removed from beads prior to washes, and 'washes' contain TBST buffer retained from wash steps. FlagKDM4A = 122 kDa, GAPDH = 36 kDa.

Protein	Molecular Weight / kDa	Log(I)	rI	Log(e)	pI
Cationic trypsin	25.4	7.04	281	-398	8.2
Serine protease 1	28.1	5.9	18	-19.4	6.9
KDM4A	120.6	3.75	3	-13.1	5.6
Keratin	62	3.98	1	-8.8	5.1
Ribosomal protein 27a	18	3.71	3	-4.5	9.7

Table 3.1 Proteins recovered from a ≈140 kDa band cut out of an SDS-PAGE protein gel loaded with B-1 (1 µM) pulldown sample, from HEK293T cell lysate overexpressing FlagKDM4A (Figure 3.5). When expected contaminants are discounted (grey text), KDM4A is the top protein identified in the gel band. Log(I) = sum of raw spectrum intensities, rI = number of peptides found, Log(e) = probability of finding the protein stochastically, pI = isoelectric point for the intact gene product. One biological replicate was performed.

Specificity of the **B-1** pulldown was tested by competing **B-1** (1 µM) with 100-fold **CP2** (100 µM) (**Figure 3.5**B). Western blot analysis showed that KDM4A was not recovered when **CP2** competition was used, suggesting that **B-1** and **CP2** occupy the same KDM4A binding site, and that KDM4A enrichment is **B-1** driven and specific. This experiment was crucial in validating that the mode of action of **B-1** is the same as the parent peptide **CP2**. Similar inhibitory activity was identified for both **B-1** and **CP2** (Section 2.2.5.2), which suggested that the conjugation of biotin to **CP2** does not significantly alter the activity, further evidence for this is found in these experiments demonstrating **B-1** FlagKDM4A target engagement in cell lysates.

Taken together, **B-1** was able to affinity purify overexpressed FlagKDM4A from HEK293T cell lysates. These results suggest that **B-1** may be able to enrich KDM4A in other biological systems, which was assessed in HEK293T cell lysates expressing endogenous levels of proteins (Section 3.2.1.1.3). Future work could investigate whether **B-1** can engage KDM4B/C in cell lysates, as these enzymes are also inhibited by **CP2** (the parent peptide).

3.2.1.1.3 Pulldown of KDM4A from endogenous HEK293T cell lysates

Having established conditions to confirm **B-1** target engagement for KDM4A using a highly selective monoclonal Flag antibody[255] (section 3.2.1.1.2) in FlagKDM4A overexpressed cell lysates, next, **B-1** was tested as a pulldown probe for capturing endogenous KDM4A. (Untransfected) HEK293T cell lysates were used for endogenous target engagement studies (**Figure 3.6**), as KDM4A is expressed in HEK293T cells.[264,265] In these experiments pulldown was assessed using western

blots probed with a polyclonal KDM4A antibody (Bethyl Laboratories, A300-861A), and silver stained SDS-PAGE protein gels. Note that polyclonal antibodies suffer greater batch-to-batch variability compared to monoclonal antibodies (such as the Flag antibody used in Section 3.2.1.1.2). Polyclonal antibodies are formed of a combination of multiple immunisation-dependent antibodies, whereas monoclonal antibodies contain only one sequence.[266] Therefore a monoclonal antibody was used in the initial **B-1** pulldown optimisation assays (Section 3.2.1.1.2).

HEK293T cells were seeded in 10 cm cell culture dishes (2.2×10^6 cells per dish), and grown to 90% confluency for 48–72 hours. Cells were then washed (analogously to FlagKDM4A overexpressing cell lysates (Section 3.2.1.1.1, Section 6.8.2.1)) and lysed using IGEPAL buffer (**Table 3.2**).

B-1 was tested at two concentrations (1 µM and 10 µM) in initial endogenous pulldown assays, and the samples were analysed using silver stained SDS-PAGE protein gels and western blots probed with KDM4A antibody. Using western blots, KDM4A bands (expected KDM4A MW = 121 kDa) could be found from **B-1** (but not biotin) pulldown at both concentrations, indicating target engagement (**Figure 3.6**A). A stronger band for KDM4A pulldown was found at 10 µM **B-1** compared to 1 µM, suggesting that **B-1** is a limiting factor under these conditions. Therefore 10 µM **B-1** was used in most subsequent experiments. No band corresponding to KDM4A was visible on silver-stained SDS-PAGE protein gels, likely due to lower levels of KDM4A in these cell lysates compared to the overexpressed system (Section 3.2.1.1.2).

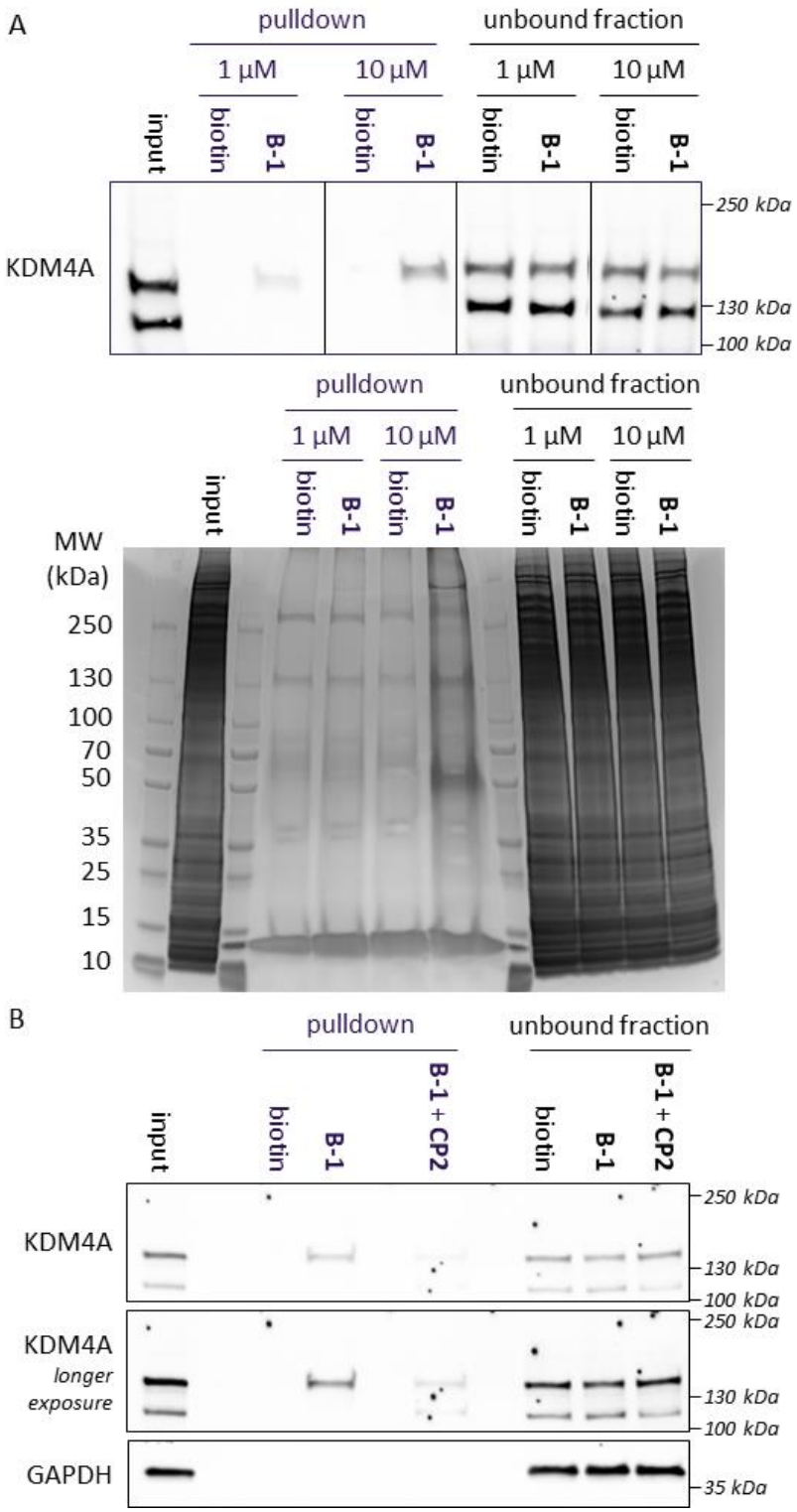


Figure 3.6 B-1 captures KDM4A from HEK293T cell lysates expressing endogenous levels of proteins. (A) Analysis using a western blot probed with KDM4A antibody shows **B-1** (1/10 µM) can pulldown KDM4A from HEK293T cell lysates. More KDM4A is recovered when using 10 µM **B-1** compared to 1 µM. Representative data from two biological replicates. (B) Analysis using a western blot probed with KDM4A antibody for pulldowns using **B-1** (10 µM), biotin (10 µM), or **B-1** + 10-fold **CP2** competition (10 µM **B-1** + 100 µM **CP2**). Lower KDM4A protein levels were detected in the competition sample, indicating that endogenous **B-1** pulldown is specific for the same KDM4A binding site as **CP2**. Representative western blot from five biological replicates. 'Input' contains unmodified cell lysate used in the experiment, 'pulldown' (purple text) comprises samples eluted from affinity capture beads, 'unbound fraction' contains lysate removed from beads prior to washes. KDM4A = 121 kDa, GAPDH = 36 kDa

Having established **B-1** target engagement in endogenous cell lysates, a competition experiment with **CP2** was performed to investigate the specificity of the pulldown (**Figure 3.6**B). Here, only 10-fold greater **CP2** concentration (100 μM) compared to **B-1** (10 μM) could be used, as **CP2** aggregated at higher concentrations tested. Competition of **B-1** with **CP2** greatly reduced the amount of KDM4A recovered from the pulldown, suggesting the pulldown is specific. However, 10-fold competition of **B-1** with **CP2** was not able to totally stop visible KDM4A enrichment when assessed by western blots probed with KDM4A antibody. A lower level of stoichiometric competition for **CP2** compared to **B-1** was used for enrichment from endogenous lysates (10-fold), compared to overexpressed lysates (100-fold, **CP2** (100 μM) compared to **B-1** (1 μM)) in Section 3.2.1.1.2. An alternative negative control (**B-2**, **Figure 3.3**) was synthesised (Section 2.2.3) and tested in future experiments (Section 3.2.1.1.4), the parent peptide for **B-2** is **CP2-R6A**, a variant of **CP2** with much lower KDM4A inhibitory activity (**CP2** IC_{50}(KDM4A) = 42 nM, **CP2-R6A** IC_{50}(KDM4A) = 2700 nM).[40] **B-2** has much greater similarity to **B-1** than biotin, so it was hoped more non-specific interactions could be identified through comparing **B-1** pulldown to **B-2**.

Lysate samples analysed by western blot probed with KDM4A antibody identified two bands for KDM4A in these experiments (**Figure 3.6**, approximately 120 kDa and 150 kDa, expected KDM4A MW = 121 kDa). The molecular weights observed are consistent with published western blots probed with this KDM4A antibody (Bethyl A300-861A), which report KDM4A bands at either 150 kDa,[267–269] or 125 kDa.[270] Multiple KDM4A bands are visible in one published paper, from Lapidot *et al.* (2021),[267] although none of the papers citing this antibody include uncropped blots. Interestingly, **B-1** could only affinity purify higher molecular weight KDM4A (approximately 150 kDa); one possible explanation could be that the lower band corresponds to degraded KDM4A, or a truncated KDM4A isoform, which **B-1** cannot bind with high affinity (*e.g.* lacking the *N*-terminal Jumonji-C demethylase domain)[271]. Greater KDM4A levels were enriched using 10 μM **B-1** compared to 1 μM (**Figure 3.6**A), conversely FlagKDM4A enrichment from lysates containing overexpressed FlagKDM4A 1 μM **B-1** was sufficient for clear enrichment (**Figure 3.5**,

Section 3.2.1.1.2). This may suggest that KDM4A is less available for **B-1** binding in endogenous lysates, perhaps a higher proportion of KDM4A molecules are engaged in large PPI complexes making the KDM4A active site less available. Endogenous KDM4A protein pulldown using **B-1** could not be observed using silver-stained SDS-PAGE protein gels, where overexpressed KDM4A pulldown was visible using this readout (**Figure 3.6** *versus* **Figure 3.5**). This is most likely due to the level of KDM4A enriched in the endogenous lysates being below the detection limit for silver-stain.[245] Significantly lower KDM4A expression is found in the endogenous lysates compared to the FlagKDM4A overexpressed lysates, which was observed when these lysates were loaded and compared on the same western blot, as the signal from the overexpressed lysate was saturating such that this lane needed to be cut from the blot to facilitate imaging of endogenous KDM4A pulldown (Appendix 3.2).

3.2.1.1.4 Work towards capture of endogenous protein-protein interaction partners of KDM4A

Having established endogenous **B-1** target engagement, attention was next turned to whether KDM4A interactors can be co-purified with KDM4A. Multiple KDM4A interactions have been identified in the literature, and these can be dependent on the cell type investigated.[33,271–273] Interaction with histone H3 is probably the most studied KDM4A interaction, and KDM4A can bind H3 through the Jumonji-C catalytic domain, or through the double Tudor reader domain (Section 1.1.2.1). The double Tudor domain of KDM4A is able to strongly bind H3K4Me$_3$ (K_d = 2.7 µM), H3K20Me$_3$ (K_d = 6.2 µM), and H3K23Me$_3$ (K_d = 2.2 µM).[274,275] The binding affinity of KDM4A's Jumonji-C catalytic domain for H3K9Me$_3$ (K_d = 110 µM) was measured in a separate study, alongside **CP2** (K_d = 30 nM).[40] **B-1** most likely binds at the H3K9/36Me$_{2/3}$ Jumonji-C binding site (Section 2.2.2.1), however it was hypothesised that it may be possible to retain and capture KDM4A in complex with histone H3, in particular *via* the H3 interactions with the double Tudor domain.

The ability of **B-1** to enrich KDM4A with H3 from endogenous cell lysates was tested using western blots probed with KDM4A and H3 antibodies. Different lysis buffers were tested as histone solubility is buffer dependent; buffers tested are summarised in **Table 3.2**. TOPEX buffer is reported to solubilise all proteins in a cell, and has been used for western blotting analysis of histones.[276] TOPEX buffer was used at room temperature for cell lysis as SDS was insoluble in this buffer at 4 °C; all other lysis buffers were used at 4 °C during cell lysis (as previously Section 3.2.1.1.1). In addition to standard TOPEX buffer containing 1% SDS,[276] a 0.1% SDS version of TOPEX buffer ('0.1% SDS', **Table 3.2**) was formulated to test whether lower SDS content would impact the pulldown. RIPA buffer was also tested, as it is a well-established lysis buffer for immunoprecipitation, and has been utilised to study both nuclear and cytoplasmic proteins.[277]

Buffer	Tris	NaCl	Detergent	SDS	Additional reagents
IGEPAL	**10 mM, pH 8**	150 mM	0.5% IGEPAL CA-630	-	-
TOPEX	**50 mM, pH 7.5**	300 mM	0.5% Triton-X-100	1%	1 mM DTT, 33.3 U·mL^{-1} benzonase nuclease
RIPA	**10 mM, pH 8**	150 mM	1% IGEPAL CA-630	0.1%	0.5% sodium deoxycholate
0.1% SDS	**50 mM, pH 7.5**	300 mM	0.5% Triton-X-100	0.1%	1 mM DTT, 33.3 U·mL^{-1} benzonase nuclease

Table 3.2 Components of lysis buffers used in KDM4A pulldown experiments. All lysis buffers contain cOmplete mini ethylenediaminetetraacetic acid (EDTA)-free protease inhibitors, used in accordance with manufacturer's instructions. SDS = sodium dodecyl sulfate, DTT = 1,4-dithiothreitol.

Western blot analysis of KDM4A enrichment by **B-1** in HEK293T cell lysates prepared using different buffers (**Figure 3.7**, **Table 3.2**) showed a clear pulldown band for KDM4A with IGEPAL buffer, but pulldown bands for KDM4A were faint or non-existent in other buffers tested (**Figure 3.7**A/B). Co-purification of H3 with KDM4A was not observed in IGEPAL buffer, likely due to low solubility of H3 in IGEPAL buffer compared to TOPEX buffer or 0.1% SDS buffer (**Figure 3.7**). **B-2** (**Figure 3.3**) was tested as a pulldown probe in these experiments, which uses **CP2-R6A** as the parent peptide, a variant of **CP2** with >50-fold lower inhibitory activity.[40] **B-2** was unable to enrich KDM4A from endogenous IGEPAL lysates, demonstrating that arginine-6 is required for KDM4A pulldown by **B-1**. **B-2** is likely a better negative control compared to biotin due to greater similarity to **B-1**. This should facilitate more accurate subtraction of non-specific binding when investigating

the identity of protein interaction partners in future work, and will be complementary to the use of **B-1** + **CP2** competition, and biotin controls.

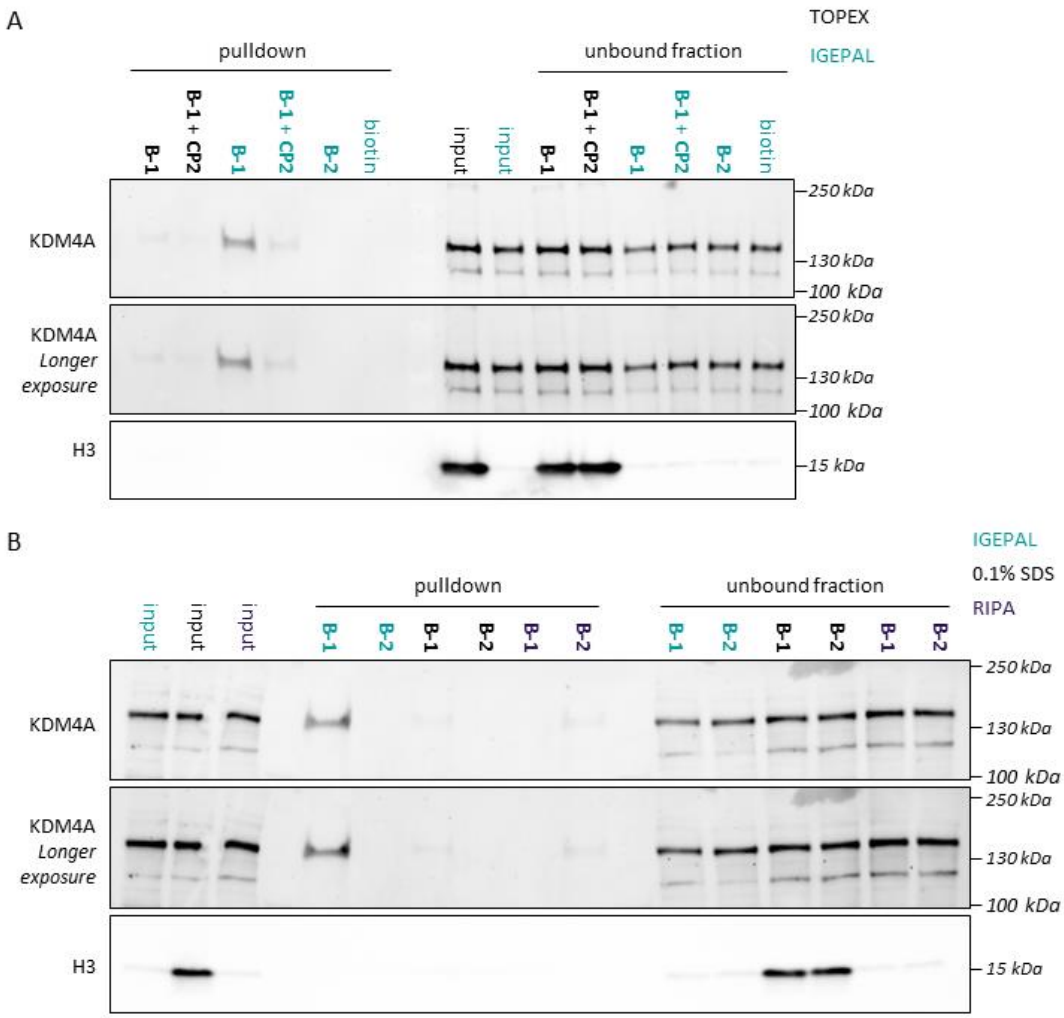


Figure 3.7 Different lysis buffers were tested to investigate if H3 was co-precipitated with KDM4A from cell lysates expressing endogenous levels of protein. (A) Analysis using a western blot probed with KDM4A and H3 antibodies showed that very little KDM4A was enriched from cell lysates prepared using TOPEX buffer (black text) compared to IGEPAL (teal text), and no H3 was observed in pulldown samples. One biological replicate was performed. (B) Analysis using a western blot probed with KDM4A and H3 antibodies for pulldown from cell lysates prepared using 0.1% SDS buffer (black text), RIPA buffer (purple text), or IGEPAL buffer (teal text). KDM4A precipitation was strongest in IGEPAL buffer. One biological replicate was performed. Both (A) and (B) test **B-2** as a negative control, which was unable to precipitate KDM4A in IGEPAL buffer in either experiment. Compound concentrations: **B-1** = 10 µM, **B-2** = 10 µM, biotin = 10 µM, **B-1** + **CP2** = 10 µM **B-1** + 100 µM **CP2**. 'Input' contains unmodified cell lysate used in the experiment, 'pulldown' comprises samples eluted from affinity capture beads, 'unbound fraction' contains lysate removed from beads prior to washes. KDM4A = 121 kDa, GAPDH = 36 kDa

Overall, KDM4A enrichment by **B-1** was only clearly observed from IGEPAL cell lysates, using western blots probed with KDM4A antibody. A key difference between IGEPAL buffer, and the others tested (**Table 3.2**) is the presence of SDS in all other buffers tested. SDS is a denaturing reducing agent, which may have encouraged misfolding of KDM4A such that **B-1** cannot bind. A

systematic follow-up of each buffer component would be required to identify precisely which buffer components increase enrichment of KDM4A by **B-1**, alongside which buffer components improve H3 solubility in search of appropriate conditions for KDM4A-H3 co-precipitation. However, if **B-1** is only able to engage properly folded KDM4A, this could be advantageous in the discovery of novel protein-interaction partners. Future work should explore the reason behind different efficiencies of target pulldown in different buffers.

*3.2.1.2 **B-1** as a MS-based proteomics probe*

Proteomics studies the large-scale analysis of proteins, and is an increasingly important technique for biological analysis in the 'post-genomic era'.[278] The main applications for MS-based proteomics are to identify protein interactions, catalogue protein identities, and to identify protein modifications in a sample (such as a cell or tissue extract).[238] Intact proteins can be analysed, often referred to as 'top down' proteomics, but this approach tends to have limited throughput as sufficient separation of many different proteins prior to MS is difficult to achieve.[279] Tandem MS (MS/MS) is often used to facilitate differentiation between molecules of the same mass but different structure, through measuring the mass of the molecule, and fragments created from the molecule in the same experiment.[238] Greater strides have been made in 'bottom up' or 'shotgun' proteomics, in which proteins are chemically/enzymatically digested into smaller peptide fragments prior to MS. Using MS/MS, the identity of the proteins from which the peptides originate can be assigned.[280] Overall, MS-based proteomics is increasingly relied upon in pre-clinical drug discovery, including for: target validation, lead compound profiling, and the identification of safety hazards.[83,214]

Preliminary proteomics experiments were performed to identify proteins enriched by **B-1** in affinity pulldowns from endogenous HEK293T cell lysates.[d] The workflow was similar to the

[d] Proteomics experiments were performed with help from NUPPA. Sample preparation was performed by myself, and MS and protein assignments were performed by NUPPA. Further data analysis was performed by myself.

pulldown studies above (Section 3.2.1.1), except for the elution step. Detergents in SDS-PAGE gel loading dye interfere with mass spectrometers,[281] and therefore a different elution method was required for samples prepared for proteomics. In this case, a trypsin digest was performed on the beads (bound to **B-1** and any protein binding partners) to elute peptides for 'bottom up' proteomics, inspired by Mohammed *et al.* (2016).[244] C-18 was used to desalt peptides,[282] followed by concentration of the samples for MS analysis (Section 6.9.2).

An initial proteomics experiment compared proteins enriched from endogenous HEK293T cell lysates by **B-1** (2 μM), and biotin (2 μM); and KDM4A was identified as a binding partner for **B-1** but not biotin. The biotin pulldown should contain both proteins which interact with biotin, and proteins which non-specifically interact with the beads. However, a competition control (*e.g.* **B-1** + **CP2**) was not included in this experiment, which would have tested the specificity of **B-1** protein interactions identified. Also, **B-1** concentration (2 μM) was lower than the optimal concentration for KDM4A pulldown by **B-1** determined in Section 3.2.1.3 (**Figure 3.6**), and therefore a second proteomics experiments was planned.

The next proteomics experiment compared proteins enriched from endogenous HEK293T cell lysates by **B-1** (10 μM) ± **CP2** (100 μM), and biotin (10 μM) (**Table 3.3**). Proteins identified were initially ordered by the difference in the number of unique peptides found by **B-1** pulldown, which identified KDM4A as the top protein. Further data analysis compared **B-1** pulldown with and without **CP2**, based upon the percentage difference between the number of unique peptides, MS/MS count, and sequence coverage. Proteins with greater than 60% difference in each of the parameters, and ≥2 more unique peptides enriched with **B-1** alone, compared to with **CP2**, are displayed in **Table 3.2**; equal enrichment by **B-1** with/without **CP2** would have 50% difference, indicating a non-specific interaction. All the parameters tested demonstrated higher KDM4A protein levels from **B-1** pulldown in the absence of **CP2** compared to when in the presence of **CP2** (**Table 3.3**). This demonstrates specific target engagement by **B-1**, and correlates with KDM4A

antibody western blot analysis demonstrating KDM4A pulldown by **B-1** from endogenous HEK293T cell lysates (**Figure 3.6**). Lower protein enrichment across the board was observed in the biotin pulldown compared to samples containing **B-1** (with/without **CP2**), which is shown by a lower number of unique peptides in biotin samples. This suggests that biotin is not a sufficient negative control for assigning proteins non-specifically enriched by **B-1**, however it should identify proteins non-specifically enriched by interaction with biotin / streptavidin beads, and therefore should be included in future proteomics experiments. **B-2** can be included as an additional control in future proteomics experiments, as a lower affinity **B-1** variant.

Protein	Mapped gene	UniProt accession no.	B-1			B-1 + CP2			Biotin			Unique peptide difference / %	MS/MS count difference / %	Sequence coverage difference / %
			No. of unique peptides	MS/MS count	Sequence coverage	No. of unique peptides	MS/MS count	Sequence coverage	No. of unique peptides	MS/MS count	Sequence coverage			
KDM4A	KDM4A	O75164	10	13	13.3	3	1	3.4	1	0	1.1	76.92	92.86	79.64
PTPRF	PTPRF	P10586	17	19	13.4	11	9	8.4	0	0	0	60.71	67.86	61.47
WDCP	C2orf44	Q9H6R7	7	9	14.7	2	2	4.4	0	0	0	77.78	81.82	76.96
MAP4K4	MAP4K4	E7ENQ1	4	7	10.7	0	0	1.3	0	0	0	100.00	100.00	89.17
PKN3	PKN3	Q6P5Z2	6	5	10	2	2	2.9	0	0	0	75.00	71.43	77.52
VARS2	VARS2	B7ZL25	8	9	10.3	4	1	5.4	0	0	0	66.67	90.00	65.61
RBM10	RBM10	P98175	10	9	14	6	3	8.7	0	0	0	62.50	75.00	61.67
RN123	RNF123	Q5XPI4	3	3	4	0	0	0	0	0	0	100.00	100.00	100.00
RBM12B	RBM12B	B9ZVT1	4	4	4.9	1	0	1.1	0	0	0	80.00	100.00	81.67
HELZ	HELZ	P42694	5	5	4.2	2	1	1.5	0	0	0	71.43	83.33	73.68
M3K6	MAP3K6	O95382	6	5	6.7	3	3	4.2	0	0	0	66.67	62.50	61.47
LENG8	LENG8	Q96PV6	6	6	12	3	2	4.1	0	0	0	66.67	75.00	74.53
UB2G2	UBE2G2	P60604	2	2	20	0	0	0	0	0	0	100.00	100.00	100.00
DHX32	DHX32	Q7L7V1	2	2	4.7	0	0	0	0	0	0	100.00	100.00	100.00
CTU1	CTU1	Q7Z7A3	2	2	8.9	0	0	0	0	0	0	100.00	100.00	100.00
MINK1	MINK1	Q8N4C8	2	3	6	0	0	1.1	0	0	0	100.00	100.00	84.51
PI4K2B	PI4K2B	G5E9Z4	2	2	5.2	0	0	0	0	0	0	100.00	100.00	100.00
BCAS3	BCAS3	K7ESE9	2	2	4.1	0	0	0	0	0	0	100.00	100.00	100.00
SPF30	SMNDC1	O75940	3	3	14.3	1	1	4.2	0	0	0	75.00	75.00	77.30
RBM5	RBM5	P52756	3	3	4.5	1	0	1.6	0	0	0	75.00	100.00	73.77
ZN598	ZNF598	Q86UK7	3	3	4.2	1	0	1	0	0	0	75.00	100.00	80.77
SELO	SELO	Q9BVL4	3	5	4.9	1	0	1.2	0	0	0	75.00	100.00	80.33
NSE4A	NSMCE4A	Q9NXX6	3	3	8.8	1	0	2.6	0	0	0	75.00	100.00	77.19
RFOX2	RBFOX2	O43251	3	6	13.1	1	1	3.8	2	4	10.5	75.00	85.71	77.51
HEAT6	HEATR6	Q6AI08	4	4	6.4	2	1	3.1	0	0	0	66.67	80.00	67.37
INT8	INTS8	Q75QN2	4	4	4.9	2	1	1.7	0	0	0	66.67	80.00	74.24
PRAME	PRAME	P78395	4	3	8.8	2	1	4.7	1	1	2.2	66.67	75.00	65.19
GANP	MCM3AP	O60318	5	3	3.4	3	3	2.1	0	0	0	62.50	50.00	61.82
ASNS	ASNS	P08243	5	4	10.3	3	2	6.8	0	0	0	62.50	66.67	60.23
KIF14	KIF14	Q15058	5	7	3.9	3	1	2.4	0	0	0	62.50	87.50	61.90
FIP1	FIP1L1	Q6UN15	5	5	12.5	3	1	7.4	0	0	0	62.50	83.33	62.81
MIRO1	RHOT1	Q8IXI2	5	6	9.9	3	4	4.4	0	0	0	62.50	60.00	69.23

Table 3.3 Proteomics data profiling proteins affinity purified using B-1 (10 µM), biotin (10 µM), or B-1 + 10-fold CP2 competition (10 µM B-1 + 100 µM CP2). Comparing **B-1**, and **B-1** + **CP2** samples, all proteins displayed had ≥2 peptides difference in unique peptides, and greater than 60 % proportional difference between **B-1** ± **CP2** for unique peptides, MS/MS count, and sequence coverage. MS and protein assignments were performed by NUPPA.

To look for known PPI networks within the list of proteins generated in **Table 3.3**, proteins were input to the STRING database[283] and an interaction network was generated (**Figure 3.8**). Enrichment of known protein interactions would increase confidence that the proteins identified are enriched through **B-1**/KDM4A interactions, and would suggest that **B-1** pulldown could be used to profile KDM4A interactions in other, more disease relevant, cell/tissue extracts. The PPI network did not find any known KDM4A interactions from the proteins identified in this proteomics experiment. Furthermore, many of the proteins identified are not engaged in known protein interactions. However, proteins identified could include novel KDM4A interactors, and to validate if this is the case more biological repeats are required. For good statistical confidence in these types of proteomics experiments greater than three repeats are needed. Protein interactions identified should be validated using alternative techniques, *e.g.* BRET-based assays.[121]

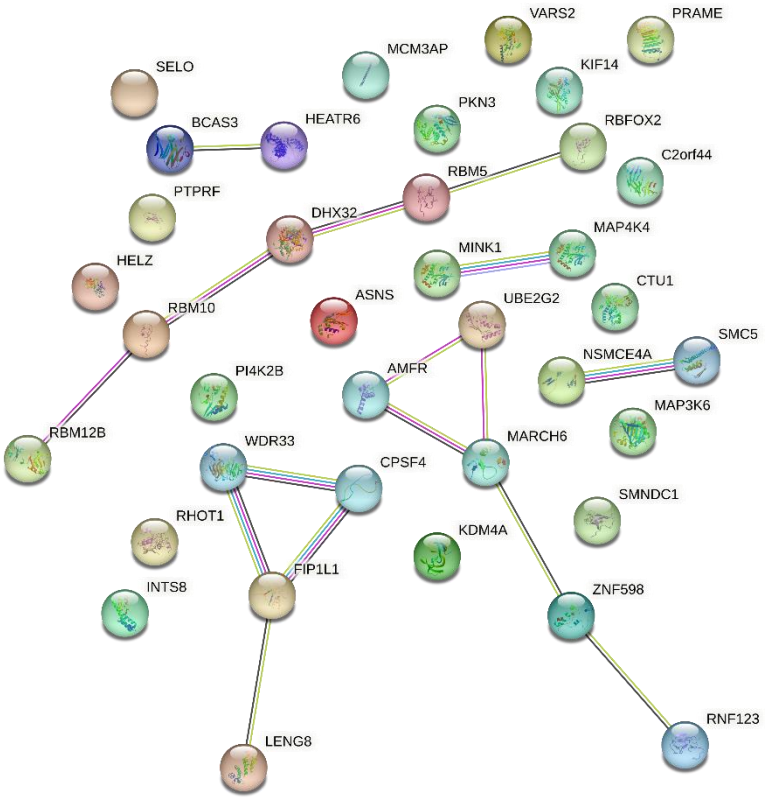


Figure 3.8 A STRING database protein interaction network[283] was made using the proteins displayed in Table 3.3. Many of the proteins identified, including KDM4A, are not part of known interaction networks. However, more repeats of this experiment are required to confirm these results. Figure created using https://string-db.org/

Future work should further optimise the proteomics workflow to ensure it is as biologically relevant as possible, for which a number of parameters can be studied, such as: protein stability in the lysis buffer, and disease relevance of the samples. Alternative lysis buffers could be tried, which may solubilise different proteins (as was observed in Section 3.2.1.1.4), and thus capture different KDM4A interactions to those found in this experiment. Other cell lines, or even patient derived tissues, could offer more disease relevant extracts;[284] KDM4s are implicated in several cancers, including leukaemia, and prostate/breast cancer, for which several immortal cell lines are available.[36] Additionally, future work could try treating whole cells with biotinylated macrocyclic peptides prior to cell lysis and pulldown, capturing protein interactions prior to cell lysis may preserve more interactions than are present in the lysate.[284] For target engagement inside cells a biotinylated version of **CP2.3** may be better than **B-1**, as **CP2.3** is more cell active than **CP2** (**CP2** EC_{50} ≈ 1µM, **CP2.3** EC_{50} < 100 nM);[40] however, **B-1** should also be treated on whole cells prior to lysis, as **CP2**-based protein degraders were cell active where a **CP2.3**-based protein degrader was not (Section 4.2.2.2.3). Many other stages of the proteomic workflow could also be investigated: different wash buffers, pre-clearing the lysate to remove non-specific bead interactors, and the trypsin digest. Furthermore, covalent linkage between the probe and target is often employed for chemoproteomics experiments which treat whole cells with the probe; this ensures the probe and target remain engaged during cell lysis.[214] In light of this, photoreactive cyclic peptide variants were tested for covalent target engagement in Section 3.2.2.

3.2.1.3 TET pulldown from cell lysates overexpressing TET1

Having established KDM4A pulldown using **B-1** (Section 3.2.1.1), pulldown of TET1 using peptide-biotin conjugates based upon cyclic peptide inhibitor **TiP1** (ref. [55]) were assessed.[e] These experiments were used to establish if **CP2**-based KDM4A pulldown methodology was transferable to other targets. TETs, like KDM4A, are epigenetic modifiers and members of the 2OG oxygenase

[e] TET pulldown studies were performed by Valeria Cobiltean under my supervision.

super family of enzymes (Section 1.1.2.2). Two **TiP1**-biotin conjugates were made, **B-3** and **B-4** (Section 2.2.3.2, **Figure 3.3**), which used tyrosine-6 and serine-13 as the attachment points to $(PEG)_4$-biotin respectively. The attachment points chosen maintained **TiP1** inhibitory activity in an alanine scan,[230] and multiple attachment points were tested to increase the chance of finding an accessible location for the reporter group, in the absence of more detailed structural information for **TiP1**-TET binding. This was not necessary for **CP2**-based **B-1**, as a co-crystal structure and extensive structure activity relationship studies indicated that threonine-13 would be accessible.

Cell lysates overexpressing, and not overexpressing, FlagTET1 were prepared from a doxycycline-inducible U2OS cells line.[f] Doxycycline-inducible expression systems overexpress the protein encoded by the plasmid when cells are treated with doxycycline,[285] and U2OS cells used for these experiments had been stably transduced with a doxycycline-inducible FlagTET1 plasmid. This allowed overexpression of FlagTET1 upon induction with doxycycline, or much lower expression levels when cells were not treated with doxycycline. For these experiments, lysates were prepared from the doxycycline-inducible FlagTET1 U2OS cells that had been induced with doxycycline, as well as cells that had not been induced. Cell lysates were produced using IGEPAL buffer (**Table 3.2**), doxycycline induced cells contain overexpressed FlagTET1, and non-induced cells contain close to endogenous TET1 levels.

FlagTET1 pulldown by **B-3** and **B-4** (**Figure 3.3**, 1 μM) from U2OS cell lysates was tested and analysed using western blots probed with Flag and TET1 antibodies, as well as silver stained protein gels (**Figure 3.9**). Western blot analysis using a Flag antibody of **B-3**/**B-4** affinity pulldown from U2OS lysates overexpressing FlagTET1 (**Figure 3.9**(A)) showed that **B-4** successfully enriched a Flag-containing protein between 100-130 kDa, most likely FlagTET1. The theoretical molecular weight of TET1 is 235 kDa, however TET1 is known to be expressed in multiple isoforms. Evidence

[f]Doxycycline-inducible FlagTET1 stably transduced U2OS cell line was generated by Prof Peter Radcliffe's laboratory, University of Oxford. Doxycycline-induced and not induced FlagTET1 U2OS cell lysates were produced with Dr Hilal Sarac.

suggests a short isoform of TET1 (≈150 kDa) is predominantly found in the cytoplasm while the full length TET1 isoform (≈235 kDa) is more commonly found in the nucleus.[286,287] The predominant FlagTET1 detected in the input sample was between 100-130 kDa (**Figure 3.9**(A)), suggesting that FlagTET1 was predominantly expressed in a truncated isoform in these U2OS cell lysates overexpressing FlagTET1, thus it can be predicted that **B-4** enriched a truncated isoform of FlagTET1. The FlagTET1 overexpressed **B-3**/**B-4** pulldown samples were also analysed using silver-stained protein gels (**Figure 3.9**(B)), and western blots probed with a polyclonal TET1 antibody (Genetex, GTX125888), however, no clear bands were found at 100-235 kDa (the expected molecular weight of different TET1 isoforms). This suggests that little FlagTET1 was enriched by **B-4**, as only one method of detection could identify enrichment. U2OS cell lysates which were not doxycycline induced were also tested for affinity purification of endogenous levels of TET1 by **B-3**/**B-4**, but no TET1 pulldown was observed when analysed by western blot probed with TET1 antibody or using a silver stained SDS-PAGE protein gel (**Figure 3.9**(B)). It is unsurprising that endogenous TET1 engagement could not be observed, as the detection methods used were unable to determine enrichment from the overexpressed system, containing higher TET1 concentrations.

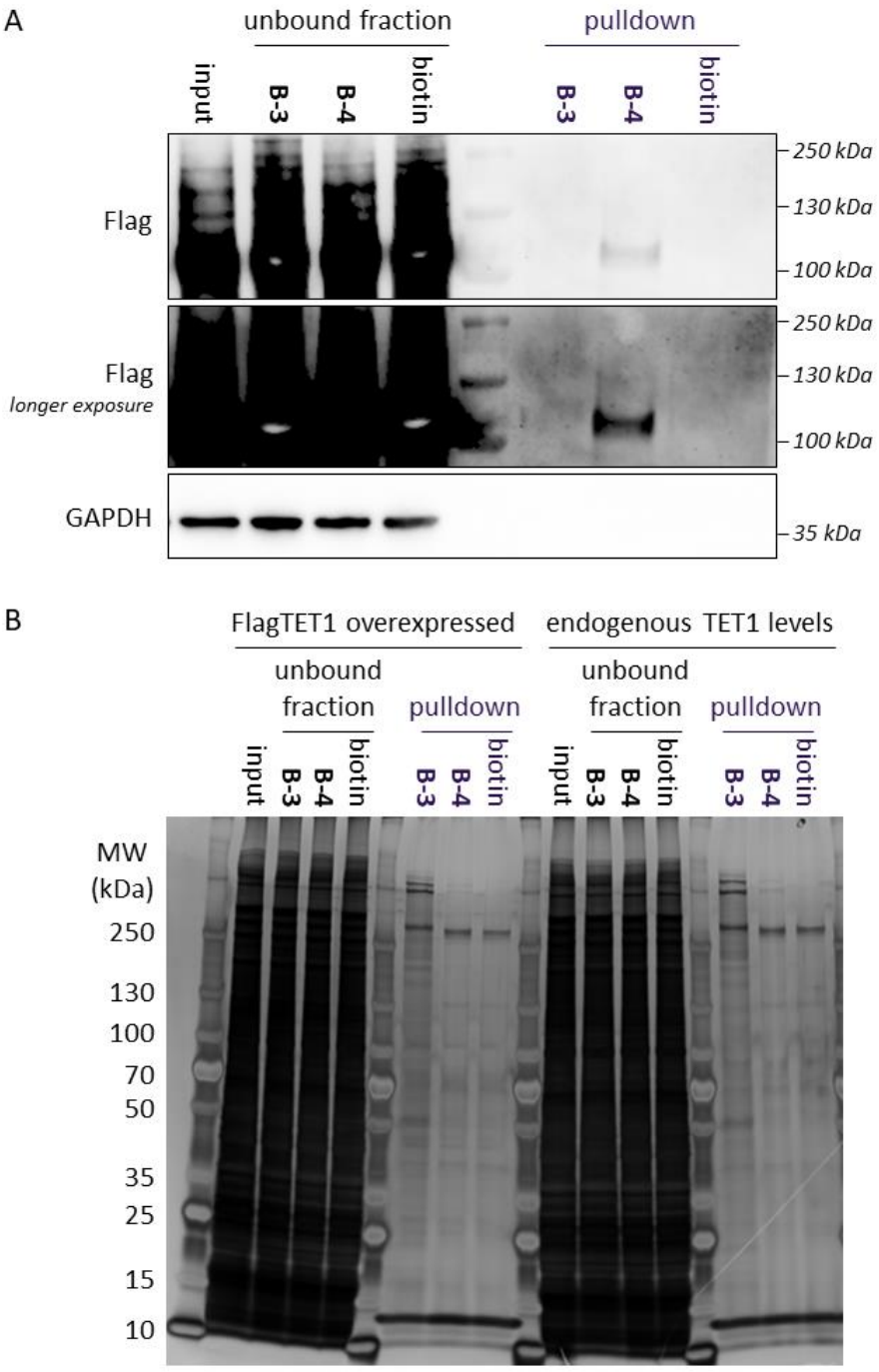


Figure 3.9 B-4 captures FlagTET1 from U2OS cell lysates overexpressing FlagTET1. (A) Analysis using western blot probed with Flag antibody and (B) silver stained SDS-PAGE protein gel, from pulldown using **B-3** (1 μM), **B-4** (1 μM), or biotin (1 μM). For **B-4** pulldown, a band between 100–130 kDa can be observed in the Flag western blot, suggesting enrichment of a Flag-containing protein, likely to be FlagTET1. A western blot probed with TET1 antibody was also performed, however successful pulldown was not observed. One biological replicate was performed. 'Input' contains unmodified cell lysate used in the experiment, 'pulldown' (purple text) comprises samples eluted from affinity capture beads, 'unbound fraction' contains lysate removed from beads prior to washes. FlagTET1 ≈ 100-235 kDa, GAPDH = 36 kDa.

The specificity of **B-4** pulldown of truncated FlagTET1 was tested using a competition experiment, **B-4** (1 μM) was competed with 100-fold **TiP1** (100 μM), the parent peptide of **B-4** (**Figure 3.10**B).

The pulldown was assessed using a western blot probed with Flag antibody, and FlagTET1 was not recovered when **TiP1** competition was used, indicating that **B-4** uses the same TET1 binding site as **TiP1** (**Figure 3.10**B). This experiment also replicated enrichment of a truncated version of FlagTET1 by **B-4** alone, showing that the initial pulldown is repeatable. This experiment was important to show that FlagTET1 protein enriched by **B-4** is obtained using the same site as **TiP1**-TET1 binding.

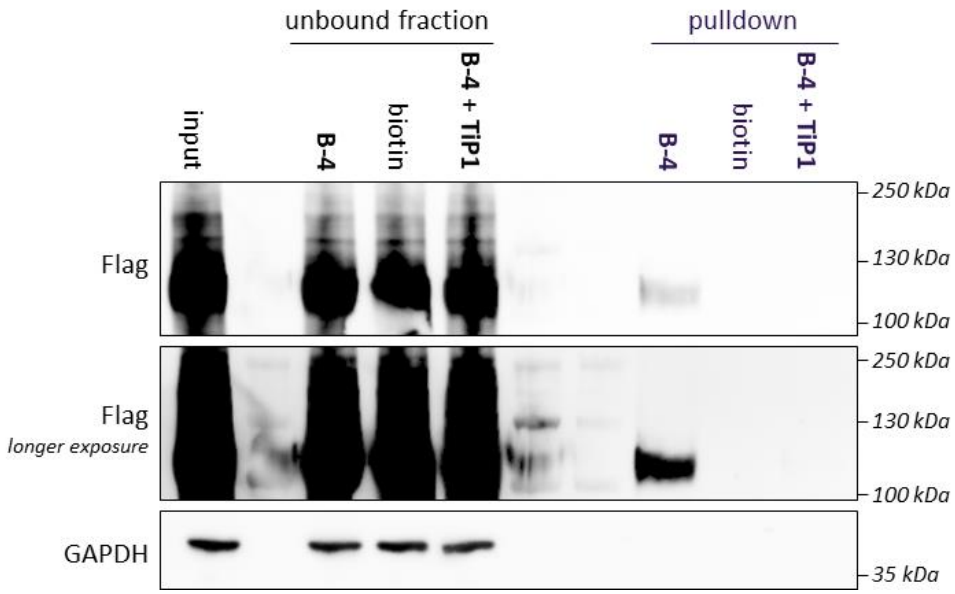


Figure 3.10 B-4 FlagTET1 pulldown from U2OS cell lysates overexpressing FlagTET1 is specific. Analysis using a western blot probed with Flag antibody from pulldown using **B-4** (1 µM), biotin (1 µM), or **B-4** + 100-fold-**TiP1** (1 µM **B-4** + 100 µM **TiP1**). FlagTET1 was not detected when competitor was included, indicating that **B-4** pulldown is specific. One biological replicate was performed. 'Input' contains unmodified cell lysate used in the experiment, 'pulldown' (purple text) comprises samples eluted from affinity capture beads, 'unbound fraction' contains lysate removed from beads prior to washes. FlagTET1 ≈ 100-235 kDa, GAPDH = 36 kDa.

Overall, **B-4** (1 µM) was able to enrich Flag-tagged protein, likely truncated TET1 from U2OS cell lysates overexpressing FlagTET1 (**Figure 3.9**), but enrichment could not be observed from cell lysates expressing close to endogenous TET1 levels. A competition experiment with 100-fold **TiP1** (100 µM) demonstrated that **B-4** (1 µM) pulldown is specific. Small levels of FlagTET1 protein enrichment were observed, this can be seen in the western blots probed with Flag antibody, where the pulldown band is much fainter than the input or unbound fractions. The same level of exposure was not required to image **B-1** KDM4A pulldown by western blots probed with either Flag antibody or KDM4A antibody in Section 3.2.1.1.2–4. Further evidence that low levels of

FlagTET1 were enriched is found from not being able to observe enrichment by western blot probed with TET antibody or silver-stained SDS-PAGE protein gel. The polyclonal TET antibody (Genetex, GTX125888) is likely less sensitive than the Flag antibody (can detect <10 ng protein, Merck, F1804),[258] although minimum protein detection limits for the TET antibody were not reported in the product information. Expanding upon this work, development of a TET1 isoform selective pulldown probe could be explored, based on **B-4**. The resulting probe would be extremely useful, as the different TET1 isoforms have been found to have contradictory roles in breast cancer,[286] as well as different roles in memory formation within the brains of mice.[287]

3.2.2 Photoreactive cyclic peptide probes

Reversible **B-1** and **B-4** peptides could enrich their targets from cell lysates, however it may be possible to improve protein recovery, as it is clear that high levels of target proteins remain in the unbound fractions (**Figure 3.5**, **Figure 3.6**, & **Figure 3.9**). This could be due to sub-optimal stoichiometry of the peptide-biotin conjugates *versus* their target; or it could also be due to the reversible nature of the cyclic peptide binding interaction to the target, where wash steps can reduce target recovery.[214] Therefore, formation of covalent linkages between the target and cyclic peptides was explored, using photocrosslinking.[84] Photoreactive probes can covalently bind their target upon irradiation with specific wavelengths of light (Section 2.1.1).[78] Covalent linkage can improve affinity pulldowns, by ensuring the target and probe do not dissociate during sample processing.[214] Photoreactive cyclic peptides would facilitate comparison of reversible/irreversible peptide affinity probes, to establish whether target-probe dissociation impacts cyclic peptide affinity pulldown.

Photoreactive peptide design can be aided by photoreactive amino acids which are structurally similar to natural amino acids: for example, a photoreactive cyclodepsipeptide variant of a cotranslational translocation inhibitor replaced leucine with photo-leucine, to produce a probe that was used for target deconvolution.[288] Photoreactive probes must retain affinity for the target,

while positioning the photoreactive group such that it can covalently react with the target upon irradiation, therefore detailed structure activity relationship data is usually necessary for the development of photoreactive probes.[289] For both **CP2** and **TiP1**, benzophenone or diazirine photoreactive groups were installed in multiple positions to increase the likelihood of identifying protein crosslinkers (**Figure 3.11**). The design and synthesis of **CP2**- and **TiP1**-based photoreactive cyclic peptide probes were described in Section 2.2.4.

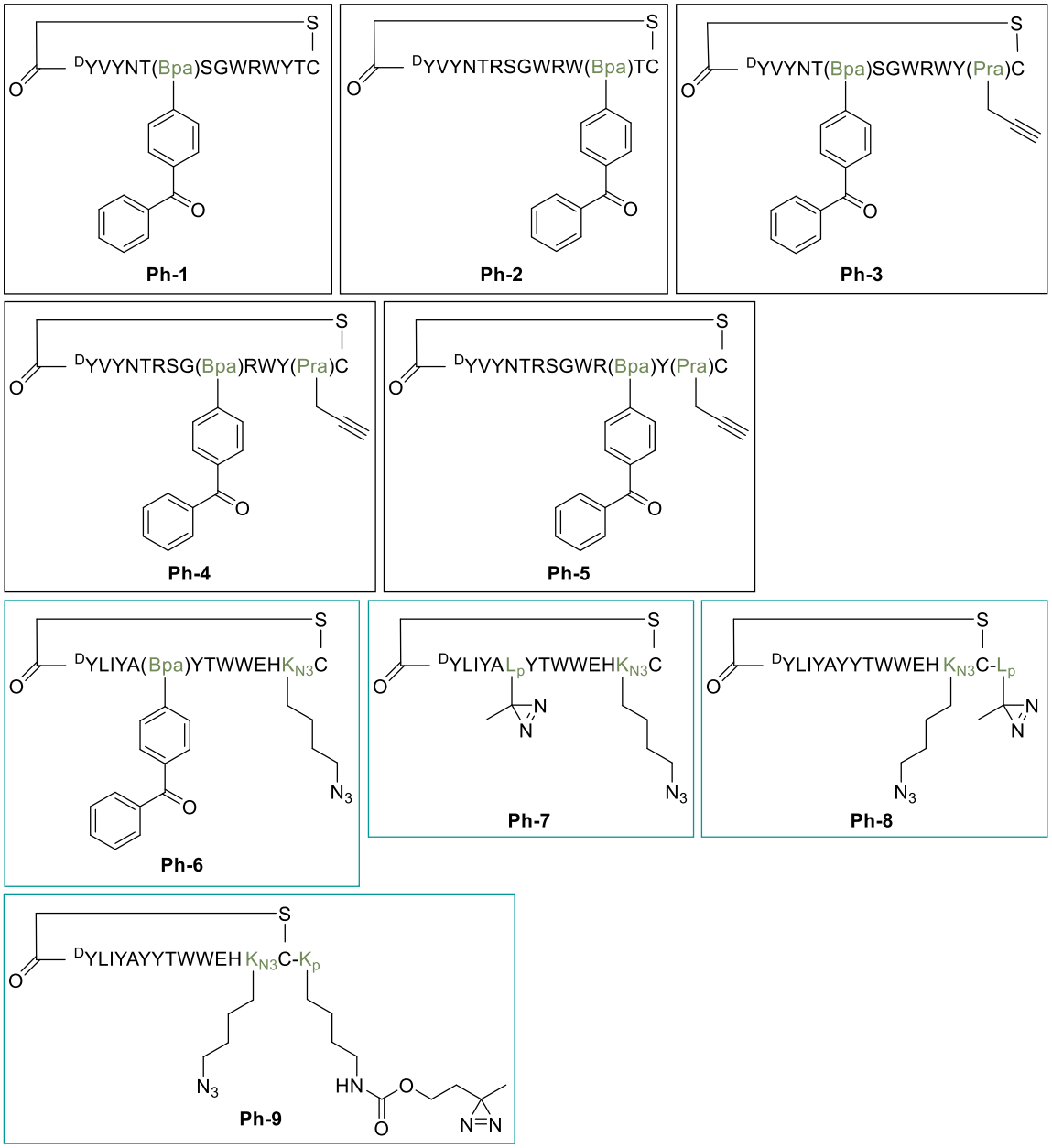


Figure 3.11 Structures of photoreactive macrocyclic peptides tested in Section 3.2.2 Ph-1 – **Ph-5** are **CP2**-based (black boxes), and **Ph-6** – **Ph-9** are **TiP1**-based (teal boxes). **Ph-3** – **Ph-5** were synthesised by Dr Jacob Bush, **Ph-6** – **Ph-9** were synthesised by Valeria Cobiltean under my supervision.

3.2.2.1 CP2 photoreactive probes

Multiple benzophenone positions were tested in **Ph-1** – **Ph-5**,[g] which included active site targeting arginine-6, as well as tryptophan-9/11 and tyrosine-12 which are situated further away from the KDM4A active site (**Figure 3.11**).[40] **Ph1** – **Ph5** were tested as photocrosslinkers to recombinant KDM4A, and intact liquid chromatography-mass spectrometry (LC-MS) was initially used to identify whether crosslinking had occurred. Successful peptide crosslinking can be observed by protein LC-MS from the presence of a species at the combined molecular weight of KDM4A + photoaffinity peptide. The proportion of protein crosslinked to peptide can be assessed by comparing the area of the crosslinked protein peak, compared to the combined area of the peaks containing KDM4A (**Figure 3.12**A).

For the initial crosslinking study, **Ph-1** – **Ph-5** (10 µM or 100 µM) were crosslinked with recombinant biotinylated His_6-$KDM4A_{1-359}$ (1 µM).[h] Samples were pre-incubated at 4 °C for 15 minutes prior to crosslinking, and subsequently irradiated at 302 nm for 10 minutes at 4 °C (conditions inspired by Fallon *et al.* (2021)[290]). The proportion of protein crosslinked to peptides was analysed using intact protein LC-MS (**Figure 3.12**). A 10% crosslinking threshold was set as the minimum for follow up studies, and both **Ph-4** and **Ph-5** were found to crosslink greater than 10% of protein, but not **Ph-1** – **Ph-3** (**Figure 3.12**A). Subsequently, several timepoints for **Ph-4** and **Ph-5** (10 µM or 100 µM) crosslinking to biotinylated His_6-$KDM4A_{1-359}$ (1 µM) was measured using intact protein LC-MS. Following a 15-minute pre-incubation of the peptides with KDM4A, samples were irradiated at 302 nm for 0.6–20 minutes, all at 4 °C (**Figure 3.12**B). Maximal crosslinking was observed within 5–10 minutes for both **Ph-4** and **Ph-5**, and thus 5 minutes irradiation time was used for remaining experiments testing photocrosslinking dose-response and competition with

[g] **Ph-3** – **Ph-5** were synthesised by Dr Jacob Bush.
[h] Biotin His_6-$KDM4A_{1-359}$ produced during my masters project.[321]

CP2 (**Figure 3.14** & **Figure 3.15**). Protein was not detectable using intact protein LC-MS when **Ph-5** was irradiated for 20 minutes, which may be due to protein aggregation.

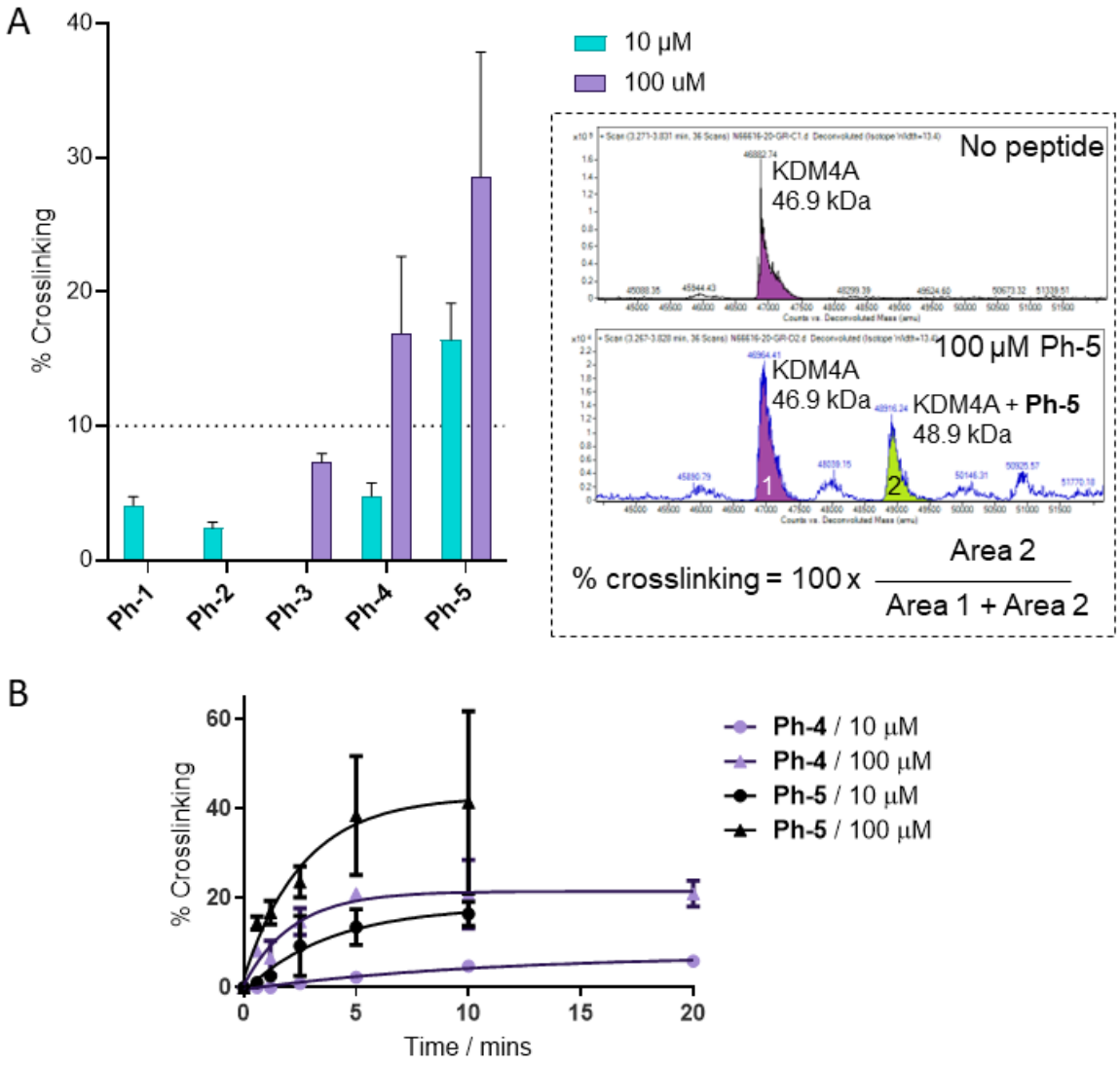


Figure 3.12 Crosslinking of a panel of CP2-based photoreactive peptides with recombinant KDM4A. (A) **Ph-4** and **Ph-5** were able to crosslink >10% KDM4A after irradiation at 302 nm for 10 minutes at 4 °C. Percentage crosslinking was measured using intact protein LC-MS, and example spectra from samples with and without peptide are shown, along with the equation used to calculate % crosslinking. Conditions: 10/100 µM peptides, 1 µM biotinylated His_6-$KDM4A_{1-359}$. (n = 2, mean ± standard deviation shown). (B) **Ph-4** and **Ph-5** crosslinking at 302 nm was tested at multiple timepoints at 4 °C, percentage crosslinking was calculated using intact protein LC-MS. Conditions: 10/100 µM peptides, 1 µM biotinylated His_6-$KDM4A_{1-359}$. (n = 2, mean ± standard deviation).

Intact protein LC-MS was relatively low throughput, requiring 10 minutes per sample. The length of LC-MS machine time required capped the number of samples that can be investigated in one experiment, because of protein degradation within the LC-MS samples over time which will eventually make the sample unmeasurable. Biotinylated His_6-$KDM4A_{1-359}$ stability over ≈40 hours

in the LC-MS injector (10 °C) was tested, by measuring the area of the protein peak for samples injected at multiple timepoints from ≈0.5–40 hours. The KDM4A peak area remained high, but the area of the peak reduced by up to 35% over the 40-hour window tested (Appendix 4). An alternative method of sample analysis was sought which would allow all samples to be analysed simultaneously. The two photoreactive peptides which were able to crosslink KDM4A both contained terminal alkyne handles, allowing CuAAC click chemistry to be used for conjugation of a reporter group to the peptide after crosslinking (Section 1.4). Peptides were coupled to Cy5-azide fluorescent dye after crosslinking, and samples were resolved using SDS-PAGE (**Figure 3.13**). Fluorescence imaging of the SDS-PAGE protein gel showed whether peptide crosslinking had occurred, through presence of a fluorescent band at approximately the molecular weight of KDM4A. Fluorescently imaged gels were subsequently stained using Coomassie protein stain, which allowed imaging of all proteins present in the gel, acting as a loading control. The CuAAC click reaction used an excess of Cy5-azide, which encouraged unbiased labelling of all peptides whether covalently linked to protein or not. Crosslinking of 0–50 µM **Ph-4**/**Ph-5** was analysed using both intact protein LC-MS, and fluorescently imaged gels after Cy5 attachment (**Figure 3.14**). Up to 10 µM peptide, there was correlation between Cy5 fluorescence and percentage crosslinking measured by intact protein LC-MS. At 50 µM photoreactive peptide a non-specific band was observed at approximately 28 kDa, which is likely due to protein degradation since protein is not visible using Coomassie stain (**Figure 3.14**).

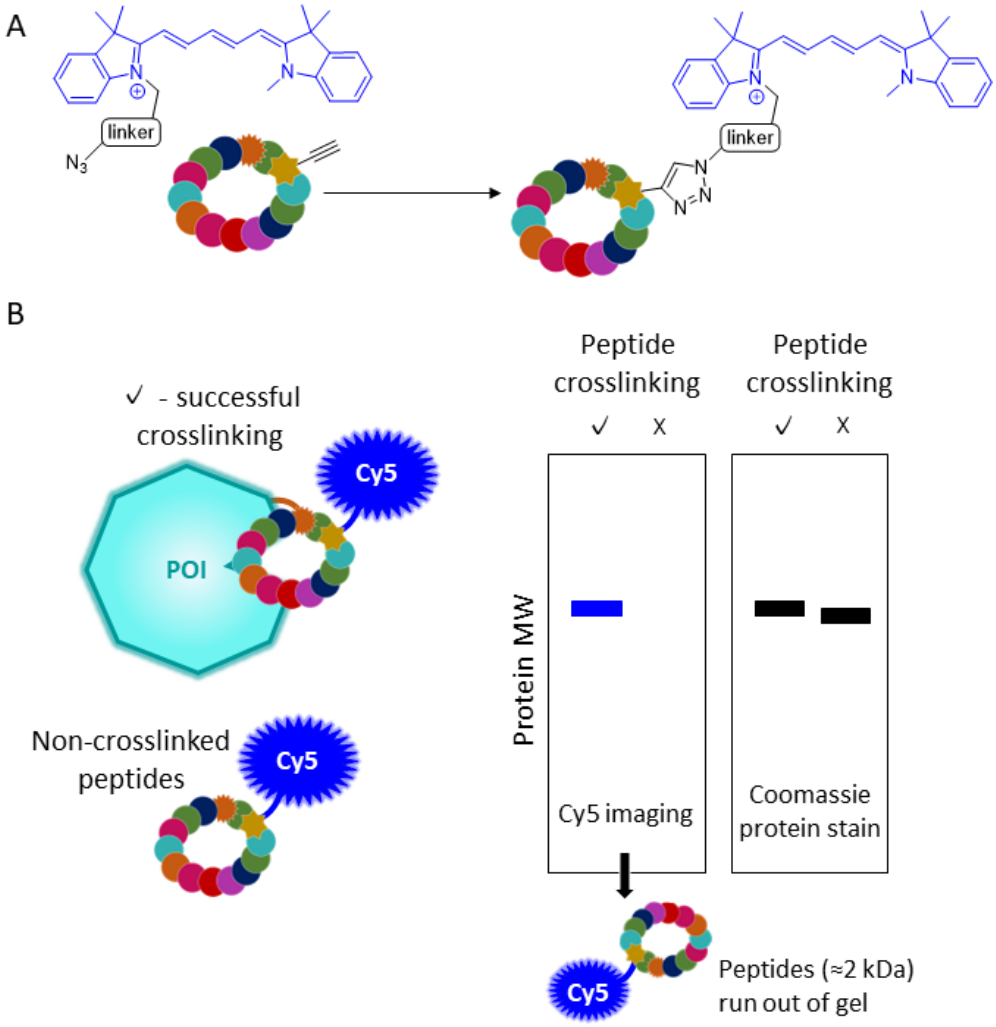


Figure 3.13 Schematic representation of fluorescent gel photocrosslinking analysis. (A) CuAAC click chemistry is used to link peptides (crosslinked or not) to Cy5-azide dye. (B) If photocrosslinking has occurred, Cy5 dye is covalently linked to the protein *via* the peptide, excess peptides are also Cy5 labelled. Samples are run on SDS-PAGE protein gels, and fluorescent imaging can be used to assess photocrosslinking to protein target. Coomassie protein stain is subsequently performed as a loading control.

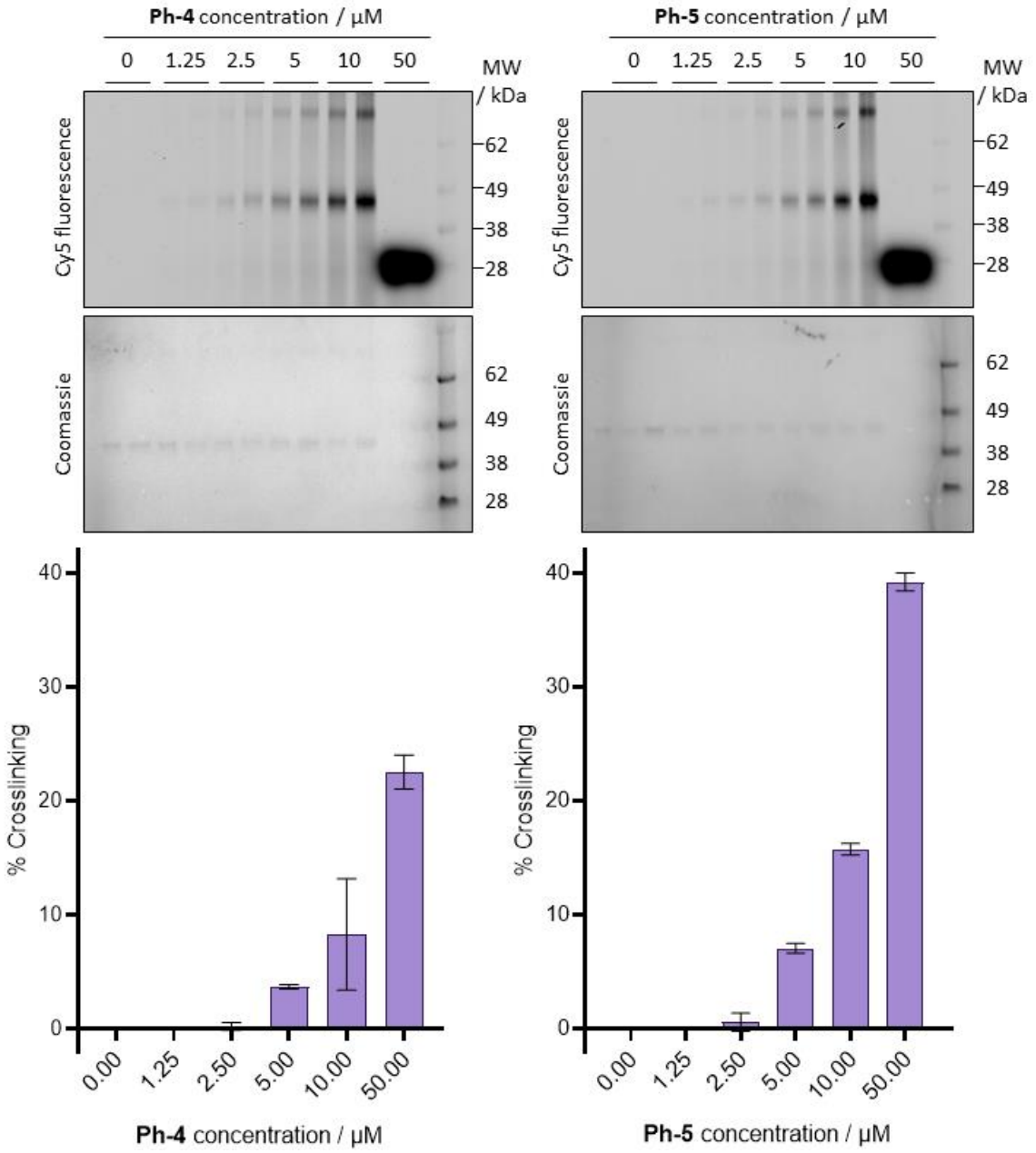


Figure 3.14 Comparison of fluorescent protein gel imaging (top) and intact protein LC-MS (bottom) for analysing Ph-4 and Ph-5 photocrosslinking to recombinant KDM4A. Crosslinking can be observed using fluorescent gel imaging up to 10 µM photoreactive peptides, to biotinylated His_6-$KDM4A_{1-359}$ (1 µM). Samples were irradiated at 302 nm for 5 minutes at 4 °C. Biotin-KDM4A = 47 kDa, biotin-KDM4A + peptide = 49 kDa. (n = 2, mean ± standard deviation shown).

A competition study between **Ph-4**/**Ph-5** and **CP2** was used to analyse specificity of the photocrosslinking observed. Photoreactive peptide crosslinking to the active site of KDM4A should be reduced as **CP2** competitor concentration is increased. The specificity of crosslinking observed was tested by competing **Ph-4**/**Ph-5** (10 µM) with **CP2** (0–100 µM), both cyclic peptides were pre-incubated with recombinant KDM4A for 15 minutes prior to crosslinking. Initial experiments used

1 µM KDM4A, but the protein could not be observed using intact protein LC-MS. High concentrations of cyclic peptides in the sample may have encouraged aggregation of the protein, such that the concentration of protein left in solution was below the detection limit for intact protein LC-MS. Therefore, the experiment was repeated using 2 µM KDM4A, with **Ph-4**/**Ph-5** (10 µM) and **CP2** (0–100 µM). KDM4A within these samples could be observed, but crosslinking to KDM4A was not detected using LC-MS. **Ph-4**/**Ph-5** were present at 5-fold excess to KDM4A in these samples. One to five stoichiometry might not have been enough to observe significant crosslinking at this KDM4A concentration, as previous experiments all used 1 µM KDM4A (**Figure 3.12** & **Figure 3.14**). Samples were also analysed through conjugation of Cy5-azide using protein gel fluorescence, however peptide crosslinking did not visibly decrease as **CP2** concentration increased – suggesting that crosslinking by **Ph-4** and **Ph-5** is non-specific (**Figure 3.15**). Non-specific crosslinking suggests the peptides covalently bind regions of KDM4A that are away from the active site, and therefore that peptide-KDM4A engagement is not needed for the crosslinking observed.

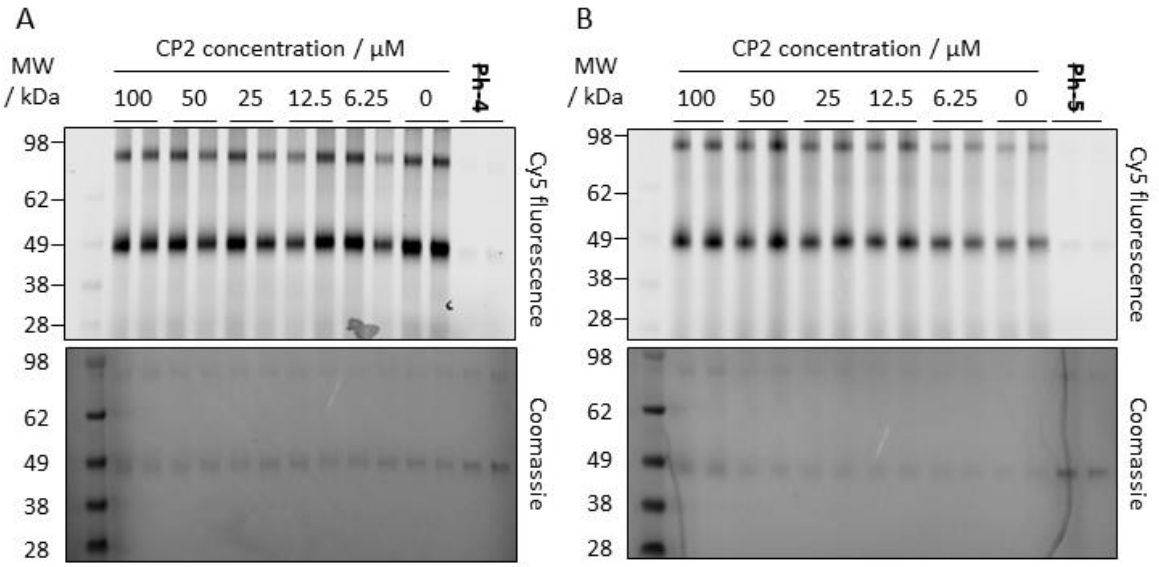


Figure 3.15 Photocrosslinking of Ph-4/Ph-5 to recombinant KDM4A with variable CP2-competition concentrations. Fluorescent protein gel imaging of photocrosslinking to biotin His_6-$KDM4A_{1-359}$ (2 µM) by **Ph-4** (10 µM, (A)) and **Ph-5** (10 µM, (B)) with variable **CP2**-competition (0–100 µM). Competition with **CP2** does not reduce crosslinking under the conditions tested. LC-MS analysis of this experiment did not identify peptide-protein crosslinking, so fluorescent gel imaging was the only experiment readout available. The right hand two lanes of each gel contain KDM4A without either photoreactive peptide or CP2. Samples were irradiated at 302 nm for 5 minutes at 4 °C. Biotin-KDM4A = 47 kDa, biotin-KDM4A + peptide = 49 kDa.

Further experiments to see if there are conditions at which **Ph-4**/**Ph-5** can specifically crosslink KDM4A are required. Irradiation at 365 nm was tested, but crosslinking was not observed; this contrasts with other studies whereby higher wavelengths increased benzophenone crosslinking efficiency.[291] Alternative photoreactive groups may also alter crosslinking by these peptides, due to different reactivity of the intermediates formed upon activation with UV light.

*3.3.2.2 **TiP1** photoreactive probes*

Photoreactive variants of TET1 inhibitor, **TiP1**, were synthesised (**Ph-6** – **Ph-9**, **Figure 3.11**, Section 2.2.4), despite **CP2**-based photocrosslinkers being non-specific (Section 3.2.2.1). **TiP1**-based **B-4** pulldown of FlagTET1 (**Figure 3.9**) was weaker than **CP2**-based **B-1** pulldown of FlagKDM4A/endogenous KDM4A (**Figure 3.5** & **Figure 3.6**), and thus it was hypothesised that **TiP1**-based affinity pulldown would greatly benefit from covalent TET1 attachment. For **TiP1**-based photoreactive probes diazirine photoreactive groups were tested in addition to benzophenones (**Figure 3.11**). Testing different photoreactive groups can improve chances of producing a successful photocrosslinking probe, as the different groups have different reactivity.[290]

Initially, a single dose study was used to test photocrosslinking of **Ph-6** – **Ph-9** (25 µM) to recombinant TET2 (2 µM).[i] TET2 was used for these experiments as TET1 was not available in recombinant form; TET2 is an isozyme of TET1, which can be highly expressed recombinantly, however **TiP-1** has lower TET2 affinity than TET1 (IC_{50}(TET1) = 0.2 µM, IC_{50}(TET2) = 14.5 µM).[55,230] **Ph-6** – **Ph-9** were pre-incubated with TET2 for 15 minutes at 4 °C, prior to irradiating at 365 nm for 20 minutes at 4 °C.[j] Crosslinking was analysed using intact protein LC-MS, and by clicking to Cy5-alkyne dye followed by fluorescently imaged SDS-PAGE protein gels (**Figure 3.13**). Crosslinking was not observed under the conditions tested using either analysis method (**Figure 3.16**), despite including two different photoreactive groups in these peptides. The fluorescently imaged protein

[i] TET photocrosslinking studies performed by Valeria Cobiltean under my supervision.
[j] Photocrosslinking conditions based upon unpublished work by Dr Klemensas Simelis

gel has bands at the mass of recombinant TET2 (51.2 kDa) for all samples loaded, including DMSO vehicle control samples, as well as samples which were not irradiated (**Figure 3.16**A). This suggests Cy5-alkyne dye non-specifically interacts with TET2, and thus this detection method is inappropriate for analysis of crosslinking to recombinant TET2. However, as crosslinking could not be observed using intact protein LC-MS, it is unlikely that **Ph-6** – **Ph-9** are able to crosslink recombinant TET2 under the conditions tested.

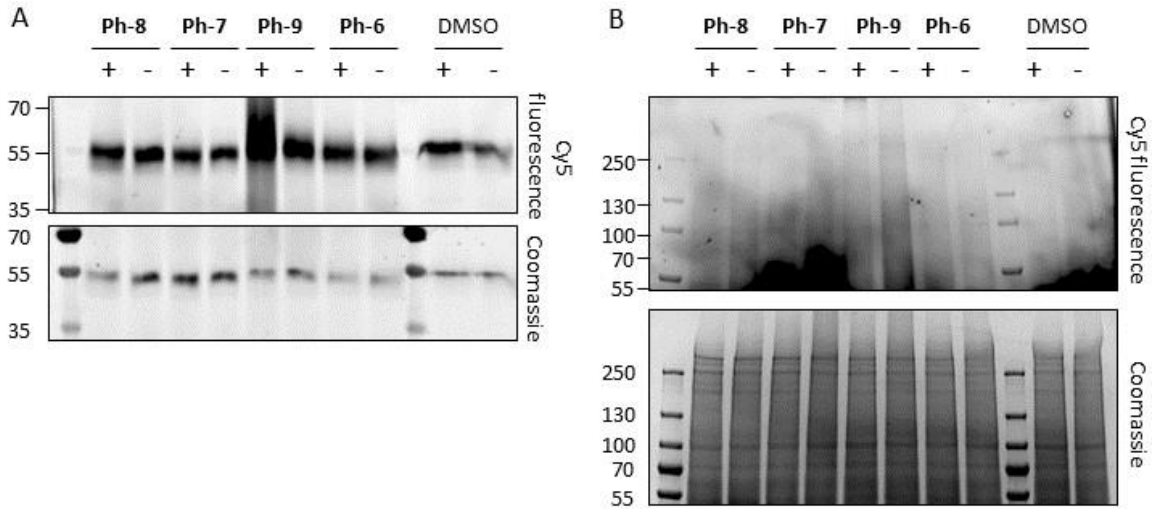


Figure 3.16 Photocrosslinking of a panel of TiP1-based photoreactive cyclic peptides was tested against TET1 and TET2 proteins. (A) Fluorescently imaged SDS-PAGE protein analysis of **Ph-6** – **Ph-9** (25 μM) crosslinking to recombinant TET2 (2 μM). Cy5 non-specifically binds TET protein, this is demonstrated by fluorescent bands for DMSO samples. (B) Fluorescent protein gel imaging analysis showing crosslinking of **Ph-6** – **Ph-9** in U2OS cell lysate containing overexpressed FlagTET1, no clear bands showing crosslinking can be seen. For both experiments irradiation was performed at 365 nm for 20 minutes at 4 °C. Recombinant TET2 construct = 51 kDa, recombinant TET2 + peptide = 53 kDa, variable lengths of TET1 may be expressed in cell lysate ≈ 100-235 kDa (**Figure 3.9**).

Since photocrosslinking **Ph-6** – **Ph-9** with recombinant TET2 was unsuccessful, another system was devised to test **Ph-6** – **Ph-9** with TET1, which is challenging to express recombinantly. TET1 production was achieved using cell lysates from U2OS cells overexpressing FlagTET1 (Section 3.2.1.3), and **Ph-6** – **Ph-9** (25 μM) were tested as photocrosslinkers at a single concentration in these cell lysates. Intact protein LC-MS was inappropriate for use with **Ph-6** – **Ph-9** TET1 cell lysate samples, as they contained proteins from the full proteome, and relatively pure samples are needed for intact protein LC-MS.[279] Therefore, Cy5-dye was conjugated to photocrosslinking samples and they were run on SDS-PAGE protein gels, gels were fluorescently imaged prior to staining with Coomassie blue. No specific fluorescent bands were detected across

the region that TET1 protein may be found (100-235 kDa, Section 3.2.1.3), demonstrating that detectable crosslinking did not occur (**Figure 3.16**B).

Overall, photocrosslinking was not observed for **Ph-6** – **Ph-9** to recombinant TET2 protein, or to overexpressed FlagTET1 protein in U2OS cell lysates (**Figure 3.16**). Future work could include systematic profiling of photocrosslinking by **Ph-6** – **Ph-9**, as these studies only tested one peptide concentration (25 μM), with one irradiation wavelength (365 nm), and one irradiation time length (20 minutes). Testing lower irradiation wavelengths may be beneficial, despite reports that 365 nm should be appropriate for both these groups,[292] as efficient photocrosslinking has been observed for both diazirines and benzophenones at 302 nm.[78,290] Many other factors can be considered during photocrosslinking optimisation for these probes, including: the irradiation times, probe concentration, and using recombinant TET1–3. Furthermore, the click reaction conditions (Section 6.4.2) used should be optimised, to reduce non-specific labelling of TET2 by Cy5-alkyne. CuAAC click reaction conditions were maintained from **Ph-1** – **Ph-5** (Section 3.2.2.1), but the alkyne and azide have been switched around in **Ph-6** – **Ph-9**, which may have affected the reaction. Another line of investigation could test whether TET2 non-specifically interacts with Cy5, here background fluorescent labelling of TET2 could be compared to that produced with KDM4A (which had minimal background fluorescence in the absence of cyclic peptides, **Figure 3.14** & **Figure 3.15**).

3.3 Discussion

This chapter focussed on affinity cyclic peptide probes. Reversibly binding peptide-biotin conjugates against KDM4A (**B-1**) and TET1 (**B-4**) proteins were able to engage their targets in cell lysates (Section 3.2.1), demonstrating the target versatility of this approach. Preliminary proteomic experiments further validated KDM4A engagement by **B-1** in cell lysates (**Table 3.3**). Photoreactive cyclic peptides were also tested, in work towards investigation of whether covalent linkage of the peptides to their targets could increase affinity pulldown efficiency. However, photoreactive peptides were unable to specifically crosslink their targets under the conditions

tested (Section 3.2.2). Follow up work can include systematic profiling of the photoreactive peptides under different crosslinking conditions.

CP2 (ref. [40]) and **TiP1** (ref. [55]) are both 14-mer peptides discovered using mRNA display selection methodology using thioether cyclisation.[293] It is interesting that the successful KDM4A and TET1 pulldown probes, **B-1** and **B-4** respectively, are both conjugated to biotin through their 13th position of 14 (**Figure 3.3**). The 13th residue within these peptides is close to the thioether cyclisation moiety formed between the 1st and 14th residues, and also next to the *C*-terminus where mRNA is conjugated during peptide selection.[169] Crystal structures of thioether macrocyclic peptides discovered using mRNA peptide selection show that the cyclisation portion of peptides discovered using this methodology often interacts minimally with the target,[40,294] likely due to the steric bulk of mRNA/DNA conjugated to peptides during the selection.[169] Biotin conjugated through the 13th residue may be more accessible to streptavidin beads in the affinity pulldown of **B-1** and **B-4**, contrasting with **B-3** where biotin is conjugated in the centre of **TiP1** (**Figure 3.3**). It would be informative to test the binding and inhibition activity of both **B-3** and **B-4** against TET1, to work towards elucidating the reason why **B-3** did not enrich TET1 at 1 μM.

TiP1 has lower TET1 affinity (K_d = 113 nM)[55] than **CP2**'s KDM4A affinity (K_d = 29.8 nM)[40], which may have contributed to a weaker TET1 pulldown by **B-4** than KDM4A pulldown by **B-1** (**Figure 3.5** & **Figure 3.9**). Future work should test **B-4** pulldown at higher concentrations, and work towards identifying suitable conditions for TET pulldown from cell lysates containing endogenous protein levels. Optimal **B-4** pulldown conditions can be used in proteomics studies, similarly to **B-1** (Section 3.2.1.3).

It is interesting that different cell lysis buffers altered **B-1** KDM4A pulldown (**Figure 3.7**). The pulldown results from different cell lysis buffers suggests the folding of KDM4A may be different in different buffers, and thus the protein interactions captured in different buffers are likely to change. These results highlight a limitation of proteomic experiments, that only proteins which

are soluble/detectable under the lysis conditions used can be identified.[260] Despite these shortcomings, proteomics is extremely valuable in drug discovery and is increasingly relied upon for target validation and pharmaceutical development.[83]

The biotinylated cyclic peptides used in this chapter could be employed in other experiments, for example, dose-response competition experiments with the parent peptides could be utilised to determine the binding kinetics of the peptides in cells / cell lysates.[76] Alternatively, biotinylated cyclic peptides could also be used to identify genomic DNA interacting with their targets, by using Chem-seq, which identifies DNA interacting with biotinylated small molecules probes. Biotinylated-JQ1 was used to profile JQ1-DNA interactions, and selectivity of the compound was profiled by comparison to ChIP-seq experiments.[88]

The workflow used in this chapter can be widely expanded to other macrocyclic peptides, including those identified using *de novo* screening, and substrate mimics. Peptides have high binding affinity for their targets, and are often much more selective than small molecules.[147] Peptides are also facile and quick to synthesise using automated Fmoc/*t*Bu SPPS.[229] Peptide synthesis is highly reproducible, meaning peptides do not suffer from the batch-to-batch variation which complicates the use of antibodies.[147] These advantages of peptides over antibodies and small molecules gives them the potential to become crucial parts of target validation in drug discovery, complementing or replacing the role of antibodies in this process.

Chapter 4 Macrocyclic peptide protein degradation

4.1 Introduction

4.1.1 Mode of action of PROTACs

PROTACs were first reported by Sakamoto *et al.* (2001),[105] and the field has grown enormously since (Section 1.2.2.3). Recently, PROTACs have translated into the clinic, the first of which was in 2019 by Arvinas therapeutics,[95] and many more PROTACs have entered clinical trials since.[295]

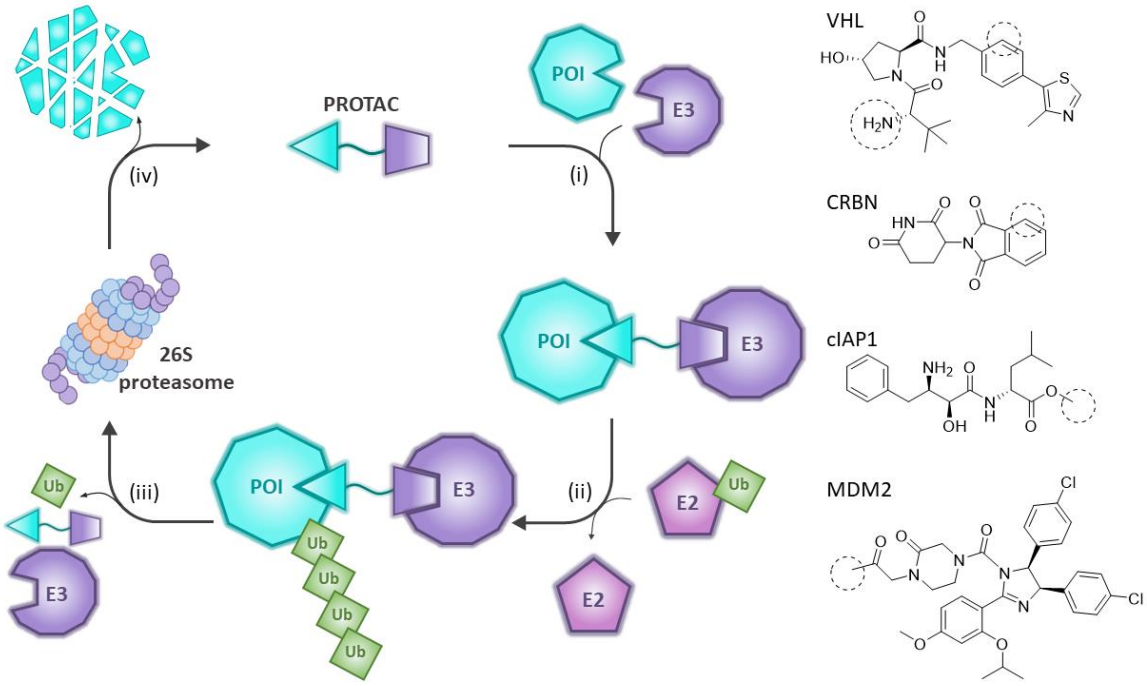


Figure 4.1 Schematic of protein degradation by PROTACs, and structures of commonly employed E3 ligase binders. PROTAC degraders consist of a ligand targeting the protein of interest (POI) (triangle, cyan), linked to an E3 ligase ligand (trapezium, purple). (i) A ternary complex is formed between PROTAC, POI, and E3 ligase. (ii) Ubiquitin-conjugated E2, forms an E2-E3 complex, and transfers ubiquitin to the POI. Subsequently, the ubiquitin chain is extended. (iii) The polyubiquitinated POI enters the 26S proteasome. Ubiquitin, PROTAC, and E3 ligase are released prior to protein degradation, allowing the PROTAC to degrade another POI. (iv) The POI is degraded. Right hand side shows structures of some commonly used E3 ligase ligands, with potential linker attachment points circled. VHL = von-Hippel Lindau protein, CRBN = cereblon, cIAP2 = cellular inhibitor of apoptosis protein 2, MDM2 = mouse double minute 2 homolog. Figure adapted from Cromm *et al*. (2020).[296]

PROTACs consist of a ligand targeting a POI, connected to a ligand for an E3 ligase, *via* a linker, facilitating formation of a ternary complex between the POI, PROTAC, and E3 ligase (**Figure 4.1**).[104] E3 ligases are part of the UPS, through which polyubiquitinated proteins are degraded. Ubiquitination of the POI occurs through a protein complex formed of an E3 ubiquitin-ligase, with an E2 ubiquitin-conjugating enzyme (activated by an E1 ubiquitin-activator), which mediates the

transfer of ubiquitin from E2 to the POI. A polyubiquitin chain is usually built by attaching the *C*-terminus of a new ubiquitin molecule to a lysine residue within the previous ubiquitin in the ubiquitinated-POI, and polyubiquitinated proteins are generally degraded by the 26S proteasome complex (Section 1.2.2.3).[297] Furthermore, with reversibly binding PROTACs, the PROTACs, E3 ligases and ubiquitin are released, allowing them to induce the degradation of new POI molecules.[101]

VHL and CRBN are the E3 ligases most commonly used in the development of PROTACs, although other E3 ligases have been employed in PROTAC development, such as MDM2, and inhibitors of apoptosis proteins (IAP) (different E3 ligase ligands shown in **Figure 4.1**).[107] Between 2010–12 major breakthroughs in the PROTAC field took place, with the discovery of the mode of action of thalidomide against CRBN, along with the development of small molecule peptidomimetic VHL binders.[296,298–300] VHL and CRBN have remained the best studied E3 ligases targeted by PROTACs, although future PROTAC design is likely to employ more of the numerous E3 ligases available (>600).[107]

Protein degraders do not require inhibitors/activators, nor a target with catalytic activity, a PROTAC target simply needs to be ligandable; allowing PROTACs to target more proteins than traditional inhibition strategies.[301] As a therapeutic strategy PROTACs show great promise, since they mimic knockdown/knockout studies used in target validation much more closely than inhibition strategies, as the entire POI is removed, rather than just modulating the activity of the target.[206] A disadvantage of PROTACs is that they do not obey traditional drug-like properties (Lipinski's rule-of-five).[145] This is because PROTACs are often large molecules (>700 Da),[191] with many hydrogen bond donors and acceptors.[191] It is therefore surprising that many PROTACs are able to elicit a biological effect inside cells, something which is partly due to their catalytic mode of action, which allows multiple equivalents of the same protein to be degraded by one PROTAC molecule.[104] Also, the catalytic mode of action allows lower compound doses to be used, which is

promising for drug discovery as this can reduce side effects for patients. PROTAC drugs may not need to be administered as frequently as traditional inhibitor drugs, since their effect can be maintained after clearance, while protein levels recover to normal.[104] Furthermore, some PROTACs can degrade their targets even when they contain drug-resistant mutations, where traditional drugs cannot.[104] The altered mechanism used by PROTACs often allows molecules with poor cell activity, or mediocre affinity for their target, to be developed into selective degraders – aided by opportunities for improved selectivity in the linker chemistry, E3 ligase targeted, and the derivatisation points of the molecules being linked.[302] However, altered pharmacology of PROTACs, compared to traditional small molecule drugs, could delay their path to approval as finding the optimal dosing pattern may be more challenging. The hook effect in particular may complicate dosing, where higher doses of PROTAC compounds may reduce degradation activity, through a preference for formation of PROTAC-POI/PROTAC-E3 dimers over the productive E3-PROTAC-POI ternary complex.[301]

4.1.2 Peptide-based PROTACs

The first PROTAC, PROTAC-1 (**Figure 4.2**), used a phosphopeptide ligand for the E3 ligase Skp1-Cullin-F box complex, ligated to a small molecule for methionine aminopeptidase-2, but it was not cell permeable.[105] The first cell permeable PROTAC, PROTAC-4, was developed in 2004; PROTAC-4 comprised of a FK506-binding protein ligand, a VHL recruiting peptide, and was conjugated to poly-D-arginine in order to improve cell permeability.[303] Several peptide-based PROTACs have been developed since, targeting a range of diseases, including: multiple cancers, neurodegenerative diseases, and the X protein of hepatitis B.[304,305] Many peptide PROTACs developed to date have included CPPs, such as PROTAC-4, to improve cell permeability. However, Liao *et al.* (2020)[306] did not require a CPP for a β-catenin targeting PROTAC (xStAx-VHLL, **Figure 4.2**), instead a stapled peptide against β-catenin, was conjugated to a VHL binding peptide using 6-aminocaproic acid. xSTAx-VHLL suppressed Wnt signalling, and even inhibited growth of patient derived colorectal cancer tumour organoids. PPI substrate mimics have often been employed in

peptide PROTACs to date,[305] such as the peptide used to recruit the Skp1-Cullin-F box complex in PROTAC-1. Conversely, Ma *et al.* (2022)[307] employed a novel peptide binder for androgen receptor splice variant-7, identified using artificial intelligence aided design (Au-AR pep-PROTAC). Au-AR pep-PROTAC was delivered to cells by ultrasmall gold nanoparticles.

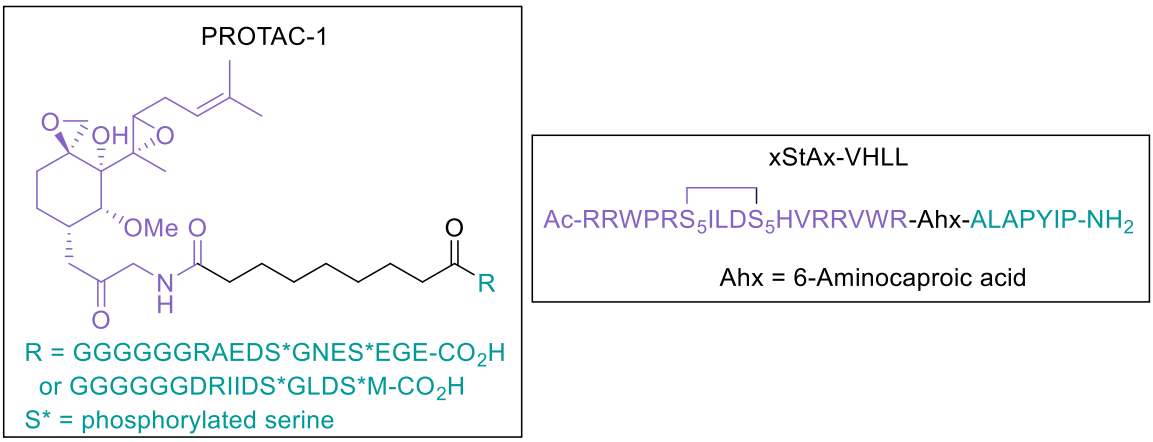


Figure 4.2 Structures of selected peptide-based PROTACs. PROTAC-1 was the first PROTAC developed, against methionine aminopeptidase-2, recruiting E3 ligase Skp1-Cullin-F box complex.[105] xStAx-VHLL is a cell permeable peptide PROTAC, which targets β-catenin and suppresses Wnt signalling.[306] POI binders in purple, E3 ligase binders in teal, linkers in black.

Nanobodies have also been used to induce protein degradation inside cells, in a novel approach called the Affinity-Directed Protein Missile (AdPROM) system. AdPROMs are currently introduced by viral transduction, but the system has been successful at degrading 'undruggable' K-Ras in cells.[308]

4.1.3 Epigenetic PROTACs

PROTACs are tools with great potential for investigating the roles of epigenetic proteins, and they have been successfully developed to study readers, writers, and erasers within epigenetics.[309,310] The high sequence homology between proteins within epigenetic families makes the development of selective compounds a challenge (Section 1.2.1),[196] to this end PROTACs offer additional routes to selectivity through linker design, E3 ligase choice, and the kinetics of ternary complex formation.[103] Furthermore, many epigenetic PROTACs developed so far have had an effect similar

to knockdown/knockout studies, where corresponding epigenetic inhibitors have not been able to.[309]

BET family proteins are epigenetic readers, and PROTACs targeting BET proteins have been extensively studied.[154,311–314] **MZ1** (**Figure 4.3**), consists of **JQ1** (a pan BET bromodomain inhibitor) linked to a VHL binder, and it has shown selective BRD4 degradation over BRD2/3.[312] A co-crystal structure of the **MZ1**-BRD4-VHL complex showed how the presence of new PROTAC induced PPIs were responsible for the stability of the ternary complex, and the selectivity towards BRD4.[141]

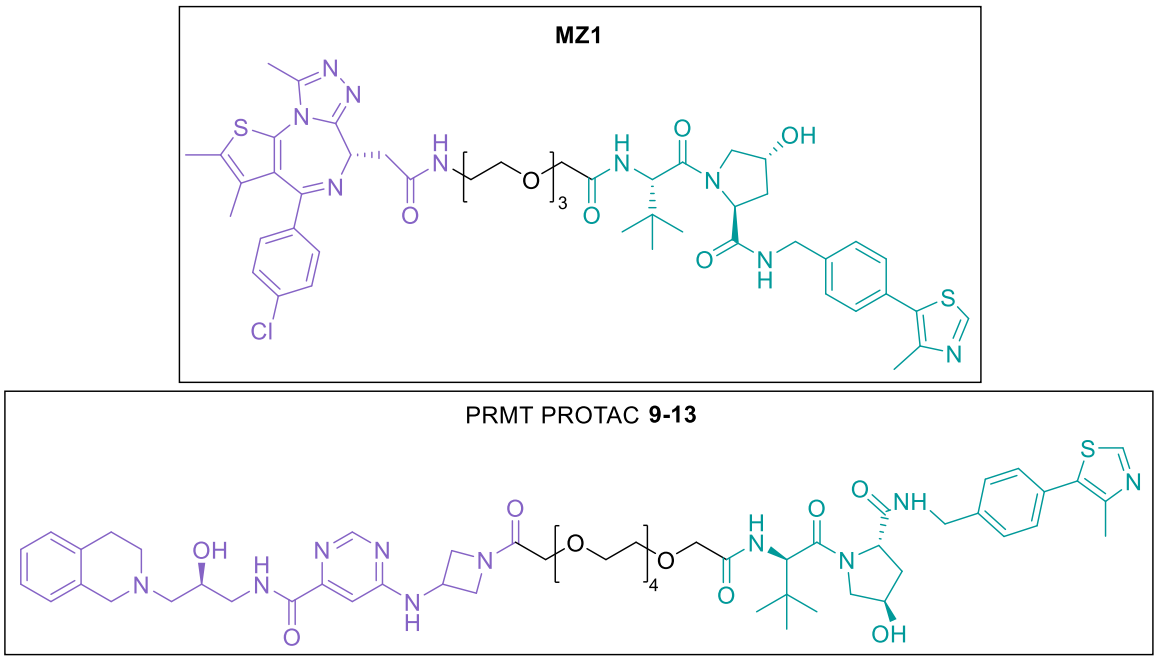


Figure 4.3 Structures of selected epigenetic PROTACs. MZ1 selectively degrades BRD4 and has been extensively studied.[312] PRMT5 PROTAC **9-13** is a highly selective degrader of methyltransferase PRMT5, established using global proteomics.[315] POI binders in purple, E3 ligase binders in teal, linkers in black.

First-in-class degraders of arginine methyltransferase 5 (PRMT5) were identified (**Figure 4.3**), and exquisitely high selectivity was demonstrated by global proteomics data showing that the level of only three of the proteins detected were reduced. Furthermore the PRMT5 PROTAC was able to induce degradation of a PRMT5 protein binding partner, alongside PRMT5.[315] Methyltransferase targeting PROTACs have also been developed, which degrade the core subunit components of polycomb repressive complex 2 (PRC2): EED, EZH2, and SUZ12. PRC2 hypermethylates H3K27, and drives tumorigenicity in many cancers. The PRC2 PROTACs are capable of inhibiting proliferation

of PRC2-dependent cancer cells; the mode of action may combat resistance to catalytic inhibitors, as well as allow regulation of PRC2-dependent cancer cells that are independent on the catalytic function of PRC2.[192]

HDACs are the only erasers for which PROTACs have been reported so far, although HDAC protein degradation has received significant attention.[316–318] A panel of 48 HDAC targeting PROTACs were tested using chemoproteomics, identifying a range of selectivity profiles and also showing degradation of protein complex partners in some instances.[316] Degradation of protein interaction partners within a complex could be an advantage of PROTACs compared to inhibitors for epigenetic targets,[310] this was observed with PRMT5, PRC2, and HDAC PROTACs.

4.1.4 The potential of macrocyclic peptides PROTACs

The highly selective and potent nature of macrocyclic peptides[147] make them excellent options for selective protein degradation, and the number of reports of peptide PROTACs is expected to increase.[319] Peptides identified using genetically encoded selection methodology have a growth point for developing heterobifunctional molecules, without extensive structure activity relationship studies, by extending the peptide at the point at which genetic information was ligated to the peptide during selection. Furthermore, PROTACs require binders of a POI, as opposed to modulators, which are rapidly identified using genetically encoded selection methodology.[162] It is possible that the catalytic mode of action of PROTACs might be more favourable for macrocyclic peptides than inhibition strategies, as peptides suffer poor cell permeability, peptide protein degraders may require less compound for a biological effect.[103] Furthermore, many PROTACs already fall outside of the parameters of traditional small molecule druggable space, but are still able to modulate their targets inside cells.[191]

4.1.5 Chapter aims

This chapter aims to investigate KDM4A protein degradation using macrocyclic peptides functionalised with E3 ligase binders. Initial work will establish that KDM4A can be ubiquitinated

by E3 ligases and degraded in cells, before turning the attention towards the development of **CP2**-based macrocyclic peptide degraders for KDM4A. The work is broken down into the following objectives:

- Clone a construct for mammalian cell expression of HaloTag KDM4A.
- Use HaloPROTACs to establish conditions in which transiently overexpressed KDM4A degradation can be observed.
- Assess the in-cell activity of the developed panel of macrocyclic peptides conjugated to E3 ligase binders as PROTACs (Chapter 2).
- Study macrocyclic peptide protein degradation in detail.

4.2 Results

4.2.1 Establishing KDM4A protein degradation using HaloTag targeting PROTACs

HaloTag-targeting PROTACs, known as HaloPROTACs, can be used to degrade HaloTag fusion proteins, without needing a specific ligand for the POI.[193] HaloTags are derived from bacterial dehalogenases, they covalently react with hexyl chloride, and a wide variety of molecules have been conjugated to HaloTag ligands for protein labelling, and for protein analysis.[119] HaloPROTACs offer a very versatile system to explore degradability of target proteins, however the irreversible covalent linkage used means they do not benefit from the catalytic mode of action of reversibly binding PROTACs.

The first HaloTag targeting PROTACs were developed by Buckley *et al.* (2015),[193] who developed a number of small molecule VHL recruiting PROTACs against HaloTag-green fluorescent protein; HaloPROTAC1 (up to 20% degradation, 2.5 µM, 24-hour dose) and HaloPROTAC3 (DC_{50} = 19 nM, 24-hour dose) are shown in **Figure 4.4**, and use different VHL ligand attachment points. HaloPROTAC3 has an isoindolinone where the attachment is in HaloPROTAC1. Inspired by this work Tovell *et al.* (2019)[194] developed HaloPROTAC-E (DC_{50} = 3-10 nM, 24-hour dose) based upon HaloPROTAC3, but employing a cyclopropylcyanoacetic moiety in place of the isoindolinone group

to improve VHL binding affinity, and hopefully PROTAC activity. Interestingly, HaloPROTAC3 and HaloPROTAC-E do not display a hook effect (dosing up to 10 µM), but HaloPROTAC1 has quite a small window of activity between single and double digit micromolar.[193,194]

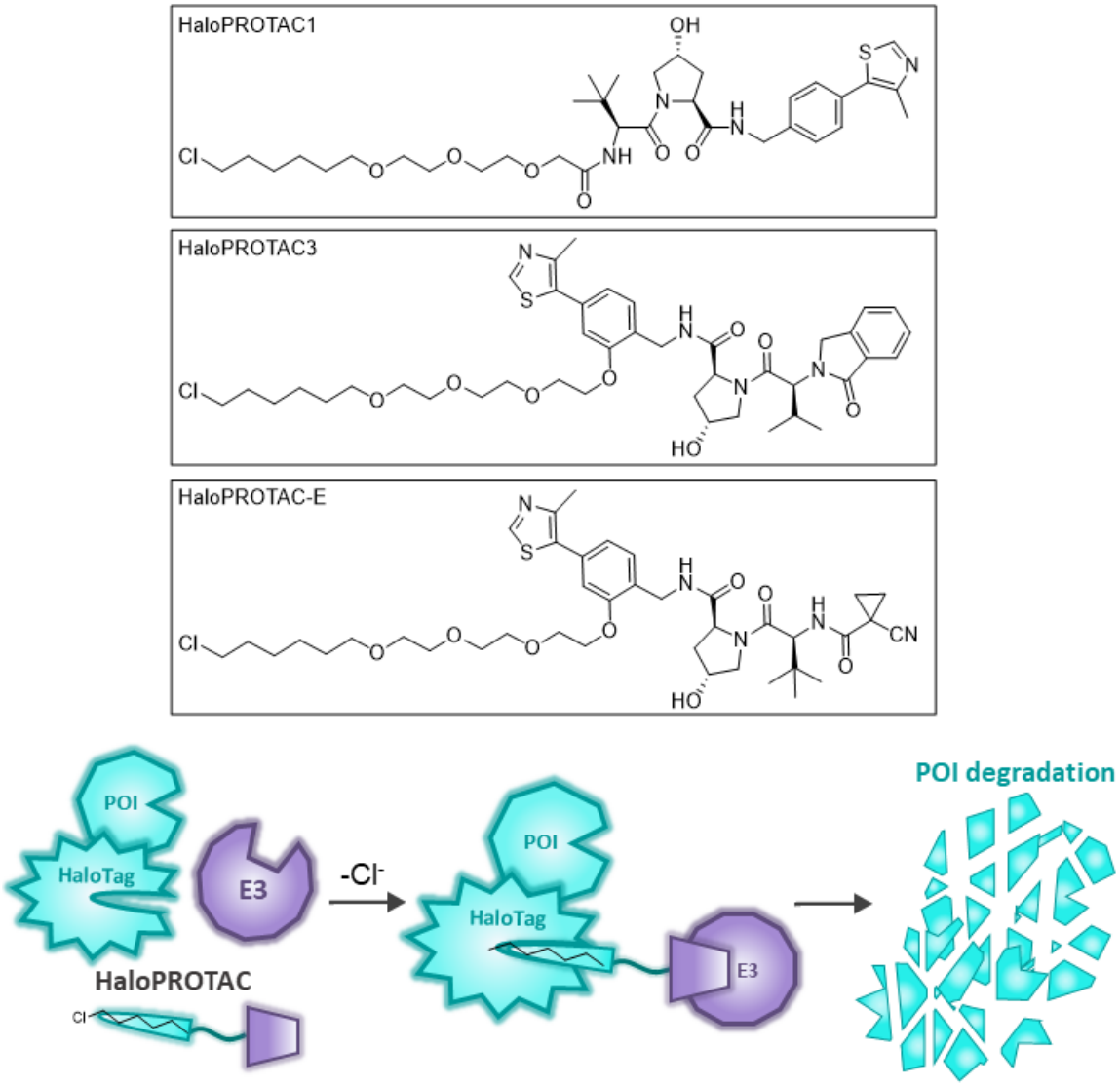


Figure 4.4 HaloPROTAC mode of action, and structures of HaloPROTAC compounds discussed.[193,194] HaloPROTACs covalently bind HaloTag fusion proteins, and the HaloTag protein dechlorinates the HaloTag ligand. E3 ligases are positioned in close proximity to the fusion protein, allowing POI degradation *via* the UPS.

In this section the degradability of KDM4A protein was investigated using HaloPROTACs. HaloPROTAC1 from Buckley *et al.* (2015)[193] (MedChemExpress) and HaloPROTAC-E from Tovell *et al.* (2019)[194] (gifted from Ciulli laboratory, University of Dundee) were employed to establish KDM4A protein degradation assays using cells transiently overexpressing HaloTag and 3 X Flag-tag KDM4A (HaloFlagKDM4A). These compounds have two different length linkers (PEG_2/PEG_3), and

different linker attachment points, so it was hoped that one of these compounds would successfully degrade HaloTag KDM4A protein in cells.

4.2.1.1 Cloning FlagKDM4A into a pFN21A vector

Initially *KDM4A* with an *N*-terminal triple Flag-tag sequence was cloned into a pFN21A HaloTag® CMV Flexi® Vector (Promega), which can be used for mammalian expression of a POI with an *N*-terminal HaloTag (**Figure 4.5**, Section 6.11). This yielded mammalian cell expression of HaloFlagKDM4A, which consists of an *N*-terminal HaloTag, 3 X Flag-tag, and full length KDM4A, under the CMV promoter. A Flag sequence was included to allow highly specific Flag antibodies to be used to quantify KDM4A protein levels by western blot, since antibodies against KDM4A have been found to be less specific and more expensive. Flag/KDM4A antibodies were preferred over HaloTag antibodies as Flag-detection had been established for FlagKDM4A in Section 3.2.1.

This cloning strategy uses positive selection pressure, replacing the lethal barnase gene with an insert of interest (*FlagKDM4A*).[320] The barnase gene is flanked by SgfI and PmeI restriction enzyme cutting sites, which are rare, and not contained in *KDM4A*. An SgfI cutting site was ligated to the *N*-terminus of the insert, and a PmeI cutting site to the *C*-terminus. The vector and insert were then combined, and only plasmids which have had the barnase gene replaced with the insert survived when transformed into *E. coli.* Successful cloning of plasmids for wild type (wt) *KDM4A* and a catalytically inactive mutant (mut) (H188A) was confirmed by Sanger sequencing of the constructs. The mutation in KDM4A prevents Fe(II) binding by the His/Glu catalytic triad, which in turn inactivates the enzyme.[33]

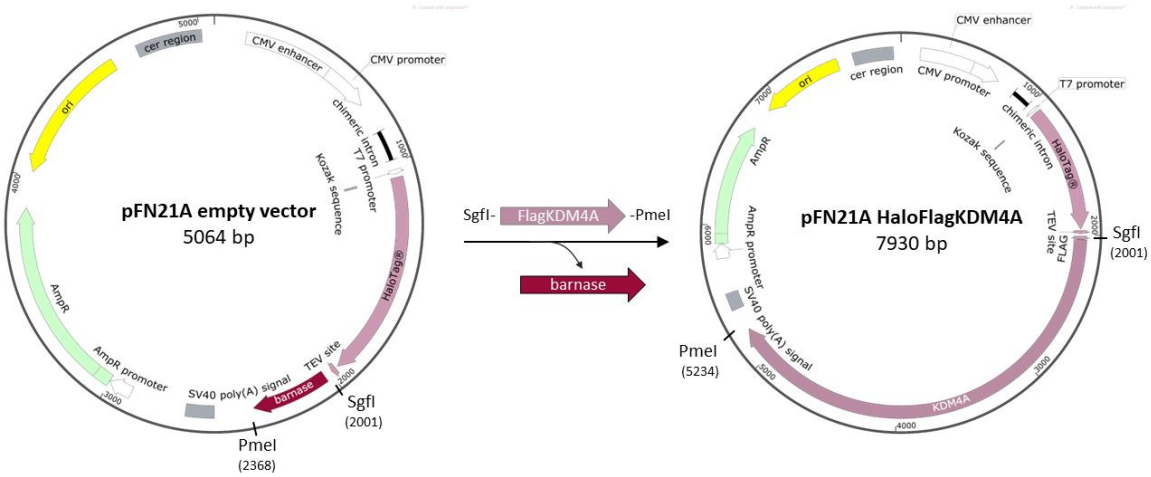


Figure 4.5 Cloning strategy for production of *HaloFlagKDM4A* vector. *N*-terminally triple Flag-tagged *KDM4A* was inserted into a pFN21A HaloTag® CMV Flexi® Vector, allowing mammalian expression of HaloFlagKDM4A protein. Plasmid maps generated in SnapGene Viewer 5.4.2

Overexpression of HaloFlagKDM4A protein was evaluated in HEK293T cell extracts, lysed 48 hours after transfection with the pFN21A *HaloFlagKDM4Awt/mut* construct. Subsequently, the cells were washed and lysed, and the cell lysates were analysed using western blotting, probing the membrane for Flag *versus* GAPDH as a loading control (**Figure 4.6**A). Expression levels of both HaloFlagKDM4Awt/mut protein were relatively similar (as quantitated using western blots probed with Flag antibody), and no Flag signal was detected in the 'no DNA' control. For the 'no DNA' control, transfection reagent without plasmid DNA was applied to cells, analogously to the transfected samples. This indicated that the system was appropriate for quantitating HaloFlagKDM4A protein levels and for testing PROTAC compounds. It is worth noting that immunofluorescence experiments investigating FlagKDM4A (without HaloTag) in U2OS cells have variable expression of KDM4A protein across the cells (**Figure 4.6**B).[321] This is common with transiently transfected cells, and is a benefit of using cell lines stably expressing the construct of interest, which express the POI more homogeneously as they are generated from a single cell.[322] However, as a proof-of-concept study, protein degradation was initially tested in transiently transfected cells.

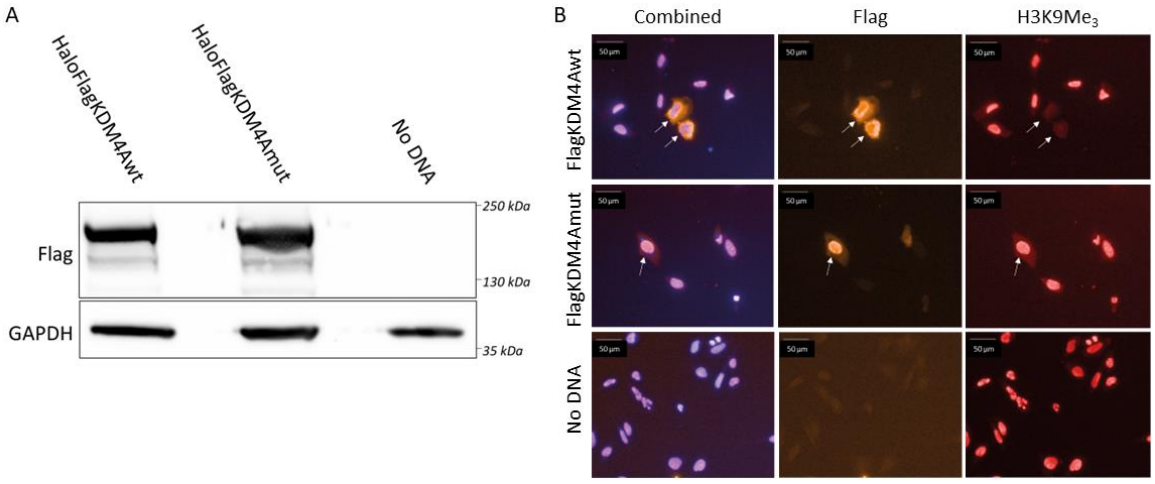


Figure 4.6 Transient overexpression of KDM4A constructs. (A) Western blot analysis of HEK293T cells transiently transfected with *HaloFlagKDM4A*. *HaloFlagKDM4Awt/mut*, or no DNA and incubated for 48 hours prior to cell lysis; Flag antibody was used to detect HaloFlagKDM4A and GAPDH was used as a loading control. (B) Representative immunofluorescence analysis of U2OS cells transiently transfected with FlagKDM4A. Data shows that not all cells express KDM4A at the same level, likely due to different transfection efficiency across the cells. DNA was transiently transfected, and the media was changed to 1% DMSO 4 hours post-transfection. Cells were fixed 16 hours after media change. HaloFlagKDM4A = 157 kDa, GAPDH = 36 kDa. Section B adapted from Roper (2017).[321]

4.2.1.2 HaloPROTAC protein degradation

Optimal conditions for degradation of KDM4A protein were investigated using HaloPROTACs dosed on HEK293T cells transiently overexpressing HaloFlagKDM4A. Several factors were considered including length of time post-transfection before compounds were added, length of compound treatment, and compound concentration.

4.2.1.2.1 HaloPROTAC treatment 48 hours post-transfection

HaloPROTACs were dosed on HEK293T cells 48 hours post-transfection, with this transfection length chosen as cells lysed 48 hours post-transfection had maximal FlagKDM4A protein levels in HEK293T cell lysates generated in Section 3.2.1.1.1.

An initial time course experiment dosing HaloPROTAC1, and HaloPROTAC-E (1 µM) for 1–24 hours showed no KDM4A protein degradation at any of the time points tested (**Figure 4.7**A). Next, HaloPROTAC compounds were tested at a range of concentrations, since the PROTACs may have been displaying the hook effect,[104] or cells may have needed higher compound concentrations to observe a biological effect. A PROTAC concentration above and below 1 µM was selected, 10 µM was used as the highest dose for both compounds, and the lowest concentrations were based on

the degradation activity in the respective original publications: 500 nM for HaloPROTAC1,[193] and 300 nM for HaloPROTAC-E.[194] HaloPROTACs were added to cells for 4 hours, which should be short enough to prevent secondary effects of protein degradation (such as increased protein translation), and was used extensively tested by Tovell *et al*. (2019).[194] HaloPROTAC1 (0.5, 1, and 10 µM), and HaloPROTAC-E (0.3, 1, and 10 µM) were tested across three biological replicates, however HaloFlagKDM4A protein degradation was not observed under these conditions (**Figure 4.7**B/C).

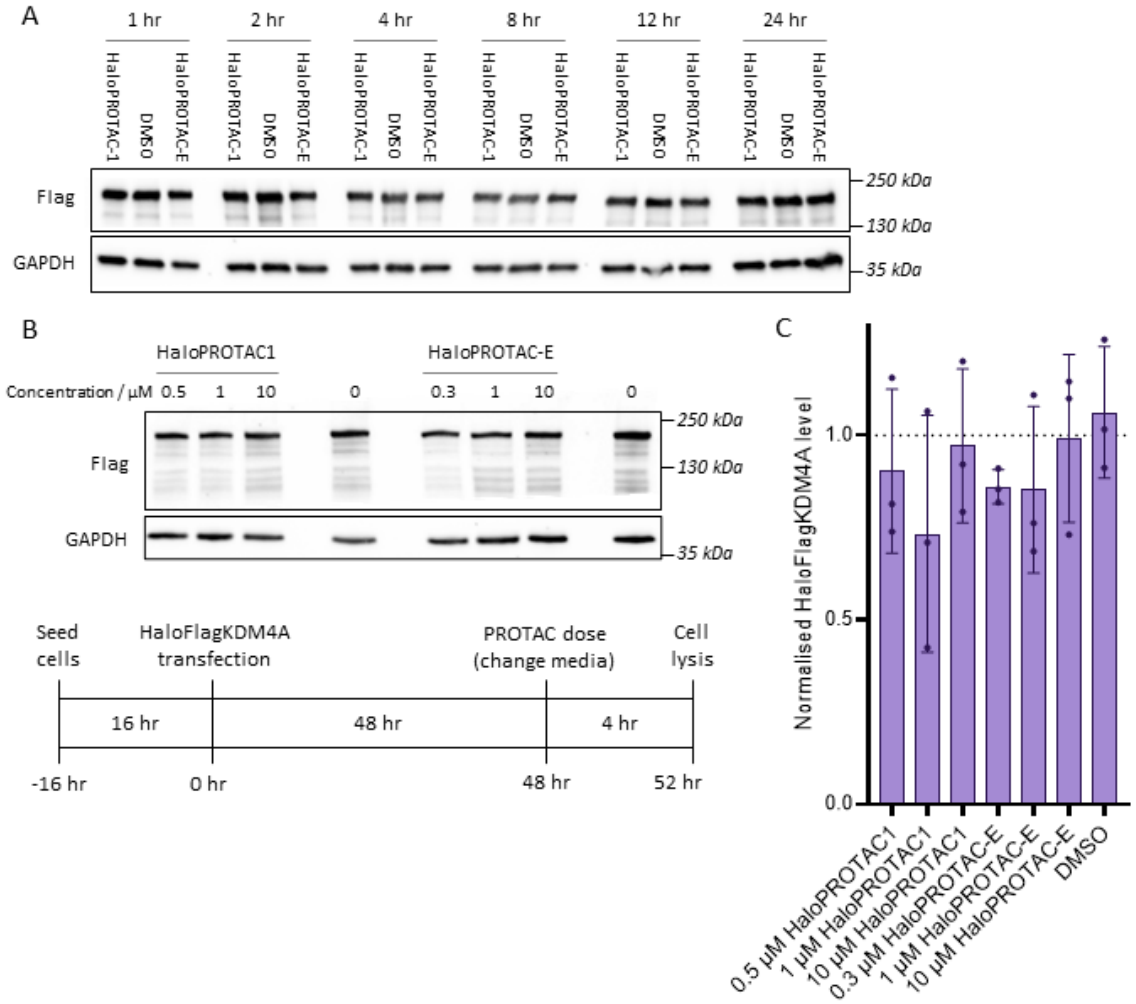


Figure 4.7 HaloPROTACs are not active when dosed on HEK293T cells transfected with *HaloFlagKDM4A* 48 hours before treatment. (A) Western blot analysis for cells treated with 1 µM HaloPROTAC1/HaloPROTAC-E, lysing cells 1, 2, 4, 8, 12, and 24 hours after compound addition. (B) Western blot analysis of cells treated with varying concentrations of HaloPROTAC1 or HaloPROTAC-E for 4 hours. Representative data shown from three biological replicates. (C) Quantification of HaloFlagKDM4A protein level normalised to GAPDH across three biological repeats from (B) using densitometry. A One-Way ANOVA test with multiple comparisons to DMSO control (GraphPad Prism 9.4.0) found no significant changes in normalised KDM4A protein level when exposed to HaloPROTAC(1/-E) compared to DMSO control under these conditions. Protocol scheme shows length of treatments inside the box, and time point within the experiment below. (n = 3, mean ± standard deviation) HaloFlagKDM4A = 157 kDa, GAPDH =36 kDa.

HaloFlagKDM4A protein degradation was not observed when different HaloPROTAC incubation times were tested, or when different compound concentrations were used. It was hypothesised that KDM4A protein levels may be saturating, such that low levels of degradation cannot be observed when HaloPROTACs were treated on cells 48 hours post-transfection, and thus additions of HaloPROTACs a shorter time after transfection was tested in Section 4.2.1.2.2.

4.2.1.2.2 HaloPROTAC treatment 4 hours post-transfection

Next, HEK293T cells were transfected with *HaloFlagKDM4A* 4 hours before treatment with HaloPROTACs for 4 hours, using the same doses as in **Figure 4.7**B. Significant degradation was observed at 1/10 µM HaloPROTAC-E across three biological replicates (**Figure 4.8**). Greater than 60% protein degradation was seen on average with 10 µM HaloPROTAC-E, and at 1 µM HaloPROTAC-E approximately 50% protein degradation was observed. Cells treated with 500 nM HaloPROTAC1 showed variable levels of degradation, which are not significant across three biological replicates. β-actin was used as the loading control, rather than GAPDH in this experiment, after variable levels of GAPDH were observed in separate experiments increasing DMSO concentration to 1% from 0.1%. DMSO concentration is 0.1% in all experiments unless otherwise noted.

These results showed that KDM4A protein can be degraded using a VHL ligand in HEK293T cells under these conditions, and that KDM4A is a tractable target for protein degradation studies. These conditions were therefore used to test other potential protein degraders in cells transiently overexpressing KDM4A in subsequent cyclic peptide PROTAC work, and HaloPROTAC-E (10 µM) was used as a positive control for protein degradation. It is possible that treating cells with lower concentrations of HaloPROTAC1 may have resulted in consistent degradation, however this was not tested, and Buckley *et al.* (2019) achieved less than 20% degradation in their publication.[193]

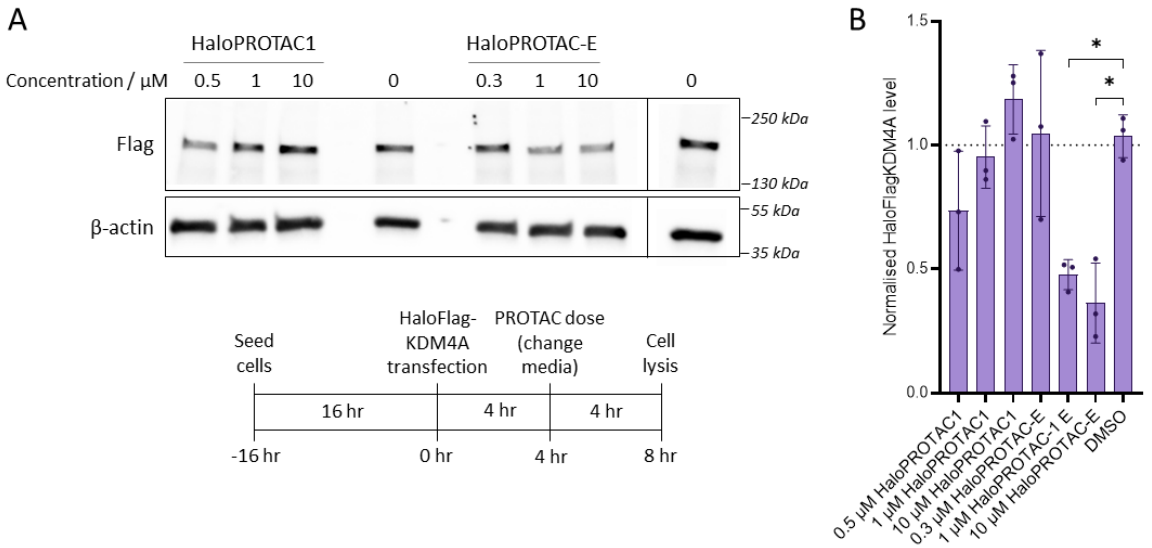


Figure 4.8 HaloPROTAC-E consistently degrades KDM4A protein in HEK293T cells transfected with *HaloFlagKDM4A* 4 hours before treatment with HaloPROTACs. (A) Western blot analysis of cells dosed with varying concentrations of HaloPROTAC1 and HaloPROTAC-E for 4 hours. Representative data shown from three biological replicates. (B) Quantification of HaloFlagKDM4A protein level normalised to β-actin across three biological repeats from (A) using densitometry. A One-Way ANOVA test with multiple comparisons to DMSO control (GraphPad Prism 9.4.0) found significant changes in normalised KDM4A protein level when incubated with 1 µM (P* = 0.0492) and 10 µM (P* = 0.0365) HaloPROTAC-E compared to DMSO control under these conditions, unlabelled bars were not significant. Dosing time scheme shows length of treatments inside the box, and time point within the experiment below. (n = 3, mean ± standard deviation) HaloFlagKDM4A = 157 kDa, β-actin = 42 kDa.

4.2.1.2.3 Investigation of HaloPROTAC-E mode of action

HaloPROTAC-E showed promising protein degradation activity in cells overexpressing HaloFlagKDM4A. To validate that the reduction in KDM4A protein level was due to proteasomal degradation as opposed to other mechanisms, further control experiments were carried out using proteasome inhibitors and VHL inhibitors.

PROTACs may be co-incubated with proteasome inhibitors to demonstrate that degradation is not observed when the proteasome has reduced activity. MG132 is a proteasome inhibitor (**Figure 4.9**), which is frequently used in PROTAC studies.[323] MLN4924 is a neddylation inhibitor, and neddylation is a process related to the UPS.[324] Neddylation is necessary for the function of cullin-RING E3 ligases, and inhibition of neddylation with MLN4924 (**Figure 4.9**) has been used to demonstrate that protein degradation is occurring *via* the UPS for PROTACs which recruit cullin-RING E3 ligases, including VHL/CRBN,[325] similarly to MG132.[326] Thus, MG132 and MLN4924 were used as controls in these assays. Further, HaloPROTAC-E was also co-incubated with VHL

ligands as competitors for VHL, which should reduce degradation by out competing HaloPROTAC-E from binding to VHL and, as such, lowering the number of ternary complexes formed. Both VH298 (K_d = 90 nM)[327] and VH032 (K_d = 185 nM)[328] were tested, which are potent VHL inhibitors, (**Figure 4.9**). These control compounds were added 30 minutes prior to HaloPROTAC treatment, to ensure they had entered the cells before HaloPROTAC-E, as different rates of cell permeation may impact results. The final DMSO concentration used in these experiments was 0.2%.

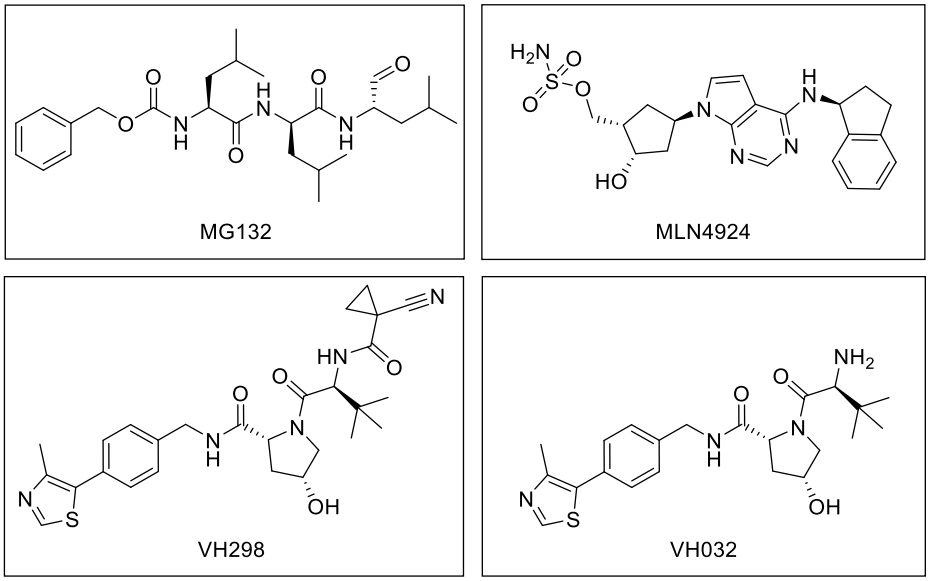


Figure 4.9 Structures of proteasome inhibitors and VHL inhibitors used to validate HaloFlagKDM4A protein degradation. MG132 is a proteasome inhibitor,[323] MLN4924 a neddylation inhibitor,[324] and VH298/VH032 are VHL inhibitors.[327,328]

MG132 significantly reduced HaloFlagKDM4A protein level (**Figure 4.10**A) when co-incubated with 1/10 µM HaloPROTAC-E. However, similar reduction in KDM4A protein levels were observed in cells incubated with MG132 alone, with levels <20% relative to cells incubated with the DMSO control. Thus, co-incubation of PROTACs with MG132 was not an appropriate control for validating that protein degradation is occurring through the proteasome in this system. Co-incubation of MLN4924 and HaloPROTAC-E inhibited KDM4A protein degradation compared to the DMSO control (**Figure 4.10**A-C), which suggests the UPS is required for HaloPROTAC-E degradation. However, these results are complicated as MLN4924 alone stabilised HaloFlagKDM4A protein

levels, which may also suggest that MLN4924 is not an appropriate tool for validating the mode of action of HaloPROTAC-E as degradation *via* the UPS.

The effect of VH298 and VH032 on HaloFlagKDM4A protein levels varied across three independent experiments (**Figure 4.10**A-C), in some instances there was little KDM4A protein degradation, but in other cases degradation was observed. In most cases, when 10 µM HaloPROTAC-E was co-incubated with VH298/VH032, HaloFlagKDM4A protein level remained the same compared to the VHL inhibitors alone, and a reduction in degradation was observed compared to HaloPROTAC-E treatment alone **Figure 4.10**A-B. A One-Way ANOVA analysis (GraphPad Prism 9.4.0) did not find a significant difference in KDM4A protein level between cells treated with VH298 followed by either DMSO or 10 µM HaloPROTAC-E across the three biological replicates shown in **Figure 4.10**9A-C. Reduced degradation when HaloPROTAC-E was co-incubated with VHL binding ligand (10-fold) suggests that HaloPROTAC-E is acting *via* recruitment of VHL.

MLN4924 co-incubation suggests degradation is occurring *via* the UPS, and VH298/VH032 results suggest VHL is recruited by HaloPROTAC-E for degradation activity. However, the results from these validation experiments are complicated by alterations in the level of HaloFlagKDM4A protein when the proteasome inhibitors and VHL inhibitors were added alone. This could indicate that KDM4A interacts with the proteasome and/or VHL; there are number of papers reporting a role for KDM4A in regulation of protein levels by multiple mechanisms: interactions with E3 ligases to promote protein degradation,[329] protein stabilisation interactions,[330] interaction with deubiquitinases,[331] and regulation of protein synthesis.[272] There is much more to discover about the role of KDM4A in cells, and its role appears to be context dependent, illustrated by the observation that KDM4A has different roles in the cytoplasm and nucleus.[272]

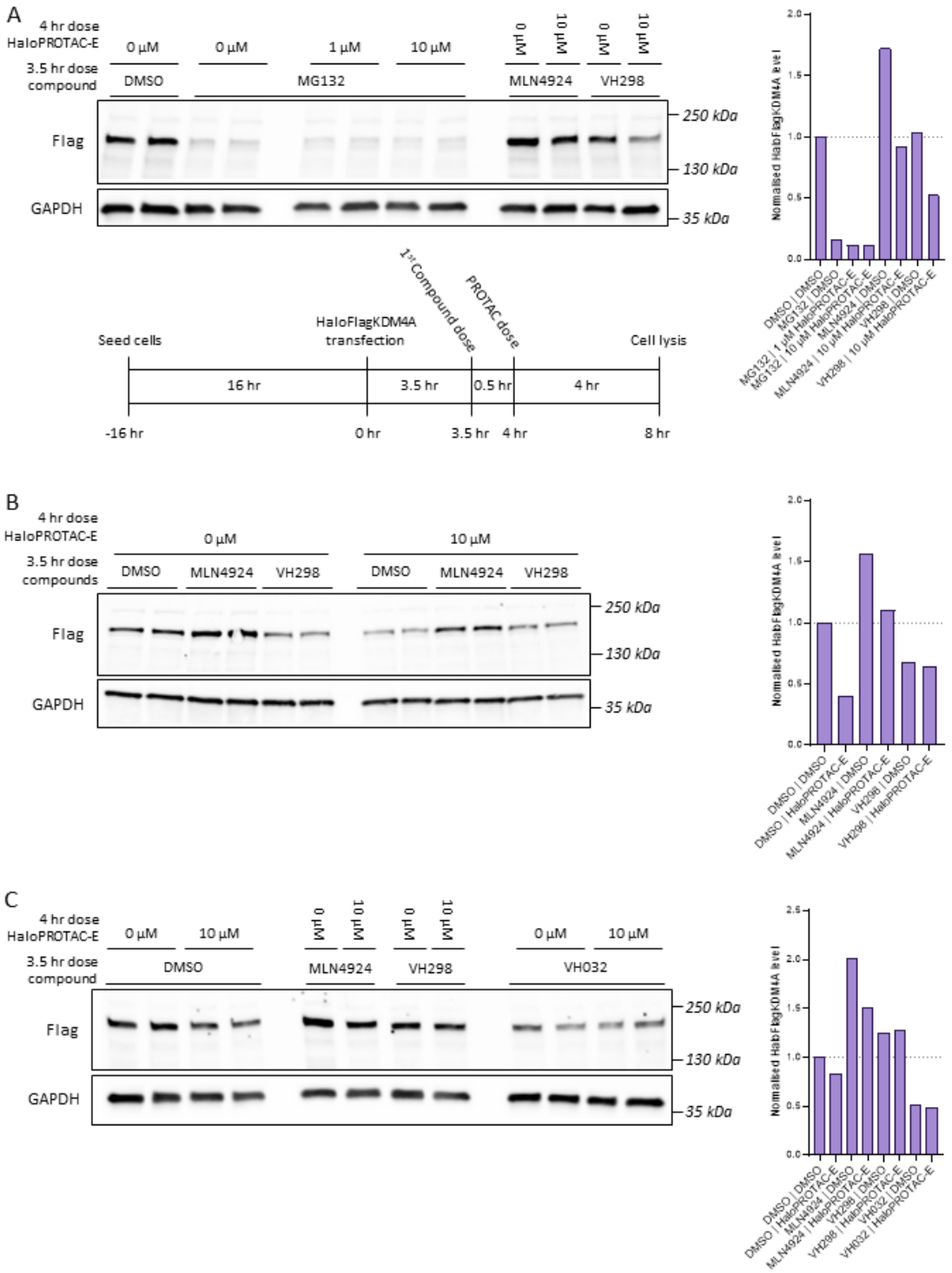


Figure 4.10 HaloPROTAC-E validation experiments, co-dosing proteasome inhibitors and VHL inhibitors. HEK293T cells overexpressing HaloFlagKDM4A protein were incubated with MG132/MLN4924/VH032/VH298 3.5 hours after transfection, followed by HaloPROTAC-E treatment 30 minutes later, cells were lysed 4 hours after PROTAC treatment. (A) Western blot and quantification data from HEK293T cells treated with MG132, MLN4924, or VH298 followed by HaloPROTAC-E. Compound end concentrations: 50 µM MG132, 3 µM MLN4924, VH298 100 µM, 1/10 µM HaloPROTAC-E. Protocol scheme used in these experiments shows length of treatments inside the box, and time point within the experiment below. (B) Western blot and quantification data from HEK293T cells dosed with MLN4924, or VH298, followed by HaloPROTAC-E. Compound doses: 3 µM MLN4924, 100 µM VH298, 10 µM HaloPROTAC-E. (C) Western blot and quantification data from HEK293T cells treated with MLN4924, VH298, or VH032, followed by HaloPROTAC-E. Compound concentrations: 3 µM MLN4924, 100 µM VH298, 100 µM VH032, 10 µM HaloPROTAC-E. (Graphs: n = 1) HaloFlagKDM4A = 157 kDa, GAPDH = 36 kDa.

4.2.2 Cyclic peptide protein degradation

CP2 is highly selective for KDM4A-C, but has limited evidence of cell activity.[40] Linking to an E3 ligase binder will increase the size of the molecule (≈2,500 Da), further beyond Lipinski's rule-of-five.[145] However, many cell active PROTACs fall outside traditional drug-like parameters, and often have cell activity at lower doses compared to the parent inhibitor used for PROTAC development.[103] Therefore it was hypothesised that protein degradation could be effective at improving the cell activity of **CP2**. Furthermore, establishing protein degradation *via* use of cyclic peptides in PROTAC molecules may allow a general route to cell activity for both active site targeting, and allosteric site targeting, peptides.

CP2 was linked to ligands for both VHL and CRBN using a range of linkers, to select for optimal macrocyclic peptide-based protein degraders (**Figure 4.11**). For this work VHL and CRBN were used because the ligands targeting them are the best established in the field.[107] GSK kindly provided VHL and CRBN binders attached to alkynes with 1–10 carbon saturated alkyl chains, an additional VHL ligand attached to an alkyne by PEG_2 was also procured (Biotechne Europe Ltd). There are no general rules for linkers in PROTACs, and testing different lengths and linker types is often necessary to develop a successful PROTAC.[205] All carbon linkers allow connection of the molecules without additional oxygen/nitrogen heteroatoms, preventing additional hydrogen bonding interactions which can reduce cell permeability, and the associated increased lipophilicity can benefit cell permeability. A PEG linker was also tested as PEG can improve compound solubility, and is a common linker in the development of PROTACs.[332,333] Successful PROTACs form stable ternary complexes, but predicting PROTAC structures to achieve this is challenging;[334] trial and error is normally necessary to develop protein degraders, therefore the panel of PROTACs with varied E3 ligase binders and linkers were developed (Section 2.2.3.2). In total, twelve bifunctional cyclic peptide-E3 ligase binder peptide conjugates were synthesised, using CuAAC click chemistry to form the final products.

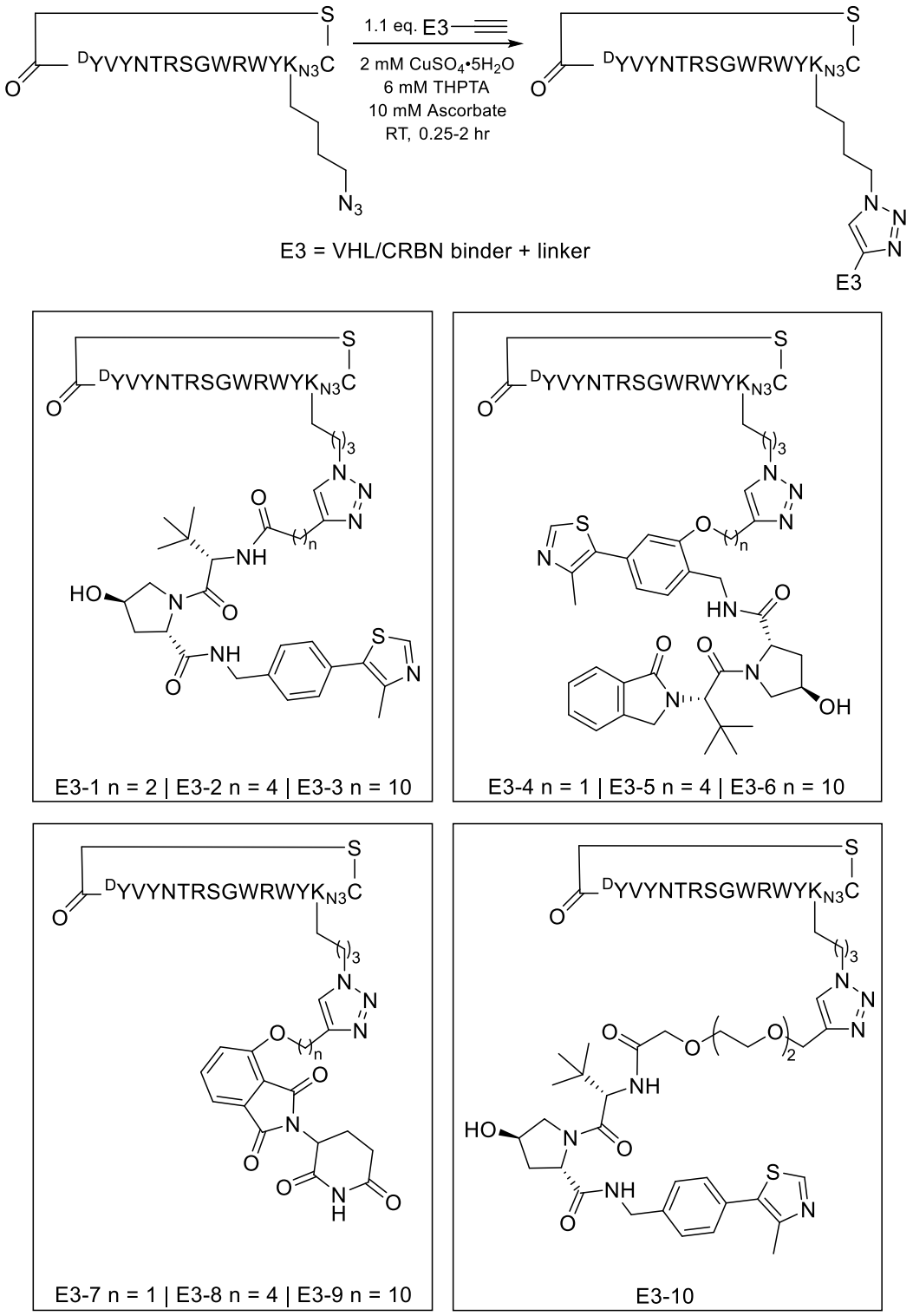


Figure 4.11 Click chemistry was used to make a library of potential CP2-PROTACs. Bottom left box shows CRBN targeting PROTACs (thalidomide based), and the remainder are VHL targeting (VH032 based).

4.2.2.1 Testing a panel of macrocyclic peptide protein degraders

A panel of ten **CP2**-PROTACs, **E3-1** – **E3-10**, were initially synthesised, using click chemistry to link **CP2-T13K$_{N3}$** to E3 ligase binders connected to a terminal alkyne *via* variable length linkers. Compounds were obtained at greater than 95% purity (analytical HPLC, Appendix 1). The panel was tested at 10 μM on HEK293T cells transiently overexpressing HaloFlagKDM4A, using conditions established in Section 4.2.1.2.2. KDM4A protein levels were compared to HaloPROTAC-E, unmodified **CP2**, and DMSO controls, incubating cells with compounds for 4 hours, 4 hours post-transfection.

Across four biological replicates three compounds consistently degraded KDM4A (**Figure 4.12**) – **E3-8** (mean KDM4A protein level = 33%), **E3-9** (mean KDM4A protein level = 35%), and **E3-10** (mean KDM4A protein level = 29%). **E3-8** and **E3-9** are both CRBN recruiting PROTACs, and **E3-10** recruits VHL. Many of the other cyclic peptide degraders showed variable activity across the replicates. HaloFlagKDM4A protein levels were not significantly changed when treated with **CP2**, demonstrating that conjugation to an E3 ligase binder is necessary for **CP2**-PROTACs to reduce KDM4A protein levels. HaloPROTAC-E was able to consistently degrade HaloFlagKDM4A protein, suggesting that it is an appropriate positive control.

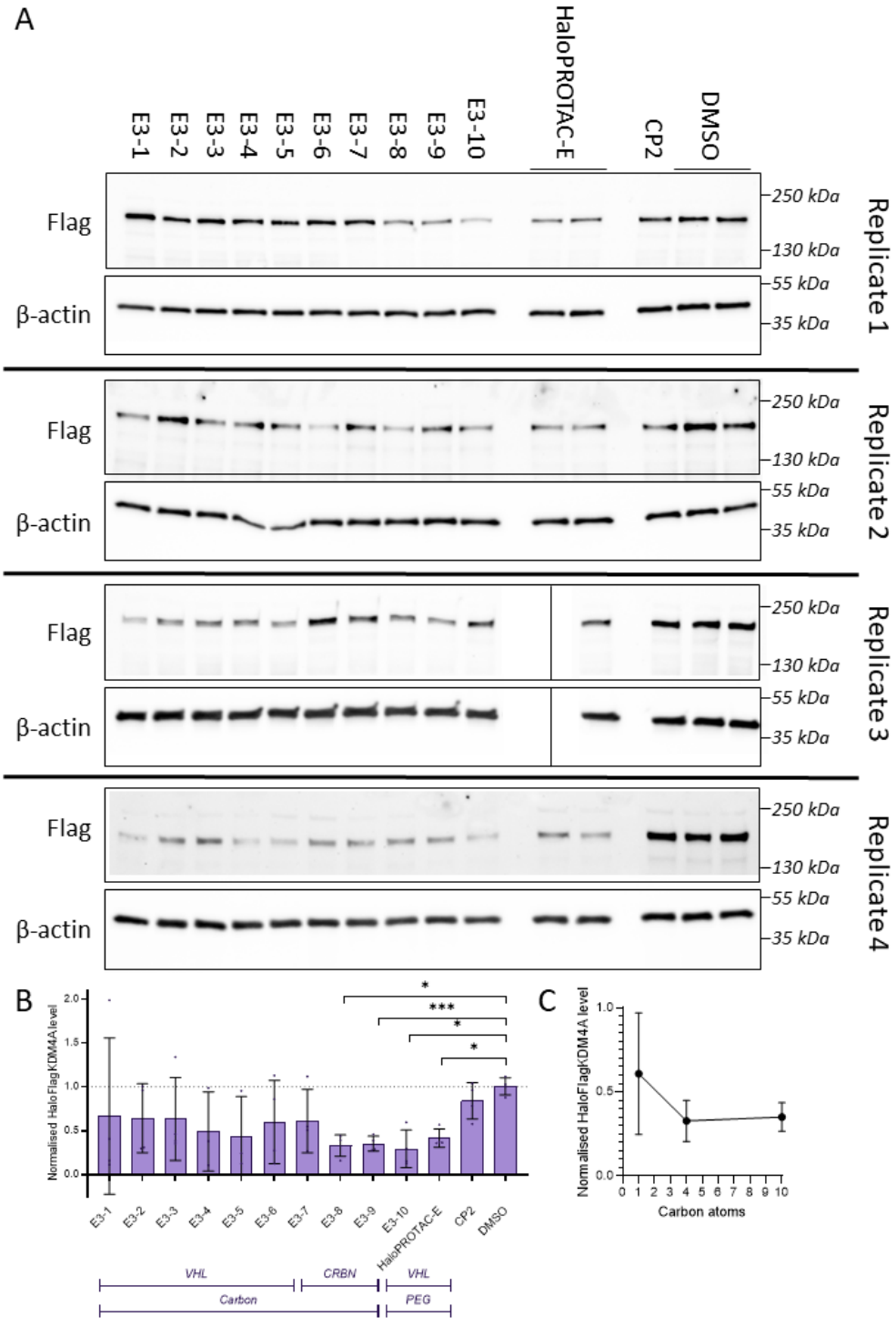


Figure 4.12 KDM4A protein degradation by the panel of CP2-PROTACs in cells overexpressing HaloFlagKDM4A. (A) Western blot analysis of four biological replicates, treating HEK293T cells with 10 µM **CP2**-PROTACs for 4 hours, 4 hours post-transfection with HaloFlagKDM4A. **CP2** concentration was 12.4 µM in replicate 1, and 10 µM in subsequent replicates. (B) Quantification of HaloFlagKDM4A protein level normalised to β-actin across four biological replicates using densitometry. Replicate 2 for **E3-4** and **E3-5** were excluded due to an issue with the loading control for these samples. Row statistics with multiple comparisons to DMSO control (GraphPad Prism 9.4.0) found significant changes in normalised HaloFlagKDM4A protein level when cells were treated with **E3-8** (P* = 0.0246), **E3-9** (P*** = 0.0007), **E3-10** (P* = 0.0294), and HaloPROTAC-E (P* = 0.0162) compared to DMSO control under these conditions, unlabelled bars indicate no significance. One DMSO/HaloPROTAC-E lane per experiment was chosen for this analysis (n = 3-4, mean ± standard deviation). (C) Correlation of linker length compared to HaloFlagKDM4A protein level for **E3-7** – **E3-9**, CRBN targeting PROTACs (n = 4, mean ± standard deviation). HaloFlagKDM4A = 157 kDa, β-actin = 42 kDa.

Successful **CP2**-PROTACs had medium and long length linkers, suggesting that the shorter linkers did not leave enough space to form a stable ternary complex between KDM4A, PROTAC, and E3 ligase. **E3-10** is the only PROTAC which has a PEG linker, as opposed to an all carbon linker. Interestingly, **E3-3** and **E3-10** differ only in the linker length (1 atom) and linker composition, but average KDM4A protein level in cells treated with **E3-3** were reduced to 63%, compared to 29% for **E3-10**. Possible reasons for the difference seen could be: lower cell permeability of **E3-3** compared to **E3-10**, productive linker-protein interactions with **E3-10** improving the stability of the ternary complex, or differing cellular stability. However, **E3-8** and **E3-9** have a 6 atom difference in linker length, but very little difference in mean KDM4A protein level across the experiments (HaloFlagKDM4A protein level with **E3-8** and **E3-9** was 33% and 35% respectively, **Figure 4.12**C). With the CRBN targeting PROTACs the most important factor appears to be including sufficient distance between **CP2** and the E3 ligase ligand to allow a stable ternary complex to be formed. These results highlight the importance of linkers in PROTACs.[139] Overall, KDM4A protein degradation was observed using PROTACs targeting VHL and CRBN, but different linker chemistry was required for the PROTACs recruiting different E3 ligases. It is also interesting that within the same VHL ligand, the linker positioning had an effect (**E3-3**, **E3-6**, **E3-10, Figure 4.11** & **Figure 4.12**).

For many compounds within the **CP2**-PROTAC panel, the level of KDM4A protein degradation varied across the repeats. However, attempts to test the panel at higher concentrations (17–20 μM) were unsuccessful due to solubility problems which remained when DMSO concentration was increased from 0.1% to 1%. DMSO concentrations were capped at 1% in these investigations due to cytotoxicity of DMSO.[335,336] It is possible that transient transfection increases the variability between experiments compared to using stably expressing cell lines, as each transfection will have slightly different efficiency. However, the conditions optimised using transient transfection were successful at identifying cyclic peptide PROTACs.

4.2.2.2 Detailed investigation of E3-10

Out of all the VHL-recruiting PROTACs tested, **E3-10** was the most efficient protein degrader. This meant that it can be used alongside HaloPROTAC-E as a control for the same E3 ligase pathway. **E3-10** also had slightly higher solubility than compounds **E3-8** and **E3-9**, thus **E3-10** was selected for further investigation. A range of experiments were performed to characterise the compound in greater detail, including dose-response, quantification of mRNA levels in cells treated with the compound, and synthesis of additional PROTACs based upon **E3-10**.

4.2.2.2.1 E3-10 dose-response

HEK293T cells transiently overexpressing HaloFlagKDM4A were treated with different concentrations of **E3-10** (0.125 µM, 1.25 µM, 6.25 µM and 12.5 µM) for 4 hours, 4 hours post-transfection (**Figure 4.13**). Across three biological replicates, reduction in HaloFlagKDM4A protein levels relative to DMSO control were observed with increased **E3-10** concentration. However, 12.5 µM **E3-10** was the only concentration where degradation of KDM4A protein was statistically significant. Degradation of KDM4A protein by 12.5 µM **E3-10** was at a similar level to HaloPROTAC-E, and across these experiments, degradation by **E3-10** was more consistent than degradation by HaloPROTAC-E which was not statistically significant. This demonstrates robust cellular KDM4A protein degradation by a **CP2**-based PROTAC, suggesting that a cell-active **CP2** -variant was generated from a parent peptide which had little cell activity.[40]

When using high concentrations of PROTACs the hook effect may be observed, where binary PROTAC-POI/PROTAC-E3 complexes are favoured over the ternary complex required for degradation.[104] Attempts were made to test **E3-10** at higher concentrations to explore whether the hook effect was seen with this compound; however, solubility issues were encountered when increasing the compound concentration beyond 12.5 µM, as described in Section 4.2.2.1; thus higher concentrations could not be tested in cells.

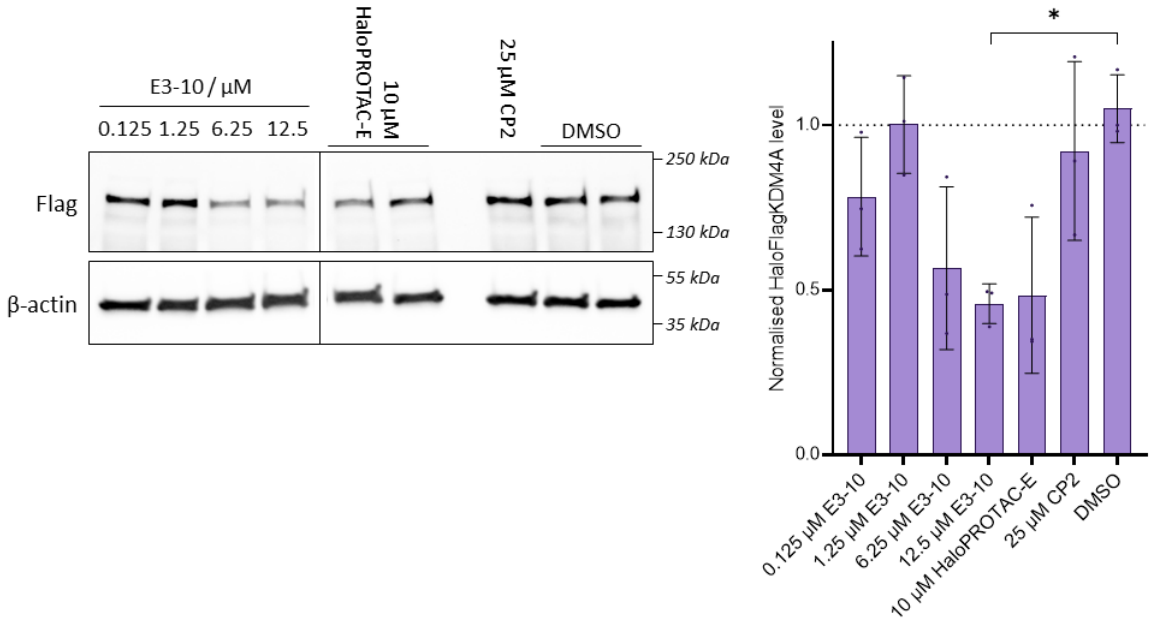


Figure 4.13 E3-10 dose-response experiment. Representative western blot analysis from three biological replicates is shown. Quantification of HaloFlagKDM4A protein level normalised to β-actin using densitometry shown in graph. A One-Way ANOVA test with multiple comparisons to DMSO control (GraphPad Prism 9.4.0) found a significant difference in KDM4A protein level compared to DMSO at only 12.5 µM **E3-10** (P* = 0.0167), no significant change was found for unlabelled bars. One DMSO/HaloPROTAC_E lane per experiment was chosen for this analysis (n=3, mean ± standard deviation). HaloFlagKDM4A = 157 kDa, β-actin = 42 kDa

4.2.2.2.2 Quantification of mRNA expression in cells treated with E3-10 and HaloPROTAC-E

Levels of mRNA for KDM4A *versus* β-actin in cells treated with **E3-10** and HaloPROTAC-E were quantified to establish whether reduction in KDM4A protein levels upon treatment with these compounds was due to reduced mRNA (and therefore translation) for KDM4A, or if the reduction was happening at another level in the cell. **CP2** and VH032 (**Figure 4.9**) were included as controls, as they are individual components of **E3-10**, which bind KDM4A and VHL respectively.

RNA extraction was performed on HEK293T cells treated for 4 hours, 4 hours post-transfection, with 10 µM HaloPROTAC-E, 12.5 µM **E3-10**/**CP2**/VH032, or 0.1% DMSO. A non-transfected control was also included, in which transfection reagent but no DNA were added to cells. Reverse transcription was used to convert mRNA into the complementary DNA, allowing a real time-quantitative polymerase chain reaction assay (RT-qPCR) to be used to quantify the amount of KDM4A/β-actin DNA contained within each sample. As DNA for KDM4A/β-actin was amplified, SYBR green fluorescence increased, which facilitated quantification of the amount of

KDM4A/β-actin DNA within the samples. Comparison of the level of KDM4A to β-actin as an internal standard allowed comparison of the relative levels of mRNA in the samples.

Across three biological replicates, there was no significant change in mRNA levels for cells treated with PROTACs or control compounds relative to the DMSO-treated cells (**Figure 4.14**). This suggests that the reduction in KDM4A protein levels observed in cells treated with these compounds occurs after translation, and therefore can likely be attributed to degradation *via* the UPS.[337] Furthermore, there was >100-fold less KDM4A mRNA for non-transfected cells compared to the DMSO control, demonstrating that the endogenous KDM4A protein levels in HEK293T cells are much lower.

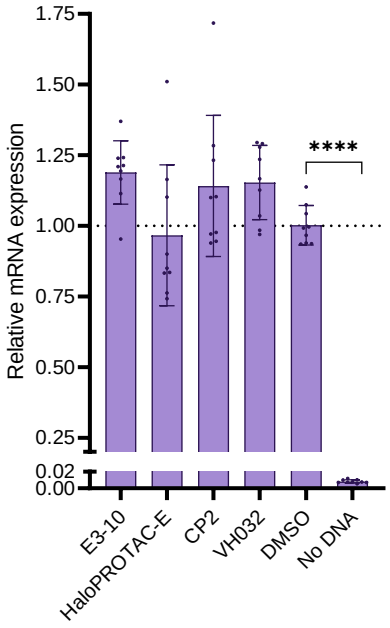


Figure 4.14 Quantification of KDM4A mRNA *versus* β-actin mRNA. Graph shows data from three biological replicates, with three technical replicates used in each RT-qPCR experiment. A Two-Way ANOVA test with multiple comparisons to DMSO control (GraphPad Prism 9.4.0) found a significant difference in KDM4A mRNA compared to DMSO for the no DNA control (P**** < 0.0001), and no significant change for all other conditions (unlabelled bars). (n = 3, mean ± standard deviation).

4.2.2.2.3 E3-11 – a CP2.3-based PROTAC?

A variant of **E3-10** was synthesised using **CP2.3** as the macrocyclic peptide (**E3-11**) to test if it would be more efficient at degrading KDM4A (**Figure 4.15**), since **CP2.3** is a more cell-active variant of **CP2**.[40] The incorporation of *N*-methyl groups in the backbone of **CP2.3** shields some hydrogen

bond interactions, increasing the likelihood that the peptide will be cell permeable.[225] Also, the addition of more non-proteinogenic amino acids should increase the proteolytic stability of the peptide, which may improve cellular activity.[338] **E3-11** was synthesised using the same method as other PROTACs within the **CP2**-PROTAC library, but the purity was only 76.8% (analytical HPLC, Appendix 1).

E3-11 was tested alongside **E3-10**, treating cells with compounds for 4 hours, 4 hours post-transfection (**Figure 4.15**). Surprisingly, **E3-11** was not able to degrade KDM4A protein under these conditions. One possible reason for this could be due to the lower affinity of **CP2.3** (IC_{50} = 110 nM) for KDM4A compared to **CP2** (IC_{50} = 42 nM),[40] which may reduce the stability of the ternary complex between the PROTAC, KDM4A, and VHL. This result highlights the importance of trying multiple PROTAC variations, as **E3-11** was predicted to degrade KDM4A protein at least as potently as **E3-10**. Interestingly, the ability of **E3-10** to degrade KDM4A protein was not statistically significant in this experiment, which is likely due to a lower maximum dose (10 µM instead of 12.5 µM in **Figure 4.13**). HaloPROTAC-E was found to significantly reduce the level of KDM4A protein in this experiment (**Figure 4.15**), where it had not in **Figure 4.13** – this shows variability in level of degradation across the experiments.

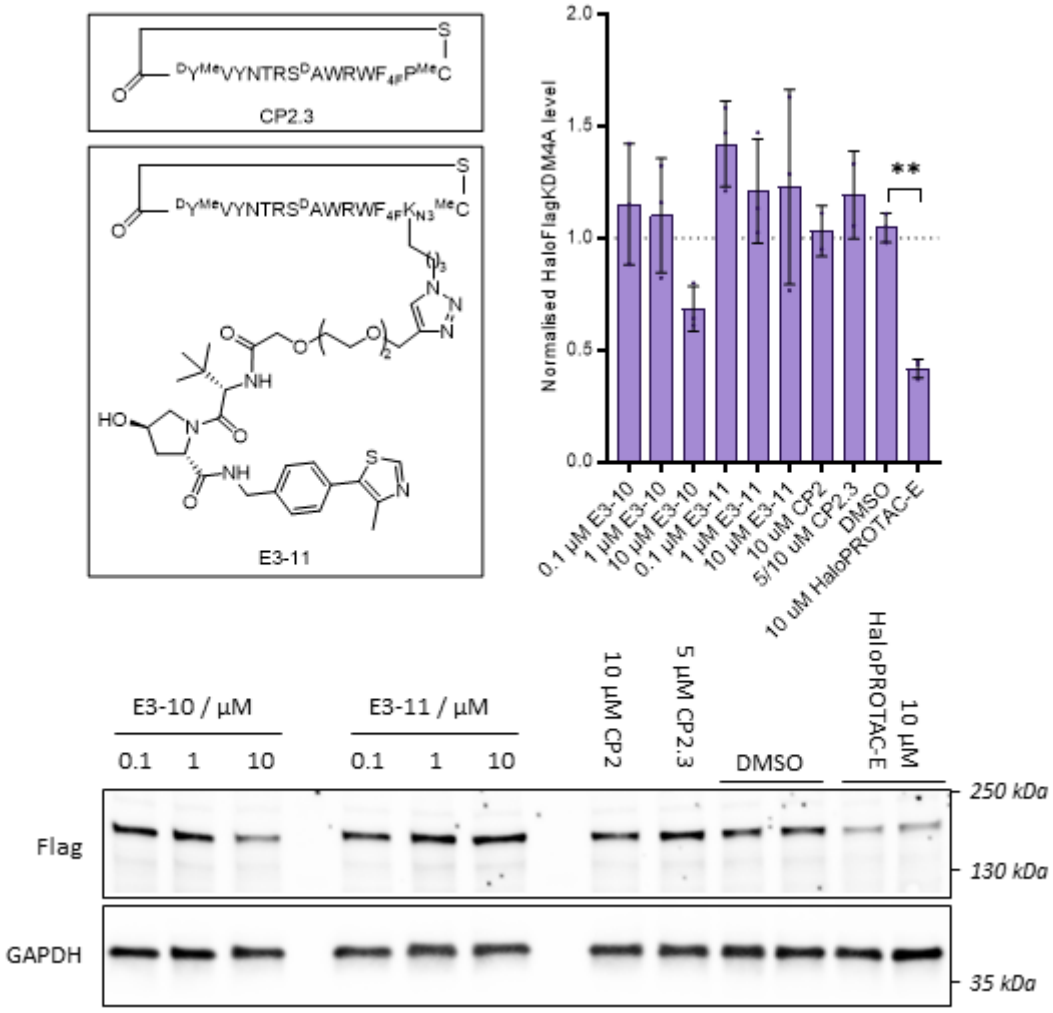


Figure 4.15 E3-11 did not degrade KDM4A protein when added to HEK293T cells overexpressing HaloFlagKDM4A. Structure of **E3-11** is shown. Cells were incubated with compounds for 4 hours, 4 hours post-transfection. Representative western blot analysis from three biological replicates shown. Quantification data of KDM4A protein level normalised to β-actin using densitometry across three biological replicates shown in graph. A One-Way ANOVA test with multiple comparisons to DMSO control (GraphPad Prism 9.4.0) found a significant difference in KDM4A protein level between only HaloPROTAC-E and DMSO (P** = 0.0021), no significant change was found for unlabelled bars. One DMSO/HaloPROTAC-E lane per experiment was chosen for this analysis. (n = 3, mean ± standard deviation). HaloFlagKDM4A = 157 kDa, GAPDH = 36 kDa

4.2.2.2.4 Protein degradation is not observed when E3-10 or E3-11 are added to non-transfected HEK293T cells

Having established degradation in cells overexpressing KDM4A, **E3-10** and **E3-11** were next tested for endogenous KDM4A degradation by treating non-transfected HEK293T cells (**Figure 4.16**). Both a range of concentrations and a range of time points were tested. **E3-11** was tested endogenously despite not degrading KDM4A in HEK293T cells overexpressing HaloFlagKDM4A as it had been expected to be more cell permeable/active than **E3-10**.

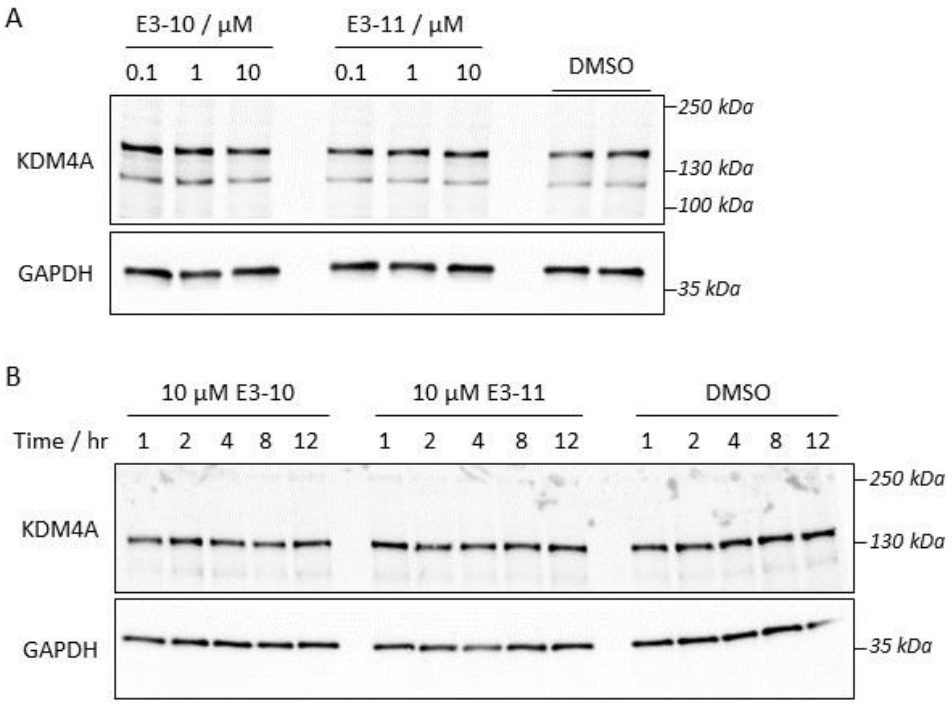


Figure 4.16 E3-10 and E3-11 were unable to degrade endogenous KDM4A in HEK293T cells. (A) Western blot analysis from 4-hour treatment with **E3-10** and **E3-11** at a range of concentrations, using similar conditions as overexpressing cells (Section 4.2.2.1). Representative data from two biological replicates shown. (B) Western blot analysis of cells treated with 10 µM **E3-10** and **E3-11** for 1–12 hours. KDM4A = 121 kDa, GAPDH = 36 kDa

Initially, a 4-hour incubation was tested, as this had been successful in cells transiently overexpressing HaloFlagKDM4A (**Figure 4.16**A). The levels of KDM4A protein was analysed using western blots probed with KDM4A antibody, but no changes in the levels of KDM4A protein was observed. Next, multiple time points from 1–12 hours were tested, using 10 µM **E3-10**/**E3-11**, but again no degradation was observed (**Figure 4.16**B). The reason behind **E3-10** not degrading endogenous KDM4A protein is not clear. It is possible that, in the overexpressed system, the newly synthesised KDM4A in the cytoplasm is predominantly being targeted for degradation by VHL. Thus, endogenous KDM4A, which will be mainly resident in the nucleus,[272] may be less prone to degradation. In addition, it is also possible that KDM4A is not expressed sufficiently in HEK293T cells to be detectable for degradation by western blot (KDM4A is expressed >100-fold in transfected cells relative to non-transfected cells, **Figure 4.14**). It would be informative to investigate other cell lines, with higher levels of KDM4A protein in the future, to investigate if endogenous levels of KDM4A can be degraded by **E3-10**. Longer treatment times should also be

tested. Higher doses of **E3-10** may degrade KDM4A protein, however, attempts to use >12.5 µM **E3-10** were unsuccessful due to solubility issues (Section 4.2.2.1). **E3-11** did not degrade endogenous KDM4A, in line with previous experiments in cells overexpressing HaloFlagKDM4A.

4.2.2.2.5 Work towards validation of the mode of action of E3-10

PROTACs are often validated through co-incubation with proteasome/neddylation inhibitors, or with E3 ligase binder competition. Many of these were tested with HaloPROTAC-E in Section 4.2.1.2.3, but most of the compounds tested altered KDM4A protein levels when treated on cells alone. Therefore, alternative approaches were tested towards validating that the mode of action of **E3-10** is *via* protein degradation using the UPS.

Several experiments within this chapter have included **CP2** (without an E3 ligase binder) to demonstrate that connection to VH032 is necessary for protein degradation (**Figure 4.12** & **Figure 4.14**). Whilst this has been informative, **CP2** is chemically very different to **E3-10** with greater than 600 $g \cdot mol^{-1}$ difference in molecular weight of the molecules. Substitution of arginine-6 to alanine in **CP2** results in a mutant (**CP2-R6A**) with greater than 50-fold lower inhibitory activity for KDM4A than **CP2** (**CP2** IC_{50}(4A)= 42 nM, **CP2-R6A** IC_{50}(4A) = 2700 nM).[40] Therefore, an equivalent **CP2-R6A**-based protein degrader, **E3-12**, was synthesised (> 95% purity (analytical HPLC, Appendix 1)).

E3-12 was assayed alongside **E3-10** at 6.25/12.5 µM, across three biological repeats (two biological repeats for 6.25 µM **E3-12**). Protein degradation was not observed when cells were treated with **E3-12** (**Figure 4.17**), which suggests that a potent KDM4A binder is essential for the activity of **E3-10**. Although, there is also potential that **E3-12** has reduced cell permeability compared to **E3-10**, despite the R6A mutation being quite a small change. Furthermore, a statistically significant difference was not obtained for 12.5 µM **E3-10** (or 10 µM HaloPROTAC-E) compared to DMSO in this experiment, therefore further repeats are required to confirm this result.

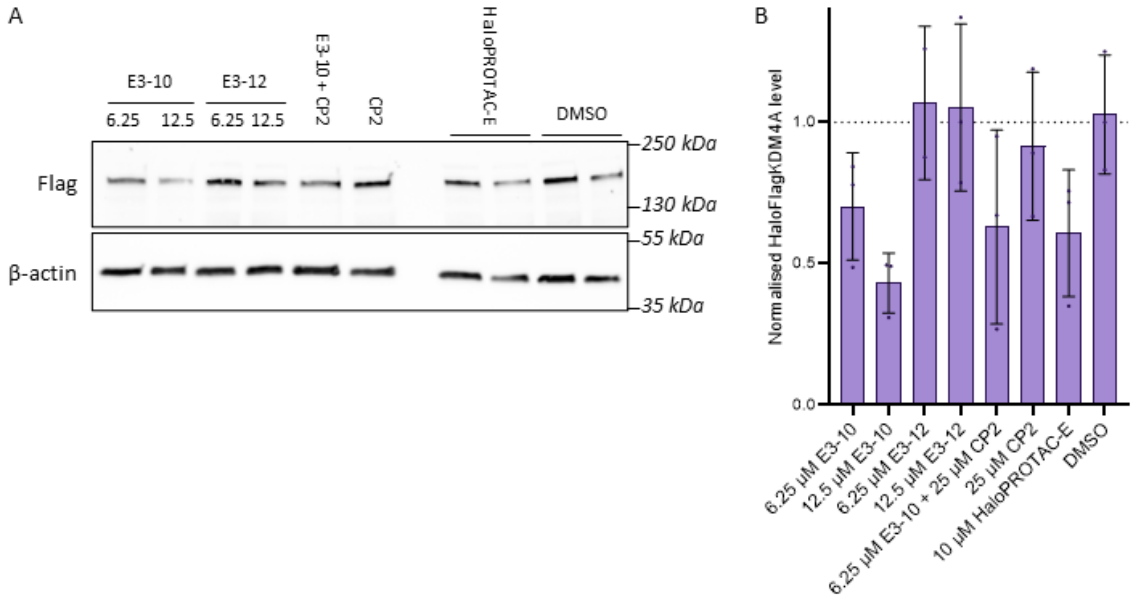


Figure 4.17 Experiments towards validation of the mode of action of E3-10 in HEK293T cells overexpressing HaloFlagKDM4A. (A) Western blot analysis for cells treated with 12.5 µM **E3-12**, compared to **E3-10** and **CP2** competition. Compounds were treated on cells for 4 hours, 4 hours post-transfection. Representative data from three biological repeats shown (two biological repeats for 6.25 µM **E3-12**). (B) Quantification of KDM4A protein level normalised to β-actin using densitometry shows variability across the replicates. A One-Way ANOVA test with multiple comparisons to DMSO control (GraphPad Prism 9.4.0) found no significant change was found for any condition tested (n = 3, mean ± standard deviation). HaloFlagKDM4A = 157 kDa, β-actin = 42 kDa

The competition experiment between **E3-10** and **CP2** did not produce conclusive results (**Figure 4.17**); demonstrated by overlapping standard deviation error bars for 6.25 µM **E3-10**, 6.25 µM **E3-10** + 25 µM **CP2**, and 25 µM **CP2**, and no statistically significant difference was found between these conditions. Differences between these samples are hard to measure, and quantification of KDM4A protein levels for 6.25 µM **E3-10** compared to DMSO were also not statistically significant in previous experiments (**Figure 4.13**), despite consistently observing a reduction in KDM4A protein level by western blot. More conclusive results may be obtained using higher concentrations of **E3-10**, but a lower dose was chosen to allow a greater degree of competition, since **CP2** has limited solubility. It is also possible that the concentration of **CP2** inside the cells was significantly lower than the concentration the cells were treated with, due to cell permeability issues with peptides.[147] In this experiment all compounds were added at the same time, thus it will be worth repeating the experiment adding competitors 30 minutes prior to adding PROTACs, as in Section 4.2.1.2.3. Future work should test the most successful strategies for validating

HaloPROTAC-E, these are: VH298 competition to test whether recruitment of VHL is necessary for PROTAC activity, and co-incubation with MLN4924 to test if degradation is occurring *via* the UPS.

4.3 Discussion

In this chapter, **CP2**-variants functionalised with E3 ligase binders were tested, yielding cell-active macrocyclic peptide PROTACs. Click chemistry allowed a small library of potential protein degraders to be generated quickly. Click chemistry has been demonstrated to aid the development of PROTACs,[175] since optimisation is generally required to find the best linker and/or E3 ligase for each system.[95] Appropriate conditions to test the small library of cyclic peptide PROTACs were identified by using HaloPROTAC-E, which demonstrated KDM4A protein degradation in cells transiently overexpressing HaloFlagKDM4A. The best VHL targeting **CP2**-PROTAC, **E3-10**, was further characterised, and two cyclic peptide variants were synthesised. KDM4A mRNA levels were unchanged when cells were treated with **E3-10** (**Figure 4.14**), indicating that the reduction in KDM4A levels is at a protein level; however, further work is needed to fully validate this PROTAC. In addition to experiments described in this chapter, **E3-10** was also co-incubated with a protein translation inhibitor (cycloheximide) across a range of time points. This experiment aimed to test if **E3-10** alters the translation of KDM4A, which could be an alternative explanation for the reduction in KDM4A protein levels upon PROTAC exposure. However, the concentration of cycloheximide used was not able to reduce KDM4A translation, so future work will involve repeating these experiments using higher cycloheximide concentrations.

Using HaloPROTACs, it was found that HaloFlagKDM4A protein degradation could be observed when compounds were added 4 hours after transfection, but not when compounds were added 48 hours after transfection. Transient transfection was favoured over generation of stably expressing cell lines because several weeks are needed to establish each stably expressing cell line, and success is not guaranteed. However, there will be a greater difference in KDM4A protein levels in separate experiments using transient transfection, due to both slightly variable transfection efficiency in each experiment, and due to different cells expressing the POI at a

different level (**Figure 4.6**). Transfection efficiency differs across experiments because it is affected by several factors, including cell confluency and passage number,[322] which vary slightly between experiments.

It was somewhat surprising to see protein degradation activity for **CP2**-based PROTACs, since **CP2** has little cell activity.[40] The catalytic nature of protein degradation may allow a greater biological effect at lower compound concentrations than the parent inhibitor. It would be interesting to study if this strategy would be more broadly applicable to other peptides, as it could present an attractive route for facile development of biologically active peptide conjugates. It is also possible that modification of **CP2** with an E3 ligase binder and linker improved cell permeability; an altered cellular uptake mechanism may be used by **CP2**-PROTACs compared to **CP2**, perhaps endocytosis is used for **CP2**-PROTACs where passive diffusion may be used for **CP2**.[339] Investigating cell permeability of the panel of PROTACs compared to **CP2** would show whether differences in activity observed between the different compounds is due to the amount that enters the cell, or differing ability to modulate KDM4A protein between the compounds. This could be investigated using either a cell target engagement assay (such as a BRET-based assay), cell permeability assay (such as a parallel artificial membrane permeability assay or a Caco-2 cell permeability assay),[340] or by using a MS-based assay to quantitate intracellular PROTACs.

While the majority of the results presented in this chapter use western blots to analyse protein degradation, it would be informative to try alternative degradation assays such as BRET or fluorescence based cell imaging. Future experiments could include imaging live or fixed cells which have been stained with a fluorescent HaloTag ligand to probe for HaloFlagKDM4A protein levels. Also, there are a number of different NanoBRET protein degradation assays available (*e.g.* from Promega) to investigate degradation of the POI, formation of a ternary complex, and target ubiquitination.[341] For these assays a new KDM4A construct would be needed, with a NanoLuc instead of a HaloTag. Many of the commercial protein degradation assays are also designed for

stably expressing cell lines, or cells which have had their genome edited to express NanoLuc-POI fusion, so several months would be needed to set up these assays for each cell line investigated.

Investigation of the effect of cyclic peptide PROTACs on global histone methylation status would be informative, to investigate if KDM4A protein degradation has a knock-on effect. The levels can also be compared to cells overexpressing catalytically inactivate KDM4A (HaloFlagKDM4Amut, Section 4.2.1.1). Attempts were made to quantify histone H3K9Me_3 levels from cell lysates obtained in these experiments using western blots (data not shown); however, variability in the histone H3 loading control was observed between samples. Variability observed is likely due to limited histone solubility in the lysis buffer used, it does not contain SDS which was found to greatly improve histone solubility when different lysis buffers were tested (Section 3.2.1.1.4). Global histone methylation status can be investigated using immunofluorescence, as in Kawamura *et al.* (2017),[40] although for these experiments compounds had to be incubated with the cells for 72 hours to observe changes.

Chapter 5 Discussion and future work

The utility of multifunctional probes in chemical biology has been widely demonstrated, particularly for preclinical target validation.[69,70,83] Among other multifunctional probe applications, PROTAC probes have chemically mimicked genetic knockdown/knockout studies, and improved the selectivity of parent inhibitors;[312] tagged-multifunctional probes enabled affinity purification of POIs from biological settings using both small molecule affinity probes, and antibodies;[82,132] and multifunctional labelling has produced fluorescent and radioisotope imaging agents.[113,116] Within epigenetics, there is an unmet need for selective probes for many epigenetic regulators,[13] which multifunctional probe development may aid. Many enzymes are involved in regulating covalent epigenetic modifications to chromatin, which alter gene expression, and dysregulation of epigenetic proteins is associated with several diseases, including cancer.[9,13,14,342] Many small molecule inhibitors of epigenetic enzymes suffer low selectivity for their targets, partly due to high structural similarity within epigenetic protein subfamilies (including the Jumonji-C KDMs, which has over 20 members).[23,343] Peptides offer greater probe-target binding surface area, and a subsequent increase in selectivity, and potency.[159] Peptide probes selective for KDM4A-C (**CP2** (ref. [40])), and TET1 (**TiP1** (ref. [55])), were used in this thesis to explore how additional functionality could be harnessed to enrich the targets from cell lysates, as well as for targeted protein degradation inside cells.

Multifunctional **CP2**- and **TiP1**-based cyclic peptide variants were synthesised in Chapter 2. Azide click chemistry handles were installed in the peptides, to allow both biotin, and E3 ligase binders to be conjugated to the peptides. Other functional groups conjugated to an alkyne can be installed in future work, provided they are stable under the click conditions used. Functional groups could include another protein binder, for the development of a dimerization probe; or fluorescent groups, for imaging applications. Other click reactions could also be utilised, in particular SPAAC, which can be performed in biological systems.[174] It would be interesting to test in-cell assembly of

the peptide conjugates profiled in Chapters 3 and 4, and to see if this impacts probe activity and cell permeability. Photoreactive cyclic peptides were also synthesised, installing both benzophenone and diazirine groups, to work towards covalent target engagement by the cyclic peptides. Other photoreactive groups could be installed, including aryl azides,[203] although with this photoreactive group the azide click handle would need to be switched to an alkyne handle.

Chapter 3 focused on affinity cyclic peptide probes, for both KDM4A, and TET1. Here reversibly binding peptide-biotin conjugates were found to successfully engage their targets in cell lysates containing overexpressed Flag-tagged protein. Additionally, a biotinylated **CP2**-variant (**B-1**) could engage endogenous KDM4A in cell lysates; this was confirmed using both western blots probed with KDM4A antibody, and preliminary MS-based proteomics experiments. Interestingly, **B-1** KDM4A engagement in cell lysates required a non-denaturing lysis buffer. This suggests that folded KDM4A is required for **B-1** target engagement, and could be harnessed to produce a selective activity-based cyclic peptide probe. Testing the cell permeability of these affinity probes would be useful, as selective cell permeable affinity agents for these proteins would offer an improvement on immunoprecipitation techniques used to study these targets.[51,331] Significant target protein remained in the unbound fractions in pulldown studies using reversibly binding peptide-biotin conjugates, and therefore irreversible cyclic peptide-target binding was explored using photocrosslinking. Under the conditions tested neither **CP2**-based or **TiP1**-based photoreactive probes could specifically crosslink to the parent peptide binding site, however, more comprehensive follow-up work is required to fully characterise the photoreactive cyclic peptides synthesised. Overall, more detailed biological characterisation of these targets can be performed using optimised reversible or irreversible binding cyclic peptides, expanding information on their protein interactions, and the genomic DNA interacting with these targets. Future work should apply these techniques to peptides which bind little studied proteins. Allosteric site, non-inhibitory binders would be particularly useful for epigenetic proteins, to capture active site epigenetic interactions. Although, expansion to cell membrane or cytoplasmic

targets, instead of nuclear targets, may aid issues with cell permeability associated with peptides.[147] Further into the drug discovery process, tumour homing peptides or antibodies could be conjugated to probes for delivery of probes to desired tissues within the body.[212]

KDM4A protein degradation was tested in Chapter 4. Initially, HaloTag targeting PROTACs (HaloPROTACs)[193,194] were employed to establish conditions under which HaloTagged and Flag-tagged KDM4A (HaloFlagKDM4A) was degradable in HEK293T cells transiently overexpressing HaloFlagKDM4A. Two HaloPROTACs were tested, and HaloPROTAC-E was able to degrade HaloFlagKDM4A protein when treated on cells for 4 hours, 4 hours post-transfection. Although it was unable to detectably degrade HaloFlagKDM4A protein when treated on cells 48 hours post-transfection. Following this work, a panel of ten **CP2**-PROTACs were tested using optimal degradation conditions for HaloPROTAC-E. Three **CP2**-based HaloFlagKDM4A protein degraders were identified at 10 μM in cells transiently overexpressing HaloFlagKDM4A, which included both VHL and CRBN recruiting PROTACs. Two different length all carbon linkers were tolerated for CRBN recruiting degraders, **E3-8** and **E3-9**. However, a PEG-linker was needed for VHL recruiting **E3-10**, and a related all carbon linked PROTAC with only 1 atom difference in linker length was not able to degrade HaloFlagKDM4A protein, **E3-3**.

More detailed analysis of VHL-recruiting **E3-10** was performed, which demonstrated concentration-dependent degradation of HaloFlagKDM4A protein from 0.125–12.5 μM **E3-10**. KDM4A mRNA levels were quantified from cells treated with **E3-10** or HaloPROTAC-E, and no change compared to the DMSO vehicle control was observed. This suggested that the change in protein level seen is due to degradation rather than decreased protein translation, however, further studies to elucidate the mode of action were performed. **E3-12** was synthesised, which is a **CP2-R6A** analogue of **E3-10**, likely to have lower KDM4A binding affinity as **CP2-R6A** has greater than 50-fold lower KDM4A inhibitory activity compared to **CP2**.[40] **E3-12** was unable to degrade HaloFlagKDM4A protein, which may suggest that arginine-6 is required for the degradation activity

of **E3-10**, however altered cell permeability or stability could not be ruled out. Therefore, further studies to validate the mode of action of **E3-10** were initiated, co-incubating **E3-10** with protease inhibitors or VHL ligands. It would be interesting to investigate the binding affinity for recombinant KDM4A and VHL, of the **CP2**-PROTAC series, and to establish if there is a link between the enzymatic affinity and protein degradation observed in cells. For this work bio-layer interferometry or surface plasmon resonance could be used, which have low sample requirements.[344,345] Furthermore, *in vitro* assays to assess the stability of PROTAC ternary complexes are also available,[334] and could give valuable insight into why **E-8**, **E3-9**, and **E3-10** were the best degraders.

E3-10 was unable to degrade endogenous KDM4A in HEK293T cells, when treated for 1–12 hours at 10 μM. However, longer treatment times should be tested, as well as alternative compound concentrations. The KDM4s have been implicated in many cancers,[33] and testing endogenous KDM4A protein degradation activity of **E3-10** in cancer cell lines would be interesting. THP-1 cells, from a monocytic leukaemia patient, have the highest KDM4A expression level of any of the cell lines in the Human Protein Atlas,[264,265] so would be an interesting cell line to begin these studies with. If KDM4A degradation can be observed in cancer cell lines, cell proliferation can also be tested, in work towards target validation of KDM4A.

CP2-PROTACs demonstrated cell activity, where the parent peptide had little cell activity,[40] despite increased molecular weight taking the products further beyond Lipinski's rule-of-five.[145] It would be interesting to test more peptide series, to see if multifunctionalisation through protein degradation, or other means, can improve bioactivity for other peptides. The modifications to peptides used are applicable both to novel peptides discovered using screening, and natural product derivatives. An alanine scan was sufficient to identify appropriate sites for the conjugation of reporter groups for **TiP1**, demonstrating that detailed structural information (such as a peptide-target co-crystal structure) was not needed for peptide functionalisation. The parent peptides

used in this project were discovered using mRNA display technology, and the peptide conjugates were formed from a position close to the mRNA ligation point in the selection. It is possible to include click handles in the peptide selection for some genetically encoded selection technologies,[169] which may produce a multifunctionalisation-ready peptide. However, this would not guarantee the click handle is orientated to allow the conjugation of reporter groups without significantly impacting the binding affinity of the parent peptide. Therefore, the strategy of growing peptides conjugates close to the genetic information attachment point may offer a more general functionalisation route than including a click handle in the selection.

Overall, multifunctional cyclic peptide probes were demonstrated to engage their targets when used both as affinity probes, and as protein degraders. CuAAC click chemistry facilitated efficient synthesis of peptide conjugates, which were screened in cells and cell lysates. Peptide-biotin conjugates using two different parent peptides were able to enrich their targets from cell lysates. Also, cell activity was observed in protein degradation work, despite low cell activity of the parent peptide, **CP2**.[40] Future work will involve further experiments to confirm the mechanism of action is protein degradation, as well as to investigate binding affinity of **E3-10** for KDM4A and VHL. It would be interesting to investigate whether multifunctionalisation of other peptides, particularly by adding protein degradation functionality, may also improve cell activity of these parent peptides as it could offer an attractive route to increased cell activity. Cyclic peptides are an ideal choice for multifunctionalisation, as they often have improved probe properties compared to linear peptide derivatives.[147] Additionally, the modularity of peptide synthesis offers facile routes to peptide diversification.[168]

Epigenetic proteins are excellent targets for multifunctional probes, to elucidate their roles within multiprotein complexes, in addition to modulating target protein activity.[76] Targeted protein degradation has been suggested as a promising therapeutic strategy for epigenetic targets.[73] Degradation may enable protein binding partners to be degraded alongside the target, which may

improve efficacy.[192,315] DUBTACs are likely to complement this approach, as they can be used to stabilise a particular target.[208] Neither DUBTACs or PROTACs require active site binding, and allosteric binding may offer more routes to selectivity for epigenetic enzymes where individual families often have high sequence homology within the active site.[13] Furthermore, the conjugation of another protein binder offers additional routes for selectivity.[207]

Chapter 6 Materials and methods

6.1 General information

Purified water with conductivity of greater than 18.2 MΩ·cm^{-1} was used for all experiments. Water was purified using a Millipore Elix® 10 system and passed through a 0.22 µm filter using a Millipore Milli-Q system; or water was purified using an Avidity Ultrapure Water system prior to passing through a 0.22 µm filter.

The pH of buffers was adjusted using a pH meter (Jenway™ or Mettler Toledo™ FiveEasy™), calibrated with fresh standards at pH 4.01, pH 7.0, and pH 10.01. Either hydrochloric acid or sodium chloride were used to adjust the pH.

Common reagents and organic solvents were used as received from Merck, Fisher Scientific, Alfa Aesar, Cambridge Reagents, or Fluorochem. Details are provided for reagents where the provenance/batch of reagent may impact the experimental outcome.

6.2 Peptide Synthesis

Peptide synthesis grade reagents were used for all peptide synthesis, including *N,N*-dimethylformamide (DMF), 20% (v/v) piperidine in DMF, coupling reagents, Fmoc-protected amino acids, and TFA.

6.2.1 9-Fluorenylmethoxycarbonyl (Fmoc) solid-phase peptide synthesis

Linear precursors to cyclic peptides were synthesised by Fmoc/*t*Bu SPPS on a 0.05 or 0.1 mmol scale using a Liberty Blue (CEM Corporation) or PurePep Chorus (Gyros Protein Technologies) automated peptide synthesiser. Histone peptides were synthesised analogously, on a 0.1 mmol scale using a Liberty Blue (CEM Corporation) peptide synthesiser. Resin was swelled for at least 20 min in DMF prior to starting peptide synthesis, and the peptide *N*-terminus was deprotected at the end of synthesis unless otherwise indicated. Fmoc SPPS conditions for the two automated synthesisers used are detailed in **Table 6.1**. Although, photo-lysine/photo-leucine (Section 2.2.4, **Table 2.5**) were coupled using 1.5 eq. amino acid; subsequently capped with 10% (v/v) acetic

anhydride, and 5% (v/v) pyridine in 10 mL DMF (15 min at 60 °C, shaking); prior to Fmoc deprotection.

Rink amide 4-methylbenzhydrylamine 100–200 mesh resin was used, typically with a loading of 0.52 mmol·g^{-1}, for 0.1 mmol scale synthesis 192 mg resin was used, and 96 mg resin was used for 0.05 scale. For 0.1 mmol scale peptide synthesis using the Liberty Blue, 5 eq. of amino acid, 10 eq. Oxyma Pure, 1 eq. *N,N*-diisopropylethylamine, and 2.5 eq. *N,N'*-diisopropylcarbodiimide were used. For 0.05 mmol scale peptide synthesis using the Liberty Blue 5 eq. of amino acid, 20 eq. Oxyma Pure, 2 eq. *N,N*-diisopropylethylamine, and 5 eq. *N,N'*-diisopropylcarbodiimide were used. For both 0.05 and 0.1 mmol scale peptide synthesis using the PurePep Chorus 5 eq. each of amino acid, Oxyma Pure, and *N,N'*-diisopropylcarbodiimide were used.

Condition	**Liberty Blue (CEM corporation)**	**PurePep Chorus (Gyros Protein Technologies)**
Resin	Rink amide 4-methylbenzhydrylamine 100–200 mesh	Rink amide 4-methylbenzhydrylamine 100–200 mesh
Activator (in DMF)	1 M Oxyma Pure & 100 mM *N,N*-Diisopropylethylamine	250 mM Oxyma Pure
Activator base (in DMF)	500 mM *N,N'*-diisopropylcarbodiimide	250 mM *N,N'*-diisopropylcarbodiimide
Amino acids (in DMF)	200 mM	250 mM
Activator, activator base, and amino acid ratio in coupling	0.05 mmol – 1:0.5:1.25 0.1 mmol – 1:0.5:2.5	1:1:1
Standard coupling	2 min, 90 °C	2.75 min, 90 °C
Arginine coupling	2 × (2 min, 75 °C)	2 × (6 min, 75 °C)
Histidine/Cysteine coupling	10 min, 50 °C	10 min, 50 °C
Deprotection	20% (v/v) piperidine in DMF 50 s, 90 °C, × 2	20% (v/v) piperidine in DMF 2.75 min, 50 °C, × 2

Table 6.1 Liberty Blue (CEM corporation) and PurePep Chorus (Gyros Protein Technologies) automated peptide synthesiser conditions. Details of L-amino acids used in SPPS are detailed in Appendix 7, and details of non-proteinogenic amino acids detailed in Section 6.2.1.1.

Next, linear precursors to cyclic peptides were capped with a chloroacetyl group, incubating the resin with 200–250 mmol chloroacetic anhydride in 5 mL DMF, or 1:1 DMF:dichloromethane. Chloroacetylation reactions were either incubated for 3 hr at room temperature, intermittently agitating the sample; or incubated for 30 min at 60 °C, agitating constantly.

For all peptides, cleavage from the resin and deprotection was performed in 5 mL of deprotection cocktail (92.5% (v/v) TFA, 2.5% (v/v) 1,3-dimethoxybenzene, 2.5% (v/v) triisopropylsilane, and 2.5% (v/v) H_2O), for 3 hr at room temperature. Crude peptides were precipitated from cold diethyl ether, and pelleted by centrifugation at 3000 × *g*, for 10 min at 4 °C. Ether solution was decanted off, and peptides were dissolved in water and acetonitrile (MeCN) prior to freeze drying, resulting in peptides as white/yellow solids. The yield of crude peptides is reported in **Table A1.3**.

6.2.1.1 Fmoc-protected non-proteinogenic amino acids used

Structures of non-proteinogenic Fmoc-protected amino acids used are shown in **Figure 6.1**, and their suppliers are shown in **Table 6.2**.

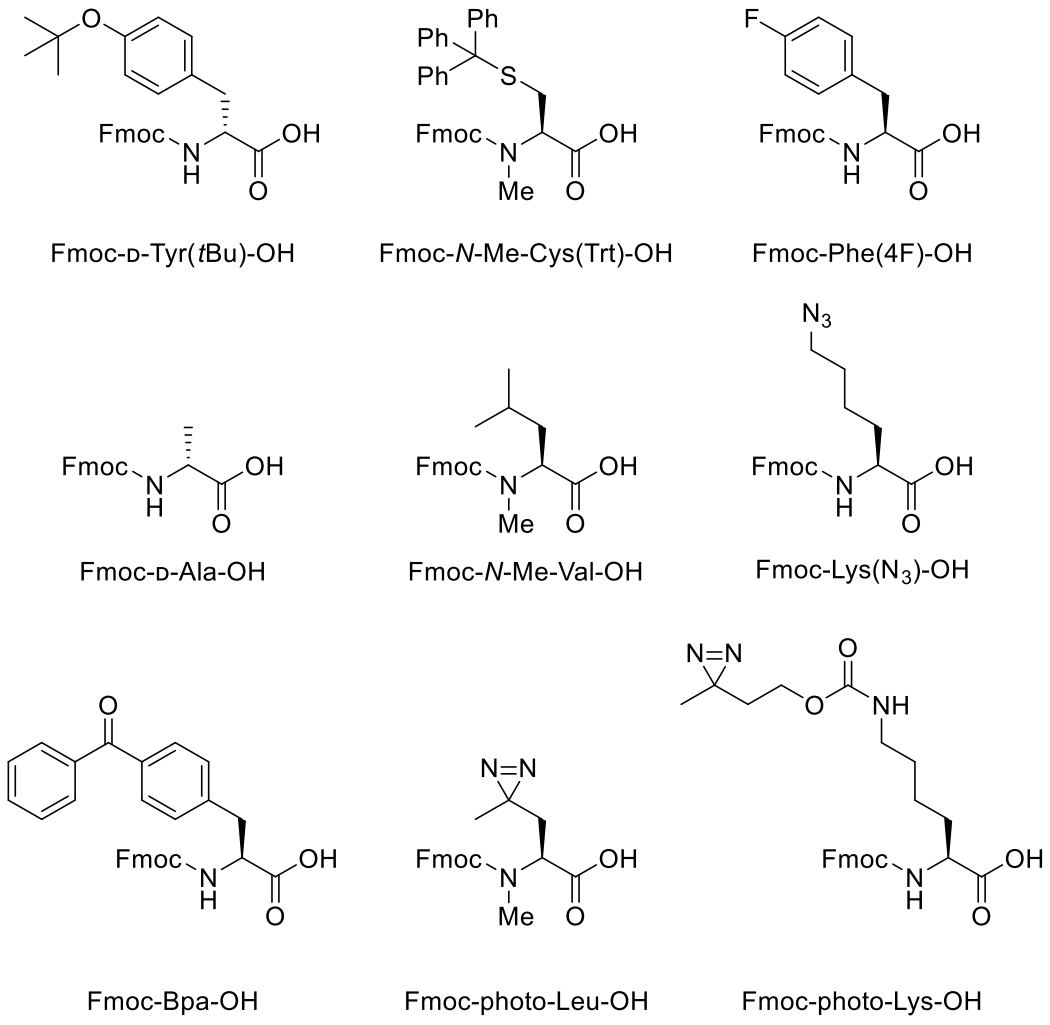


Figure 6.1 Structures of Fmoc-protected non-proteinogenic amino acids.

Amino acid	Source	Catalogue number	Cas number
Fmoc-D-Tyr(*t*Bu)-OH	Iris Biotech	FAA1341	118488-18-9
Fmoc-*N*-Me-Cys(Trt)-OH	Amatek	NA-1058	944797-51-7
Fmoc-Phe(4F)-OH	Merck (Novabiochem®)	8521950005	169243-86-1
Fmoc-D-Ala-OH	Merck (Novabiochem®)	8521420005	79990-15-1
Fmoc-*N*-Me-Val-OH	Fluorochem	M03379-5G	84000-11-3
Fmoc-Lys(N_3)-OH	Merck (Novabiochem®) Iris Biotech	8522140005 FAA1793	159610-89-6
Fmoc-Bpa-OH	Merck	09774	117666-96-3
Fmoc-photo-Leu-OH	Iris Biotech	FAA4590	1360651-24-6
Fmoc-photo-Lys-OH	Iris Biotech	FAA4600	2250437-42-2

Table 6.2 Non-proteinogenic Fmoc-protected amino acid procurement details.

6.2.2 CP2.3-P13K$_{N3}$ peptide synthesis

Extensive optimisation was required to synthesise **CP2.3-P13K$_{N3}$** at sufficient purity for further reactions (Section 2.2.2.1.1), details of the methods used are described here.

6.2.2.1 Manual amino acid coupling

Manual amino acid coupling was performed on a 0.1 mmol scale, in a 15 mL ISOLUTE® fritted reservoir (Biotage). All steps were performed at room temperature with end-over-end rotation. The resin was swelled with DMF prior to coupling where necessary, and the coupling cycle began with Fmoc removal. To remove Fmoc, the resin was incubated with 8 mL of 20% (v/v) piperidine in DMF for 1 hr. Meanwhile, the amino acid was activated, by incubating 5 eq. Fmoc-protected amino acid, 4.5 eq. HATU, and 4.5 eq. Oxyma Pure in 20% (v/v) *N*-methylmorphiline in DMF for 15 min. The resin was washed three times in 8 mL DMF for 5 min. Activated amino acid was incubated with the resin for 1.5–12 hr, until a TNBS test (Section 6.2.2.2) indicated absence of primary amine and/or a *p*-Chloranil test indicated presence of a secondary amine (Section 6.2.2.3). The resin was subsequently washed and filtrated three times in 8 mL DMF for 5 min. The peptide was capped in 5–8 mL of 3:2 acetic anhydride:pyridine for 30 min. Finally, the resin was washed four times in 8 mL DMF for 5 min.

6.2.2.2 TNBS test for primary amines

TNBS test protocol was based on Hancock and Battersby (1976).[346] A few resin beads were removed from the peptide synthesis, and placed in an LC-MS vial. An equal volume of 10% (v/v)

N,N-diisopropylethylamine in DMF, and 1 M aqueous TNBS solution (also called picrylsulfonic acid solution) were mixed, and 100 µL added to the resin. Beads were incubated for 10 min at room temperature, prior to analysis under a microscope. A positive test for primary amines results in orange/red beads, and a negative test results in colourless beads.

6.2.2.3 p-Chloranil test for secondary amines

p-Chloranil test protocol was based on Chan and White (1999).[347] A few resin beads were removed from the peptide synthesis, and placed in an LC-MS vial. An equal volume of 2% (v/v) acetaldehyde in DMF, and 2% (w/v) *p*-chloranil in DMF were mixed, and 100 µL added to the resin. Beads were incubated for 5 min at room temperature, prior to analysis under a microscope. A positive test for secondary amines results in blueish beads, and a negative test results in yellowish beads.

6.2.2.4 Synthesis of CP2.3-P13K$_{N3}$ involving manual N-methylation

CP2.3 was synthesised on a 0.05 mmol scale, *N*-methyl amino acids were formed by mono-methylation of the peptide backbone after coupling unmodified cysteine/valine. SPPS was performed as described in Section 6.2.1 up to and including the amino acid to be *N*-methylated, and Fmoc was removed from the *N*-terminus prior to *N*-methylation. Manual *N*-methylation was performed in a 15 mL ISOLUTE® fritted reservoir (Biotage), and all steps were performed at room temperature with end-over-end rotation.

First, nosyl protection of the free amine was performed by incubating the resin with 2-nitrobenzenesulfonyl chloride (44 mg, 0.2 mmol) and 2,4,6-collidine (66 µL, 0.5 mmol) in 1 mL 1-methyl-2-pyrrolidinone (NMP) for 15 min. The resin was subsequently washed five times in 3–5 mL NMP for 5 min. Subsequently, *N*-methylation was performed. The initial, unsuccessful *N*-methylation attempt was performed by incubating the resin with DBU (45 µL, 0.3 mmol) and dimethylsulfate (95 µL, 1 mmol) in 1 mL NMP for 5 min, and little *N*-methylation was detected. The successful *N*-methylation procedure was performed by incubating the resin with DBU (45 µL, 0.3 mmol) in 1 mL NMP for 3 min; subsequently dimethylsulfate (95 µL, 1 mmol) was added to the

DBU solution for 4 min (Section 2.2.2.1.1). For both *N*-methylation procedures, the mixture was drained, and the *N*-methylation step was repeated. Next, the resin was washed five times in 3–5 mL NMP for 5 min. Finally, nosyl deprotection was performed by incubating the resin with DBU (37.5 µL, 0.25 mmol) and 2-mercaptoethanol (35 µL, 0.5 mmol) in 1 mL NMP for 5 min. This mixture was drained, and the nosyl deprotection step was repeated, prior to five washes in 3–5 mL NMP for 5 min.

Resin was subsequently transferred to the peptide synthesiser reaction vessel. The subsequent amino acid was coupled with a double/triple coupling, and the Fmoc deprotection stage was removed from the cycle. After the final amino acid coupling, Fmoc was left on the *N*-terminus and the peptide was capped using either: (i) 5 mL 3:2 acetic anhydride:pyridine in DMF, in a 15 mL ISOLUTE® fritted reservoir (Biotage) (30 min at room temperature, with end-over-end rotation); or (ii) 10% (v/v) acetic anhydride, 5% (v/v) pyridine in 10 mL DMF, on the PurePep Chorus (Gyros) peptide synthesiser (30 min at 60 °C, with shaking). The resin was washed 4–5 times with DMF prior to Fmoc deprotection as usual (**Table 6.1**). **CP2.3-P13K$_{N3}$** was treated in the same way as other peptides when chloroacetylated, cleaved from the resin / deprotected (Section 6.2.1), and cyclised (Section 6.2.3).

6.2.3 Cyclisation

Chloroacetylated crude linear peptides were dissolved in either an MeCN/H_2O mix, or a DMSO/MeCN/H_2O mix. The ratio of solvents was tailored to each peptide, for maximum solubility. Subsequently, peptides were cyclised at pH ≈ 9, by addition of triethylamine. Cyclisation reactions were left for 1 hr at 37 °C, and MALDI-MS was used to confirm full cyclisation prior to re-acidification to pH ≈ 6 by addition of TFA.

6.2.4 Peptide purification

Cyclised peptides were purified by reverse-phase HPLC on JASCO, Shimadzu, or Agilent systems, with solvent A (degassed H_2O supplemented with 0.1% (v/v) TFA) and solvent B (MeCN

supplemented with 0.1% (v/v) TFA) as the solvents. A Gemini® 5 µm NX-C18 110 Å, AXIA™ packed column at either 250 × 30 mm or 250 × 21.2 mm was used, with flow rate 18–20 mL·min^{-1}. A customised gradient was used, from 2–98% solvent B (histone peptides) or 15–98% solvent B (cyclic peptides / peptide conjugates), over approximately 30 min for each peptide. Fractions containing pure peptide were identified by MALDI-MS (Section 6.3), combined, and freeze dried. The yield of cyclised and purified peptides is reported in **Table A1.3**.

6.2.5 Purified peptide characterisation

The purity of peptides and peptide conjugates was measured using analytical HPLC. An Agilent 1220 Infinity II analytical RP-HPLC system was used, with solvent A (degassed H_2O supplemented with 0.1% (v/v) TFA) and solvent B (MeCN supplemented with 0.08% (v/v) TFA) as the solvents, 99+% TFA for HPLC was used (Fisher Scientific). A BioZen XB-C18 (2.6 µm, 150 × 4.6 mm) column was used, with a flow rate of 1 mL·min^{-1}. Samples were analysed over a linear gradient of 2-98% solvent B over 40 min. The area of the observed peaks compared to a blank was used to determine the percentage purity of the samples.

The identity of the peptides was confirmed using high-resolution MS (HRMS), on a Thermo Exactive High-Resolution Orbitrap FTMS[k] or a Thermo RSLC[l]. HRMS using the Thermo RSLC was performed analogously to intact protein LC-MS (Section 6.10.2).

The concentration of most peptide or peptide conjugate DMSO stocks was determined using a NanoDrop One Spectrophotometer, for which the peptide extinction coefficients were calculated using ExPASy ProtParam.[348] Histone peptide stock concentration was determined using NMR; peptide stocks in H_2O (5 µL) and 3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt (TSP) in D_2O (5 µL, 630 µM stock) were mixed with 450 µL of D_2O, and analysed on a Bruker AVII 500 NMR. The NMR spectrum analysed using Mnova (Mestrelab V.14) software. Comparison of the area per H

[k] HRMS from Thermo Exactive High-Resolution Orbitrap FTMS was performed by submission to University of Oxford, Department of Chemistry MS service.
[l] MS performed by myself, supported by NUPPA.

atom under assigned amino acid side chain peaks was compared to the TSP proton peak, in order to calculate the concentration of peptide in the original stock.

6.3 MALDI-MS

MALDI-MS was used to assess whether desired peptides were synthesised, to identify pure HPLC fractions, and to monitor peptide conjugate click reactions. HPLC fractions were analysed without dilution, other samples were usually diluted 1:100 in H_2O prior to MALDI-MS analysis.

α-Cyano-4-hydroxycinnamic acid (**CHCA) MALDI matrix:** matrix was prepared in a black 1.5 mL tube, by suspending 10 $mg \cdot mL^{-1}$ CHCA in 50% (v/v) MeCN supplemented with 0.01% (v/v) TFA. Matrix was sonicated for 10 min, and briefly spun down in a microcentrifuge. Matrix was stored at room temperature, and used for up to 1 month.

MALDI spots were made on 96-target or 384-target Bruker ground steel MALDI plates; plating MS sample (0.8 µL), followed by CHCA MALDI matrix (0.8 µL). Spots were left to dry prior to analysis. MS spectra were obtained on either a Bruker Microflex® LRF or Bruker Autoflex® maX, using the positive ion reflectron mode. Data analysis was performed using FlexAnalysis v3.4 (Bruker), and spectra were smoothed, baseline subtracted, and the peaks labelled prior to exporting spectra.

6.4 Click reactions

6.4.1 Synthesis of peptide conjugates

Click reactions were used to form peptide conjugates from azide-peptides and alkyne-linked reporter groups in Section 2.2.3. Reactions were performed at room temperature, using $Cu(II)SO_4 \cdot 5H_2O$ as the copper source, ascorbate to reduce Cu(II) to Cu(I), and THPTA as a Cu(I) binding ligand. The optimised reaction is described below, during optimisation the peptide solvent varied as shown in **Table 2.3**, Section 2.2.3.

Azide-peptides (1.9 mM) and biotin-$(PEG)_4$-alkyne / alkyne-linked E3 ligase binders (all 100 mM) were dissolved in DMSO. $Cu(II)SO_4 \cdot 5H_2O$ (20 mM) and sodium ascorbate (100 mM) were dissolved in H_2O. THPTA (60 mM) was dissolved in PBS buffer. Following compound dilutions, azide-peptides

(490 µL) were mixed with alkynes (10 µL). Separately, Cu(II), sodium ascorbate, and THPTA were mixed 1:1:1. The reaction was initiated by mixing 500 µL Cu mix with the azide-peptide/alkyne mixture (500 µL), the percentage conversion to peptide conjugates was assessed using MALDI-MS (diluting reactions 1:100 prior to spotting). Reactions were quenched upon full conversion, or once they had been left for 2 hr, by addition of MeCN.

Reaction conditions: 0.95 mM **CP2-T13KN$_3$**, 1 mM alkyne, 2 mM Cu(II)$SO_4 \cdot 5H_2O$, 6 mM THPTA, and 10 mM ascorbate in a variety of solvents (Section 2.2.3, **Table 2.3**).

The peptide conjugates were characterised as described in Section 6.2.5.

6.4.2 Conjugation of Cy5 dye to peptides after photocrosslinking

Click chemistry was used to connect Cy5 dye with an azide/alkyne tag, to an alkyne/azide handle in photoreactive cyclic peptides respectively, so that it could be fluorescently imaged in SDS-PAGE protein gels. Reagents and methodology used were adapted from Speers *et al.* (2003),[349] using the same reagents but mixing the click reagents prior to addition to the photocrosslinking samples. Reactions were performed at room temperature, using Cu(II)$SO_4 \cdot 5H_2O$ as the copper source, tris(2-carboxyethyl)phosphine hydrochloride (TCEP) to reduce Cu(II) to Cu(I), and either tris[(-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA) or THPTA as the Cu(I) binding ligand.

Cy5-azide/alkyne was dissolved in DMSO (1.25 mM). Cu(II)$SO_4 \cdot 5H_2O$ (50 mM), TCEP (50 mM), and THPTA (1.7 mM, where used) were dissolved in H_2O. TBTA (1.7 mM, where used) was dissolved in 1:4 DMSO:*t*BuOH. Cy5 dye, Cu(II), TCEP, and TBTA/THPTA were mixed in a 3:1:1:1 ratio respectively. After photocrosslinking, peptides were diluted 2-fold in H_2O, and 50 µL placed in a tube for the click reaction. An appropriate volume of click reagent mixture was added to peptides, such that the click reaction was performed with 4 × Cy5 dye concentration to the azide/alkyne peptide. The reaction was left for 1 hr at room temperature, prior to SDS-PAGE protein gel analysis.

Click reagent mixture: 210 µM Cy5-azide/alkyne, 8.33 mM $Cu(II)SO_4 \cdot 5H_2O$, 8.33 mM TCEP, 850 µM TBTA/THPTA.

6.5 General chemical biology methods

6.5.1 Protein SDS-PAGE gels

Production of gel samples: Gel samples were mixed with 4 × NuPAGE™ lithiumdodecyl sulphate (LDS) sample buffer supplemented with 1,4-dithiothreitol (final concentration 50 mM) or 10 × NuPAGE™ sample reducing agent (final concentration 1 ×). Samples were incubated for 5–10 min at 95 °C. Photocrosslinking samples were not heat treated prior to gel loading, as some dyes degrade when heated. All gel samples were run alongside PageRuler™ (Plus) pre-stained protein ladder (Thermo Scientific™) or Seeblue™ plus2 pre-stained protein ladder (Invitrogen™).

6.5.1.1 Hand-cast SDS-PAGE protein gels

Production of hand-cast gels: SDS-PAGE protein gels were cast using Mini-PROTEAN® Tetra Handcast Systems (BIO-RAD) apparatus, using 1 mm spacer plates. The resolving gel (**Table 6.3**) was poured to 1.5 cm from the top of the glass, ethanol was poured on top to level the gel, and the gel was left to set. Ethanol was poured off prior to addition of the stacking gel (**Table 6.3**). A 15-lane ladder was inserted, and the gel was left to set for ≥30 min. Resolving and stacking gels were poured immediately after addition of ammonium persulfate.

Per gel	Resolving gel (7.5 mL)	Stacking gel (3 mL)
Tris-HCl 1.5 M, pH 8.6	1.75 mL	-
Tris-HCl 2 M, pH 6.8	-	186 µL
H_2O	3.05 mL	2.48 mL
30% (w/v) acrylamide	2.8 mL	378 µL
20% (w/v) SDS	35 µL	3 µL
10% (w/v) APS	37.5 µL	25.5 µL
TEMED	7.5 µL	3 µL

Table 6.3 Components of 12% acrylamide resolving and stacking gels. SDS = sodium dodecyl sulfate APS = ammonium persulfate, TEMED = N,N,N′,N′-tetramethylethylenediamine

Electrophoresis running buffer: 25 mM Tris, 192 mM Glycine, 0.1% (w/v) SDS, pH 8.8.

Running and staining hand-cast gels: Gels were run at 200 V in electrophoresis running buffer (see above), using a Mini-PROTEAN® Tetra Vertical Electrophoresis Cell (BIO-RAD). Electrophoresis was stopped once the dye front reached the bottom of the gel (≈30 min). Hand-cast gels were stained with InstantBlue™ (15–60 min, rocking), and destained with H_2O (2 × 5–10 min, rocking).

6.5.1.2 Running pre-cast SDS-PAGE protein gels

Running and staining pre-cast gels: Gels were run according to **Table 6.4**, and subsequently transferred to western blotting membrane (Section 6.5.2) or stained. Most Coomassie stained gels were stained with InstantBlue™ (15–60 min, rocking), and destained with H_2O (2 × 5–10 min, rocking). KDM4A photocrosslinking gels were stained with Imperial™ Protein Stain (Thermo Scientifc™), and destained in H_2O according to the manufacturer's instructions. SilverQuest™ (Invitrogen™) silver stain was performed according to manufacturer's instructions.

Gel	Gel running system	Running buffer	Voltage
4–15% Mini-PROTEAN® TGX™ Precast Protein Gels (BIO-RAD)	Mini-PROTEAN® Tetra Vertical Electrophoresis Cell (BIO-RAD)	25 mM Tris, 192 mM Glycine, 0.1% (w/v) SDS, pH 8.3 buffer	120 V
4–15% Criterion™ TGX™ Precast Midi Protein Gel (BIO-RAD)	Criterion™ Vertical Electrophoresis Cell (BIO-RAD)	25 mM Tris, 192 mM Glycine, 0.1% (w/v) SDS, pH 8.3 buffer	140 V
NuPAGE™ 4 to 12%, Bis-Tris, 1.0 mm, Mini Protein Gels (Invitrogen™)	XCell SureLock™ Mini-cell electrophoresis system (Invitrogen™)	1 × MOPS-SDS running buffer	150 V

Table 6.4 Pre-cast gel running conditions.

6.5.2 Western blotting

Transfer buffer: 25 mM Tris, 192 mM glycine, 20% (v/v) methanol.

TBST: 20 mM Tris, 150 mM NaCl, pH 7.4, 0.1% (v/v) Tween20.

Western blots were loaded with cell lysates, of which the total protein content was measured using a BCA/Bradford assay (Section 6.8.2.1). Gel samples were prepared to contain the same concentration of protein (usually 10 µg protein was loaded per lane). Lysates were mixed with water and SDS-PAGE loading buffer, as described in Section 6.5.1, such that each sample contained the same concentration of protein, and so each gel lane was loaded with the same

volume. Samples were run on pre-cast gradient gels as described in Section 6.5.1.2. Gels were trimmed to remove the comb at the top, and blue dye band at the bottom of the gel, prior to transfer to a nitrocellulose membrane (0.45 µm pore size, Cytiva) submerged in transfer buffer (see above). Transfers were performed in a Mini Trans-Blot® Cell or a Trans-Blot® Cell (both BIO-RAD) at 80 V for 80 min. Upon transfer, the tank was cooled with ice, or the transfer was performed in a cold room (4–6 °C). Ponceau S dye was used to check successful transfer, and the blot was washed twice with H_2O, prior to blocking.

For all the following steps, solutions were added such that the full blot was just submerged, and the blots were rocked. Blots were blocked in 5% (w/v) skimmed milk powder in TBST (45–90 min at room temperature), and washed three times in TBST (5 min at room temperature). Blots were incubated with primary antibodies (overnight at 4 °C), made up as indicated in **Table 6.5**, and washed three times in TBST (10 min at room temperature). Primary antibody solutions were stored and re-used up to 6 times, antibodies that were made up in 5% (w/v) milk were stored at -20 °C, and those made up in 5% (w/v) bovine serum albumin (BSA) were stored at 4 °C. Horse radish peroxidase (HRP)-conjugated secondary antibodies were made up as described in **Table 6.5** and incubated with blots (1 hr at room temperature), prior to washing three times (10 min at room temperature).

Blots were visualised using either Amersham ECL detection reagent (Cytiva) or Clarity ECL western blotting substrate (BIO-RAD), according to the manufacturer's instructions. Blots were imaged using the chemiluminescence and colorimetric channels of a ChemiDoc MP Imaging System (BIO-RAD).

6.5.2.1 Antibodies used to probe western blots

Details of antibodies used to probe western blots are detailed in **Table 6.5**, peroxidase antibodies were used as secondary antibodies to allow the blots to be imaged using chemiluminescence.

Antibody	Details	Dilution	Dilution buffer
Flag	Monoclonal Anti-FLAG® M2 antibody produced in mouse (Merck, F3165)	1:1000	5% (w/v) skimmed milk powder in TBST + 0.02% (w/v) NaN_3
KDM4A	Polyclonal Anti-JMJD2A (KDM4A) antibody produced in rabbit (Bethyl laboratories, A300-861A)	1:500	5% (w/v) BSA in TBST + 0.02% (w/v) NaN_3
GAPDH	Monoclonal Anti-GAPDH (14C10) antibody produced in rabbit (Cell Signalling Technology, 2118)	1:3000	5% (w/v) BSA in TBST + 0.02% (w/v) NaN_3
B-actin	Monoclonal Anti-β-actin antibody produced in mouse (Merck, A1978)	1:2000	5% (w/v) BSA in TBST + 0.02% (w/v) NaN_3
H3	Monoclonal Anti-Histone H3 (1B1B2) antibody (ChIP grade) produced in mouse (abcam, ab195277)	1:1000	5% (w/v) BSA in TBST + 0.02% (w/v) NaN_3
$H3K9Me_3$	Polyclonal Anti-Histone H3 (tri methyl K9) antibody (ChIP grade) produced in rabbit (abcam, ab8898)	1:1000	5% (w/v) BSA in TBST + 0.02% (w/v) NaN_3
Mouse, peroxidase	Peroxidase conjugated Anti-Mouse IgG Antibody (H+L) produced in horse (Vector laboratories, PI-2000-1)	1:2222	5% (w/v) skimmed milk powder in TBST + 0.02% (w/v) NaN_3
Rabbit, peroxidase	Peroxidase conjugated Anti-Rabbit IgG Antibody (H+L) produced in goat (Vector laboratories, PI-1000-1)	1:2222	5% (w/v) skimmed milk powder in TBST + 0.02% (w/v) NaN_3

Table 6.5 Details of antibodies used for probing western blots.

6.5.3 DNA agarose gel electrophoresis

TAE buffer: 40 mM Tris, 1 mM Ethylenediaminetetraacetic acid (EDTA), 20 mM glacial acetic acid.

TBE buffer: 90 mM Tris-borate, 2 mM EDTA.

Agarose gels were used to visualise DNA. Agarose (1% (w/v)) was dissolved in TAE/TBE buffer by heating in a microwave. This was allowed to cool a little before addition of 1 in 10000 (v/v) SYBR™ Safe DNA gel stain (Invitrogen™). Gels were cast in an H2-SET gel tray to a depth of 0.7 cm with a comb inserted at one end of the gel, and allowed to cool at room temperature for the gel to set. Samples were prepared by mixing with 6 × DNA Gel Loading Dye (final concentration 1 ×), and run at 150 V alongside a DNA ladder. Gels were imaged using the SYBR™ safe channel of a ChemiDoc MP Imaging System (BIO-RAD).

6.5.4 Transformation of *E. coli*

Luria-Bertani (LB) media (1 L): 10 g NaCl, 10 g tryptone, and 5 g yeast extract were mixed with 1 L water. Media was either autoclaved directly, or used for agar preparation (see below).

Agar plate preparation: Agar was prepared by addition of 2% (w/v) agar to LB media, and autoclaved. Agar was left to cool to approximately 60 °C before being supplemented with antibiotics, plates were poured in a biological safety cabinet. Liquid agar was added to 10 cm Petri dishes until the base was covered with agar, and left to set at room temperature.

Antibiotic preparation: A final concentration of 100 $\mu g \cdot mL^{-1}$ ampicillin sodium salt, or 30–50 $\mu g \cdot mL^{-1}$ kanamycin sulfate was used. Stocks at 1000-fold final concentration were dissolved in water, and sterile filtered (0.2 µm) prior to being aliquoted, and stored at -20 °C for up to 3 months. Antibiotic aliquots were discarded after two freeze thaw cycles.

Transformation: Bacteria was handled in a biological safety cabinet using sterile pipette tips. The plasmid (≈50 ng) / cloning product, and bacteria (20–50 µL) were incubated on ice for 20 min. Heat shock was performed according to manufacturer's instructions. For XL10-Gold (Agilent) or NEB® 5-alpha *E. coli*, the tube was incubated for 30 s in a 42 °C water bath, followed by 2 min on ice. Super optimal broth with catabolite (S.O.C.) media (Invitrogen™, 500 µL) was added to each tube, and they were incubated for 45–60 min at 37 °C, shaking at 225 RPM. Bacteria was spread on agar plates with antibiotic selection pressure, and incubated overnight at 37 °C. After transformation, agar plates were stored at 4 °C, agar side up, and wrapped with parafilm, to avoid condensation. Plates were not left more than 48 hr before picking colonies.

6.5.5 DNA plasmid preparation

Mini preps were performed using a Nucleospin Plasmid (Macherey-Nagel™) kit, or a GeneJET Plasmid MiniPrep Kit (Thermo Scientific™), according to manufacturer's instructions.

Midi preps were performed to yield sufficient plasmid for mammalian cell transfection. A NucleoBond Xtra Midi (Macherey-Nagel™) kit, or PureLink™ HiPure Plasmid Filter Maxiprep Kit (Invitrogen™) was used, according to manufacturer's instructions.

6.6 Production of recombinant KDM4A

Biotin-KDM4A$_{1\text{-}359}$ was recombinantly produced and purified during my masters project.[321] The protein was expressed using a pNIC-Bio3 plasmid encoding for *N*-terminally His$_6$-tagged and *C*-terminally Avi-tagged *KDM4A$_{1\text{-}359}$* expressed with biotin supplement, as described.[40]

6.6.1 Expression of *N*-His$_6$-KDM4A$_{1\text{-}359}$ in *E. Coli*

N-terminally His$_6$-tagged KDM4A$_{1\text{-}359}$ was expressed from a pNIC28-Bsa4 plasmid, the plasmid was a gift from Nicola Burgess-Brown (Addgene plasmid number 38846). Bacteria was handled in a biological safety cabinet, using autoclaved pipette tips and sterile consumables, until the final pellet was prepared for centrifugation on the benchtop.

2 × TY media: 9.6 g tryptone, 6 g yeast extract, 3 g NaCl, 600 mL water, per flask. Media was autoclaved in 2 L culture flasks.

BL21 DE3 cells were transformed with a pNIC28-Bsa4 plasmid encoding for *N*-His$_6$-KDM4A$_{1\text{-}359}$ by heat shock (Section 6.5.4). A single colony was picked and used to inoculate an overnight starter culture (100 mL 2 × TY, supplemented with 30 μg·mL^{-1} kanamycin), which was incubated for 16 hr at 37 °C, shaking at 180 RPM. Starter culture was divided across twelve flasks, each containing 600 mL 2 × TY media supplemented with 30 μg·mL^{-1} kanamycin, the cells were incubated at 37 °C, shaking at 180 RPM, until an OD_{600} of 0.75 was reached (≈2.5 hr). At this point expression of KDM4A was induced by the addition of 300 μL of 500 mM isopropyl β-D-thiogalactoside solution per flask. Flasks were incubated for 18 hr at 18 °C, 180 RPM. Cells were harvested by centrifugation at 9,650 × *g* for 10 min at 4 ˚C, discarding the supernatant. A total of 22.4 g cell pellet was recovered and stored at -80 °C.

6.6.2 *N*-His_6 $KDM4A_{1-359}$ purification

Cell pellet was suspended in 110 mL binding buffer (buffer compositions are detailed in **Table 6.6**) supplemented with 2–5 mg each of DNase, phenylmethylsulfonyl fluoride, and lysozyme. Cell solution was lysed by sonication (Sonics vibracell, 60% amplitude; 3 s on, 3 s off for a total sonication time of 9 min). Lysed cells were pelleted, and supernatant was filtered using a 0.45 µm syringe filter, before loading onto a HisTrap HP, 5 mL column (GE Healthcare LifeSciences). After the protein was loaded, the column was washed with 5 column volumes (CV) binding buffer, followed by 10 CV wash buffer. Next, the protein was eluted across a linear gradient of 0–100% elution buffer over 5 CV, followed by 2 CV of 100% elution buffer. Fractions containing protein (assessed by SDS-PAGE) were pooled and concentrated to 1 mL using a 30 kDa Amicon MWCO filter (Merck). The concentrated protein was further purified by size exclusion chromatography using an ÄKTA FPLC system with a Superdex S200 300 mL column (GE Healthcare). Purified protein fractions (assessed by hand-cast SDS-PAGE protein gels, Section 6.5.1.1) were pooled and concentrated, yielding 2.5 mL of 228 µM protein. Protein was aliquoted before being snap frozen in liquid nitrogen; RapidFire MS was used to confirm the mass of the protein,[m] and an FDH enzyme coupled fluorescence assay was used to confirm that it was active (Section 2.2.5.2).

Recombinant protein purification buffers		**Component**			
		HEPES, pH 7.5	NaCl	Glycerol	Imidazole
HisTrap	Binding buffer	50 mM	500 mM	5% (v/v)	5 mM
	Wash buffer	50 mM	500 mM	5% (v/v)	30 mM
	Elution buffer	50 mM	500 mM	5% (v/v)	250 mM
Size exclusion	Gel filtration buffer	10 mM	300 mM	5% (v/v)	-

Table 6.6 Components of buffers used for KDM4A protein purification. HEPES = 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

6.7 Formaldehyde dehydrogenase coupled His_6-$KDM4A_{1-359}$ assays

FDH buffer: 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, pH 7.5), 0.01% (v/v) Tween®20.

[m] RapidFire MS performed by Dr Leah Taylor-Kearney.

Activity assay conditions: 200 µM 2OG, 100 µM ascorbate, 10 µM Fe(II), 500 µM NAD^+, 15 µM $H3K9Me_3$(1–15) or 30 µM $H3K36Me_3$(28–42), 0.05 U FDH and 1 µM KDM4A, 0% or 1% (v/v) DMSO (50 µL reaction).

Inhibition assays conditions: 200 µM 2OG, 100 µM ascorbate, 10 µM Fe(II), 500 µM NAD^+, 15 µM $H3K9Me_3$(1–15) or 30 µM $H3K36Me_3$(28–42), 0.05 U FDH, 200 nM KDM4A, 1% (v/v) DMSO (50 µL reaction). Enzymes were pre-incubated with inhibitor for 15 min before the assay was initiated.

2OG/NAD^+ were dissolved in FDH buffer, as well as subsequent dilutions of protein and histone peptide. An Fe(II) stock was made by dissolving iron(II) ammonium sulfate hexahydrate to 100 mM in 20 mM HCl, followed by dilution to 1 mM in H_2O, and subsequent dilutions in FDH assay buffer. An ascorbate stock was made by dissolving sodium ascorbate to 100 mM in H_2O, followed by subsequent dilutions in FDH assay buffer. Experiments were performed in a 384-well, black, µCLEAR®, non-binding plate (Greiner Bio-One). Fluorescence was measured every 30 s with 20 flashes per well for 30–60 min on a BMG PHERAstar FS or FSX plate reader (FI 350 460). Enzymes (and inhibitors where used) were plated out (25 µL), and histone peptide/co-factors were added (25 µL) to initiate the assay; the plate was put in the PHERAstar, and fluorescence measurements commenced immediately.

Data processing to give IC_{50} values: Percentage inhibition was calculated, comparing the normalised slope of NADH fluorescence over time, in the linear range of KDM4A activity, to normalised DMSO vehicle control. Percentage inhibition was plotted *versus* log(inhibitor concentration) in GraphPad Prism 5.0/9.4.0, and dose-response curves were fitted using nonlinear regression to give IC_{50} values.

6.8 Mammalian cell culture

All mammalian cell culture was performed using sterile technique, in a biological safety cabinet. Sterile, culture grade plasticware was procured, or plasticware was autoclaved prior to use. Equipment/consumables were sprayed with 70% (v/v) ethanol prior to being introduced to the

biological safety cabinet. All cell culture media, supplements, PBS, and trypsin were from Gibco™ (Fisher Scientific), unless otherwise stated. Cell lines were maintained in sterile Cellstar® tissue culture coated flasks (Greiner Bio-One).

6.8.1 Cell line maintenance and seeding

Cells were incubated at 37 °C, 5% CO_2, with high humidity, and not grown above 90% confluency. PBS (pH 7.4) was used in all wash steps in cell culture experiments.

HEK293T and U2OS cells were obtained from ATCC, and maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% (v/v) foetal bovine serum (FBS), and 1% (v/v) L-glutamine. Doxycycline-inducible FlagTET1 U2OS cells were generated in Prof Peter Radcliffe's laboratory, and maintained in DMEM supplemented with 10% (v/v) TET system approved FBS, and 1% (v/v) L-glutamine.[n]

Passage of cells and seeding cell culture plates/dishes: Culture medium was removed from cells, and cells were washed, prior to addition of trypsin-EDTA (0.05% (v/v)). Trypsin-EDTA was incubated with cells until they detached from the flask. U2OS cells were placed in a cell culture incubator for ≈2 min until cells detached, and HEK293T cells were kept at room temperature for <1 min until cells detached. Cells were gently suspended in culture media, and pelleted by centrifugation at 200 × *g* for 5 min. Cells were resuspended in culture medium, and 10 µL removed for cell counting. Cells (10 µL) were mixed with trypan blue (10 µL), loaded on to a dual chamber counting slide, and counted using a TC20™ Automated Cell Counter (BIO-RAD). An appropriate number of cells were seeded in flasks for cell line maintenance, or dishes/plates for experiments.

Transient transfection: *N*-terminally triple Flag-tagged full length KDM4A wild type (wt) or H188A catalytically inactive mutant (mut) were expressed using pcDNA-flag-hKDM4A-polyA and pcDNA-flag-hKDM4A(H188A)-polyA plasmids respectively. These plasmids were a gift from Yi Zhang

[n] U2OS cells stably overexpressing doxycycline-inducible TET1 were maintained with Dr Hilal Sarac.

(Addgene plasmid numbers 101051 and 101052 for wt/mut respectively).[350] *N*-terminally HaloTag and triple Flag-tagged full length KDM4A wt or H188A mut were expressed using a pFN21A HaloTag® CMV Flexi® Vector, cloning is described in Section 6.11. Cells were transfected using 1:3 DNA:FuGENE HD transfection reagent (Promega). DNA, transfection reagent, and OptiMEM media were incubated for 10 min, prior to dropwise addition across the well/dish of cells. For a 6-well plate: 2 µg DNA, 6 µL FuGENE, and 170 µL OptiMEM media were used. When transfecting cells in other sizes of dish/plate, the amount of reagents added were scaled with the change in culture surface area. For transfection optimisation experiments, DNA concentration is described in the main text (Section 3.2.1.1.1).

Doxycycline treatments: Doxycycline-inducible Flag-TET1 U2OS cells were seeded (2×10^6 cells in 10 cm dishes) 24 hr prior to doxycycline treatment. Cells were induced with 1 $\mu g \cdot mL^{-1}$ doxycycline (final concentration), for which 4 $\mu g \cdot mL^{-1}$ doxycycline in 5 mL medium was added to cells (15 mL media in dish). Cells not treated with doxycycline had 5 mL medium added to cells, without doxycycline.

6.8.2 Cell lysis

6.8.2.1 Whole cell lysate production

IGEPAL cell lysis buffer: 10 mM Tris (pH 8.0), 150 mM NaCl, 0.5% (v/v) IGEPAL-CA-630, one cOmplete™ mini EDTA-free protease inhibitor tablet (Roche) per 10 mL buffer.

For all cells, cell lysis was performed on ice, on a lab bench top. Cells were washed three times with ice cold PBS, scraping the cells with a cell scraper on the final wash, and collecting into cold microcentrifuge tubes. Cells were pelleted at 400 × *g* for 5 min at 4 °C. PBS was aspirated, and cells were suspended in ice cold lysis buffer. Cells were incubated for 20 min on ice, prior to clearing the lysate using centrifugation at 20,000 × *g* for 20 min at 4 °C. Without disturbing the pellet, lysates were aliquoted into cold tubes, including a 5 µL aliquot for protein quantification

(Section 6.8.2.3). Unless otherwise stated in the main text, IGEPAL lysis buffer was used. Wash and lysis buffer volumes used are displayed in **Table 6.7**.

Culture plate/dish	Culture area	PBS wash volume	Lysis buffer volume
6-well plate/35 mm dish	9.6 cm^2/8.8 cm^2	1 mL per well/dish	30 µL per well/dish
10 cm dish	56.7 cm^2	5 mL for wash 1 and 2, 2.9 mL for final wash	180 µL per dish
15 cm dish	145 cm^2	10 mL for wash 1 and 2, 5 mL for final wash	600 µL per dish

Table 6.7 Cell lysis wash and lysis buffer volumes used.

6.8.2.2 Nuclear cell lysate production

HEK293T cells for nuclear cell lysate production were grown in 10 cm dishes, and nuclear cell lysates were produced using a Nuclear Extract Kit (Active Motif, 40010) according to the manufacturer's instructions.

6.8.2.3 Cell lysate protein quantification

Total protein content of cell lysates was analysed using a bicinchoninic acid (BCA) assay, or a Bradford assay. For all protein degradation lysates, BCA assays were used.

For both protein quantification assays, 5 µL cell lysate was diluted 5-fold or 10-fold in H_2O. A BSA standard curve was produced in 20%/10% (v/v) cell lysis buffer in H_2O, supplemented with 0.02% (w/v) NaN_3, such that the standard curve was in the same buffer as the cell lysate used in the assay. Pierce™ BSA standard ampules 2 mg·mL^{-1} (Fisher Scientific, 23209) were used, and the top BSA concentration was undiluted, serial dilutions were performed to give a standard curve of: 2000, 1500, 1000, 750, 500, 250, 125, and 0 µg·mL^{-1} BSA. BSA standard curves were stored at 4 °C, and used for up to 1 month.

All protein quantification measurements were performed in duplicates, in clear, flat bottom UV-Star® 96-well plates (Greiner Bio-One). Standard curve or protein samples (10 µL) were added to wells, followed by protein detection solution (200 µL, BCA/Bradford assay protein detection

solution details below). Plates were sealed, shaken for 30 s using a vortex, briefly spun down, and incubated as detailed below.

BCA assay: A Pierce™ BCA protein assay kit (Fisher Scientific, 23227) was used, the detection agent was produced by mixing 50 parts solution A with 1 part solution B. Prior to imaging, the assay plate was incubated for 30 min at 37 °C, and let cool to room temperature. The plate was spun down briefly, and the plate seal was removed prior to imaging. A BMG PheraStar FSX was used to image the assay, using the pre-programmed BCA absorbance assay.

Bradford assay: Pierce™ Coomassie plus (Bradford) assay reagent (Fisher Scientific, 23238) was added directly to the assay plate. Prior to imaging, the assay plate was incubated for 10 min at room temperature. Plate was spun down briefly, and the plate seal was removed prior to imaging. A BMG PHERAStar FS or FSX was used to image the assay, using the pre-programmed Bradford absorbance assay.

Data analysis: For both protein quantification assays, the standard curve was calculated using BMG MARS data analysis software, fitting the curve to a linear, or four-parameter, fit.

6.9 Pulldown assays and MS-based proteomics experiments

6.9.1 Affinity pulldown using peptide-biotin conjugates

All procedures were performed in Protein LoBind tubes (Eppendorf). Beads were separated when required using a magnetic tube holder, and kept in buffer for as long as possible to prevent them from drying out. All pipette tips used to mix beads were cut to widen the end prior to use (≈2 mm cut off).

Peptide-biotin conjugates (with and without parent peptide), or biotin (variable concentrations (Section 3.2.1), 1 µL) were pre-incubated with cell lysate (50 µL, 250–500 µg protein, Section 6.8.2) for 30 min on ice. Meanwhile 50–100 µL Dynabeads™ M-280 streptavidin magnetic beads (Invitrogen™) per sample were washed twice in TBST (150–300 µL), a final wash was performed in

cell lysis buffer (150–300 µL). Affinity pulldown was performed by incubating compound/lysate mixtures with washed beads for 1 hr at 4 °C, with end-over-end rotation. Beads were subsequently washed three times in TBST (150–300 µL), moving the beads to a fresh tube on the second wash. Samples were eluted in 50 µL of 1:1 TBST:SDS-PAGE sample loading buffer (4 × NuPAGE™ LDS sample buffer, supplemented with NuPAGE™ sample reducing agent), beads were heated for 10 min at 95 °C, and sample removed from the beads prior to storage. Pulldown samples were analysed using silver stained SDS-PAGE protein gels (Section 6.5.1) and western blots (Section 6.5.2).

For global proteomics, an alternative sample elution step was used (Section 6.9.2).

6.9.2 Affinity pulldown for global MS-based proteomic analysis

Following affinity pulldown, a trypsin digest was performed on beads to elute samples for 'bottom up' proteomics, using methodology based upon Mohammed *et al.* (2016).[244] Beads were washed twice in fresh, ice cold ammonium bicarbonate (100 mM), moving the beads to a fresh tube on the second wash, and removing as much solution as possible prior to the trypsin digest. Trypsin (300 ng) in ammonium bicarbonate (100 mM) was incubated with the beads (6 µL solution total) at 37 °C for 4 hr, gently agitating the beads in a Thermomixer™ (Eppendorf™). Additional trypsin (75 ng) in ammonium bicarbonate (100 mM) was added to the beads (1.5 µL solution added), and left at 37 °C overnight, with gentle agitation. After the trypsin digest, solution was moved to a fresh tube, and acidified with 99+% formic acid to a final concentration of 5% (v/v) formic acid. For all mass spectrometry, 99+% formic acid was used (Optima™ LC/MS grade, Fisher Chemical™).

Samples were desalted using home packed C18 stage tips, as in Rappsilber *et al.* (2003).[282] Two pieces of Empore™ C18 extraction disk (Fisher Scientific, 13-100-018) were suspended in a 10 µL pipette tip, which was placed in a small hole in a 1.5 mL tube lid. Solutions were added to the top of the pipette tip, and eluted in a microcentrifuge, at 2000 × *g* for 1 min at room temperature. The C18 stage tip was conditioned with two × 70 µL 80% (v/v) MeCN / 0.02% (v/v) formic acid, prior to

equilibration with four × 70 µL 0.1% (v/v) formic acid. The C18 stage tip was placed in a fresh 1.5 mL tube, and the acidified peptides from the trypsin digest were next loaded to the C18 stage tip, and flowed through; peptides were reloaded three times to maximise peptide binding. The C18 stage tip was subsequently washed four times with 70 µL 0.1% (v/v) formic acid. Finally the C18 stage tip was placed in a fresh 1.5 mL collection tube, and the peptides eluted with two × 50 µL 50% (v/v) MeCN / 0.05% (v/v) formic acid. Eluted peptides were concentrated using a freeze dryer, and stored at -80 °C.

Data Dependent Acquisition nano LC-MS Thermo: Acquisitions were carried out on a Thermo RSLC Nano coupled to a Thermo Exploris 480 mass spectrometer. Peptide sample (≈1 µg per LC-MS run) was loaded/desalted onto a 300 µm × 5 mm C18 PepMap C18 trap stage tip in 0.1% (v/v) formic acid at 10 $\mu L \cdot min^{-1}$ for 3 min, and then further separated on a 75 µm × 50 cm C18 column (Thermo EasySpray - C18 2 µm) with integrated emitter at 250 $nL \cdot min^{-1}$. Peptides were separated over the course of 70 min with a 3–40% nonlinear gradient of solvent B (80% (v/v) MeCN supplemented with 0.1% (v/v) formic acid) to solvent A (H_2O supplemented with 0.1% (v/v) formic acid). The eluent was directed to a Thermo Orbitrap Exploris 480 mass spectrometer through the EasySpray source at a temperature of 320 °C, spray voltage 1900 V. The total LC-MS run time was 90 min. Orbitrap full scan resolution was 60,000, RF lens 50%, normalised AGC Target 300%. Precursors for MS/MS were selected via a top 20 method. MIPS set to peptide, intensity threshold 5.0 e^3, charge state 2–7 and dynamic exclusion after 1 × for 35 s 10 ppm mass tolerance. ddMS2 scans were performed at 15,000 resolution, HCD collision energy 27%, first mass 110 m/z, AGC Target Standard.

Protein identification: The acquired Data Dependent Acquisition data was searched against the protein sequence database available from (https://www.uniprot.org/proteomes/), concatenated to the Common Repository for Adventitious Proteins v.2012.01.01 (cRAP, ftp://ftp.thegpm.org/fasta/cRAP), using MaxQuant v1.6.10.43, parameters used: cysteine

alkylation: iodoacetamide, digestion enzyme: trypsin, parent mass error of 20 ppm, fragment mass error of 30 ppm. Output .txt files were further processed using Perseus (v1.6.15.0).[o]

6.9.3 Silver stain SDS-PAGE protein gel band extraction, destaining, and MS-analysis

Mass spectrometry was used in order to confirm the presence of KDM4A in a silver stained gel band from **B-1** (1 µM) affinity pulldown from cell lysates containing overexpressed FlagKDM4A (Section 3.2.1.1.2). The desired gel band, and a molecular weight marker as a loading control, were each excised from the silver stained gel (Section 6.5.1.2) using a fresh scalpel. Gel bands were destained using a SilverQuest™ (Invitrogen™) silver stain kit, following the manufacturer's instructions. Gel bands were submitted to NUPPA for MS analysis.

Each band was diced into 1×1×1 mm cubes and transferred to a clean microcentrifuge tube. Gel pieces were destained by addition of excess 50 mM ammonium bicarbonate in 50% (v/v) MeCN. The destain buffer was removed and exchanged three times, or until the gel pieces were clear. Proteins were reduced with 10 mM dithiothreitol for 30 min at 60 ˚C, to break disulphide bridges. This was followed by alkylation with 50 mM iodoacetamide for 30 min at room temperature, in the dark, to prevent disulphide reformation. Gel pieces were washed in 50 mM ammonium bicarbonate and then dehydrated with three washes of 100 µL MeCN. Residual moisture was removed from gel pieces in a vacuum drier. Proteins were digested by the addition of trypsin, added at a ratio of 30:1 (protein:trypsin), buffered with 50 mM ammonium bicarbonate, and incubated for 16 hr at 37 ˚C. The digest was quenched by the addition of 10% (v/v) TFA to a final concentration of 0.5%, and shaken for 30 min at 750 RPM. The liquid containing hydrophilic peptides was transferred to a fresh microcentrifuge tube. The gel pieces were then incubated with 80% (v/v) MeCN / 2% TFA (v/v) and shaken for 30 min at 750 RPM. This dehydrates the gel pieces and removes hydrophobic peptides from the gel. The solution containing hydrophobic peptides was pooled with the hydrophilic peptide mix. The peptide solution was dried in a centrifugal

[o] LC-MS data acquired and protein identification performed by Dr Andrew Porter, NUPPA.

evaporator, peptides were dissolved in 3% (v/v) MeCN supplemented with 0.1% (v/v) TFA. The resulting peptide solutions were desalted using home packed C18 stage tips as in Section 6.9.2. The sample was dissolved in 50 µL of 3% (v/v) MeCN supplemented with 0.1% (v/v) TFA, giving a final concentration of ≈1 µg/µL.[p]

6.10 Photoreactive protein labelling

6.10.1 Recombinant KDM4A photocrosslinking

Recombinant KDM4A photocrosslinking buffer: 1 × PBS, 20 µM $MnCl_2$

Crosslinking reactions were carried out in clear, V-bottom 96-well plates (Greiner Bio-One) or white, Small Volume™ HiBase 384-well plates (Greiner Bio-One). Peptides (variable concentrations, Section 3.2.2) were pre-incubated with recombinant biotin-$KDM4A_{1-359}$ for 15 min on ice, prior to irradiation on ice. Variable crosslinking times were used in a UVP CL-1000M 302 nm crosslinker or a UVP CL-1000L 365 nm crosslinker. Lamps were pre-warmed for 2 min prior to irradiation.

Intact protein LC-MS: Data was obtained on a 6224 TOF LC/MS (Agilent) coupled to a 1200 Series autosampler (Agilent) using MassHunter LC/MS Data Acquisition B.04.00 software. A PLRP-S reversed-phase BioHPLC column (Agilent, 1000 Å, 5 µm, 50 × 1.0 mm) was used. Injections of 10 µL were used at 70 °C. Flow rate was set to 0.5 $mL \cdot min^{-1}$ with a variable gradient method between solvent A (HPLC grade H_2O supplemented with 0.2% (v/v) formic acid) and solvent B (MeCN supplemented with 0.2% (v/v) formic acid) shown in **Table 6.8**. Eluent was sent to MS from 0.5–6 min, which was set to: capillary voltage of 4200 V; gas temperature of 340 °C; gas flow of 8.0 $L \cdot min^{-1}$. Analysis was performed using Agilent MassHunter Qualitative Analysis B.06.00. MS spectrum was extracted from the protein peak in the LC trace, spectrum was deconvoluted from 35–55 kDa across the envelope in the MS spectrum using the Maximum Entropy algorithm.

[p] Trypsin digest and mass spectrometry performed by Dr Andrew Porter, NUPPA.

Time / min	% Solvent A	% Solvent B
0.00	90	10
0.50	70	30
5	42	58
5.7	35	65
5.71	0	100
6.2	0	100
6.21	90	10
8	90	10
8.1	90	10

Table 6.8 KDM4A photocrosslinking LC-MS sample gradient.

6.10.2 TET photocrosslinking

Recombinant TET2 photocrosslinking buffer: 20 mM HEPES, 150 mM NaCl, 20 µM $MnCl_2$.

Crosslinking reactions were carried out in clear, flat bottom UV-Star® 96-well plates (Greiner Bio-One), or 0.5 mL tubes. Peptides (25 µM photoreactive peptides, with and without 250 µM **TiP1**) were pre-incubated with recombinant TET2 for 15 min on ice, prior to irradiation on ice. Alternatively, photoreactive peptides (25 µM) were pre-incubated with doxycycline-induced FlagTET1 overexpressing U2OS cell lysates (130 µg protein, Section 6.8.2) for 15 min on ice, prior to irradiation on ice. For all samples, crosslinking was performed for 20 min at 365 nm under a 230V, 50Hz 3UV Lamp (Fisher Scientfic, 10531194). The lamp was pre-warmed for 2–5 min prior to irradiation.

Intact protein LC-MS: Samples were acquired using a Thermo RSLC coupled to an ABSciex 6600 QTof. Chromatographic separations were achieved using a Fortis C8 column, 100 mm × 2.1 mm, 3 µm particle size, 45 °C at a flow rate of 250 $\mu L \cdot min^{-1}$. The loading pump was used at high flow to run linear chromatography gradients. The buffer system comprised of solvent A (HPLC grade H_2O supplemented with 0.1% (v/v) formic acid) and solvent B (MeCN supplemented with 0.1% (v/v) formic acid). The equivalent of 1–5 µM (1–5 $pmol \cdot \mu L^{-1}$) of protein sample was injected into the system per run, loading was optimised per sample. Samples were loaded onto the column and desalted online for 1 min, with eluent diverted to waste. A valve switch directed flow to the mass

spectrometer. Proteins were then eluted from the column using a linear gradient from 3–70% solvent B over 12 min. Total run time was 17 min. The eluent was directed into an ABSciex 6600 QTof, operated in positive mode. Source conditions were: temperature 180 °C, GS1 25, GS2 15, ISFV 4500v. Data was acquired in MS scan mode between 300–2500 m/z. Resolution of the instrument was 45000 at m/z 829.54 ($[M+H]^+$).[q]

Data for each chromatographic peak was averaged into an MS spectrum and was exported as a .txt file and imported into UniDec v4.4.1 for mass deconvolution.

6.11 Cloning pFN21A HaloFlagKDM4Awt and HaloFlagKDM4Amut plasmids

Triple *N*-terminal Flag-tagged full length *KDM4A* wt and H188A catalytically inactive mut, were cloned into a pFN21A HaloTag® CMV Flexi® Vector (Promega, G2821), for mammalian expression of HaloFlagKDM4A. DNA for *FlagKDM4Awt* and *FlagKDM4Amut* was amplified from pcDNA-flag-hKDM4A-polyA and pcDNA-flag-hKDM4A(H188A)-polyA plasmids respectively (Section 6.8.1).

6.11.1 Appending AsiSI and PmeI cutting sites to FlagKDM4Awt/mut

FlagKDM4Awt/mut DNA was amplified, and an amino terminal AsiSI (an isochizomer of SgfI) and carboxy terminal PmeI cutting site were added, using polymerase chain reaction (PCR). The reaction mixture is shown in **Table 6.9**, and the following PCR parameters were used: 95 °C for 2 min, 25 × [95 °C for 20 s, 55 °C for 20 s, 72 °C for 3 min], 72 °C for 3 min. DNA purity was analysed by loading 5 µL PCR reaction (mixed with loading dye) to an agarose gel (Section 6.5.3), and found to be of sufficient purity to proceed. The PCR reaction was purified using a NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel™) kit, before analysing the concentration and purity of DNA using a NanoDrop One Spectrophotometer.

[q] MS performed by myself, supported by NUPPA.

Reagent	Stock Concentration	Volume / µL	Final concentration
DNA plasmid	50 ng·µL^{-1}	2	2 ng·µL^{-1}
Forward primer	10 µM	1	0.2 µM
Reverse primer	10 µM	1	0.2 µM
dNTP mix	10 mM	1.2	250 µM
PfuUltra II fusion HD DNA polymerase		1	
PfuUltra II reaction buffer	10 ×	5	1 ×
H_2O		39.8	

Table 6.9 PCR reaction used to amplify *FlagKDM4Awt/mut* DNA, and to append SgfI and PmeI cutting sites.

6.11.2 Restriction digest of cloning inserts and acceptor pFN21A vector

Restriction digests for the PCR products and pFN21A vector were performed in tandem, reactions were set up as shown in **Table 6.10** & **Table 6.11**. Restriction digests were incubated at 37 °C for 60 min, then heated to 80 °C for 20 min to inactivate the restriction enzymes. The PCR product restriction digest was purified using PCR clean up with a NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel™) kit. The concentration of both restriction digests was subsequently tested using a NanoDrop One Spectrophotometer.

PCR products reactions **Reagent**	**Stock Concentration**	**Volume / µL**	**Final concentration**
Purified PCR product	200 ng·µL^{-1}	2.5	10 ng·µL^{-1}
AsiSI	10 U·µL^{-1}	0.5	0.1 U·µL^{-1}
PmeI	10 U·µL^{-1}	0.5	0.1 U·µL^{-1}
Cutsmart buffer	10 ×	5	1 ×
H_2O		41.5	

Table 6.10 *FlagKDM4Awt/mut* PCR products restriction digest.

Acceptor Flexi vector reaction **Reagent**	**Stock Concentration**	**Volume / µL**	**Final concentration**
Acceptor pFN21A vector	100 ng·µL^{-1}	2	6.7 ng·µL^{-1}
AsiSI	10 U·µL^{-1}	0.75	0.15 U·µL^{-1}
PmeI	10 U·µL^{-1}	0.75	0.15 U·µL^{-1}
Cutsmart buffer	10 ×	3	1 ×
H_2O		22.5	

Table 6.11 Acceptor pFN21A vector restriction digest.

6.11.3 Ligation of PCR products and Flexi Vector, and transformation

The ligations for both wt and mut plasmids were set up as shown in **Table 6.12**, and incubated for 1 hr at room temperature.

Reagent	Stock Concentration	Volume / µL	Final concentration
Digested PCR product	20 ng·µL^{-1}	5	100 ng
Digested Acceptor Vector	22 ng·µL^{-1}	2.2	50 ng
T4 DNA ligase	20 U·µL^{-1}	1	1 U·µL^{-1}
Ligase buffer	10 ×	2	1 ×
H_2O		9.5	

Table 6.12 Ligation of PCR products into pFN21A vector.

Ligation reactions were transformed into NEB® 5-alpha competent *E. coli* cells. Ligation reaction (5 µL) was mixed with NEB® 5-alpha cells (25 µL) for transformation using heat shock (Section 6.5.4). Agar plates were supplemented with ampicillin (100 µg·mL^{-1} final concentration).

6.11.4 Colony screening

Having obtained colonies for both pFN21A *KDM4Awt* and *KDM4Amut* cloning reactions, eight colonies per plasmid were grown in overnight starter cultures, to screen for the desired clones. Overnight starter cultures were prepared in 3 mL LB media supplemented with ampicillin (100 µg·mL^{-1} final concentration), for each culture a single colony was picked from the agar plate using a 10 µL pipette tip which was then left in the starter culture. Starter cultures were incubated overnight at 37 °C, shaking at 180 RPM. DNA was extracted from the starter cultures using minipreps (Section 6.5.5).

A restriction digest was performed to identify which colonies had the cloning insert within the purified plasmid. A master mix was made containing 2 × final concentration of restriction enzymes and Cutsmart buffer, so 5 µL of this master mix and 5 µL plasmid DNA (20 ng·µL^{-1}) could be mixed. Restriction digests were incubated at 37 °C for 1 hr. Subsequently, restriction digest samples were prepared directly for agarose gel electrophoresis (Section 6.5.3). A band at approximately 5 kb (plasmid) and 3 kb (insert) was observed where *FlagKDM4Awt/mut* DNA had successfully been inserted to the plasmid. For *KDM4Awt*, 7 of 8 colonies were successful, and for *KDM4Amut*, 4 of 8 colonies were successful.

Reagent	Stock Concentration	Volume / µL	Final concentration
Plasmid	20 ng·µL^{-1}	5	10 ng·µL^{-1}
AsiSI	10 U·µL^{-1}	0.1	0.1 U·µL^{-1}
PmeI	10 U·µL^{-1}	0.1	0.1 U·µL^{-1}
Cutsmart buffer	10 ×	1	1 ×
H_2O		3.8	

Table 6.13 Restriction digest used to identify colonies which contained the cloning insert.

Sanger sequencing using a T7 primer was performed on multiple colonies to establish that the *KDM4A* DNA was inserted in the correct orientation. Following this, a single colony for both *KDM4Awt* and *KDM4Amut* DNA was fully sequenced, and confirmed to contain the expected DNA. A Midi prep was performed on the final plasmids (Section 6.5.5).

6.12 HEK293T protein degradation experiments

Protein degradation experiments were performed using HEK293T cells seeded in 6-well plates or 35 mm dishes. The details of the compound concentration used in each protein degradation experiment is described in the main text, and all experiments used 0.1% (v/v) DMSO unless explicitly stated in the main text.

PROTAC treatment 48 hr post-transfection: Cells were seeded (0.25×10^6 cells per well) 24 hr prior to transfection, as described in Section 6.8.1. Compounds were added to cells by changing the media 48 hr post-transfection, and cells were lysed 1–24 hr later.

PROTAC treatment 4 hr post-transfection: Cells were seeded (0.6×10^6 cells per well) 16 hr prior to transfection, as described in Section 6.8.1. Compounds were added to cells by changing the media 4 hr post-transfection, and cells were lysed 4 hr later.

Where protease inhibitor and VHL ligands were co-incubated with HaloPROTAC-E, these were added 30 min prior to addition of PROTAC. Here media was not changed when dosing cells, and 20 × concentration compound stocks in media were added to cells (100 µL compound added to cells in ≈2 mL media).

PROTAC treatment on untransfected cells: Cells were seeded (0.6×10^6 cells per well) 20 hr prior addition of compounds to cells by changing the media, and cells were lysed 4 hr later.

Western blot densitometry to calculate relative KDM4A protein level: Densitometry analysis was performed on western blots to calculate the relative level of KDM4A protein in each sample. Images of non-saturated Flag/KDM4A and β-actin/GAPDH blots were exported from ImageLab (BIO-RAD) as .png files, and were opened in ImageJ. A rectangle was drawn to fit each western blot band as it moved across the blot, with minimal excess white space. Once the rectangle was correctly sized, the lanes were labelled using the 'Analyze->Gels' function. Next, the lanes were plotted (within the 'Analyze->Gels' function). A straight line was drawn across the base of each plot, and the wand tracing tool was used to calculate the area under the curve. Data for both the Flag/KDM4A blot and β-actin/GAPDH loading control blot for each experiment were exported to Excel. The ratio of KDM4A protein to loading control protein was calculated by dividing the area of the Flag/KDM4A plot by the area of the β-actin/GAPDH plot for each sample. The level of KDM4A protein was then calculated, relative to the DMSO control(s) within the experiment. Where greater than one DMSO sample was included in the experiment the samples were averaged prior to calculating relative KDM4A protein level, unless stated otherwise in the main text. Data was plotted in GraphPad Prism 9.4.0.

6.12.1 mRNA extraction and reverse transcription RT-qPCR

For mRNA extraction experiments, cell were dosed as previously described ('**PROTAC treatment 4 hr post-transfection**' (Section 6.12)), but cells were lysed differently using a Nucleospin RNA (Macherey-Nagel™) kit, according to manufacturer's instructions. RNA extraction was performed using the Nucleospin RNA kit, and a NanoDrop One Spectrophotometer was used to identify the concentration of RNA in each sample. Next, cDNA was synthesised from 2 µg mRNA. cDNA was synthesised using an Applied Biosystems™ cDNA kit (Fisher Scientific), according to

manufacturer's instructions, using 20 µL reactions. cDNA samples were diluted 1 in 5 in nuclease free H_2O (80 µL added to cDNA), prior to RT-qPCR, remaining cDNA was stored at -20 °C.

Both KDM4A and β-actin cDNA was quantified, and sequences of the forward and reverse primers used are shown in **Table 6.14**. RT-qPCR reactions were performed in Hard-Shell® low profile, thin wall, skirted, white/white 96-Well PCR Plates (BIO-RAD, HSP9655). cDNA (2 µL, 40 ng) was added to the plate, followed by 18 µL KDM4A or β-actin PCR mix (KDM4A and β-actin PCR mixes were made separately, and used in separate wells). Each well in the assay contained 40 ng cDNA, 300 nM forward primer, 300 nM reverse primer, 10 µL Applied Biosystems™ PowerSYBR™ Green PCR master mix (Fisher Scientific), made up to 20 µL with nuclease free H_2O. Plates were sealed with adhesive, optical Microseal 'B' PCR Plate Sealing Film (BIO-RAD, MSB1001), spun down briefly in a centrifuge, shaken on a vortex gently for 20 s, and spun down briefly in a centrifuge again. The RT-qPCR was run on a CFX96 Touch Real-Time PCR Detection System (BIO-RAD), using the parameters shown in **Table 6.15**.

Primer	Sequence
qPCR β-actin forward	GCTGTGCTACGTCGCCCTG
qPCR β-actin reverse	GGAGGAGCTGGAAGCAGCC
qPCR KDM4A forward	GCCGCTAGAAGTTTCAGTGAG
qPCR KDM4A reverse	GCGTCCCTTGGACTTCTTATT

Table 6.14 RT-qPCR primers for quantification of KDM4A and β-actin DNA.

95 °C	95 °C	55 °C	72 °C	4 °C
5 min	10 s	30 s	30 s	∞
			Capture	
	45 × ◄			

Table 6.15 RT-qPCR parameters.

Data was extracted to excel, and quantification cycle (Cq) used for calculations. The fold change in KDM4A expression was calculated using 2^{-[Cq(KDM4A)-Cq(β-actin)]},[351] and the relative KDM4A expression compared to DMSO control was plotted in GraphPad Prism 9.4.0.

Chapter 7 References

1. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500 (2016).

2. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **23**, 781–783 (2009).

3. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).

4. Bird, A. P. & Wolffe, A. P. Methylation-Induced Repression— Belts, Braces, and Chromatin. *Cell* **99**, 451–454 (1999).

5. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **28**, 1057–1068 (2010).

6. Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, eaag0025 (2017).

7. Luger, K., Mäder, a W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–260 (1997).

8. Jenuwein, T. & Allis, C. D. Translating the Histone Code. *Science* **293**, 1074–1081 (2001).

9. Dawson, M. A. & Kouzarides, T. Cancer epigenetics: From mechanism to therapy. *Cell* **150**, 12–27 (2012).

10. Fischle, W., Mootz, H. D. & Schwarzer, D. Synthetic histone code. *Curr. Opin. Chem. Biol.* **28**, 131–140 (2015).

11. Huang, H., Sabari, B. R., Garcia, B. A., David Allis, C. & Zhao, Y. SnapShot: Histone modifications. *Cell* **159**, 458–458 (2014).

12. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).

13. Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K. & Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discov.* **11**, 384–400 (2012).

14. Nicholson, T. B., Veland, N. & Chen, T. Writers, Readers, and Erasers of Epigenetic Marks. in *Epigenetic Cancer Therapy* (ed. Grey, S. G.) 31–66 (Elsevier Inc., 2015). doi:10.1016/B978-0-12-800206-3.00003-3.

15. Patel, D. J. & Wang, Z. Readout of epigenetic modifications. *Annu. Rev. Biochem.* **82**, 81–118 (2013).

16. Cheung, P., Allis, C. D. & Sassone-Corsi, P. Signaling to Chromatin through Histone Modifications. *Cell* **103**, 263–271 (2000).

17. Suganuma, T. & Workman, J. L. Signals and combinatorial functions of histone modifications. *Annu. Rev. Biochem.* **80**, 473–499 (2011).

18. Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38 (2013).

19. Hughes, H. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18**, 517–534 (2017).

20. Flashman, E. & Schofield, C. J. The most versatile of all reactive intermediates? *Nat. Chem.*

Biol. **3**, 86–87 (2007).

21. Loenarz, C. & Schofield, C. J. Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* **4**, 152–156 (2008).

22. Ploumakis, A. & Coleman, M. L. OH, the Places You'll Go! Hydroxylation, Gene Expression, and Cancer. *Mol. Cell* **58**, 729–741 (2015).

23. Thinnes, C. C. *et al.* Targeting histone lysine demethylases — Progress, challenges, and the future. *Biochim. Biophys. Acta 1839* **1839**, 1416–1432 (2014).

24. Wang, Z. A. & Cole, P. A. The Chemical Biology of Reversible Lysine Post-translational Modifications. *Cell Chem. Biol.* **27**, 953–969 (2020).

25. Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13**, 343–357 (2012).

26. Sterling, J., Menezes, S. V., Abbassi, R. H. & Munoz, L. Histone lysine demethylases and their functions in cancer. *Int. J. Cancer* **148**, 2375–2388 (2021).

27. Brejc, K. *et al.* Dynamic Control of X Chromosome Conformation and Repression by a Histone H4K20 Demethylase Article Dynamic Control of X Chromosome Conformation and Repression by a Histone H4K20 Demethylase. *Cell* **171**, 85–102 (2017).

28. Walport, L. J. *et al.* Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat. Commun.* **7**, 11974 (2016).

29. Walport, L. J., Hopkinson, R. J. & Schofield, C. J. Mechanisms of human histone and nucleic acid demethylases. *Curr. Opin. Chem. Biol.* **16**, 525–534 (2012).

30. Labbé, R. M., Holowatyj, A. & Yang, Z. Q. Histone lysine demethylase (kdm) subfamily 4: Structures, functions and therapeutic potential. *Am. J. Transl. Res.* **6**, 1–15 (2014).

31. UniProtKB. https://www.uniprot.org/uniprotkb/B2RXH2/. (2022).

32. Liu, X. *et al.* H3K9 demethylase KDM4E is an epigenetic regulator for bovine embryonic development and a defective factor for nuclear reprogramming. *Development* **145**, dev158261 (2018).

33. Berry, W. L. & Janknecht, R. KDM4/JMJD2 Histone Demethylases: Epigenetic Regulators in Cancer Cells. *Cancer Res.* **73**, 2936–2942 (2013).

34. Hillringhaus, L. *et al.* Structural and evolutionary basis for the dual substrate selectivity of human KDM4 histone demethylase family. *J. Biol. Chem.* **286**, 41616–41625 (2011).

35. Trojer, P. *et al.* Dynamic histone H1 isotype 4 methylation and demethylation by histone lysine methyltransferase G9a/KMT1C and the jumonji domain-containing JMJD2/KDM4 proteins. *J. Biol. Chem.* **284**, 8395–8405 (2009).

36. Lee, D. H. *et al.* Advances in histone demethylase KDM4 as cancer therapeutic targets. *FASEB J.* **34**, 3461–3484 (2020).

37. Agger, K. *et al.* Jmjd2/Kdm4 demethylases are required for expression of Il3ra and survival of acute myeloid leukemia cells. *Genes Dev.* **30**, 1278–1288 (2016).

38. Wu, Q. *et al.* Recent Advances with KDM4 Inhibitors and Potential Applications. *J. Med. Chem.* **65**, 9564–9579 (2022).

39. Fang, Z. *et al.* Discovery of a potent and selective inhibitor of histone lysine demethylase

KDM4D. *Eur. J. Med. Chem.* **223**, 113662 (2021).

40. Kawamura, A. *et al.* Highly selective inhibition of histone demethylases by de novo macrocyclic peptides. *Nat. Commun.* **8**, 14773 (2017).

41. Chen, Y. K. *et al.* Design of KDM4 Inhibitors with Antiproliferative Effects in Cancer Models. *ACS Med. Chem. Lett.* **8**, 869–874 (2017).

42. Westaway, S. M. *et al.* Cell Penetrant Inhibitors of the KDM4 and KDM5 Families of Histone Lysine Demethylases. 2. Pyrido[3,4-d]pyrimidin-4(3H)- one Derivatives. *J. Med. Chem.* **59**, 1370−1387 (2016).

43. Bavetsias, V. *et al.* 8-Substituted Pyrido[3,4-d]pyrimidin-4(3H)-one Derivatives As Potent, Cell Permeable, KDM4 (JMJD2) and KDM5 (JARID1) Histone Lysine Demethylase Inhibitors. *J. Med. Chem.* **59**, 1388–1409 (2016).

44. Le Bihan, Y. V. *et al.* C8-substituted pyrido[3,4-d]pyrimidin-4(3H)-ones: Studies towards the identification of potent, cell penetrant Jumonji C domain containing histone lysine demethylase 4 subfamily (KDM4)inhibitors, compound profiling in cell-based target engagement assays. *Eur. J. Med. Chem.* **177**, 316–337 (2019).

45. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).

46. Rasmussen, K. D. & Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* **30**, 733–750 (2016).

47. Hu, L. *et al.* Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527**, 118–122 (2015).

48. Zhang, W. *et al.* Isoform Switch of TET1 Regulates DNA Demethylation and Mouse Development. *Mol. Cell* **64**, 1062–1073 (2016).

49. Jin, S. G. *et al.* Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Rep.* **14**, 493–505 (2016).

50. Liu, N. *et al.* Intrinsic and Extrinsic Connections of Tet3 Dioxygenase with CXXC Zinc Finger Modules. *PLoS One* **8**, e62755 (2013).

51. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–349 (2011).

52. Lian, H., Li, W. Bin & Jin, W. L. The emerging insights into catalytic or non-catalytic roles of TET proteins in tumors and neural development. *Oncotarget* **7**, 64512–64525 (2016).

53. Scourzic, L., Mouly, E. & Bernard, O. A. TET proteins and the control of cytosine demethylation in cancer. *Genome Med.* **7**, 9 (2015).

54. Huang, H. *et al.* TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11994–11999 (2013).

55. Nishio, K. *et al.* Thioether Macrocyclic Peptides Selected against TET1 Compact Catalytic Domain Inhibit TET1 Catalytic Activity. *ChemBioChem* **19**, 979–985 (2018).

56. Sudhamalla, B., Dey, D., Breski, M. & Islam, K. A rapid mass spectrometric method for the measurement of catalytic activity of ten-eleven translocation enzymes. *Anal. Biochem.* **534**, 28–35 (2017).

57. Guan, Y. *et al.* A Therapeutic Strategy for Preferential Targeting of TET2-Mutant and TET

Dioxygenase-Deficient Cells in Myeloid Neoplasms. *Cancer Discov.* **2**, 146–161 (2021).

58. Singh, A. K. *et al.* Selective targeting of TET catalytic domain promotes somatic cell reprogramming. *Proc. Natl. Acad. Sci.* **117**, 3621–3626 (2020).

59. Chua, G. N. L. *et al.* Cytosine-Based TET Enzyme Inhibitors. *ACS Med. Chem. Lett.* **10**, 180–185 (2019).

60. Weirath, N. A. *et al.* Small Molecule Inhibitors of TET Dioxygenases: Bobcat339 Activity Is Mediated by Contaminating Copper(II). *ACS Med. Chem. Lett.* **13**, 792–798 (2022).

61. Arrowsmith, C. H. *et al.* The promise and peril of chemical probes. *Nat. Chem. Biol.* **11**, 536–541 (2015).

62. Frye, S. V. The art of the chemical probe. *Nat. Chem. Biol.* **6**, 159–161 (2010).

63. Bunnage, M. E., Chekler, E. L. P. & Jones, L. H. Target validation using chemical probes. *Nat. Chem. Biol.* **9**, 195–199 (2013).

64. Garbaccio, R. M. & Parmee, E. R. The Impact of Chemical Probes in Drug Discovery: A Pharmaceutical Industry Perspective. *Cell Chem. Biol.* **23**, 10–17 (2016).

65. Mullard, A. A probe for every protein. *Nat. Rev. Drug Discov.* **18**, 733–736 (2019).

66. Nevola, L. & Giralt, E. Modulating protein-protein interactions: the potential of peptides. *Chem. Commun. (Camb).* **51**, 3302–15 (2015).

67. Didenko, V. V. Dna probes using fluorescence resonance energy transfer (FRET): Designs and applications. *Biotechniques* **31**, 1106–1121 (2001).

68. Kobayashi, H., Choyke, P. L. & Ogawa, M. Monoclonal antibody-based optical molecular imaging probes; considerations and caveats in chemistry, biology and pharmacology. *Curr. Opin. Chem. Biol.* **33**, 32–38 (2016).

69. Gerry, C. J. & Schreiber, S. L. Unifying principles of bifunctional, proximity-inducing small molecules. *Nat. Chem. Biol.* **16**, 369–378 (2020).

70. Dang, C. V, Reddy, E. P., Shokat, K. M. & Soucek, L. Drugging the 'undruggable' cancer targets. *Nat. Rev. Cancer* **17**, 502–508 (2017).

71. Emmerich, C. H. *et al.* Improving target assessment in biomedical research: the GOT-IT recommendations. *Nat. Rev. Drug Discov.* **20**, 64–81 (2021).

72. Tumber, A. *et al.* Potent and Selective KDM5 Inhibitor Stops Cellular Demethylation of H3K4me3 at Transcription Start Sites and Proliferation of MM1S Myeloma Cells. *Cell Chem. Biol.* **24**, 371–380 (2017).

73. Ganesan, A., Arimondo, P. B., Rots, M. G., Jeronimo, C. & Berdasco, M. The timeline of epigenetic drug discovery: From reality to dreams. *Clin. Epigenetics* **11**, 1–17 (2019).

74. Bowkett, D. *et al.* Identifying Small-Molecule Binding Sites for Epigenetic Proteins at Domain–Domain Interfaces. *ChemMedChem* **13**, 1051–1057 (2018).

75. Hau, M., Zenk, F., Ganesan, A., Iovino, N. & Jung, M. Cellular analysis of the action of epigenetic drugs and probes. *Epigenetics* **12**, 308–322 (2017).

76. Weigt, D., Hopf, C. & Médard, G. Studying epigenetic complexes and their inhibitors with the proteomics toolbox. *Clin. Epigenetics* **8**, 76 (2016).

77. Speers, A. E. & Cravatt, B. F. Activity-Based Protein Profiling (ABPP) and Click Chemistry

(CC)-ABPP by MudPIT Mass Spectrometry. in *Current Protocols in Chemical Biology* 29–41 (John Wiley & Sons, Inc, 2009).

78. Bush, J. T. *et al.* The Ugi four-component reaction enables expedient synthesis and comparison of photoaffinity probes. *Chem. Sci.* **4**, 4115–4120 (2013).

79. Pickens, C. J., Johnson, S. N., Pressnall, M. M., Leon, M. A. & Berkland, C. J. Practical Considerations, Challenges, and Limitations of Bioconjugation via Azide−Alkyne Cycloaddition. *Bioconjug. Chem.* **29**, 686–701 (2018).

80. Speers, A. E. & Cravatt, B. F. Profiling Enzyme Activities In Vivo Using Click Chemistry Methods. *Chem. Biol.* **11**, 535–546 (2004).

81. Meyer, K. & Selbach, M. Quantitative affinity purification mass spectrometry: A versatile technology to study protein-protein interactions. *Front. Genet.* **6**, 237 (2015).

82. Fang, H. *et al.* Recent advances in activity-based probes (ABPs) and affinity-based probes (AfBPs) for profiling of enzymes. *Chem. Sci.* **12**, 8288–8310 (2021).

83. Schirle, M., Bantscheff, M. & Kuster, B. Mass spectrometry-based proteomics in preclinical drug discovery. *Chem. Biol.* **19**, 72–84 (2012).

84. Cravatt, B. F., Wright, A. T. & Kozarich, J. W. Activity-based protein profiling: From enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* **77**, 383–414 (2008).

85. McAllister, T. E. *et al.* Non-competitive cyclic peptides for targeting enzyme–substrate complexes. *Chem. Sci.* **9**, 4569–4578 (2018).

86. Jin, C. *et al.* Chem-seq permits identification of genomic targets of drugs against androgen receptor regulation selected by functional phenotypic screens. *Proc. Natl. Acad. Sci.* **111**, 9235–9240 (2014).

87. Park, P. J. ChIP-seq: Advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10**, 669–680 (2009).

88. Anders, L. *et al.* Genome-wide localization of small molecules. *Nat. Biotechnol.* **32**, 92–98 (2014).

89. Maniaci, C. & Ciulli, A. Bifunctional chemical probes inducing protein–protein interactions. *Curr. Opin. Chem. Biol.* **52**, 145–156 (2019).

90. Kallen, J. *et al.* Structure of human cyclophilin and its binding site for cyclosporin A determined by X-ray crystallography and NMR spectroscopy. *Nature* **353**, 276–279 (1991).

91. Huai, Q. *et al.* Crystal structure of calcineurin-cyclophilin-cyclosporin shows common but distinct recognition of immunophilin-drug complexes. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12037–12042 (2002).

92. Andrei, S. A. *et al.* Stabilization of protein-protein interactions in drug discovery. *Expert Opin. Drug Discov.* **12**, 925–940 (2017).

93. Nemetski, S. M. *et al.* Inhibition by stabilization: Targeting the Plasmodium falciparum aldolase-TRAP complex. *Malar. J.* **14**, 324 (2015).

94. Waring, M. J. *et al.* Potent and selective bivalent inhibitors of BET bromodomains. *Nat. Chem. Biol.* **12**, 1097–1104 (2016).

95. Békés, M., Langley, D. R. & Crews, C. M. PROTAC targeted protein degraders: the past is prologue. *Nat. Rev. Drug Discov.* **21**, 181–200 (2022).

96. Dikic, I. Proteasomal and autophagic degradation systems. *Annu. Rev. Biochem.* **86**, 193–224 (2017).

97. Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **353**, aac4354 (2016).

98. Alabi, S. B. & Crews, C. M. Major advances in targeted protein degradation: PROTACs, LYTACs, and MADTACs. *J. Biol. Chem.* **296**, 100647 (2021).

99. Metzger, M. B., Hristova, V. A. & Weissman, A. M. HECT and RING finger families of E3 ubiquitin ligases at a glance. *J. Cell Sci.* **125**, 531–537 (2012).

100. Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229 (2012).

101. Livneh, I., Cohen-Kaplan, V., Cohen-Rosenzweig, C., Avni, N. & Ciechanover, A. The life cycle of the 26S proteasome: From birth, through regulation and function, and onto its death. *Cell Res.* **26**, 869–885 (2016).

102. Saeki, Y. & Tanaka, K. Assembly and Function of the Proteasome. in *Ubiquitin Family Modifiers and the Proteasome* (eds. Jürgen Dohmen Dohmen, R. & Scheffner, M.) vol. 832 315–338 (Humana Press, 2012).

103. Coleman, K. G. & Crews, C. M. Proteolysis-Targeting Chimeras: Harnessing the Ubiquitin-Proteasome System to Induce Degradation of Specific Target Proteins. *Annu. Rev. Cancer Biol.* **2**, 41–58 (2018).

104. Schapira, M., Calabrese, M. F., Bullock, A. N. & Crews, C. M. Targeted protein degradation: expanding the toolbox. *Nat. Rev. Drug Discov.* **18**, 949–963 (2019).

105. Sakamoto, K. M. *et al.* Protacs: Chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc. Natl. Acad. Sci.* **98**, 8554–8559 (2001).

106. Li, D., Yu, D., Li, Y. & Yang, R. A bibliometric analysis of PROTAC from 2001 to 2021. *Eur. J. Med. Chem.* **244**, 114838 (2022).

107. Ishida, T. & Ciulli, A. E3 Ligase Ligands for PROTACs: How They Were Found and How to Discover New Ones. *SLAS Discov.* **26**, 484–502 (2021).

108. Bondeson, D. P. *et al.* Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat. Chem. Biol.* **11**, 611–617 (2015).

109. Banik, S. M. *et al.* Lysosome-targeting chimaeras for degradation of extracellular proteins. *Nature* **584**, 291–297 (2020).

110. Takahashi, D. *et al.* AUTACs: Cargo-Specific Degraders Using Selective Autophagy. *Mol. Cell* **76**, 797–810 (2019).

111. Ji, C. H. *et al.* The AUTOTAC chemical biology platform for targeted protein degradation via the autophagy-lysosome system. *Nat. Commun.* **13**, 904 (2022).

112. Li, Z., Zhu, C., Ding, Y., Fei, Y. & Lu, B. ATTEC: a potential new approach to target proteinopathies. *Autophagy* **16**, 185–187 (2020).

113. Rudin, M. & Weissleder, R. Molecular imaging in drug discovery and development. *Nat. Rev. Drug Discov.* **2**, 123–131 (2003).

114. Prevet, H. & Collins, I. Labelled chemical probes for demonstrating direct target engagement in living systems. *Future Med. Chem.* **11**, 1195–1224 (2019).

115. Marshall, A. P., Shirley, J. D. & Carlson, E. E. Enzyme-targeted fluorescent small-molecule probes for bacterial imaging. *Curr. Opin. Chem. Biol.* **57**, 155–165 (2020).

116. Knight, J. C. & Cornelissen, B. Bioorthogonal chemistry : implications for pretargeted nuclear (PET/SPECT) imaging and therapy. *Am. J. Nucl. Med. Mol. Imaging* **4**, 96–113 (2014).

117. Magliocca, V. *et al.* Identifying the dynamics of actin and tubulin polymerization in iPSCs and in iPSC-derived neurons. *Oncotarget* **8**, 111096–111109 (2017).

118. Wang, L., Frei, M. S., Salim, A. & Johnsson, K. Small-Molecule Fluorescent Probes for Live-Cell Super-Resolution Microscopy. *J. Am. Chem. Soc.* **141**, 2770–2781 (2019).

119. Los, G. V. *et al.* HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).

120. Ratz, M., Testa, I., Hell, S. W. & Jakobs, S. CRISPR/Cas9-mediated endogenous protein tagging for RESOLFT super-resolution microscopy of living human cells. *Sci. Rep.* **5**, 9592 (2015).

121. Machleidt, T. *et al.* NanoBRET-A Novel BRET Platform for the Analysis of Protein-Protein Interactions. *ACS Chem. Biol.* **10**, 1797–1804 (2015).

122. Passioura, T., Bhushan, B., Tumber, A., Kawamura, A. & Suga, H. Structure-activity studies of a macrocyclic peptide inhibitor of histone lysine demethylase 4A. *Bioorganic Med. Chem.* **26**, 1225–1231 (2018).

123. Kurra, Y. *et al.* Two Rapid Catalyst-Free Click Reactions for in Vivo Protein Labeling of Genetically Encoded Strained Alkene/Alkyne Functionalities. *Bioconjug. Chem.* **25**, 1730–1738 (2014).

124. Fischman, A. J., Alpert, N. M. & Rubin, R. H. Pharmacokinetic Imaging: A Noninvasive Method for Determining Drug Distribution and Action. *Clin. Pharmacokinet.* **41**, 581–602 (2002).

125. Ma, H. & O'Kennedy, R. The Structure of Natural and Recombinant Antibodies. in *Peptide Antibodies* (ed. Houen, G.) 7–12 (Humana Press, 2015).

126. Saerens, D., Ghassabeh, G. H. & Muyldermans, S. Antibody technology in proteomics. *Briefings Funct. Genomics Proteomics* **7**, 275–282 (2008).

127. Chames, P., Van Regenmortel, M., Weiss, E. & Baty, D. Therapeutic antibodies: Successes, limitations and hopes for the future. *Br. J. Pharmacol.* **157**, 220–233 (2009).

128. Harris, L. J., Larson, S. B., Hasel, K. W. & McPherson, A. Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry* **36**, 1581–1597 (1997).

129. Mori, H. & Cardiff, R. D. Methods of Immunohistochemistry and Immunofluorescence: Converting Invisible to Visible. in *The Tumor Microenvironment* (eds. Ursini-Siegel, J. & Beauchemin, N.) 1–12 (Springer Nature, 2016).

130. Gao, Y., Huang, X., Zhu, Y. & Lv, Z. A brief review of monoclonal antibody technology and its representative applications in immunoassays. *J. Immunoass. Immunochem.* **39**, 351–364 (2018).

131. Mahmood, T. & Yang, P. C. Western blot: Technique, theory, and trouble shooting. *N. Am. J. Med. Sci.* **4**, 429–434 (2012).

132. Marsden, C. J. *et al.* The use of antibodies in small-molecule drug discovery. *J. Biomol. Screen.* **19**, 829–838 (2014).

133. Smith, B. E. *et al.* Differential PROTAC substrate specificity dictated by orientation of recruited E3 ligase. *Nat. Commun.* **10**, 131 (2019).

134. Tian, Y. & Lin, Q. Fitness Factors for Bioorthogonal Chemical Probes. *ACS Chem. Biol.* **14**, 2489–2496 (2019).

135. Renaud, J. P. *et al.* Biophysics in drug discovery: Impact, challenges and opportunities. *Nat. Rev. Drug Discov.* **15**, 679–698 (2016).

136. Hattori, T. & Koide, S. Next-generation antibodies for post-translational modifications. *Curr. Opin. Struct. Biol.* **51**, 141–148 (2018).

137. Yang, T., Liu, Z. & Li, X. D. Developing diazirine-based chemical probes to identify histone modification 'readers' and 'erasers'. *Chem. Sci.* **6**, 1011–1017 (2015).

138. Borsari, C., Trader, D. J., Tait, A. & Costi, M. P. Designing Chimeric Molecules for Drug Discovery by Leveraging Chemical Biology. *J. Med. Chem.* **63**, 1908–1928 (2020).

139. Cyrus, K. *et al.* Impact of linker length on the activity of PROTACs. *Mol. Biosyst.* **7**, 359–364 (2011).

140. Hendrick, C. E. *et al.* Direct-to-Biology Accelerates PROTAC Synthesis and the Evaluation of Linker E ff ects on Permeability and Degradation. *ACS Med. Chem. Lett.* **13**, 1182–1190 (2022).

141. Gadd, M. S. *et al.* Structural basis of protac cooperative recognition for selective protein degradation. *Nat. Chem. Biol.* **13**, 514–521 (2017).

142. Králová, J. *et al.* Influence of fluorophore and linker length on the localization and trafficking of fluorescent sterol probes. *Sci. Rep.* **10**, 22053 (2020).

143. Su, D. & Zhang, D. Linker Design Impacts Antibody-Drug Conjugate Pharmacokinetics and Efficacy via Modulating the Stability and Payload Release Efficiency. *Front. Pharmacol.* **12**, 687926 (2021).

144. Deng, Y., Couch, B. A., Koleske, A. J. & Turk, B. E. A Peptide Photoaffinity Probe Specific for the Active Conformation of the Abl Tyrosine Kinase. *ChemBioChem* **13**, 2510 – 2512 (2012).

145. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **64**, 4–17 (2012).

146. Maneiro, M. *et al.* PROTACs, molecular glues and bifunctionals from bench to bedside: Unlocking the clinical potential of catalytic drugs. in *Progress in Medicinal Chemistry* vol. 60 67–190 (Elsevier B.V., 2021).

147. Serrano, L., Roper, G. & Kawamura, A. Cyclic Peptides as Chemical Probes. in *The discovery and utility of chemical probes in target discovery* (eds. Brennan, P. & Vazquez Rodriguez, S.) 100–123 (Royal Society of Chemistry, 2020). doi:10.1039/9781839160745.

148. Santos, G. B., Ganesan, A. & Emery, F. S. Oral Administration of Peptide-Based Drugs: Beyond Lipinski's Rule. *ChemMedChem* **11**, 2245–2251 (2016).

149. Henninot, A., Collins, J. C. & Nuss, J. M. The Current State of Peptide Drug Discovery: Back to the Future? *J. Med. Chem.* **61**, 1382–1414 (2018).

150. Ali, A. M., Atmaj, J., Van Oosterwijk, N., Groves, M. R. & Dömling, A. Stapled Peptides Inhibitors: A New Window for Target Drug Discovery. *Comput. Struct. Biotechnol. J.* **17**, 263–281 (2019).

151. Cox, O. B. *et al.* A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem. Sci.* **7**, 2322–2330 (2016).

152. Hopkinson, R. J. *et al.* 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem. Sci.* **4**, 3110–3117 (2013).

153. Dexter, B. D. D., Disease, V., York, N. & York, N. Conformations of Penicillin G : Crystal Structure of Procaine Penicillin G Monohydrate and a Refinement of the Structure of Potassium PenicillinG. *J. Chem. Soc. Perkin Trans. 1* 185–190 (1978).

154. Winter, G. E. *et al.* Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* **348**, 1376–1381 (2015).

155. Wood, S. P. *et al.* Crystal structure analysis of deamino-oxytocin: Conformational flexibility and receptor binding. *Science* **232**, 633–636 (1986).

156. Mikol, V., Kallen, J., Pflügl, G. & Walkinshaw, M. D. X-ray structure of a monomeric cyclophilin A-cyclosporin A crystal complex at 2.1 Å resolution. *Journal of Molecular Biology* vol. 234 1119–1130 (1993).

157. Stewart, M. L., Fire, E., Keating, A. E. & Walensky, L. D. The MCL-1 BH3 helix is an exclusive MCL-1 inhibitor and apoptosis sensitizer. *Nat. Chem. Biol.* **6**, 595–601 (2010).

158. Wang, Z. *et al.* Structural insight into a matured humanized monoclonal antibody HUA21 against HER2-overexpressing cancer cells. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 554–563 (2019).

159. Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001–9 (2007).

160. Bradbury, A. & Plückthun, A. Reproducibility: Standardize antibodies used in research. *Nature* **518**, 27–29 (2015).

161. Egelhofer, T. A. *et al.* An assessment of histone-modification antibody quality. *Nat. Struct. Mol. Biol.* **18**, 91–94 (2011).

162. McAllister, T. E., Coleman, O. D., Roper, G. & Kawamura, A. Structural diversity in de novo cyclic peptide ligands from genetically encoded library technologies. *Pept. Sci.* **113**, e24204 (2021).

163. Wang, L. *et al.* Therapeutic peptides: current applications and future directions. *Signal Transduct. Target. Ther.* **7**, 48 (2022).

164. Vlieghe, P., Lisowski, V., Martinez, J. & Khrestchatisky, M. Synthetic therapeutic peptides: science and market. *Drug Discov. Today* **15**, 40–56 (2010).

165. Fouché, M. *et al.* Design and Development of a Cyclic Decapeptide Scaffold with Suitable Properties for Bioavailability and Oral Exposure. *ChemMedChem* **11**, 1048–1059 (2016).

166. Matijass, M. & Neundorf, I. Cell-penetrating peptides as part of therapeutics used in cancer research. *Med. Drug Discov.* **10**, 100092 (2021).

167. Merrifield, R. B. Solid-Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am.*

Chem. Soc **85**, 2149–2154 (1963).

168. Behrendt, R., White, P. & Offer, J. Advances in Fmoc solid-phase peptide synthesis. *J. Pept. Sci.* **22**, 4–27 (2016).

169. Hipolito, C. J. & Suga, H. Ribosomal production and in vitro selection of natural product-like peptidomimetics: The FIT and RaPID systems. *Curr. Opin. Chem. Biol.* **16**, 196–203 (2012).

170. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie - Int. Ed.* **40**, 2004–2021 (2001).

171. Tang, W. & Becker, M. L. '"Click"' reactions: a versatile toolbox for the synthesis of peptide-conjugates. *Chem Soc Rev* **43**, 7013–1039 (2014).

172. NobelPrize.org. https://www.nobelprize.org/prizes/chemistry/2022/press-release/. *The Novel Prize in Chemistry 2022* (2022).

173. Palomo, J. M. Click reactions in protein chemistry: from the preparation of semisynthetic enzymes to new click enzymes. *Org. Biomol. Chem.* **10**, 9309–9318 (2012).

174. Jewett, J. C. & Bertozzi, C. R. Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev.* **39**, 1272–1279 (2010).

175. Wurz, R. P. *et al.* A 'click Chemistry Platform' for the Rapid Synthesis of Bispecific Molecules for Inducing Protein Degradation. *J. Med. Chem.* **61**, 453–461 (2018).

176. Kolb, H. C. & Sharpless, K. B. The growing impact of click chemistry on drug discovery. *Drug Discov. Today* **8**, 1128–1137 (2003).

177. Tyler, D. S. *et al.* Click chemistry enables preclinical evaluation of targeted epigenetic therapies. *Science* **356**, 1397–1401 (2017).

178. Xi, W., Scott, T. F., Kloxin, C. J. & Bowman, C. N. Click chemistry in materials science. *Adv. Funct. Mater.* **24**, 2572–2590 (2014).

179. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise huisgen cycloaddition process: Copper(I)-catalyzed regioselective 'ligation' of azides and terminal alkynes. *Angew. Chemie - Int. Ed.* **41**, 2596–2599 (2002).

180. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046–15047 (2004).

181. Berkel, S. S. Van, Eldijk, M. B. Van & Hest, J. C. M. Van. Staudinger Ligation as a Method for Bioconjugation Angewandte. *Angew. Chemie Int. Ed.* **50**, 8806–8827 (2011).

182. Soellner, M. B., Nilsson, B. L. & Raines, R. T. Reaction Mechanism and Kinetics of the Traceless Staudinger Ligation. *J. Am. Chem. Soc.* **128**, 8820–8828 (2006).

183. Inglis, A. J. & Barner-kowollik, C. Ultra Rapid Approaches to Mild Macromolecular Conjugation a. *Macromol. Rapid Commun.* **31**, 1247–1266 (2010).

184. Tasdelen, M. A. Diels – Alder '" click "' reactions : recent applications in polymer and material. *Polym. Chem.* **2**, 2133–2145 (2011).

185. Hoyle, C. E. & Bowman, C. N. Thiol-ene click chemistry. *Angew. Chemie - Int. Ed.* **49**, 1540–1573 (2010).

186. Lau, Y. H., Wu, Y., De Andrade, P., Galloway, W. R. J. D. & Spring, D. R. A two-component 'double-click' approach to peptide stapling. *Nat. Protoc.* **10**, 585–594 (2015).

187. Wu, Y. *et al.* Toolbox of Diverse Linkers for Navigating the Cellular Efficacy Landscape of Stapled Peptides. *ACS Chem. Biol.* **14**, 526–533 (2019).

188. Sachin, K. *et al.* F-18 Labeling Protocol of Peptides Based on Chemically Orthogonal Strain-Promoted Cycloaddition under Physiologically Friendly Reaction Conditions F-18 Labeling Protocol of Peptides Based on Chemically Orthogonal Strain-Promoted Cycloaddition under Physi. *Bioconjug. Chem.* **23**, 1680−1686 (2012).

189. Wright, T. H. *et al.* Direct peptide lipidation through thiol-ene coupling enables rapid synthesis and evaluation of self-adjuvanting vaccine candidates. *Angew. Chemie - Int. Ed.* **52**, 10616–10619 (2013).

190. Lebraud, H., Wright, D. J., Johnson, C. N. & Heightman, T. D. Protein degradation by in-cell self-assembly of proteolysis targeting chimeras. *ACS Cent. Sci.* **2**, 927–934 (2016).

191. Edmondson, S. D., Yang, B. & Fallan, C. Proteolysis targeting chimeras (PROTACs)in 'beyond rule-of-five' chemical space: Recent progress and future challenges. *Bioorganic Med. Chem. Lett.* **29**, 1555–1564 (2019).

192. Hsu, J. H.-R. *et al.* EED-Targeted PROTACs Degrade EED, EZH2, and SUZ12 in the PRC2 Complex. *Cell Chem. Biol.* **27**, 1–6 (2019).

193. Buckley, D. L. *et al.* HaloPROTACS: Use of Small Molecule PROTACs to Induce Degradation of HaloTag Fusion Proteins. *ACS Chem. Biol.* **10**, 1831–1837 (2015).

194. Tovell, H. *et al.* Rapid and Reversible Knockdown of Endogenously Tagged Endosomal Proteins via an Optimized HaloPROTAC Degrader. *ACS Chem. Biol.* **14**, 882–892 (2019).

195. Smits, A. H. & Vermeulen, M. Characterizing Protein–Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. *Trends Biotechnol.* **34**, 825–834 (2016).

196. de Lera, A. R. & Ganesan, A. Two-hit wonders: The expanding universe of multitargeting epigenetic agents. *Curr. Opin. Chem. Biol.* **57**, 135–154 (2020).

197. Bond, M. J. & Crews, C. M. Proteolysis targeting chimeras (PROTACs) come of age: entering the third decade of targeted protein degradation. *RSC Chem. Biol.* **2**, 725–742 (2021).

198. Wu, M., Huang, W., Yang, N. & Liu, Y. Learn from antibody–drug conjugates: consideration in the future construction of peptide-drug conjugates for cancer therapy. *Exp. Hematol. Oncol.* **11**, 1–18 (2022).

199. Chivers, C. E., Koner, A. L., Lowe, E. D. & Howarth, M. How the biotin-streptavidin interaction was made even stronger: Investigation via crystallography and a chimaeric tetramer. *Biochem. J.* **435**, 55–63 (2011).

200. Wilchek, M. & Bayer, E. A. The avidin-biotin complex in bioanalytical applications. *J. Clin. Chem. Clin. Biochem.* **27**, 889–890 (1989).

201. Dundas, C. M., Demonte, D. & Park, S. Streptavidin-biotin technology: Improvements and innovations in chemical and biological applications. *Appl. Microbiol. Biotechnol.* **97**, 9343–9353 (2013).

202. Futran, A. S., Kyin, S., Shvartsman, S. Y. & Link, A. J. Mapping the binding interface of ERK and transcriptional repressor Capicua using photocrosslinking. *Proc. Natl. Acad. Sci.* **112**, 8590–8595 (2015).

203. Weber, P. J. A. & Beck-Sickinger, A. G. Comparison of the photochemical behavior of four different photoactivatable probes. *J. Pept. Res.* **5**, 375–383 (1997).

204. Zeng, Y. *et al.* Functional probes for cardiovascular molecular imaging. *Quant. Imaging Med. Surg.* **8**, 838–852 (2018).

205. Troup, R. I., Fallan, C. & Baud, M. G. J. Current strategies for the design of PROTAC linkers: a critical review. *Explor. Target. Anti-tumor Ther.* **1**, 273–312 (2020).

206. Farnaby, W., Koegl, M., McConnell, D. B. & Ciulli, A. Transforming targeted cancer therapy with PROTACs: A forward-looking perspective. *Curr. Opin. Pharmacol.* **57**, 175–183 (2021).

207. Bondeson, D. P. *et al.* Lessons in PROTAC Design from Selective Degradation with a Promiscuous Warhead. *Cell Chem. Biol.* **25**, 78–87 (2018).

208. Henning, N. J. *et al.* Deubiquitinase-targeting chimeras for targeted protein stabilization. *Nat. Chem. Biol.* **18**, 412–421 (2022).

209. Siriwardena, S. U. *et al.* Phosphorylation-Inducing Chimeric Small Molecules. *J. Am. Chem. Soc.* **142**, 14052–14057 (2020).

210. Yamazoe, S. *et al.* Heterobifunctional Molecules Induce Dephosphorylation of Kinases-A Proof of Concept Study. *J. Med. Chem.* **63**, 2807–2813 (2020).

211. Costales, M. G., Suresh, B., Vishnu, K. & Disney, M. D. Targeted Degradation of a Hypoxia-Associated Non-coding RNA Enhances the Selectivity of a Small Molecule Interacting with RNA. *Cell Chem. Biol.* **26**, 1180-1186.e5 (2019).

212. Hoppenz, P., Els-Heindl, S. & Beck-Sickinger, A. G. Peptide-Drug Conjugates and Their Targets in Advanced Cancer Therapies. *Front. Chem.* **8**, Article 571 (2020).

213. Schneider, A. F. L., Kithil, M., Cardoso, M. C., Lehmann, M. & Hackenberger, C. P. R. Cellular uptake of large biomolecules enabled by cell-surface-reactive cell-penetrating peptide additives. *Nat. Chem.* **13**, 530–539 (2021).

214. Meissner, F., Geddes-McAlister, J., Mann, M. & Bantscheff, M. The emerging role of mass spectrometry-based proteomics in drug discovery. *Nat. Rev. Drug Discov.* **21**, 637–654 (2022).

215. Blagg, J. & Workman, P. Chemical biology approaches to target validation in cancer. *Curr. Opin. Pharmacol.* **17**, 87–100 (2014).

216. Jessani, N. & Cravatt, B. F. The development and application of methods for activity-based protein profiling. *Curr. Opin. Chem. Biol.* **8**, 54–59 (2004).

217. Jessani, N., Liu, Y., Humphrey, M. & Cravatt, B. F. Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10335–10340 (2002).

218. Bachovchin, D. A., Brown, S. J., Rosen, H. & Cravatt, B. F. Identification of selective inhibitors of uncharacterized enzymes by high-throughput screening with fluorescent activity-based probes. *Nat. Biotechnol.* **27**, 387–394 (2009).

219. Hong, V., Presolski, S. I., Ma, C. & Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew. Chemie - Int. Ed.* **48**, 9879–9883 (2009).

220. Huisgen, R. 1,3-Dipolar Cycloadditions Past and Future. *Angew. Chemie - Int. Ed.* **2**, 565–

632 (1963).

221. Li, Y., Hoskins, J. N., Sreerama, S. G. & Grayson, S. M. MALDI-TOF mass spectral characterization of polymers containing an azide group: Evidence of metastable ions. *Macromolecules* **43**, 6225–6228 (2010).

222. Chatterjee, J., Laufer, B. & Kessler, H. Synthesis of N-methylated cyclic peptides. *Nat. Protoc.* **7**, 432–444 (2012).

223. Texido, M., Albericio, F. & Giralt, E. Solid-phase synthesis and characterization of N -methyl-rich peptides. *J. Pept. Res.* **65**, 153–166 (2005).

224. Jaradat, D. M. M. Thirteen decades of peptide synthesis: key developments in solid phase peptide synthesis and amide bond formation utilized in peptide ligation. *Amino Acids* **50**, 39–68 (2018).

225. Kelly, C. N. *et al.* Geometrically Diverse Lariat Peptide Scaffolds Reveal an Untapped Chemical Space of High Membrane Permeability. *J. Am. Chem. Soc.* **143**, 705–714 (2021).

226. Passioura, T., Liu, W., Dunkelmann, D., Higuchi, T. & Suga, H. Display Selection of Exotic Macrocyclic Peptides Expressed under a Radically Reprogrammed 23 Amino Acid Genetic Code. *J. Am. Chem. Soc.* **140**, 11551–11555 (2018).

227. Biron, E., Chatterjee, J. & Kessler, H. Optimized selective N-methylation of peptides on solid support. *J. Pept. Sci.* **12**, 213–219 (2006).

228. Pennington, M. W. & Byrnes, M. E. Procedures to improve difficult couplings. in *Methods in Molecular Biology: Peptide Synthesis Protocols* (eds. Pennington, M. W. & Dunn, B. M.) vol. 35 1–16 (Humana Press, 1994).

229. Amblard, M., Fehrentz, J. A., Martinez, J. & Subra, G. Methods and protocols of modern solid-phase peptide synthesis. *Mol. Biotechnol.* **33**, 239–254 (2006).

230. Simelis, K. Development of Chemical Tools for Ten-Eleven Translocation Enzymes. (University of Oxford, 2022).

231. Haldón, E., Nicasio, M. C. & Pérez, P. J. Copper-catalysed azide-alkyne cycloadditions (CuAAC): An update. *Org. Biomol. Chem.* **13**, 9528–9550 (2015).

232. Presolski, S. I., Hong, V. P. & Finn, M. G. Copper-Catalyzed Azide–Alkyne Click Chemistry for Bioconjugation. *Curr. Protoc. Chem. Biol.* **3**, 153–162 (2011).

233. Ng, S. S. *et al.* Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448**, 87–91 (2007).

234. Couture, J. F., Collazo, E., Ortiz-Tello, P. A., Brunzelle, J. S. & Trievel, R. C. Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat. Struct. Mol. Biol.* **14**, 689–695 (2007).

235. King, O. N. F. *et al.* Quantitative high-throughput screening identifies 8- hydroxyquinolines as cell-active histone demethylase inhibitors. *PLoS One* **5**, (2010).

236. Kawamura, A. *et al.* Development of homogeneous luminescence assays for histone demethylase catalysis and binding. *Anal. Biochem.* **404**, 86–93 (2010).

237. Krishnan, S., Collazo, E., Ortiz-Tello, P. A. & Trievel, R. C. Purification and assay protocols for obtaining highly active Jumonji C demethylases. *Anal. Biochem.* **420**, 48–53 (2012).

238. Han, X., Aslanian, A. & Yates, J. R. Mass spectrometry for proteomics. *Curr. Opin. Chem.*

Biol. **12**, 483–490 (2008).

239. Zhang, F. *et al.* Application of 3D NMR for Structure Determination of Peptide Natural Products. *J. Org. Chem.* **80**, 8713–8719 (2015).

240. Spencer, R. K. & Nowick, J. S. A Newcomer's Guide to Peptide Crystallography. *Isr. J. Chem.* **55**, 698–710 (2015).

241. Conway, S. J. Bifunctional Molecules beyond PROTACs. *J. Med. Chem.* **63**, 2802–2806 (2020).

242. Boyd, S. R. *et al.* Design and Applications of Bifunctional Small Molecules in Biology. *Biochim. Biophys. Acta - Proteins Proteomics* **1869**, 140534 (2021).

243. Drewes, G. & Knapp, S. Chemoproteomics and Chemical Probes for Target Discovery. *Trends Biotechnol.* **36**, 1275–1286 (2018).

244. Mohammed, H. *et al.* Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat. Protoc.* **11**, 316–326 (2016).

245. Chevallet, M., Luche, S. & Rabilloud, T. Silver staining of proteins in polyacrylamide gels. *Nat. Protoc.* **1**, 1852–1858 (2006).

246. Anighoto, A., Bajorath, J. & Rastelli, G. Polypharmacology: Challenges and opportunities in drug discovery and development. *J. Med. Chem.* **57**, 7874−7887 (2014).

247. Salisbury, C. M. & Cravatt, B. F. Activity-based probes for proteomic profiling of histone deacetylase complexes. *Proc. Natl. Acad. Sci.* **104**, 1171–1176 (2007).

248. Salisbury, C. M. & Cravatt, B. F. Optimization of activity-based probes for proteomic profiling of histone deacetylase complexes. *J. Am. Chem. Soc.* **130**, 2184–2194 (2008).

249. Bantscheff, M. *et al.* Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat. Biotechnol.* **29**, 255–268 (2011).

250. Dittmann, A. *et al.* The commonly used PI3-kinase probe LY294002 is an inhibitor of BET bromodomains. *ACS Chem. Biol.* **9**, 495–502 (2014).

251. Li, X. *et al.* Quantitative chemical proteomics approach to identify post-translational modification-mediated protein-protein interactions. *J. Am. Chem. Soc.* **134**, 1982–1985 (2012).

252. Fan, J., Shu, Q., Li, Y. M. & Shi, J. Efficient synthesis of terminal-diazirine-based histone peptide probes. *Tetrahedron Lett.* **102**, 153878 (2022).

253. Fu, A., Tang, R., Hardie, J., Farkas, M. E. & Rotello, V. M. Promises and Pitfalls of Intracellular Delivery of Proteins. *Bioconjug. Chem.* **25**, 1602–1608 (2014).

254. Richards, A. L., Eckhardt, M. & Krogan, N. J. Mass spectrometry-based protein–protein interaction networks for the study of human diseases. *Mol. Syst. Biol.* **17**, e8792 (2021).

255. Zhang, X. *et al.* Epitope tagging of endogenous proteins for genome-wide ChIP-chip studies. *Nat. Methods* **5**, 163–165 (2008).

256. Hopp, T. P. *et al.* A short polypeptide marker sequence useful for recombinant protein identification and purification. *Nat. Biotechnol.* **6**, 1204–1210 (1988).

257. Gerace, E. & Moazed, D. Affinity Pull-Down of Proteins Using Anti-FLAG M2 Agarose Beads. in *Methods in Enzymology* vol. 559 99–110 (Elsevier Inc., 2015).

258. Harvey, M. C. *et al. Affinity-Purified ANTI-FLAG M2 Antibody with High Specificity and Sensitivity*. *Merck* (2005).

259. Hancock, R. Investigating the effect of hypoxia on the JmjC histone lysine demethylase KDM4A. (University of Oxford, 2017).

260. Rogers, J. C. & Bomgarden, R. D. Sample Preparation for Mass Spectrometry-Based Proteomics; from Proteomes to Peptides. in *Modern Proteomics - Sample Preparation, Analysis and Practical Applications* (eds. Mirzaei, H. & Carrasco, M.) vol. 919 345–382 (Springer Nature, 2016).

261. Wei, C. T., Maly, D. J. & Fowler, D. M. *Temporal and rheostatic control of genome editing with a chemically-inducible Cas9*. *Methods in Enzymology* vol. 633 (Elsevier Inc., 2020).

262. Han, W. *et al.* Cadherin-dependent adhesion modulated 3D cell-assembly. *J. Mater. Chem. B* **10**, 4959–4966 (2022).

263. Smith, S. L. & Shioda, T. Advantages of COS-1 monkey kidney epithelial cells as packaging host for small-volume production of high-quality recombinant lentiviruses. *J. Virol. Methods* **157**, 47–54 (2009).

264. Proteinatlas.org. https://www.proteinatlas.org/ENSG00000066135-KDM4A/single+cell+type. *The Human Protein Atlas*.

265. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).

266. Lipman, N. S., Jackson, L. R., Trudel, L. J. & Weis-Garcia, F. Monoclonal versus polyclonal antibodies: Distinguishing characteristics, applications, and information resources. *ILAR J.* **46**, 258–267 (2005).

267. Lapidot, M. *et al.* Essential role of the histone lysine demethylase KDM4A in the biology of malignant pleural mesothelioma (MPM). *Br. J. Cancer* **125**, 582–592 (2021).

268. Hoyle, R. G., Wang, H., Cen, Y., Zhang, Y. & Li, J. IOX1 suppresses wnt target gene transcription and colorectal cancer tumorigenesis through inhibition of KDM3 histone demethylases. *Mol. Cancer Ther.* **20**, 191–202 (2021).

269. Jie, X. *et al.* USP9X-mediated KDM4C deubiquitination promotes lung cancer radioresistance by epigenetically inducing TGF-β2 transcription. *Cell Death Differ.* **28**, 2095–2111 (2021).

270. Zhang, Y. *et al.* An interaction between BRG1 and histone modifying enzymes mediates lipopolysaccharide-induced proinflammatory cytokines in vascular endothelial cells. *J. Cell. Biochem.* **120**, 13216–13225 (2019).

271. Verrier, L., Escaffit, F., Chailleux, C., Trouche, D. & Vandromme, M. A new isoform of the histone demethylase JMJD2A/KDM4A is required for skeletal muscle differentiation. *PLoS Genet.* **7**, e1001390 (2011).

272. van Rechem, C. *et al.* Lysine demethylase KDM4A associates with translation machinery and regulates protein synthesis. *Cancer Discov.* **5**, 255–263 (2015).

273. Dobrynin, G. *et al.* KDM4A regulates HIF-1 levels through H3K9me3. *Sci. Rep.* **7**, 11094 (2017).

274. Shin, S. & Janknecht, R. Diversity within the JMJD2 histone demethylase family. *Biochem. Biophys. Res. Commun.* **353**, 973–977 (2007).

275. Su, Z. *et al.* Reader domain specificity and lysine demethylase-4 family function. *Nat. Commun.* **7**, 13387 (2016).

276. Ohguchi, H. *et al.* The KDM3A-KLF2-IRF4 axis maintains myeloma cell survival. *Nat. Commun.* **7**, 10258 (2016).

277. Sefton, B. M. Labeling Cultured Cells with 32Pi and Preparing Cell Lysates for Immunoprecipitation. *Curr. Protoc. Cell Biol.* **3**, (1999).

278. Mallick, P. & Kuster, B. Proteomics: A pragmatic perspective. *Nat. Biotechnol.* **28**, 695–709 (2010).

279. Catherman, A. D., Skinner, O. S. & Kelleher, N. L. Top Down proteomics: Facts and perspectives. *Biochem. Biophys. Res. Commun.* **445**, 683–693 (2014).

280. Bekker-Jensen, D. B. *et al.* An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst.* **4**, 587–599 (2017).

281. Loo, R. R., Dales, N. & Andrews, P. C. The effect of detergents on proteins analyzed by electrospray ionization. in *Methods in molecular biology: Protein and Peptide Analysis by Mass Spectrometry* (ed. Chapman, J. R.) vol. 61 141–160 (Humana Press, 1996).

282. Rappsilber, J., Ishihama, Y. & Mann, M. Stop And Go Extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).

283. Szklarczyk, D. *et al.* The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).

284. Lill, J. R., Mathews, W. R., Rose, C. M. & Schirle, M. Proteomics in the pharmaceutical and biotechnology industry: a look to the next decade. *Expert Rev. Proteomics* **18**, 503–526 (2021).

285. Das, A. T., Zhou, X., Metz, S. W., Vink, M. A. & Berkhout, B. Selecting the optimal Tet-On system for doxycycline-inducible gene expression in transiently transfected and stably transduced mammalian cells. *Biotechnol. J.* **11**, 71–79 (2016).

286. Alzahayqa, M., Jamous, A., Khatib, A. A. H. & Salah, Z. TET1 Isoforms Have Distinct Expression Pattern, Localization and Regulation in Breast Cancer. *Front. Oncol.* **12**, 848544 (2022).

287. Greer, C. B. *et al.* Tet1 isoforms differentially regulate gene expression, synaptic transmission, and memory in the mammalian brain. *J. Neurosci.* **41**, 578–593 (2021).

288. MacKinnon, A. L., Garrison, J. L., Hegde, R. S. & Taunton, J. Photo-leucine incorporation reveals the target of a cyclodepsipeptide inhibitor of cotranslational translocation. *J. Am. Chem. Soc.* **129**, 14560–14561 (2007).

289. Smith, E. & Collins, I. Photoaffinity labeling in target-and binding-site identification. *Future Med. Chem.* **7**, 159–183 (2015).

290. Fallon, D. J. *et al.* One-Step Synthesis of Photoaffinity Probes for Live-Cell MS-Based Proteomics. *Chem. - A Eur. J.* **27**, 17880–17888 (2021).

291. Preston, G. W. & Wilson, A. J. Photo-induced covalent cross-linking for the analysis of biomolecular interactions. *Chem. Soc. Rev.* **42**, 3289 (2013).

292. Sumranjit, J. & Chung, S. J. Recent advances in target characterization and identification by photoaffinity probes. *Molecules* **18**, 10426 (2013).

293. Passioura, T. & Suga, H. A RaPID way to discover nonstandard macrocyclic peptide modulators of drug targets. *Chem. Commun.* **53**, 1931–1940 (2017).

294. Chowdhury, R. *et al.* Use of cyclic peptides to induce crystallization: case study with prolyl hydroxylase domain 2. *Sci. Rep.* **10**, 21964 (2020).

295. Mullard, A. Targeted protein degraders crowd into the clinic. *Nat. Rev. Drug Discov.* **20**, 247–250 (2021).

296. Cromm, P. M., Crews, C. M. & Weinmann, H. PROTAC-mediated Target Degradation: A Paradigm Changer in Drug Discovery? in *Protein Degradation with New Chemical Modalities: Successful Strategies in Drug Discovery and Chemical Biology* (eds. Weinmann, H. & Crews, C. M.) 1–13 (Royal Society of Chemistry, 2020). doi:10.1039/9781839160691-00001.

297. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479 (1998).

298. Buckley, D. L. *et al.* Small-molecule inhibitors of the interaction between the E3 ligase VHL and HIF1α. *Angew. Chemie - Int. Ed.* **51**, 11463–11467 (2012).

299. Buckley, D. L. *et al.* Targeting the von Hippel-Lindau E3 ubiquitin ligase using small molecules to disrupt the VHL/HIF-1α interaction. *J. Am. Chem. Soc.* **134**, 4465–4468 (2012).

300. Ito, T. *et al.* Identification of a primary target of thalidomide teratogenicity. *Science* **327**, 1345–1350 (2010).

301. Burslem, G. M. & Crews, C. M. Proteolysis-Targeting Chimeras as Therapeutics and Tools for Biological Discovery. *Cell* **181**, 102–114 (2020).

302. Gopalsamy, A. Selectivity through Targeted Protein Degradation (TPD). *J. Med. Chem.* **65**, 8113–8126 (2022).

303. Schneekloth, J. S. *et al.* Chemical Genetic Control of Protein Levels: Selective in Vivo Targeted Degradation. *J. Am. Chem. Soc.* **126**, 3748–3754 (2004).

304. Montrose, K. & Krissansen, G. W. Design of a PROTAC that antagonizes and destroys the cancer-forming X-protein of the hepatitis B virus. *Biochem. Biophys. Res. Commun.* **453**, 735–740 (2014).

305. Ma, S. *et al.* Non-small molecule PROTACs (NSM-PROTACs): Protein degradation kaleidoscope. *Acta Pharm. Sin. B* **12**, 2990–3005 (2022).

306. Liao, H. *et al.* A PROTAC peptide induces durable β-catenin degradation and suppresses Wnt-dependent intestinal cancer. *Cell Discov.* **6**, 35 (2020).

307. Ma, B. *et al.* De Novo Design of an Androgen Receptor DNA Binding Domain-Targeted peptide PROTAC for Prostate Cancer Therapy. *Adv. Sci.* 2201859 (2022) doi:10.1002/advs.202201859.

308. Röth, S. *et al.* Targeting Endogenous K-RAS for Degradation through the Affinity-Directed Protein Missile System. *Cell Chem. Biol.* **27**, 1151–1163 (2020).

309. Vogelmann, A., Robaa, D., Sippl, W. & Jung, M. Proteolysis targeting chimeras (PROTACs) for epigenetics research. *Curr. Opin. Chem. Biol.* **57**, 8–16 (2020).

310. Webb, T., Craigon, C. & Ciulli, A. Targeting epigenetic modulators using PROTAC degraders: Current status and future perspective. *Bioorganic Med. Chem. Lett.* **63**, 128653 (2022).

311. Lu, J. *et al.* Hijacking the E3 Ubiquitin Ligase Cereblon to Efficiently Target BRD4. *Chem. Biol.* **22**, 755–763 (2015).

312. Zengerle, M., Chan, K. H. & Ciulli, A. Selective Small Molecule Induced Degradation of the BET Bromodomain Protein BRD4. *ACS Chem. Biol.* **10**, 1770–1777 (2015).

313. Ohoka, N. *et al.* In vivo knockdown of pathogenic proteins via specific and nongenetic Inhibitor of Apoptosis Protein (IAP)-dependent protein erasers (SNIPERs). *J. Biol. Chem.* **292**, 4556–4570 (2017).

314. Testa, A., Hughes, S. J., Lucas, X., Wright, J. E. & Cuilli, A. Structure-Based Design of a Macrocyclic PROTAC. *Angew. Chemie Int. Ed.* **59**, 1727–1734 (2020).

315. Shen, Y. *et al.* Discovery of First-in-Class Protein Arginine Methyltransferase 5 (PRMT5) Degraders. *J. Med. Chem.* **63**, 9977–9989 (2020).

316. Xiong, Y. *et al.* Chemo-proteomics exploration of HDAC degradability by small molecule degraders Chemo-proteomics exploration of HDAC degradability by small molecule degraders. *Cell Chem. Biol.* **28**, 1–14 (2021).

317. Roatsch, M., Vogelmann, A., Herp, D., Jung, M. & Olsen, C. A. Proteolysis-Targeting Chimeras (PROTACs) Based on Macrocyclic Tetrapeptides Selectively Degrade Class I Histone Deacetylases 1–3. *ChemRxiv* (2020).

318. Schiedel, M. *et al.* Chemically Induced Degradation of Sirtuin 2 (Sirt2) by a Proteolysis Targeting Chimera (PROTAC) Based on Sirtuin Rearranging Ligands (SirReals). *J. Med. Chem.* **61**, 482–491 (2018).

319. Jin, J. *et al.* The peptide PROTAC modality: A novel strategy for targeted protein ubiquitination. *Theranostics* **10**, 10141–10153 (2020).

320. Blommel, P. G., Martin, P. A., Seder, K. D., Wrobel, R. L. & Fox, B. G. Flexi Vector Cloning. in *High Throughput Protein Expression and Purification. Methods in Molecular Biology* (ed. Doyle, Sharon A. (GlaxoSmithKline, Walbut Creek CA, U.) vol. 498 55–73 (Humana Press, 2009).

321. Roper, G. Development of Cyclic Peptide Probes Selective for Histone Demethylases KDM4A-C. (University of Oxford, 2017).

322. Fus-Kujawa, A. *et al.* An Overview of Methods and Tools for Transfection of Eukaryotic Cells in vitro. *Front. Bioeng. Biotechnol.* **9**, 701031 (2021).

323. Lavie, J. *et al.* Ubiquitin-Dependent Degradation of Mitochondrial Proteins Regulates Energy Metabolism. *Cell Rep.* **23**, 2852–2863 (2018).

324. Soucy, T. A. *et al.* An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* **458**, 732–736 (2009).

325. Girardini, M., Maniaci, C., Hughes, S. J., Testa, A. & Ciulli, A. Cereblon versus VHL: Hijacking E3 ligases against each other using PROTACs. *Bioorganic Med. Chem.* **27**, 2466–2479 (2019).

326. Collins, I., Wang, H., Caldwell, J. J. & Chopra, R. Chemical approaches to targeted protein degradation through modulation of the ubiquitin–proteasome pathway. *Biochem. J.* **474**, 1127–1147 (2017).

327. Frost, J. *et al.* Potent and selective chemical probe of hypoxic signalling downstream of HIF-1alpha hydroxylation via VHL inhibition. *Nat. Commun.* **7**, 13312 (2016).

328. Galdeano, C. *et al.* Structure-Guided Design and Optimization of Small Molecules Targeting the Protein−Protein Interaction between the von Hippel−Lindau (VHL) E3 Ubiquitin Ligase and the Hypoxia Inducible Factor (HIF) Alpha Subunit with in Vitro Nanomolar Affinities. *J. Med. Chem.* **57**, 8657–8663 (2014).

329. Johmura, Y. *et al.* SCF Fbxo22-KDM4A targets methylated p53 for degradation and regulates senescence. *Nat. Commun.* **7**, 10574 (2016).

330. Zhang, B. *et al.* Targeting KDM4A-AS1 represses AR/AR-Vs deubiquitination and enhances enzalutamide response in CRPC. *Oncogene* **41**, 387–399 (2022).

331. Cui, S. Z. *et al.* Targeting USP1-dependent KDM4A protein stability as a potential prostate cancer therapy. *Cancer Sci.* **111**, 1567–1581 (2020).

332. Zagidullin, A., Milyukov, V., Rizvanov, A. & Bulatov, E. Novel approaches for the rational design of PROTAC linkers. *Explor. Target. Anti-tumor Ther.* **1**, 381–390 (2020).

333. Naylor, M. R. *et al.* Lipophilic Permeability Efficiency Reconciles the Opposing Roles of Lipophilicity in Membrane Permeability and Aqueous Solubility. *J. Med. Chem.* **61**, 11169–11182 (2018).

334. Roy, M. J. *et al.* SPR-Measured Dissociation Kinetics of PROTAC Ternary Complexes In fluence Target Degradation Rate. *ACS Chem. Biol.* **14**, 361–368 (2019).

335. Verheijen, M. *et al.* DMSO induces drastic changes in human cellular processes and epigenetic landscape in vitro. *Sci. Rep.* **9**, 4641 (2019).

336. Hammoudeh, S. M., Hammoudeh, A. M. & Hamoudi, R. High-throughput quantification of the effect of DMSO on the viability of lung and breast cancer cells using an easy-to-use spectrophotometric trypan blue-based assay. *Histochem. Cell Biol.* **152**, 75–84 (2019).

337. Berg, J. M., Tymoczko, J. L., Gatto Jr, G. J. & Stryer, L. Protein tunrover and amino acid metabolism. in *Biochemistry* 681–712 (W. H. Freeman, 2015).

338. Walport, L. J., Obexer, R. & Suga, H. Strategies for transitioning macrocyclic peptides to cell-permeable drug leads. *Curr. Opin. Biotechnol.* **48**, 242–250 (2017).

339. Dougherty, P. G., Sahni, A. & Pei, D. Understanding Cell Penetration of Cyclic Peptides. *Chem. Rev.* **119**, 10241–10287 (2019).

340. Klein, V. G. *et al.* Understanding and Improving the Membrane Permeability of VH032-Based PROTACs. *ACS Med. Chem. Lett.* **11**, 1732–1738 (2020).

341. Riching, K. M. *et al.* Quantitative Live-Cell Kinetic Degradation and Mechanistic Profiling of PROTAC Mode of Action. *ACS Chem. Biol.* **13**, 2758–2770 (2018).

342. Zoghbi, H. Y. & Beaudet, A. L. Epigenetics and human disease. *Cold Spring Harb. Perspect. Biol.* **8**, a019497 (2016).

343. Wang, Y. *et al.* Targeting cancer epigenetic pathways with small-molecule compounds: Therapeutic efficacy and combination therapies. *Pharmacol. Res.* **173**, 105702 (2021).

344. Frenzel, D. & Willbold, D. Kinetic titration series with biolayer interferometry. *PLoS One* **9**, e106882 (2014).

345. Capelli, D. *et al.* Surface Plasmon Resonance as a Tool for Ligand Binding Investigation of

Engineered GPR17 Receptor, a G Protein Coupled Receptor Involved in Myelination. *Front. Chem.* **7**, 910 (2020).

346. Hancock, W. S. & Battersby, J. E. A New Micro-Test for the Detection of Incomplete Coupling Reactions in Solid-Phase Peptide sulphonic Acid. *Anal. Biochem.* **264**, 260–264 (1976).

347. Chan, W. C. & White, P. D. Basic procedures. in *Fmoc Solid Phase Peptide Synthesis: A Practical Approach* (eds. Chan, W. C. & White, P. W.) 41–76 (Oxford University Press, 1999).

348. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server Elisabeth. in *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 555–570 (Humana Press, 2005). doi:10.1385/1-59259-890-0:067.

349. Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-based protein profiling in vivo using a copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 4686–4687 (2003).

350. Chung, Y. G. *et al.* Histone Demethylase Expression Enhances Human Somatic Cell Nuclear Transfer Efficiency and Promotes Derivation of Pluripotent Stem Cells. *Cell Stem Cell* **17**, 758–766 (2015).

351. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* **25**, 402–408 (2001).

Appendix 1 Peptide and peptide conjugate characterisation

Appendix 1.1 Peptide and peptide conjugate characterisation data

Probe	Parent Peptide	Chemical Formula	Exact Mass	Expected m/z	Observed MS	Purity (Analytical HPLC)
B-1	**CP2**	$C_{111}H_{154}N_{30}O_{27}S_2$	2403.10411	$[M+2H]^{2+}$ 1202.55933 $[M+3H]^{3+}$ 802.04198	1202.5387 802.0284	93.2%
B-2	**CP2-R6A**	$C_{108}H_{147}N_{27}O_{27}S_2$	2318.04011	$[M+2H]^{2+}$ 1160.02733 $[M+3H]^{3+}$ 773.68731	1160.0091 773.6753	96.8%
B-3	**TiP1**	$C_{114}H_{153}N_{25}O_{28}S_2$	2384.08	$[M+2H]^{2+}$ 1193.04519	1193.0360	92.6%
B-4	**TiP1**	$C_{120}H_{157}N_{25}O_{28}S_2$	2460.10	$[M+2H]^{2+}$ 1231.06084	1231.0507	98.9%
E3-1	**CP2**	$C_{117}H_{153}N_{31}O_{25}S_2$	2456.10953	$[M+2H]^{2+}$ 1229.06204 $[M+3H]^{3+}$ 819.71045	1129.0445 819.6992	99.9%
E3-2	**CP2**	$C_{119}H_{157}N_{31}O_{25}S_2$	2484.14083	$[M+2H]^{2+}$ 1243.07769 $[M+3H]^{3+}$ 829.05422	1243.0585 829.0424	99.0%
E3-3	**CP2**	$C_{125}H_{169}N_{31}O_{25}S_2$	2568.23473	$[M+2H]^{2+}$ 1285.12464 $[M+3H]^{3+}$ 857.08552	1285.1049 857.0732	98.7%
E3-4	**CP2**	$C_{123}H_{155}N_{31}O_{26}S_2$	2546.12010	$[M+2H]^{2+}$ 1274.06732 $[M+3H]^{3+}$ 849.71398	1274.0490 849.7026	98.7%
E3-5	**CP2**	$C_{126}H_{161}N_{31}O_{26}S_2$	2588.16705	$[M+2H]^{2+}$ 1295.09080 $[M+3H]^{3+}$ 863.72963	1295.0740 863.7189	98.3%
E3-6	**CP2**	$C_{132}H_{173}N_{31}O_{26}S_2$	2672.26095	$[M+2H]^{2+}$ 1337.13775 $[M+3H]^{3+}$ 891.76093	1337.1182 891.7484	97.8%
E3-7	**CP2**	$C_{106}H_{131}N_{29}O_{26}S$	2257.95408	$[M+2H]^{2+}$ 1129.98431 $[M+3H]^{3+}$ 753.65864	1129.9677 753.6488	96.8%
E3-8	**CP2**	$C_{109}H_{137}N_{29}O_{26}S$	2300.00103	$[M+2H]^{2+}$ 1151.00779 $[M+3H]^{3+}$ 767.67429	1150.9916 767.6642	96.8%
E3-9	**CP2**	$C_{115}H_{149}N_{29}O_{26}S$	2384.09493	$[M+2H]^{2+}$ 1193.05474 $[M+3H]^{3+}$ 795.70559	1193.0365 795.6943	99.4%
E3-10	**CP2**	$C_{121}H_{161}N_{31}O_{28}S_2$	2560.15688	$[M+2H]^{2+}$ 1281.08571 $[M+3H]^{3+}$ 854.39290	1281.0663 854.3808	97.8%
E3-11	**CP2.3**	$C_{124}H_{166}FN_{31}O_{27}S_2$	2604.19949	$[M+2H]^{2+}$ 1303.10702 $[M+3H]^{3+}$ 869.07377	1303.0884 869.0626	76.8%
E3-12	**CP2-R6A**	$C_{118}H_{154}N_{28}O_{28}S_2$	2475.09288	$[M+2H]^{2+}$ 1238.55372 $[M+3H]^{3+}$ 826.03824	1238.5351 826.0268	98.6%
CP2		$C_{88}H_{116}N_{24}O_{22}S$	1892.84167	$[M+2H]^{2+}$ 947.42811 $[M+H]^{+}$ 1893.84895	947.4246 1893.8517	92.8%
CP2.3		$C_{92}H_{121}FN_{24}O_{20}S$	1932.88937	$[M+2H]^{2+}$ 967.45196 $[M+3H]^{3+}$ 645.30373	967.4386 645.2960	97.9%
TiP1		$C_{96}H_{117}N_{19}O_{23}S$	1935.83	$[M+2H]^{2+}$ 968.92180	968.9150	95.2%

Table A1.1 Characterisation data for peptide conjugates, and parent peptides. TiP1, **B-3**, and **B-4** were synthesised by Valeria Cobiltean under my supervision.

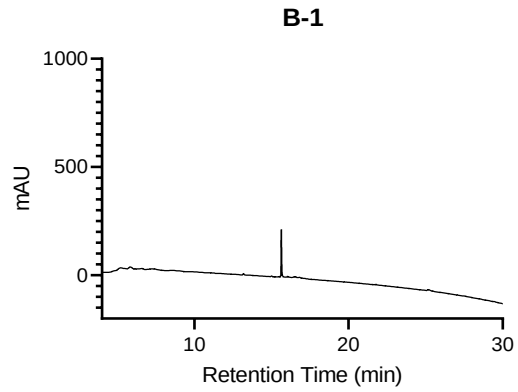
Probe	Parent Peptide	Variations to parent	Chemical Formula	Exact Mass	Expected m/z	Observed MS
Ph-1	**CP2**	R6(Bpa)	$C_{98}H_{118}O_{23}N_{21}S$	1987.84	$[M+H]^+$ 1988.84247	1988.8488
Ph-2	**CP2**	Y12(Bpa)	$C_{95}H_{121}O_{22}N_{24}S$	1980.87	$[M+H]^+$ 1981.88025	1981.8839
Ph-3	**CP2**	R6(Bpa) T13(Pra)	$C_{99}H_{116}O_{22}N_{21}S$	1981.82	$[M+H]^+$ 1982.83190	1982.8358
Ph-4	**CP2**	W9(Bpa) T13(Pra)	$C_{94}H_{118}O_{22}N_{23}S$	1952.85	$[M]^+$ 1952. 85370	1952.8575
Ph-5	**CP2**	W11(Bpa) T13(Pra)	$C_{94}H_{118}O_{22}N_{23}S$	1952.85	$[M]^+$ 1952. 85370	1952.8577
Ph-6	**TiP1**	Y6(Bpa) S13K_{N3}	$C_{106}H_{126}N_{22}O_{22}S$	2090.91	$[M+3H]^{2+}$ 1046.96025	1046.9622
Ph-7	**TiP1**	Y6L_p S13K_{N3}	$C_{95}H_{120}N_{24}O_{21}S$	1964.88	$[M+2H]^{2+}$ 983.44631	983.4381
Ph-8	**TiP1**	S13K_{N3} *C*-term L_p	$C_{104}H_{129}N_{25}O_{23}S$	2127.94	$[M+2H]^{2+}$ 1064.97797	1064.9691
Ph-9	**TiP1**	S13K_{N3} *C*-term K_p	$C_{110}H_{140}N_{26}O_{25}S$	2257.02	$[M+2H]^{2+}$ 1129.51746	1129.5161

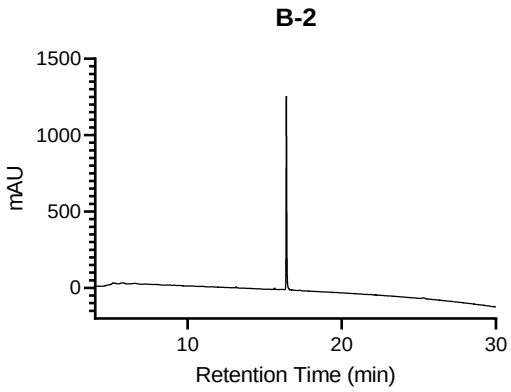
Table A1.2 Characterisation data for photoreactive peptides. Ph-3 – **Ph-5** were synthesised by Dr Jacob Bush; **Ph-6** – **Ph-9** were synthesised by Valeria Cobiltean under my supervision. Bpa = benzophenone alanine, Pra = propargylglycine, K_{N3} = azido-lysine, L_p = photo-leucine, K_p = photo-lysine (**Table 2.5**)

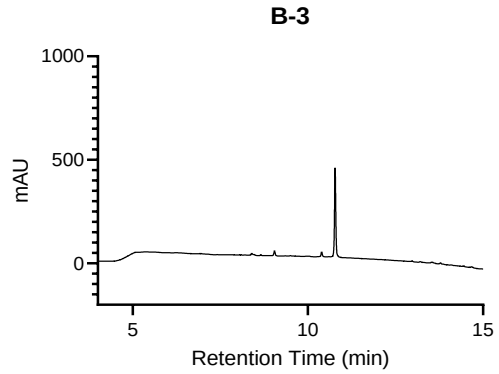
Peptide	Sequence	Synthesis scale	Crude linear peptide yield	Purified cyclic peptide yield	Total yield
CP2	^{D}YVYNTRSGWRWYTC	0.1 mmol	97.2%	5.5%	5.3%
CP2-T13K$_{N3}$	^{D}YVYNTRSGWRWY K_{N3}C	0.1 mmol	96.6%	8.0%	7.7%
CP2-R6A-T13K$_{N3}$	^{D}YVYNTASGWRWY K_{N3}C	0.1 mmol	99.1%	10.8%	10.7%
CP2.3-P13K$_{N3}$	$^{D}Y^{Me}$VYNTRS^{D}AWRWF$_{4F}$$K_{N3}$^{Me}C	0.05 mmol	95.4%	11.2%	10.7%
Ph-1	^{D}YVYNT(Bpa)SGWRWYTC	0.05 mmol	99.4%	8.1%	8.1%
Ph-2	^{D}YVYNTRSGWRW(Bpa)TC	0.05 mmol	99.2%	1.7%	1.7%
TiP1	^{D}YLIYAYYTWWEHSC	0.1 mmol	92.9%	3.7%	3.4%
TiP1-Y6K$_{N3}$	^{D}YLIYAK_{N3}YTWWEHSC	0.1 mmol	94.9%	2.6%	2.5%
TiP1-H12K$_{N3}$	^{D}YLIYAYYTWWEK_{N3}SC	0.1 mmol	87.4%	0.0%	0.0%
TiP1-S13K$_{N3}$	^{D}YLIYAYYTWWEHK_{N3}C	0.1 mmol	85.8%	11.6%	10.0%
Ph-6	^{D}YLIYA(Bpa)YTWWEHK_{N3}C	0.05 mmol	98.9%	1.1%	1.1%
Ph-7	^{D}YLIYAL_pYTWWEHK_{N3}C	0.05 mmol	54.9%	9.8%	5.4%
Ph-8	^{D}YLIYAYYTWWEHK_{N3}C-L_p	0.05 mmol	64.5%	1.5%	1.0%
Ph-9	^{D}YLIYAYYTWWEHK_{N3}C-K_p	0.05 mmol	56.5%	1.6%	0.9%

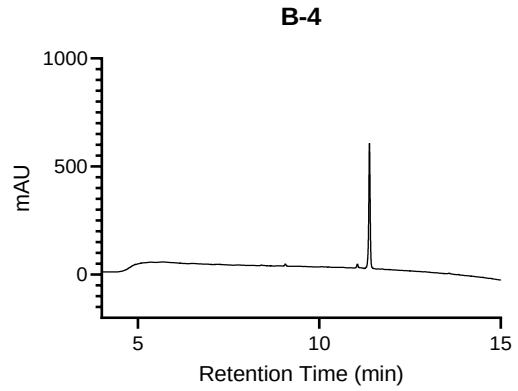
Table A1.3 Yield of peptides synthesised. Peptides were synthesised, cyclised, and purified as described in Section 6.2. **TiP1**-based peptides and **Ph-6** – **Ph-9** were synthesised by Valeria Cobiltean under my supervision. Bpa = benzophenone alanine, K_{N3} = azido-lysine, L_p = photo-leucine, K_p = photo-lysine.

Appendix 1.2 Peptide conjugates analytical HPLC data

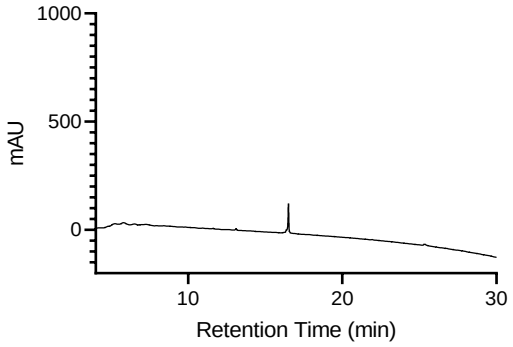




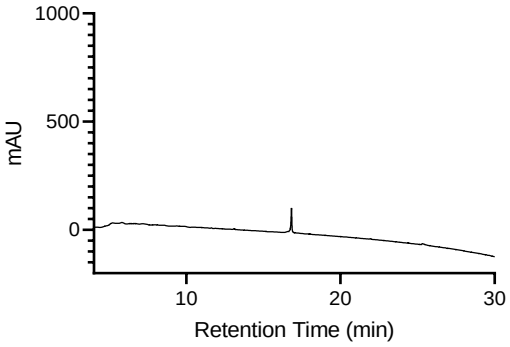




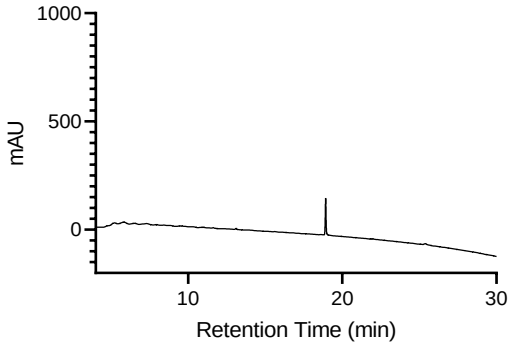
E3-1



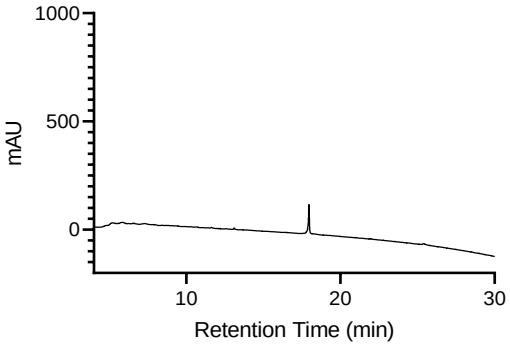
E3-2



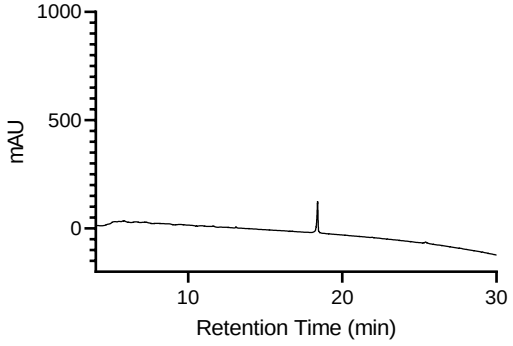
E3-3



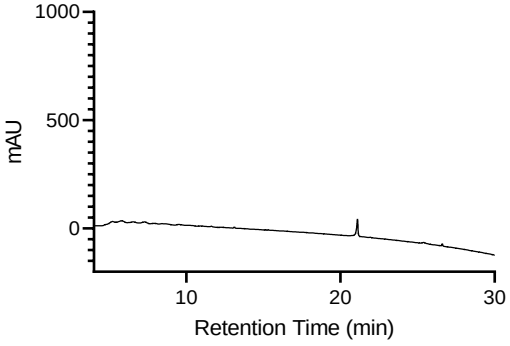
E3-4



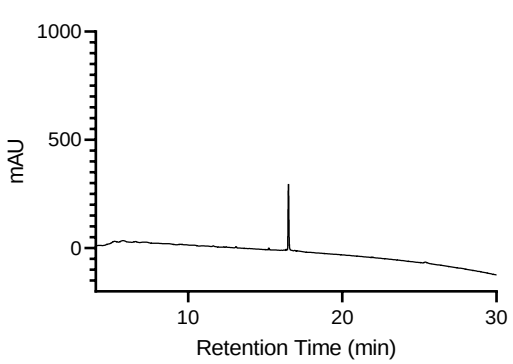
E3-5



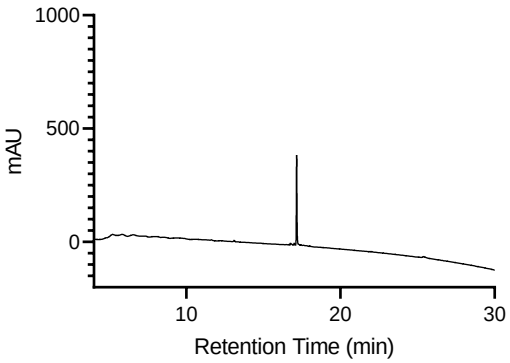
E3-6

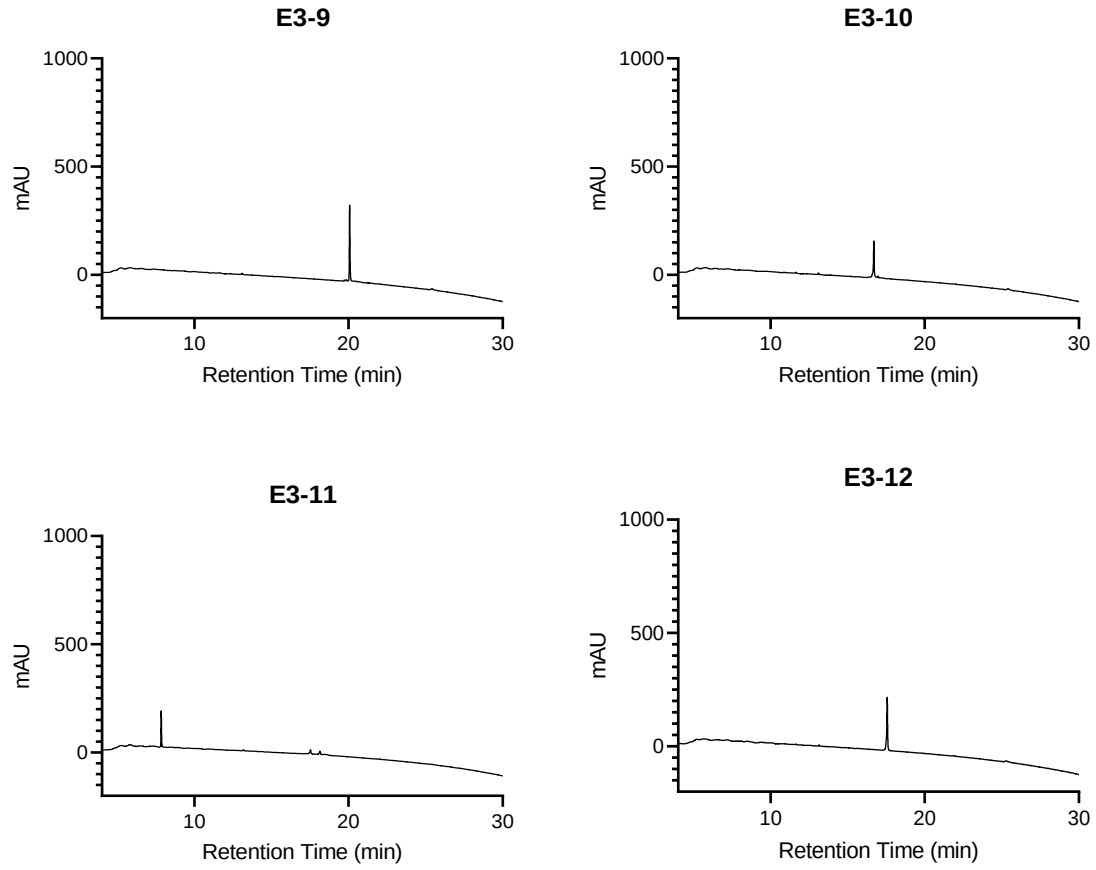


E3-7



E3-8





Appendix 1.3 Example HRMS data

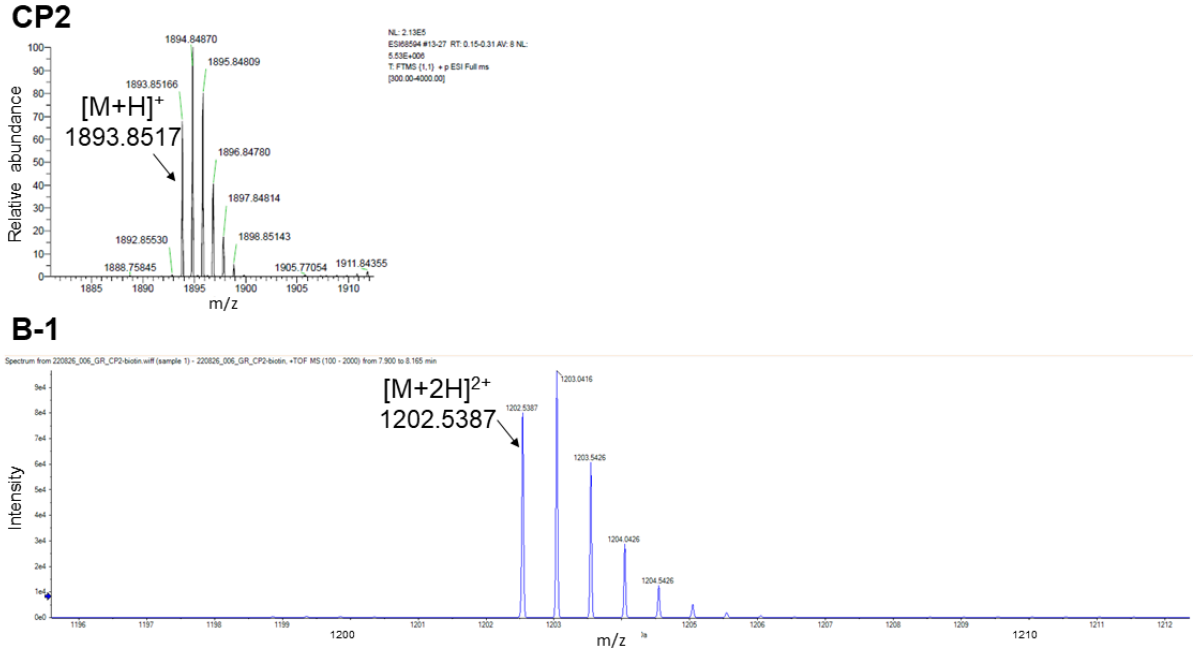


Figure A1.1 Example HRMS spectra for CP2 and B-1. CP2 was measured using the University of Oxford, Chemistry Department accurate mass service, and **B-1** was measured by myself with support from NUPPA (Section 6.2.5).

Appendix 2 FDH coupled assay linear range

An FDH coupled assay was used to determine the inhibitory activity of **CP2** and **B-1** (Section 2.2.5.2) across the linear range of KDM4A activity, using both H3K9Me$_3$(1–15) and H3K36Me$_3$(28–42) substrates. The linear range taken was 15 minutes for H3K9Me$_3$(1–15), and 30 minutes for H3K36Me$_3$(28–42) (**Figure A2.1**).

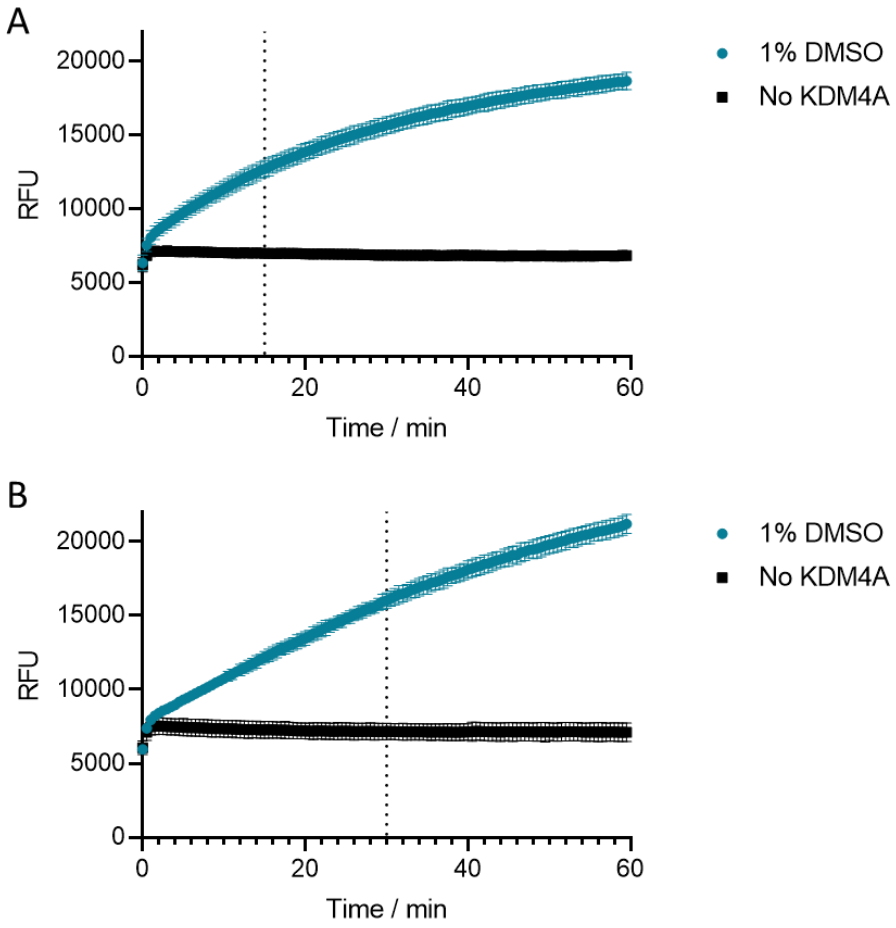


Figure A2.1 KDM4A activity with (A) H3K9Me$_3$(1–15) and (B) H3K36Me$_3$(28–42) substrates. For H3K9Me$_3$ the linear range was taken over 15 minutes and for H3K36Me$_3$ the linear range was taken over 30 minutes, time points indicated with a dotted line.

Appendix 3 Additional cell lysate blots

Appendix 3.1 Nuclear and remaining cell lysate western blot

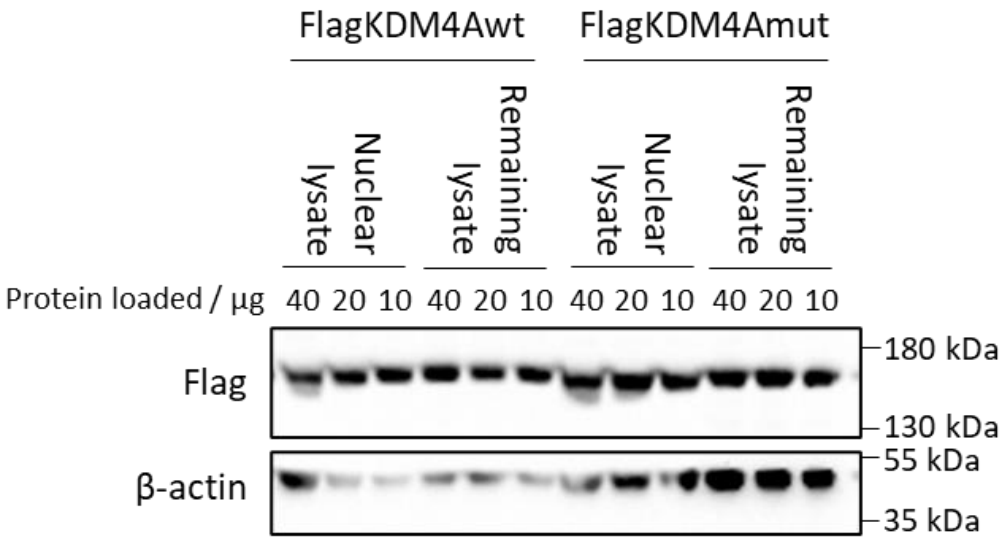


Figure A3.1 Western blot analysis of HEK293T nuclear lysate and remaining lysate using a Nuclear Extraction Kit (Active Motif). For both wildtype (wt) and mutant (mut) plasmids FlagKDM4A was found in both the nuclear lysate and remaining lysate. FlagKDM4A = 122 kDa, β-actin = 42 kDa.

Appendix 3.2 Endogenous and overexpressed KDM4 western blot

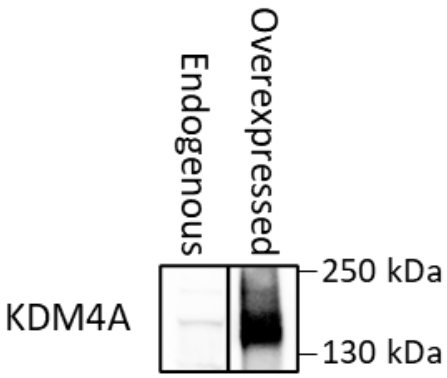


Figure A3.2 Western blot analysis of HEK293T cell lysates containing endogenous KDM4A and overexpressed FlagKDM4A. Overexpressed lysates were produced using optimised cell lysis described in Section 3.2.1.1.1.

Appendix 4 KDM4A protein stability in LC-MS injector after irradiation

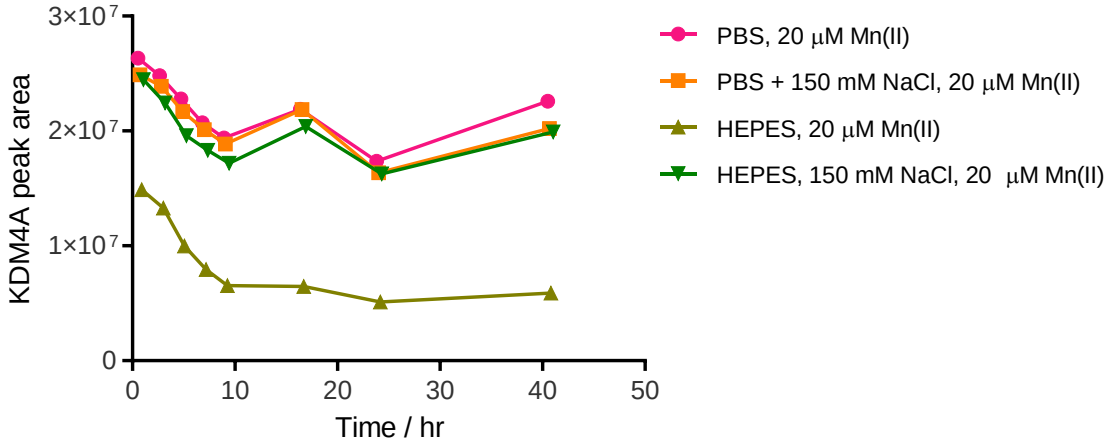


Figure A4.1 Intact protein LC-MS KDM4A peak area in different buffers over ≈40 hours in the LC-MS injector (10 °C). KDM4A (1 μM) was irradiated at 302 nm for 10 minutes at 4 °C prior to LC-MS, buffer used in remaining KDM4A photocrosslinking experiments was PBS with 20 μM Mn(II).

Appendix 5 Uncropped western blots from Chapter 3

Appendix 5.1 Figure 3.4

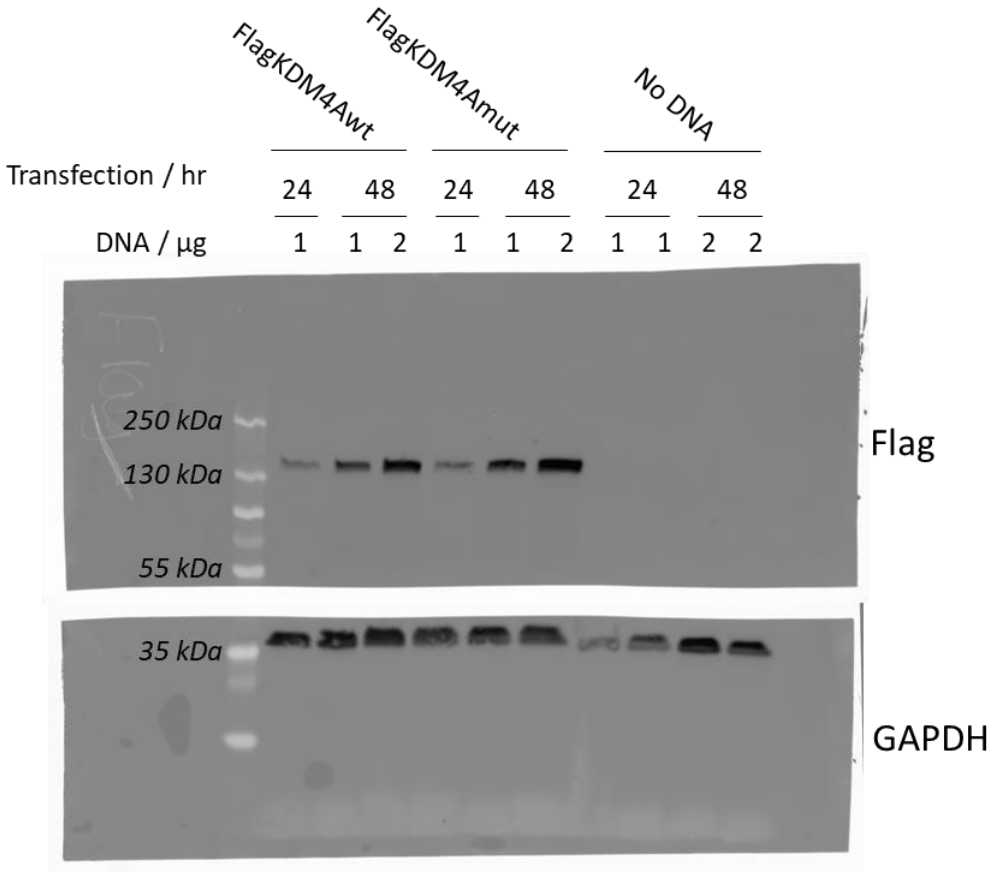


Figure A5.1 Uncropped western blot from Figure 3.4(A)

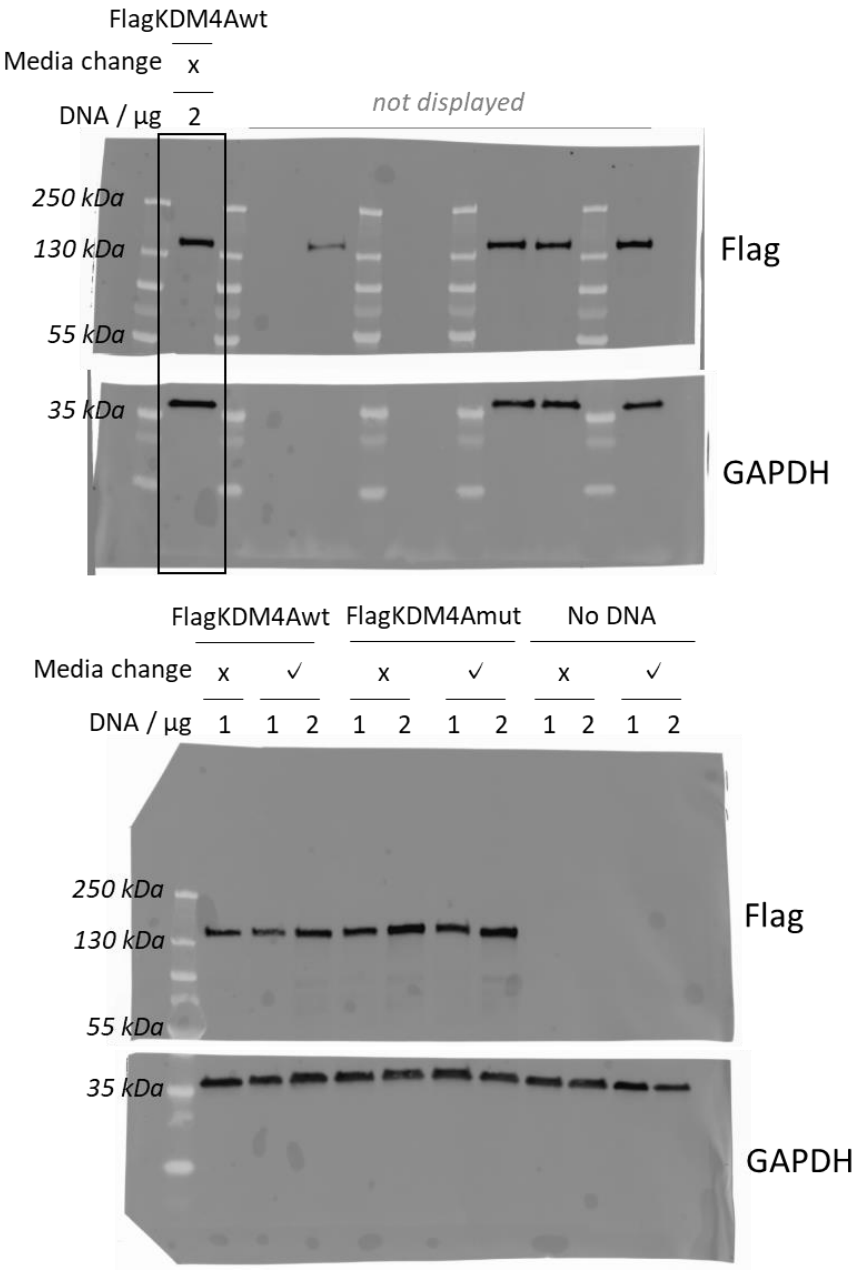


Figure A5.2 Uncropped western blot from Figure 3.4(B). Top blot shows the left hand panel in **Figure 3.4**, and bottom blot shows the main blot.

Appendix 5.2 Figure 3.5

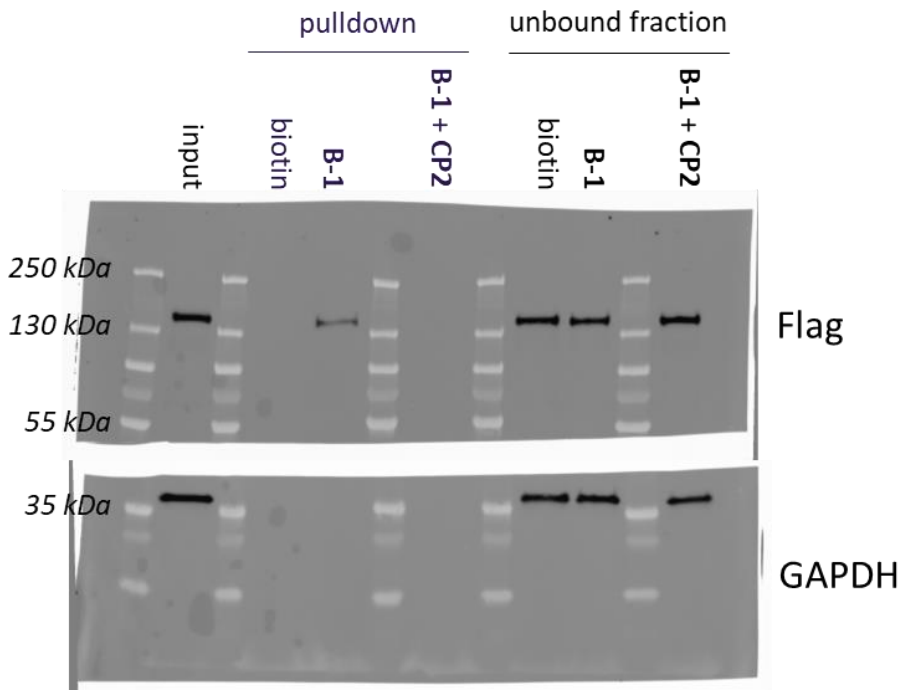


Figure A5.3 Uncropped western blot from Figure 3.5(B).

Appendix 5.3 Figure 3.6

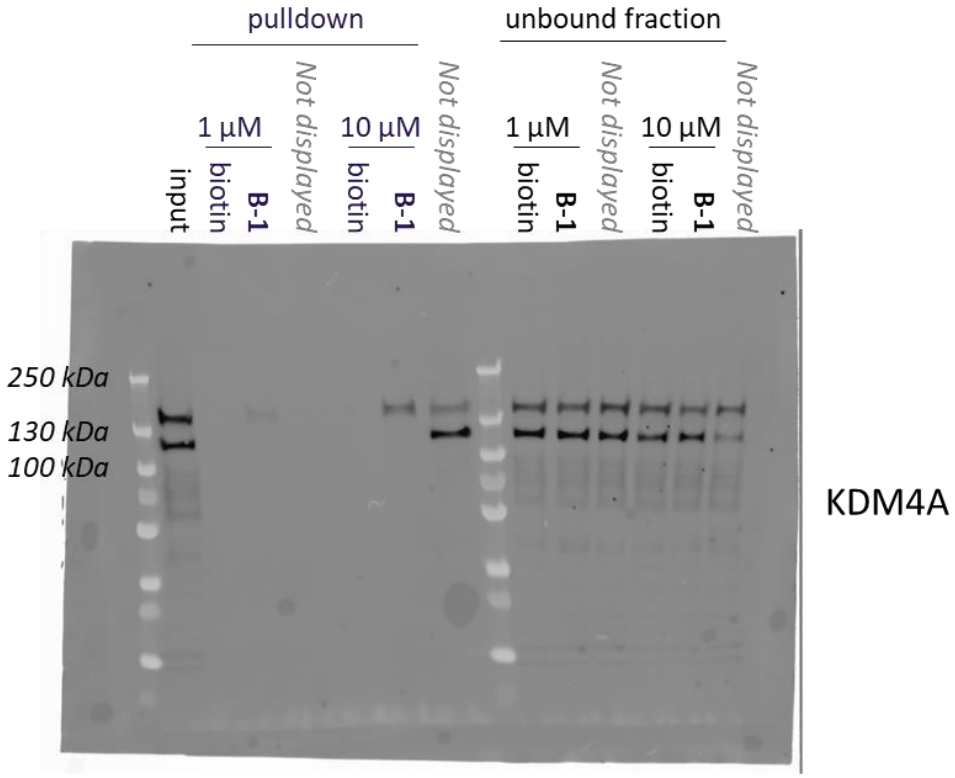


Figure A5.4 Uncropped western blot from Figure 3.6(A).

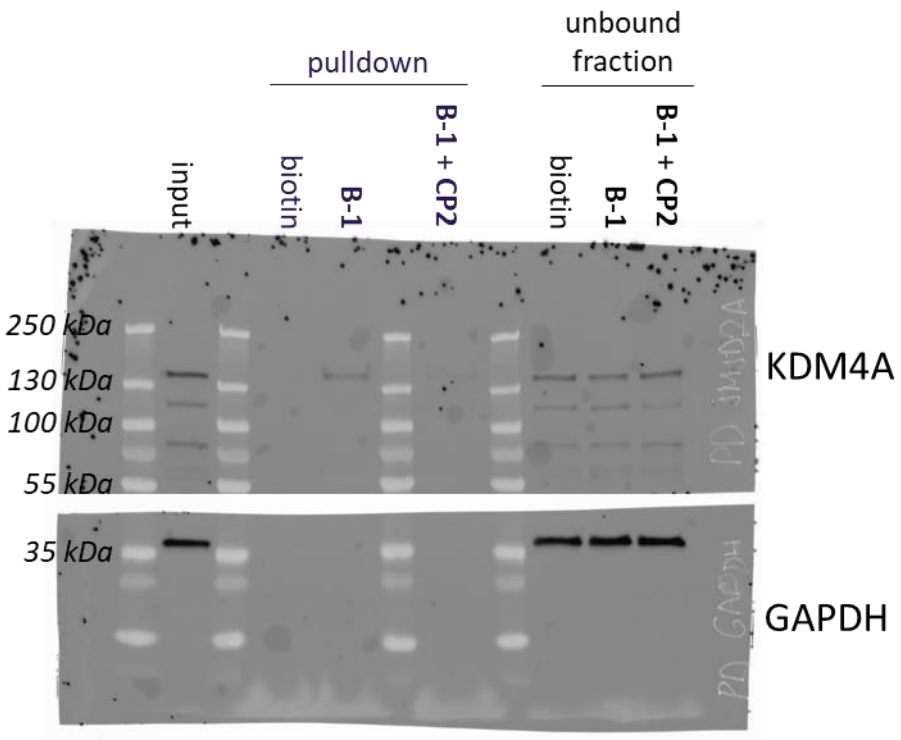


Figure A5.5 Uncropped western blot from Figure 3.6(B).

Appendix 5.4 Figure 3.7

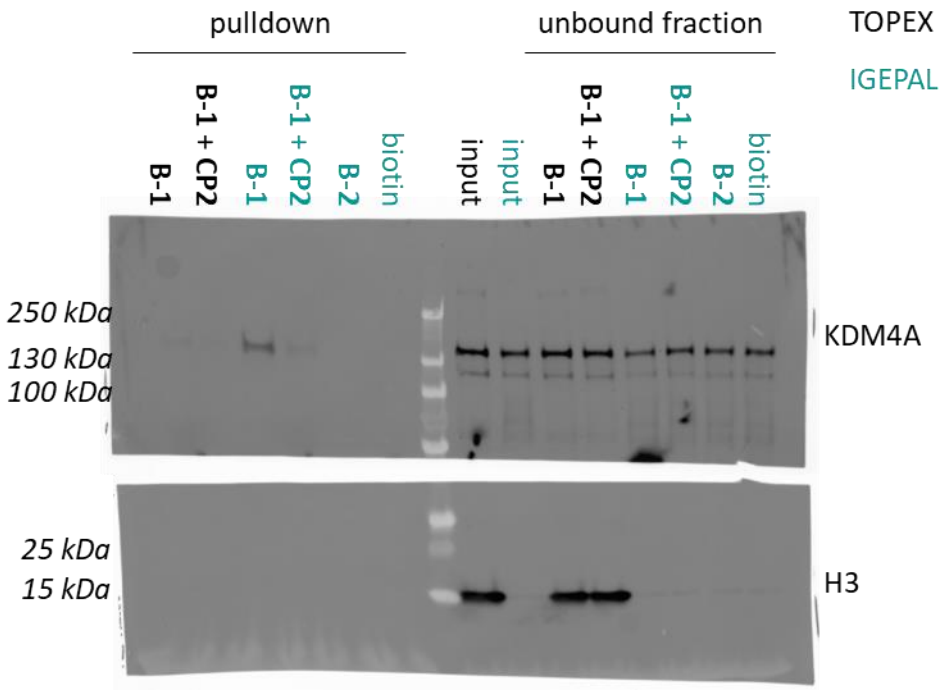


Figure A5.6 Uncropped western blot from Figure 3.7(A).

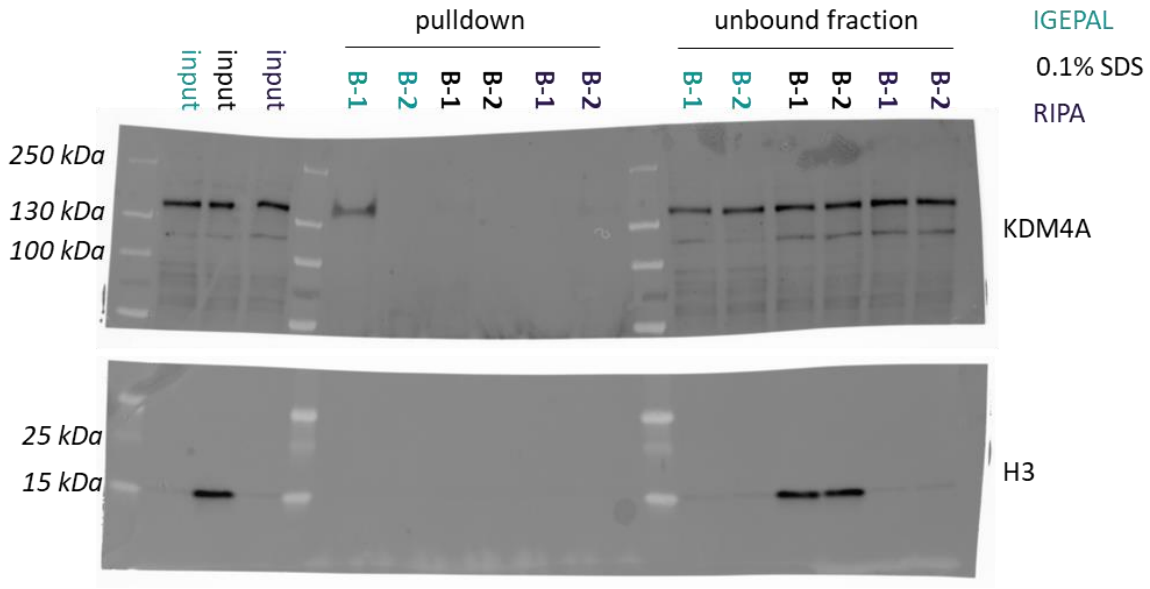


Figure A5.7 Uncropped western blot from Figure 3.7(B).

Appendix 5.5 Figure 3.9

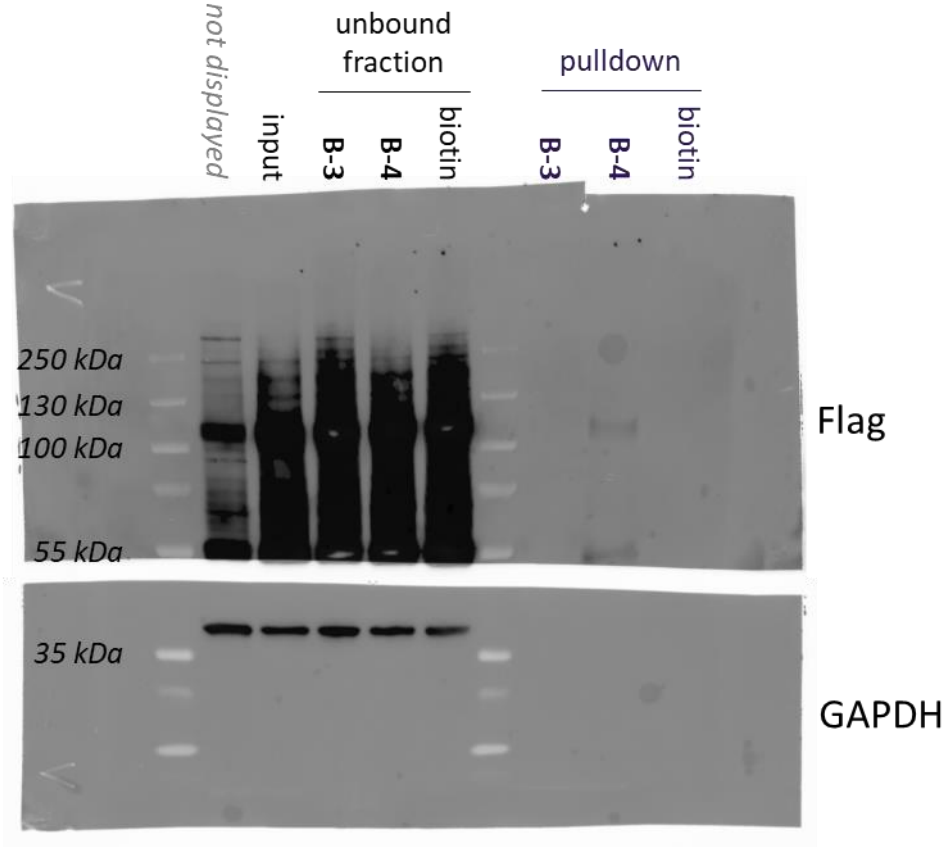


Figure A5.8 Uncropped western blot from Figure 3.9(A).

Appendix 5.6 Figure 3.10

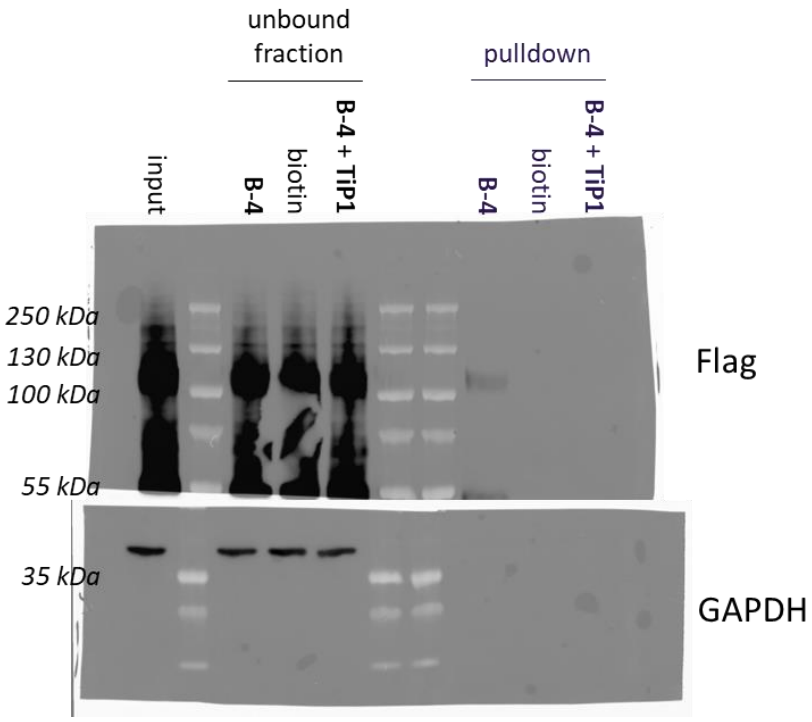


Figure A5.9 Uncropped western blot from Figure 3.10.

Appendix 6 Uncropped western blots from Chapter 4

Appendix 6.1 Figure 4.6

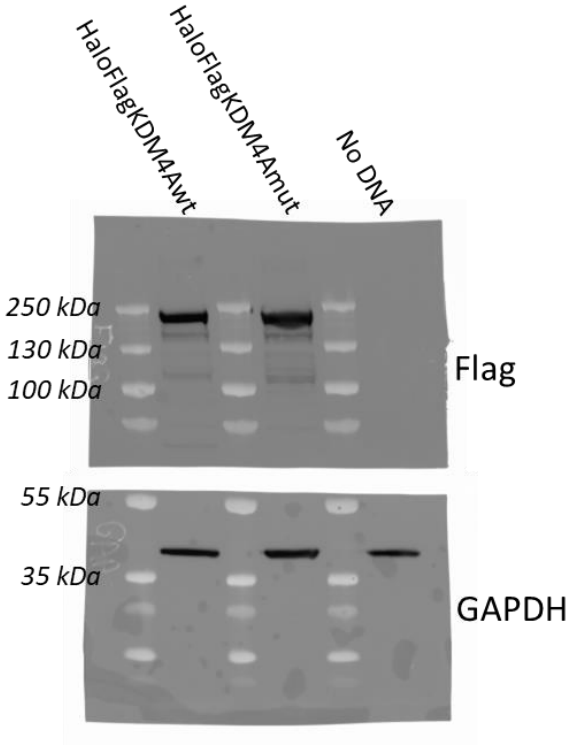


Figure A6.1 Uncropped western blot from Figure 4.6(A).

Appendix 6.2 Figure 4.7

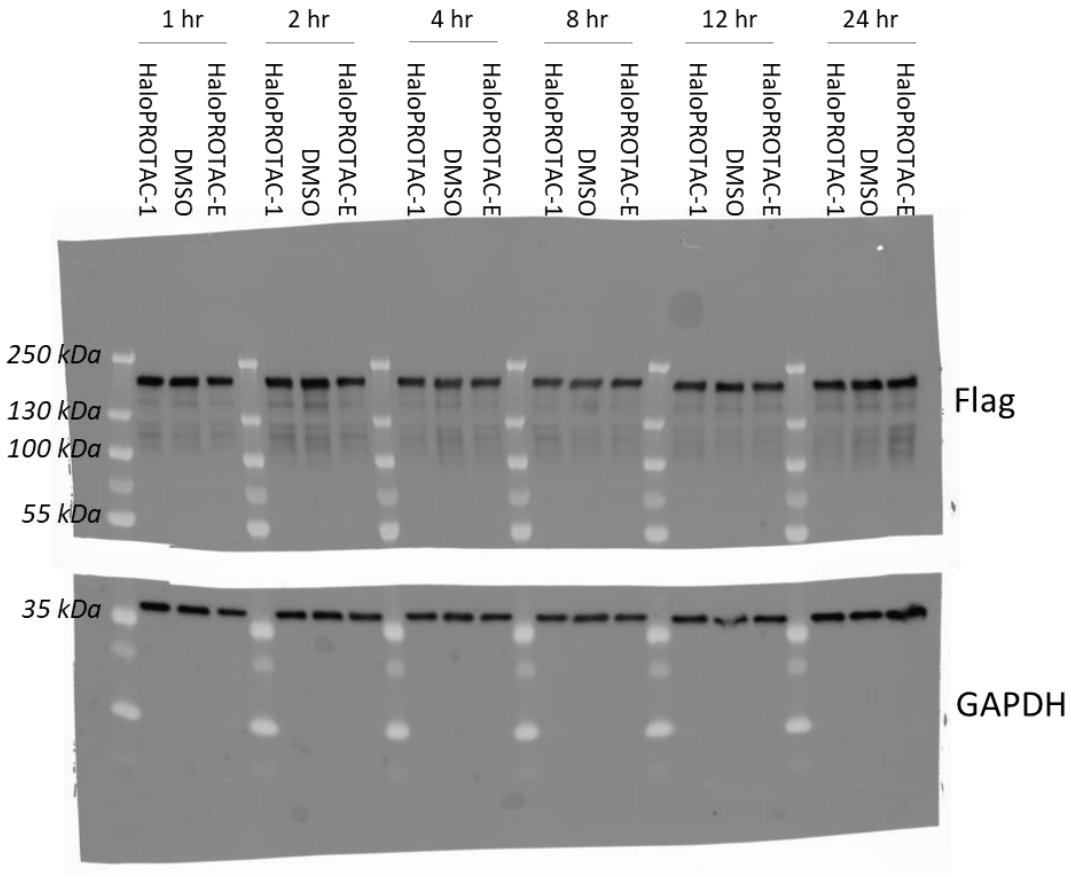


Figure A6.2 Uncropped western blot from Figure 4.7(A).

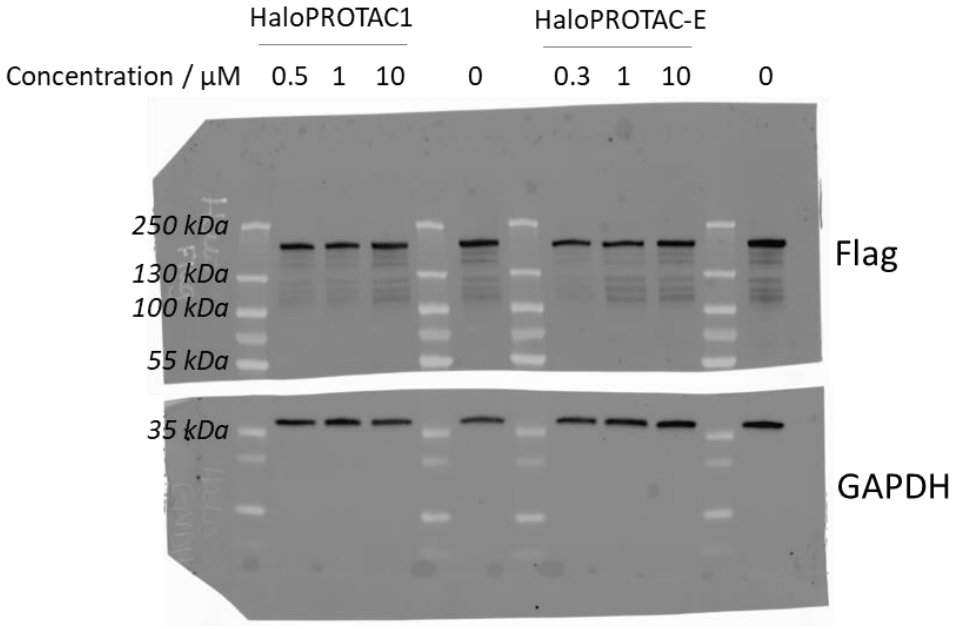


Figure A6.3 Uncropped western blot from Figure 4.7(B).

Appendix 6.3 Figure 4.8

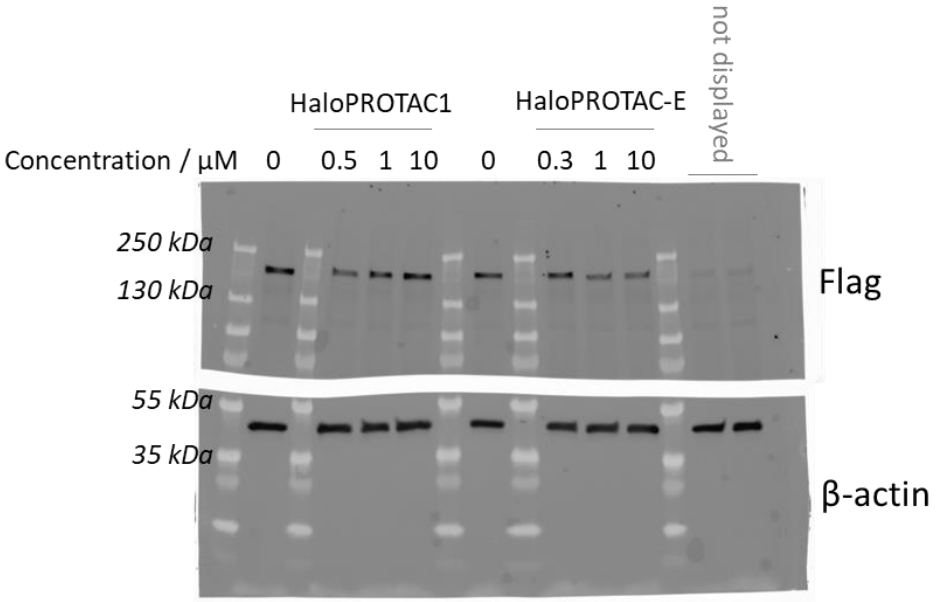


Figure A6.4 Uncropped western blot from Figure 4.8(A).

Appendix 6.4 Figure 4.10

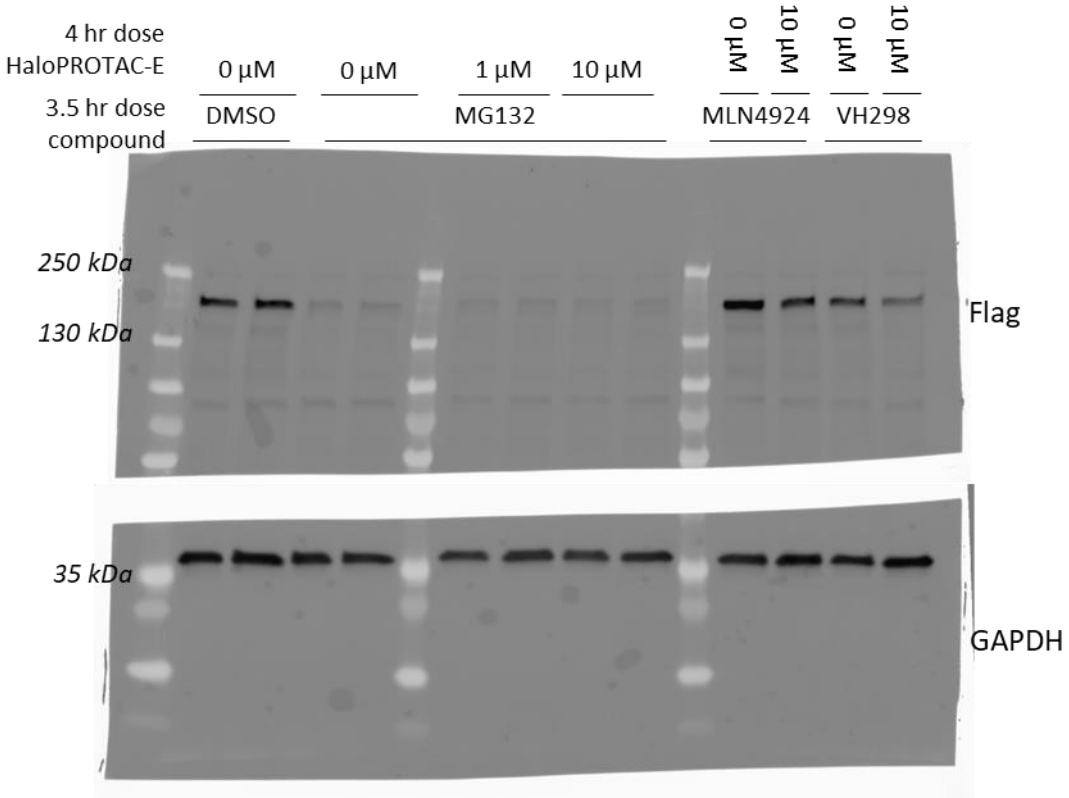


Figure A6.5 Uncropped western blot from Figure 4.10(A).

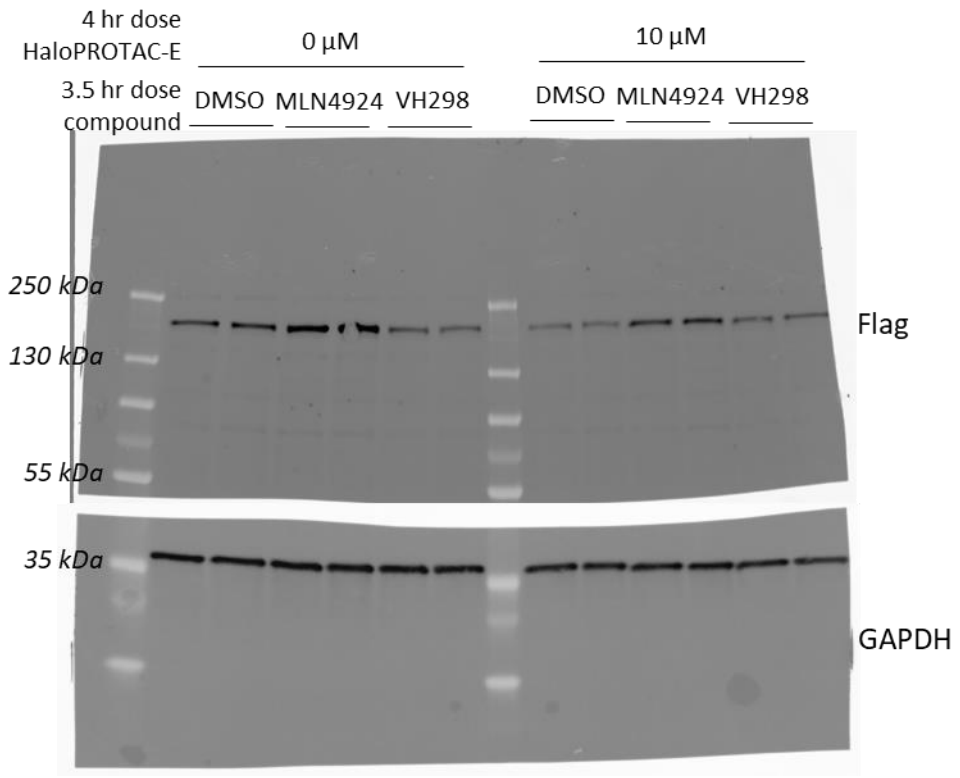


Figure A6.6 Uncropped western blot from Figure 4.10(B).

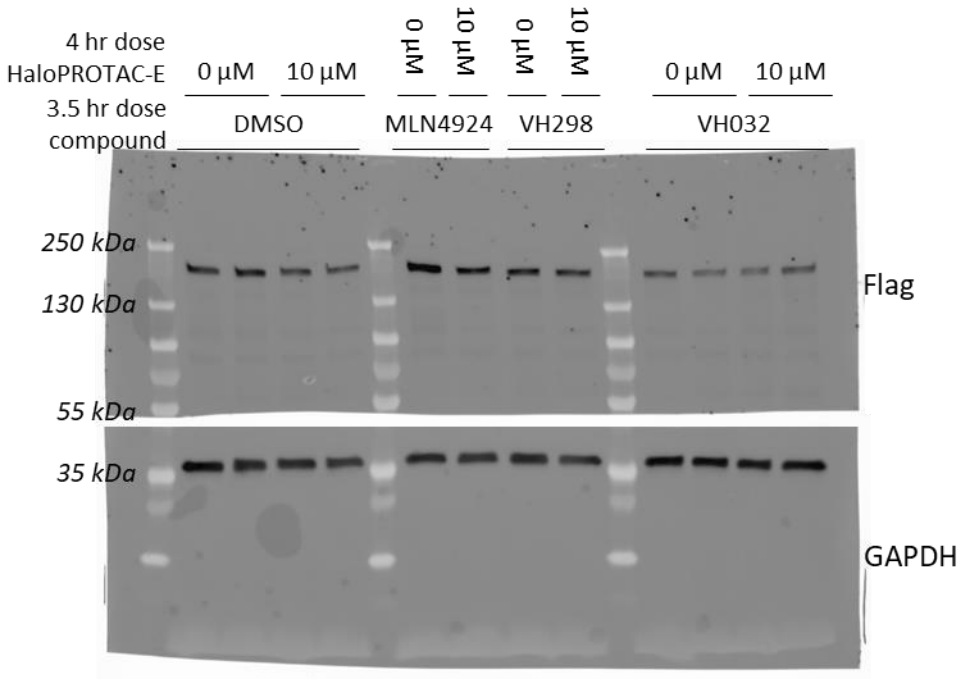


Figure A6.7 Uncropped western blot from Figure 4.10(C).

Appendix 6.5 Figure 4.12

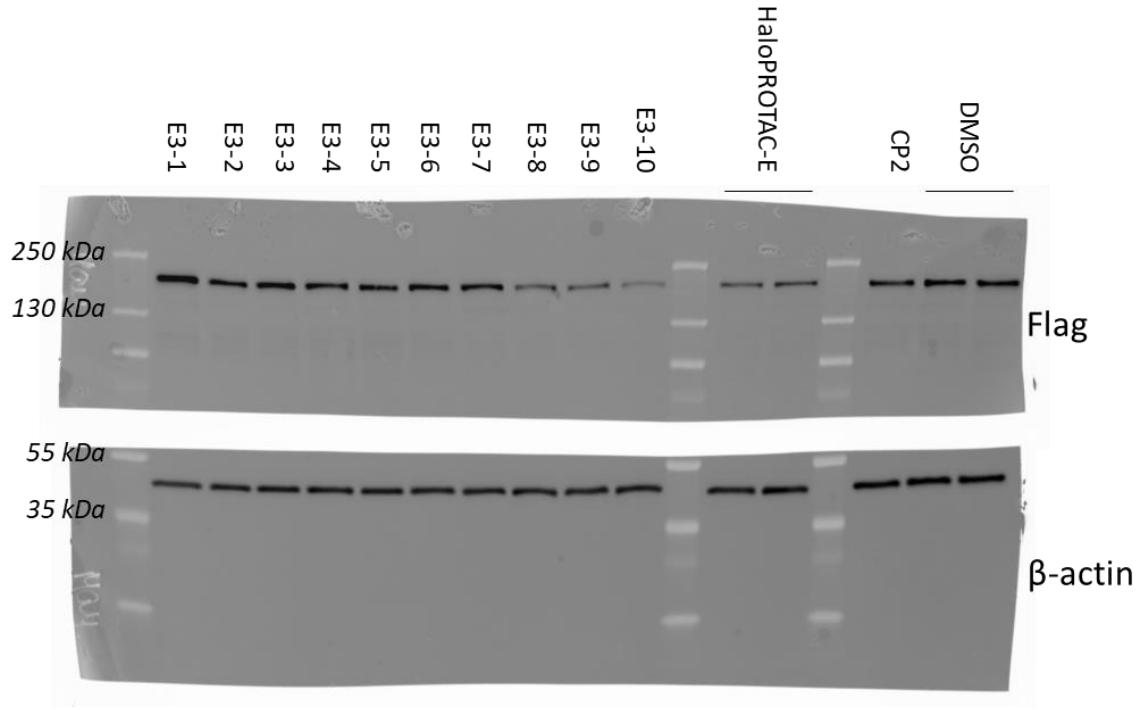


Figure A6.8 Uncropped western blot from Figure 4.12(A) Replicate 1.

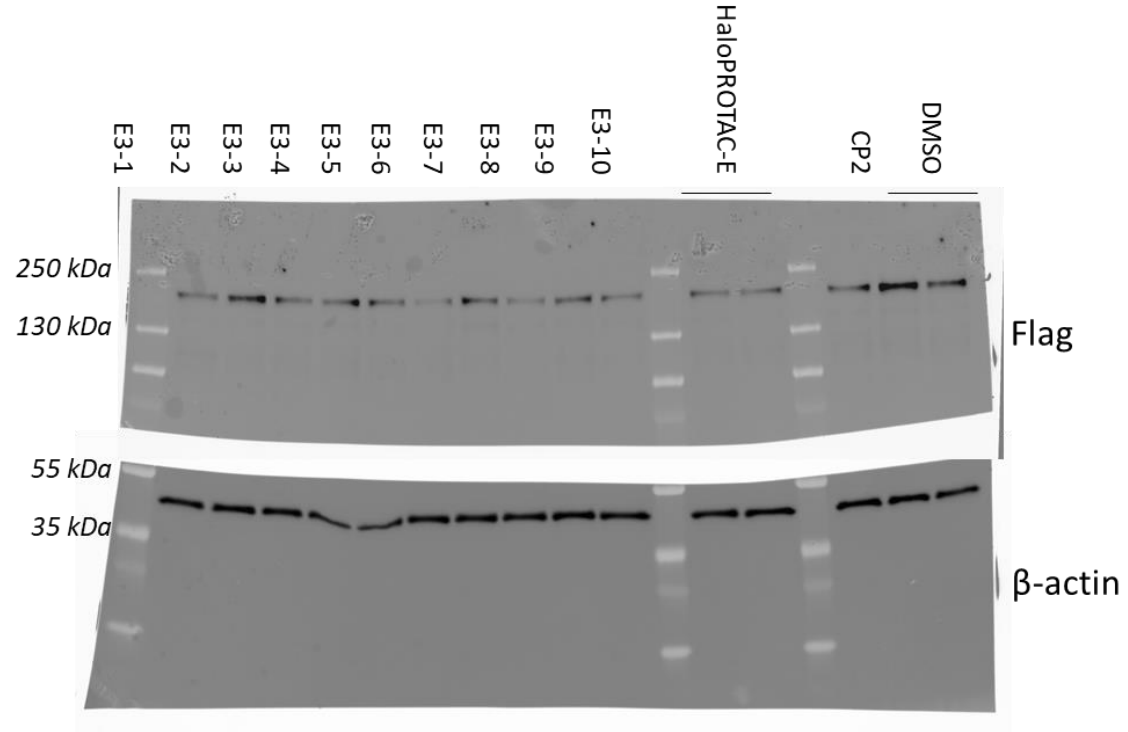


Figure A6.9 Uncropped western blot from Figure 4.12(A) Replicate 2.

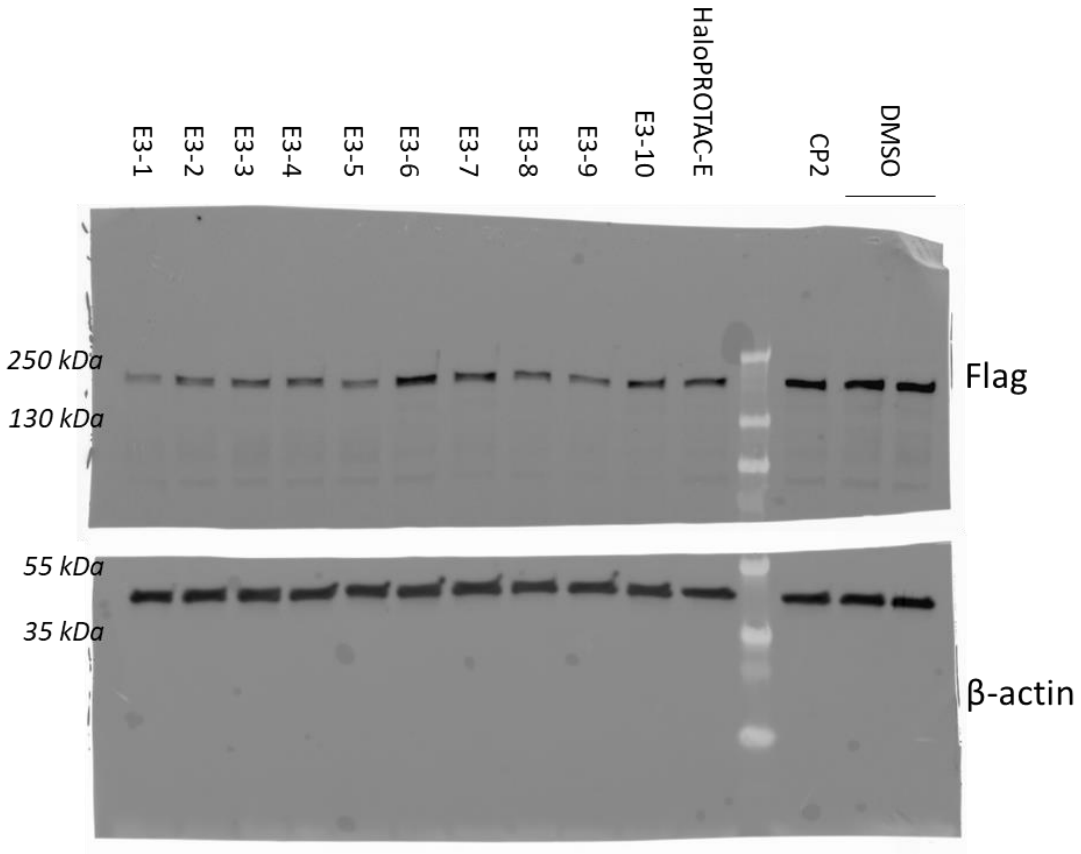


Figure A6.10 Uncropped western blot from Figure 4.12(A) Replicate 3.

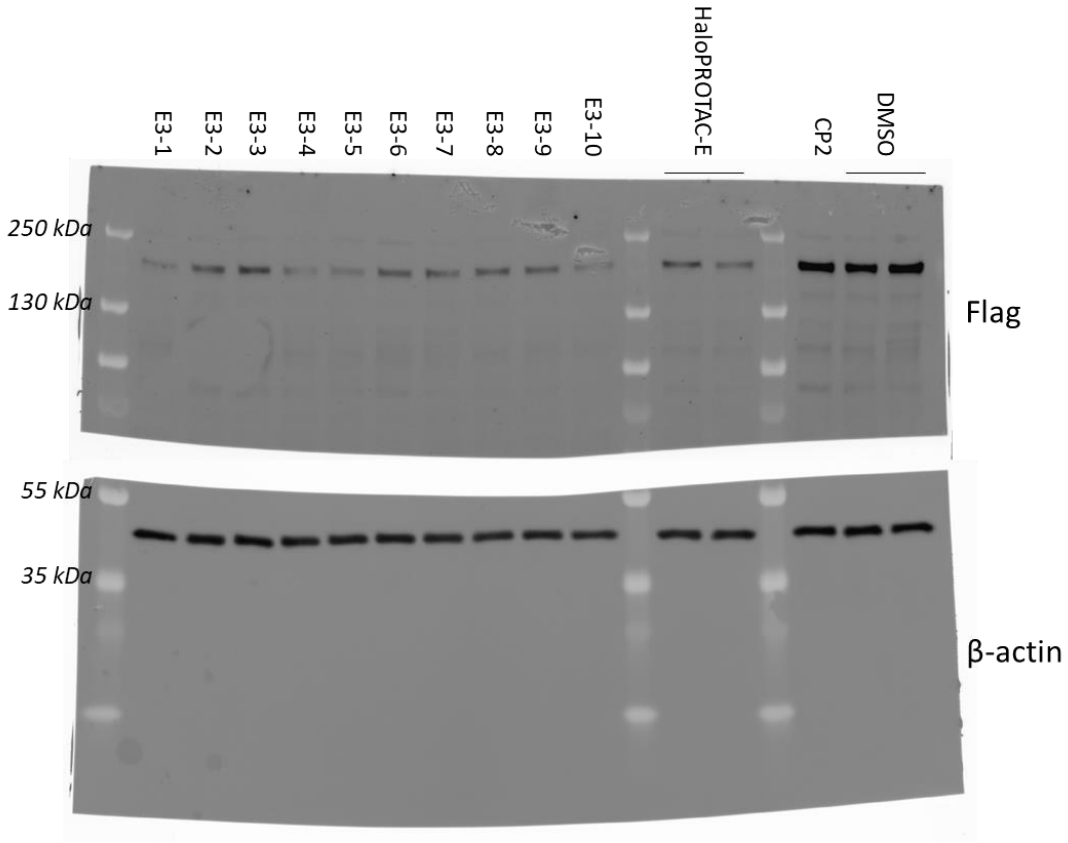


Figure A6.11 Uncropped western blot from Figure 4.12(A) Replicate 4.

Appendix 6.6 Figure 4.13

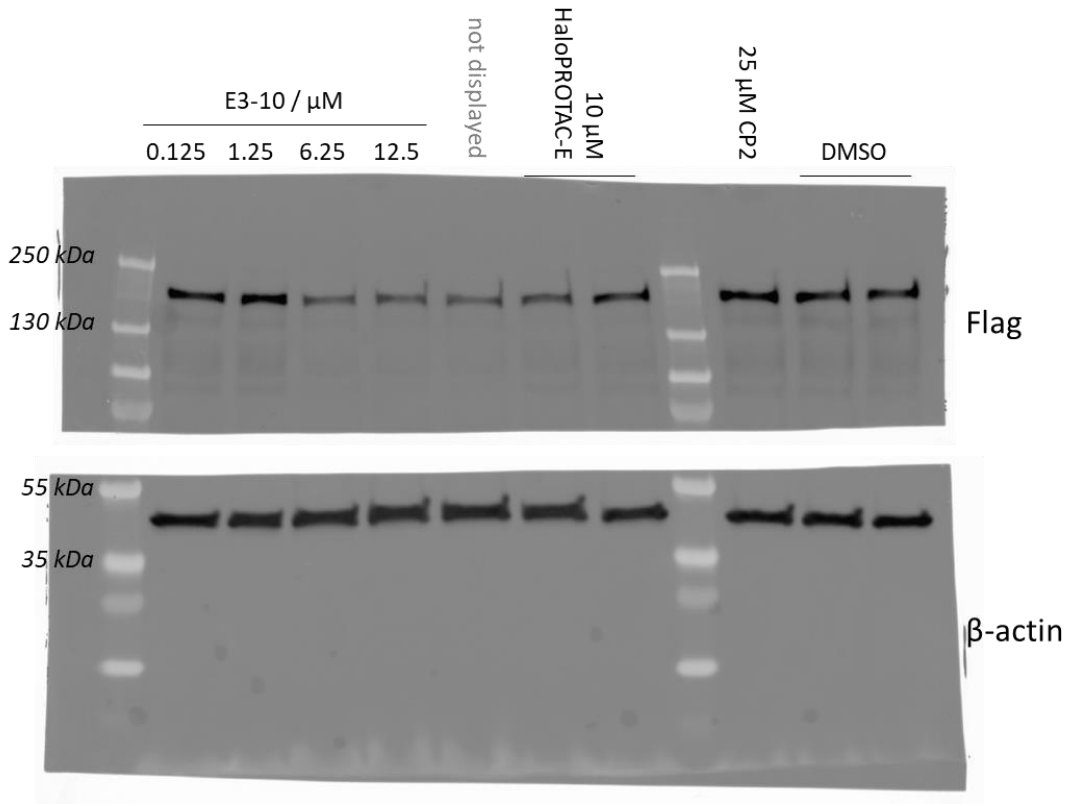


Figure A6.12 Uncropped western blot from Figure 4.13.

Appendix 6.7 Figure 4.15

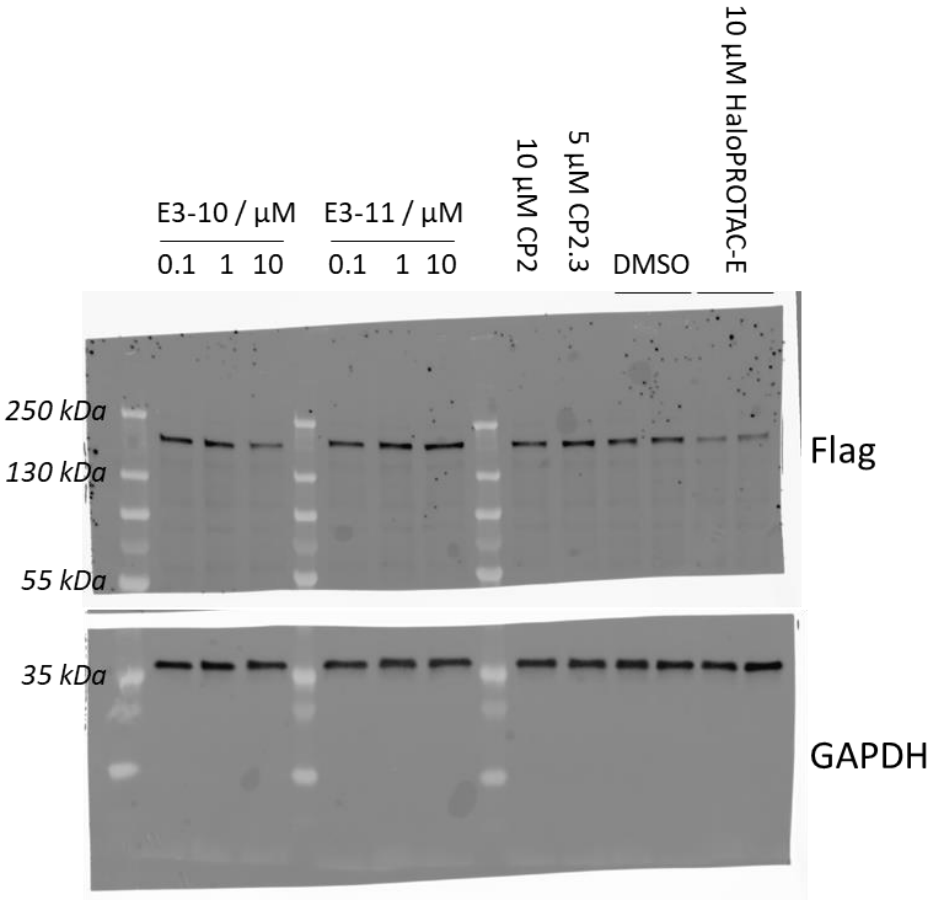


Figure A6.13 Uncropped western blot from Figure 4.15.

Appendix 6.8 Figure 4.16

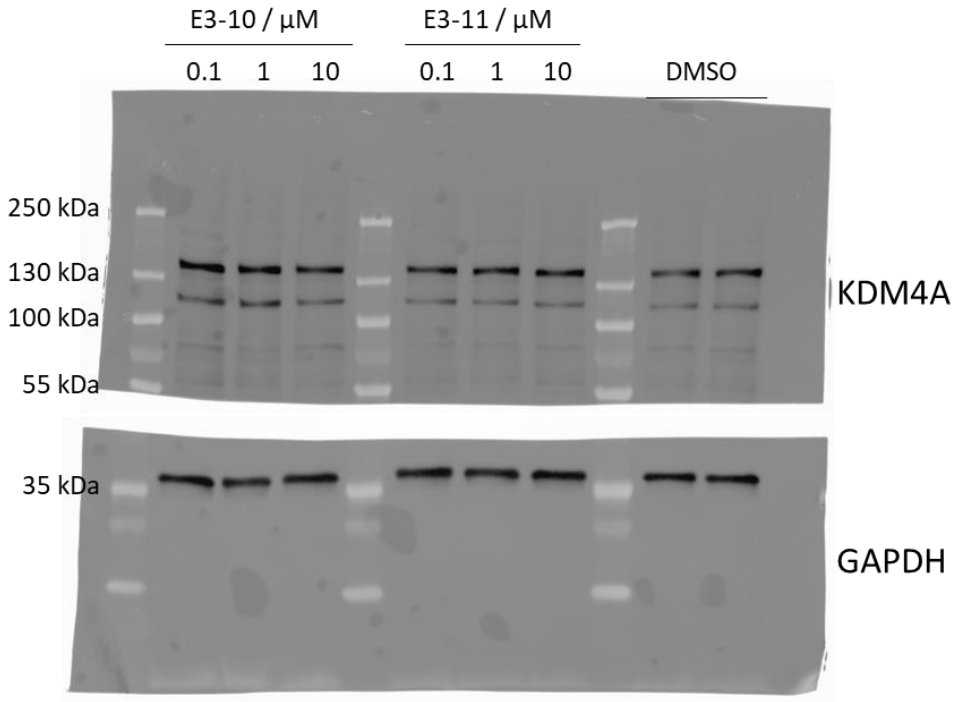


Figure A6.14 Uncropped western blot from Figure 4.16(A).

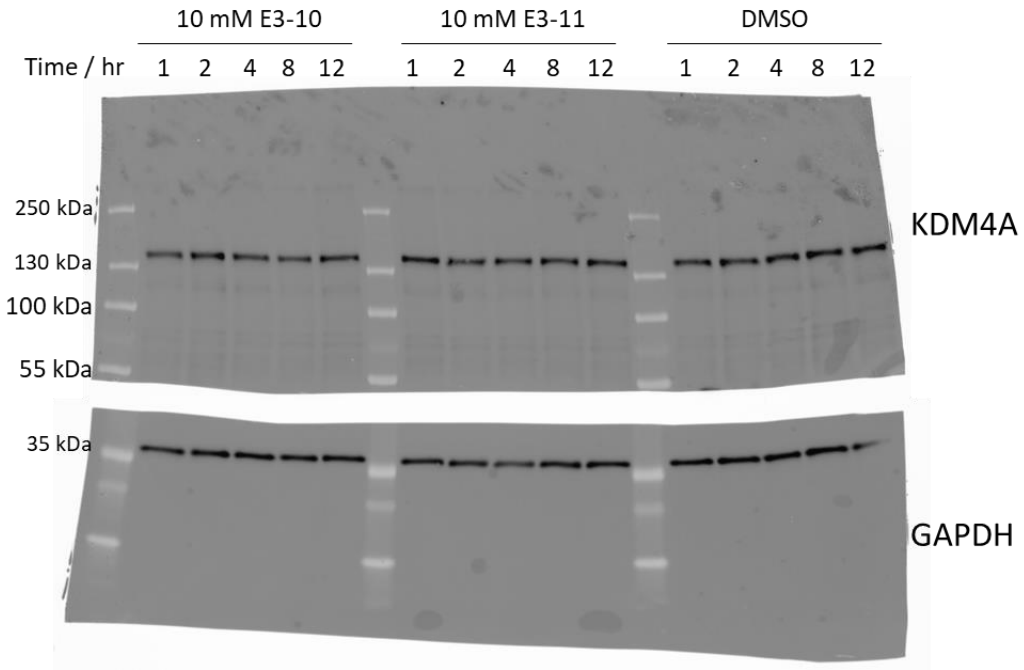


Figure A6.15 Uncropped western blot from Figure 4.16(B).

Appendix 6.9 Figure 4.17

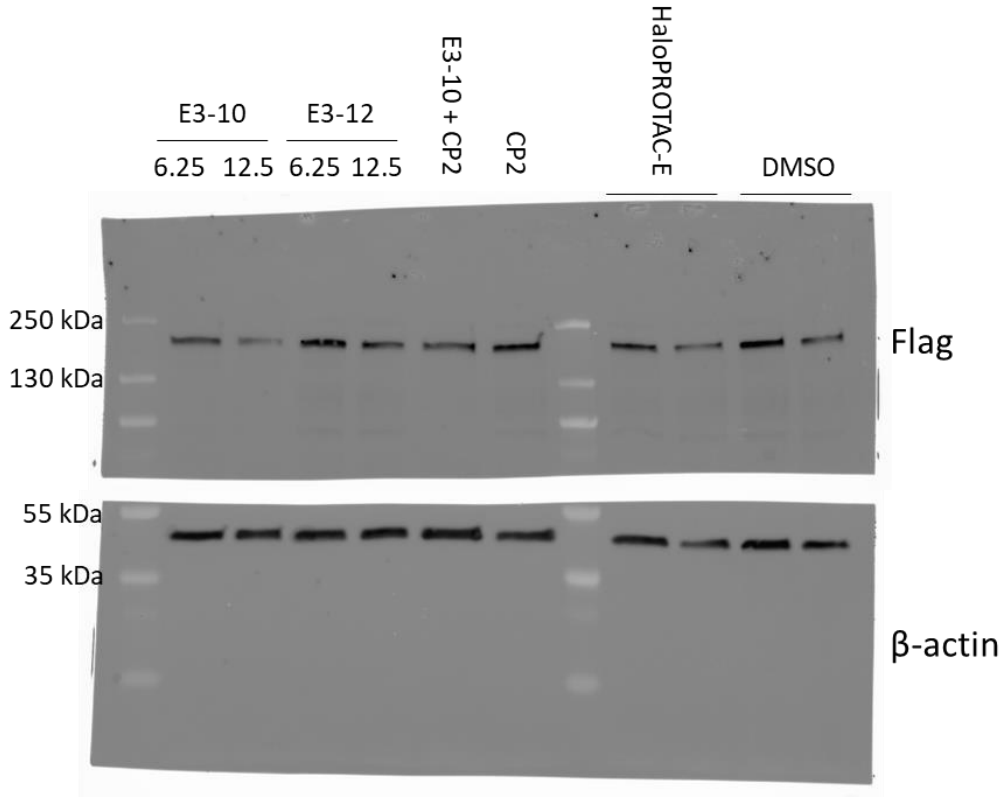


Figure A6.16 Uncropped western blot from Figure 4.17.

Appendix 7 Fmoc-protected L-amino acid details

Amino acid	Triple letter code	Single letter code	Fmoc-protected amino acid
Alanine	Ala	A	Fmoc-L-Ala-OH
Cysteine	Cys	C	Fmoc-L-Cys(Trt)-OH
Aspartic acid	Asp	D	Fmoc-L-Asp(O*t*Bu)-OH
Glutamic acid	Glu	E	Fmoc-L-Glu(O*t*Bu)-OH
Phenylalanine	Phe	F	Fmoc-L-Phe-OH
Glycine	Gly	G	Fmoc-L-Gly-OH
Histidine	His	H	Fmoc-L-His(Trt)-OH
Isoleucine	Ile	I	Fmoc-L-Ile-OH
Lysine	Lys	K	Fmoc-L-Lys(Boc)-OH
Leucine	Leu	L	Fmoc-L-Leu-OH
Methionine	Met	M	Fmoc-L-Met-OH
Asparagine	Asn	N	Fmoc-L-Asn(Trt)-OH
Proline	Pro	P	Fmoc-L-Pro-OH
Glutamine	Gln	Q	Fmoc-L-Gln(Trt)-OH
Arginine	Arg	R	Fmoc-L-Arg(Pbf)-OH
Serine	Ser	S	Fmoc-L-Ser(*t*Bu)-OH
Threonine	Thr	T	Fmoc-L-Thr(*t*Bu)-OH
Valine	Val	V	Fmoc-L-Val-OH
Tryptophan	Trp	W	Fmoc-L-Trp(Boc)-OH
Tyrosine	Tyr	Y	Fmoc-L-Tyr(*t*Bu)-OH

Table A7.1 Details of the Fmoc-protected L-amino acids used in SPPS.